Machine learning applied to species occurrence and interactions: the missing link in biodiversity assessment and modelling of Antarctic plankton distribution

Marco Grillo[1,2]* , Stefano Schiaparelli[2,3], Tiziana Durazzano[4], Letterio Guglielmo[5,6], Antonia Granata[7] and Falk Huettmann[8]

Abstract

Background Plankton is the essential ecological category that occupies the lower levels of aquatic trophic networks, representing a good indicator of environmental change. However, most studies deal with distribution of single species or taxa and do not take into account the complex of biological interactions of the real world that rule the ecological processes.

Results This study focused on analyzing Antarctic marine phytoplankton, mesozooplankton, and microzooplankton, examining their biological interactions and co-existences. Field data yielded 1053 biological interaction values, 762 coexistence values, and 15 zero values. Six phytoplankton assemblages and six copepod species were selected based on their abundance and ecological roles. Using 23 environmental descriptors, we modelled the distribution of taxa to accurately represent their occurrences. Sampling was conducted during the 2016–2017 Italian National Antarctic Programme (PNRA) 'P-ROSE' project in the East Ross Sea. Machine learning techniques were applied to the occurrence data to generate 48 predictive species distribution maps (SDMs), producing 3D maps for the entire Ross Sea area. These models quantitatively predicted the occurrences of each copepod and phytoplankton assemblage, providing crucial insights into potential variations in biotic and trophic interactions, with significant implications for the management and conservation of Antarctic marine resources. The Receiver Operating Characteristic (ROC) results indicated the highest model efficiency, for Cyanophyta (74%) among phytoplankton assemblages and *Paralabidocera antarctica* (83%) among copepod communities. The SDMs revealed distinct spatial heterogeneity in the Ross Sea area, with an average Relative Index of Occurrence values of 0.28 (min: 0; max: 0.65) for phytoplankton assemblages and 0.39 (min: 0; max: 0.71) for copepods.

Conclusion The results of this study are essential for a science-based management for one of the world's most pristine ecosystems and addressing potential climate-induced alterations in species interactions. Our study emphasizes the importance of considering biological interactions in planktonic studies, employing open access and machine learning for measurable and repeatable distribution modelling, and providing crucial ecological insights for informed conservation strategies in the face of environmental change.

Keywords Copepoda, Phytoplankton, Ross sea, Terra Nova Bay, Species Distribution Model, Marine trophic web

---

*Correspondence:
Marco Grillo
m.grillo@student.unisi.it
Full list of author information is available at the end of the article

Springer Open

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 2 of 18

Introduction

Among the most uncontaminated places on Earth are the Antarctic continent (Leihy et al. 2020) and the Southern Ocean that surrounds it. Particularly, the Southern Ocean is characterized by cold temperatures, and thanks to the onset of the Antarctic Circumpolar Current and the Antarctic Polar Front, high levels of endemism in a variety of taxa and the evolution of unique physiological adaptations (Peck 2018) have been ensured. Despite the Southern Ocean still being relatively pristine, it is currently under several potential threats, such as commercial fishing, warming, acidification, shifting of ocean fronts (Peel et al. 2019), and reduction of sea ice extent and duration (Chown et al. 2012; Turner et al. 2014). These phenomena may vastly compromise its marine ecosystems, affecting their communities from all points of view (Boyd et al. 2008; Constable et al. 2023).

In an attempt to better preserve these unique ecosystems the Ross Sea Region Marine Protected Area (RSRMPA), was established on the 1st December 2017 (Brooks et al. 2020). Wilderness and protected areas of the world represent the last key areas for maintaining biodiversity and associated ecosystem services on a large scale (Mittermeier et al. 2003; Di Marco et al. 2019). Thanks to their ecological integrity, those areas provide baselines for studying, assessing, and managing potential anthropogenic impacts, either current or future (Cole and Landres 1996). Furthermore, thanks to the "spillover effect" or the export of species on adjacent and non-protected proximity areas, they ensure a positive ecological effect and cross-habitat movement of species (Gell and Roberts 2003; Hixon et al. 2014; González-Herrero et al. 2023). The RSRMPA was specifically designed to protect and maintain the function and structure of marine ecosystems. Among these are included those areas known to be important for the life cycles of important commercial species such as the Antarctic toothfish (*Dissostichus mawsoni* Norman, 1937) (Atkinson 1998; Turner 2004; Parker et al. 2021) and the Antarctic krill (*Euphausia superba* Dana, 1850). Currently, the RSRMPA is the largest and most productive MPA of the whole Southern Ocean (Arrigo et al. 2008). However, the large geographical scales considered combined with the objective logistic difficulties of sampling at high latitudes made it difficult to study and establish solid baselines. Despite these intrinsic difficulties, it is nonetheless necessary and mandatory to base any conservation or monitoring effort on the availability of robust data. Only this kind of information will guarantee robust estimates of potential future changes in the composition and the structure of food webs up to the whole marine ecosystems (Smetacek and Nicol 2005; Clarke and Harris 2003; Arrigo et al. 2002; Fraser et al. 2023). So far, the biodiversity analyses were generally conditioned by large geographical sampling bias deriving from the nature of sampling efforts, typically sparse and patchy (e.g. Peel et al. 2019). Luckily, innovative techniques such as machine learning (ML) and artificial intelligence (AI) represent powerful tools to increase and refine our understanding of complex non-linear patterns and our predictive power (Humphries et al. 2019). In the case of Antarctic biodiversity, these innovative techniques were already used to implement Species Distribution Models (SDM) on a variety of taxa, such as copepods (Pinkerton et al. 2010a, b; Grillo et al. 2022), echinoderms (Guillaumot et al. 2016), krill (Lin et al. 2022), fishes (Yates et al. 2019), birds and mammals (Huettmann and Schmid 2014; El-Gabbas et al. 2021). In the context of SDM, it is of pivotal importance for the concept of ecological niches, or the environmental conditions necessary for a species to survive and maintain its populations in a given habitat (Colwell and Rangel 2009). Specifically, the ecological niche can be visualised as a 'hyper-volume with n-dimensions', characterised by environmental descriptors that allow for the presence of an organism in that particular habitat (Hutchinson 1957; Steiner et al. 2023; Huettmann et al. 2023).

Modelling niches in polar marine ecosystems are remarkably difficult due to the complex interplay of several complex environmental drivers. For example, the sea ice extent and seasonality (Atkinson 1998; Atkinson et al. 2004; Norkko et al. 2007; O'Driscoll et al. 2011) condition the release of food substances and regulate the available underwater light (Dayton et al. 1986; Clarke 1988), supporting sympagic communities (Garrison et al. 1986; Guglielmo et al. 2000; Granata et al. 2022; Swadling et al. 2023) and regulating the abundance of key taxa such as the Antarctic krill (e.g. Atkinson et al. 2004). The release of organic matter contained in the sea ice, in particular, fertilizes and regulates phytoplankton blooms that, in turn, trigger the zooplankton swarm in the late Austral spring and summer (Meiners et al. 2012; Schnack-Schiel and Hagen 1995; Smith Jr and Nelson 1990; Alexander and Niebauer 1981; Cau et al. 2021).

The aim of this investigation is to quantitatively explore and assess, for the first time, the possible interspecific relationships that occur between different Antarctic planktonic taxa while also considering marine environmental descriptors. This type of result also identifies species and environmental characteristics that may influence their relative community distributions.

Predictive maps of the main phytoplankton assemblages prevalent in Ross Sea waters were created for the first time and additionally, for copepod species that are common and abundant in the same marine area.

This enables to scale-up and clarify possible predictive distributions of phytoplankton and mesozooplankton in

the whole Ross Sea area, also identifying species and features that may influence their distributions. In order to achieve this we implemented the approach of Breiman (2001) to deduce the three-dimensional distributions from predictions for a large scale, instead of using only field data or model adaptations and distorted and poorly performing parsimonious models (e.g., Guthery 2008; Elith et al. 2006).

Materials and methods
Field data and study area
The study area is a large portion of the Ross Sea, spanning from near the Drygalski Ice Tongue in Terra Nova Bay (TNB) to the Central Basin (Fig. 1). Specimens were obtained in the framework of the XXXIInd Italian National Antarctic Program (PNRA) expedition "P-ROSE" (Plankton biodiversity and functioning of the Ross Sea ecosystems in a changing Southern Ocean, PNRA16_00239, Melchiori 2017). Sampling activities were carried out during the Austral Summer 2017 on the R/V "Italica". The P-ROSE research aimed at identifying signals and/or response patterns of the planktonic compartment to climate change in progress. During the expedition, a series of zooplankton and micronekton samplings down to a depth of 200 m were carried out, with vertical fishing employing a WP2 standard net for quali-quantitative analysis. Various environmental data

such as water temperature (°C), salinity (PSU), fluorescence (Chl-a µg/l), oxygen concentration (mg/l), conductivity (S/m), density (kg/m^3), potential (θ_e), depth (m) and pressure (dB) were also recorded at each of the sampling stations with the use of a CTD system. The zooplankton and micronekton samples were fixed with formalin neutralized 4% fixative solution. Specimens were later transferred from a 4% formalin fixative solution to a 96% ethanol solution and are now present in the collections of the Italian National Antarctic Museum [MNA, Section of Genoa (Schiaparelli et al. 2019)] stored in 96% ethanol. In addition to the P-ROSE data, further data on other zooplanktonic taxa such as microplankton and phytoplankton collected in the Ross Sea sector were included in this study and submitted for analysis.

Additional data from available literature were also digitalized and used in the analysis to cover other groups. Phytoplankton data were obtained from Bolinesi et al. (2020) and Cordone et al. (2022), while microzooplankton data from Monti-Birkenmeier et al. (2022). The individual datasets are described and available in the supplemental materials respectively. Both studies conducted an identification of planktonic organisms down to the lowest taxonomic level. The sampling for these identifications took place during the same Antarctic study campaign at the designated analysis stations. Details are provided in Appendix A.
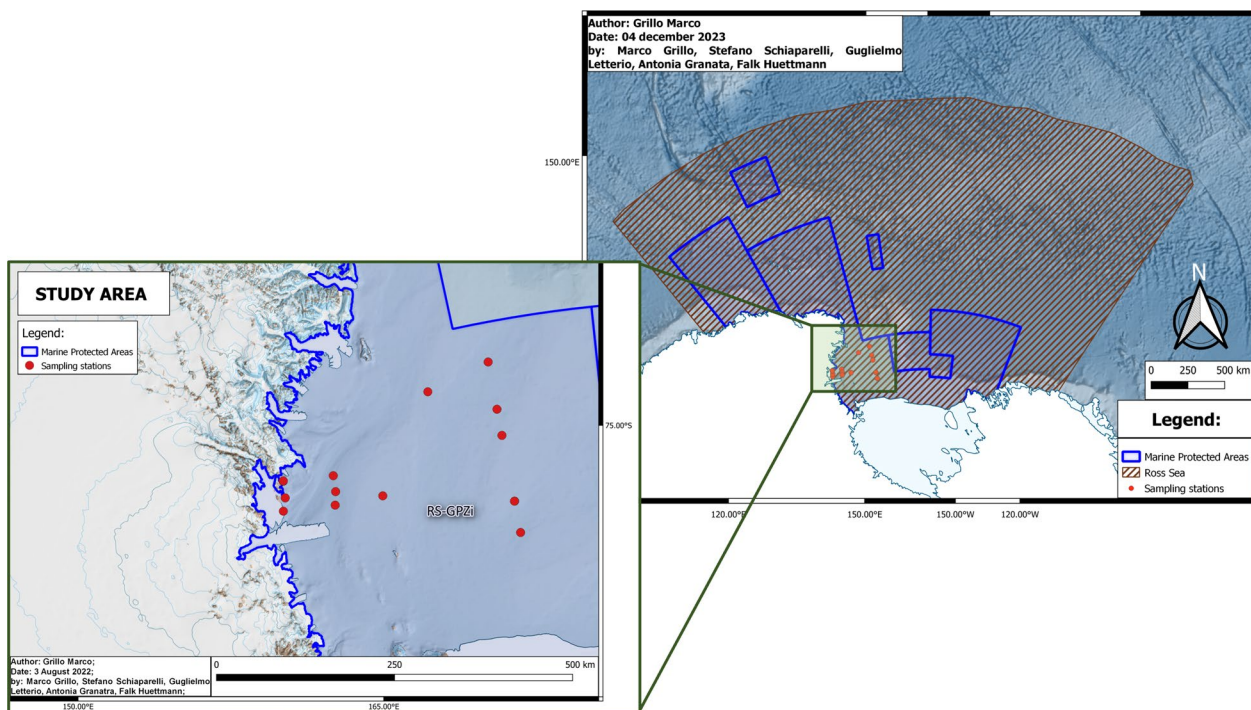

Fig. 1 Study area with stations sampled and prediction area

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 4 of 18

Data processing

We used the OpenSource R software (version 4.3.2; Windows 10) as well as OpenGIS Quantarctica layers package (Matsuoka et al. 2021) to explore, visualize, and map data as well as model predictions with basemaps. We used the projection of geographic WGS84 in decimal latitude and longitude (6 decimals), for data exploration, and the stereographic Antarctic projection WGS 84/Antarctic Polar Stereographic EPSG: 3031 for map display.

To obtain other values of the environmental descriptors present on the Polar Macroscope Layer, the points with the presence and absence data were superimposed and extracted from the attribute table using the "*extract multiple values in points*" (GIS) function. The use of these environmental layers has already been employed in the field of SDM, e.g., Huettmann and Schmid (2014), Koubbi et al. (2014), and Grillo et al. (2022). The final table ("data cube") obtained at the end of the process was used to model-predict the distribution of copepods and phytoplankton's assemblages with machine learning and its subsequent modelling and representation on Quantartica. We employed a point lattice with a resolution of 1 km (Appendix B) to establish a prediction grid. This

grid was overlaid and assessed using each of the machine learning methods, leading to the creation of the final prediction surface.

Correlation matrix: field data

To visualize possible biological interactions through correlations between phytoplanktonic, microplanktonic and zooplanktonic taxa, we used abundance data transformed into presence ('1') and absence ('0') values, together with oceanographic data recorded in situ (such as pressure, conductivity, temperature, fluorescence, salinity potential, oxygen density, etc.) and the environmental descriptors listed in Table 1, were used to perform this type of analysis. The jSDM package (version 4.3.2) (https://cran.r-project.org/web/packages/jSDM/index.html) (Warton et al. 2015) was used to perform this analysis.

Modelling: predictions

Step 1: To obtain a good and more robust signal extracted from the data cube and its selected models, "data cloning" was carried out, since the higher number of columns vs rows would have inhibited a solid fit. This technique, described by Lele et al. (2007), requires the

Table 1 Environmental descriptor sampled as OpenSource GIS Quantarctica layers, geotiff format

Descriptor	Description	Quantartica Package
Longitude	Longitude in decimal degrees	Polar macroscope layer
Latitude	Latitude in decimal degrees	Polar macroscope layer
EventID	Date of sampling	
verbatimDepth	Sampling depth	Polar macroscope layer
Temperature	Temperature of the water	Polar macroscope layer
Salinity	Salinity of the water	Polar macroscope layer
Oxygen dissolved	Oxygen dissolved in the water column	Polar macroscope layer
SeaCurrent	Mean surface current speed	Polar macroscope layer
SeaIce	Fraction of sea ice	Polar macroscope layer
GeoThermal Flux	Bottom temperature	Polar macroscope layer
Phosphate surface	Phosphate dissolved at surface	Polar macroscope layer
Phosphate 50 m	Phosphate dissolved at 50 m	Polar macroscope layer
Phosphate 200 m	Phosphate dissolved at 200 m	Polar macroscope layer
Silicate surface	Silicate dissolved at surface	Polar macroscope layer
Silicate 50 m	Silicate dissolved at 50 m	Polar macroscope layer
Silicate 200 m	Silicate dissolved at 200 m	Polar macroscope layer
NOx surface	NOx dissolved at surface	Polar macroscope layer
NOx 50 m	NOx dissolved at 50 m	Polar macroscope layer
NOx 200 m	NOx dissolved at 200 m	Polar macroscope layer
Depth	Bottom depth	Polar macroscope layer
Slope	Bottom's slope	Polar macroscope layer
NOx	Average of NOx dissolved in the water column	Polar macroscope layer
POx	Average of POx dissolved in the water column	Polar macroscope layer
N/P	Ratio of NOx to POx layers	Polar macroscope layer

These data are available Appendix B

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 5 of 18

rows of the matrix to be copied so that the actual model achieves a better fit on the additional data. This is often possible due to the subsampling methods employed in the ML methods used. In this way, the obtained signal can become more clear and robust (see Jiao et al. 2016 for an application and details). Our dataset originally had 61 rows. The structure was repeated (rows copied and pasted) three times, thus obtaining a dataset of 183 rows and 90 columns with 28 predictors.

Step 2: Salford Predictive Modeler 8 (SPM, https://www.minitab.com/en-us/products/spm/) was used to obtain the predicted distributional values.

Four predictive distribution models were created in SPM: TreeNet (TN), RandomForest (RF), CART Decision Tree (CA), and a combined model, i.e., Ensemble (EN), as previously done by other Authors (Meißner et al. 2013; Hardy et al. 2011; Huettmann and Schmid 2014).

The TN, RF, CA, and Ensemble models are widely acknowledged as among the most powerful machine-learning methods in ecology studies (Breiman 2001; Friedman 2002; Hegel et al. 2010; Mi et al. 2017). For more details on TN, RF, and CA in SPM and their performances, we refer readers to the user guide document online (https://www.salford-systems.com/products/spm/userguide), while EN models averaged values of the former three model results.

The validation method used for the selected models is V-fold cross-validation (CV) for TN and CA, while for RF, fraction of cases selected at random was used.

For CV, the entire dataset is used for learning purposes and then is partitioned into ten bins. At each fold in tenfold CV, nine bins are used as a training set and the remaining bin is used as a test set. After all tenfolds are completed, the test results from each fold are averaged to get a fair test estimate of the all-data model performance (https://www.minitab.com/en-us/products/spm/user-guides/).

The model is using approximately 20% of the data for testing. Final results are reported for both the training and testing data, and the percentage can be modified (https://www.minitab.com/en-us/products/spm/user-guides/).

From the algorithms, a Relative Index of Occurrence (RIO) was obtained for the lattice to show the suitable habitats for the copepods and phytoplankton assemblages. Furthermore, for each model, one depth assemblage was analysed 0–200 m combined.

The RIO is a relative index concerning the "occurrence" category and can assume values between low and high, usually between the range of the training data units, e.g., 0 and 1. RIO values for these analysed points were extracted to evaluate how the predictions correspond to the independent field data and for the respective depth

assemblage investigated. The accuracy of the selected models is given by the area under the curve of the Relative Operating Characteristic (ROC).

SPM also analysed the scores of the Variable Importance Plots, which represent the relative importance of variables (VIP). The importance of the predictor variables in the models are assessed by permuting the predictor variables individually in the test data. The reduction in prediction accuracy is then measured by comparing the models calculated using the permuted data with those obtained from the original data. If model accuracy decreases with the permuted variable, this indicates a strong association of that variable with response (Liaw and Wiener 2002).

Step 3: Correlations between the predicted (generalized) RIO values of copepods species and phytoplankton taxa for the entire study area were analysed. This was initially done in R (https://www.r-project.org/) with the "Performance Analytics" package (version 4.3.2) (https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html).

To obtain those large-scale community structure estimates reflecting the study area, we also described correlations in the predicted distribution values from of the Ross Sea lattice grid. We repeated this assessment with the "*varclust*" function in the HMISC library (version 4.3.2) (https://cran.r-project.org/web/packages/Hmisc/index.html). Subsequently, hierarchical clustering was performed by using the Pearson index (ρ^2) as a measure of correlation. Hierarchical clustering was achieved through the use of the "*chart.Correlation*" function of the PerformanceAnalytics library (version 4.3.2) (https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html). This clustering process was performed with the aim of analysing the RIO values and identifying the ranks formed.

Step 4: The RIO values for the Ensemble maps were obtained by averaging the respective index values (RIO) from the TN, RF and CA. It resulted into an average RIO, obtained by combining the best possible ML algorithms available. This lattice point grid was then scored in SPM using the pattern created from the points of the presence/absence matrix. Within this matrix, two values, i.e., "0" for species absence, or "1" for species presence, have been assigned based on a threshold, which represents the RIO that a given point contains as forecast of the RIO. The accuracy of the selected models is given by the area under the curve of the Relative Operating Characteristic (ROC), following the criteria of Swets (1988) and Pearce and Ferrie (2000).

Using the Inverse Distance Weighting (IDW) tool present in Quantartica, 48 predictive surfaces were generated from the scored lattice showing relative RIOs of

zooplankton and phytoplankton over the entire study area (beyond the lattice point location) in order to create a predictive surface grid. These are the first quantitative and repeatable estimates for the study area offering themselves for assessment and improvement over time.

In detail, six assemblages of phytoplankton and six copepods (Table 2) were chosen to obtain a more specific view about the plankton community. Four filter feeders (*Paralabidocera antarctica* (Thompson I.C., 1898), *Calanoides acutus* (Giesbrecht, 1902), *Metridia gerlachei* Giesbrecht, 1902, *Ctenocalanus citer* Heron & Bowman, 1971) (Hoshiai et al. 1987; Michels and Schnack-Schiel 2005), one ambush feeder *Oithona similis* Claus, 1866 (Kiørboe et al. 2009), and one predator (*Paraeuchaeta exigua* (Park 1994)) (Michels and Schnack-Schiel 2005) were analysed for copepods, while for phytoplankton we used Chlorophyta, Cryptophyta, Dinophyceae, Prymnesiophyceae, Bacillariophyceae and Cyanophyta. Appendix I includes ISO-compliant metadata, ensuring that the provided data adheres to globally recognized standards for consistency, interoperability, and usability across various research and data management systems.

Results

Training field data

First correlations between plankton were analysed using the field data. Figure 2 shows the graduated scale of correlation value that represents a proxy for biological interactions, with the intensity of colours in the correlation plot representing negative (blue, coexistence) and positive (red, biological interaction) interspecific interactions (Faust and Raes 2012). A total of 1830 values (without 1 values) were obtained, divided into: 1053 positive values (min = 0.01; max = 0.90), 762 negative values (min = −0.01; max = −0.83) and 15 zero values. Phytoplankton and microzooplankton show numerous negative values, particularly among zooplankton categories encompassing suspension feeders, filter feeders, and microplankton.

Primary consumers, such as phytoplankton, show negative values towards omnivorous, carnivorous zooplankton and microzooplankton community. Within the various phytoplankton assemblages, some positive values can be seen, such as those among Chlorophyta and Cryptophyta (0.90), while the most negative value are among Haptophyceae and for zooplankton is *Paralabidocera antarctica* (−0.75).

In the microzooplanktonic community, we observe positive values between the genera *Codonellopsis* sp. and *Laackmanniella* sp. (0.86), whereas the lowest correlation value was found *Laackmanniella* sp., *Dactyliosolen* sp., *Cymatocylis* sp. *Codonellopsis* sp. versus *Pseudo-nitzschia* spp. and *Thalassiosira* spp. (−0.83). Zooplankton show many cases of positive and negative correlations. In zooplankton, the highest values are found among the *Ctenocalanus citer* and Cyanophyta (0.76), while the lowest value is among the genus *Calanoides* sp. and Cyanophyta (−0.80). All correlation values and details can be found in Appendix C.

Predictions

We were able to compile a value-added data cube, explicitly in time and space, consisting of copepod species, phytoplankton assemblages and environmental descriptors to be used for model predictions of distributions for the wider Ross Sea wilderness area.

Most of the obtained ROCs (Table 3) showed an accuracy higher than 60%. This means that the analysed

Table 2 Taxa list

Planktonic group	Taxa	APHIA ID	Trophic group	Citation
Phytoplankton	Bacillariophycaea	148899	Mixotrophy	Villanova and Spetea (2021)
	Chlorophyta	802	Mixotrophy	Anderson et al. (2018)
	Cryptophyta	17639	Mixotrophy	McKie-Krisberg et al. (2015)
	Cyanophyta	146542	Autotrophy	Borghini et al. (2010), Van Goethem and Cowan (2019)
	Dinophyceae	19542	Mixotrophy	Zhang et al. (2013), Reguera et al. (2024)
	Prymnesiophyceae	115057	Mixotrophy	Moorthi et al. (2009), McKie-Krisberg et al. (2015), Anderson et al. (2018)
Zooplankton	*Calanoides acutus*	342434	Filter-, suspension feeder	Michels and Schnack-Schiel (2005)
	Ctenocalanus citer	342449	Filter-, suspension feeder	Michels and Schnack-Schiel (2005)
	Oithona similis	106656	Ambush feeder	Michels and Schnack-Schiel (2005)
	Paralabidocera antarctica	345234	Herbivory	Swadling and Gibson (2000)
	Metridia gerlachei	344689	Filter-, suspension feeder	Mariottini et al. (2000), Michels and Schnack-Schiel (2005)
	Paraeuchaeta exigua	345051	Predator	Michels and Schnack-Schiel (2005)

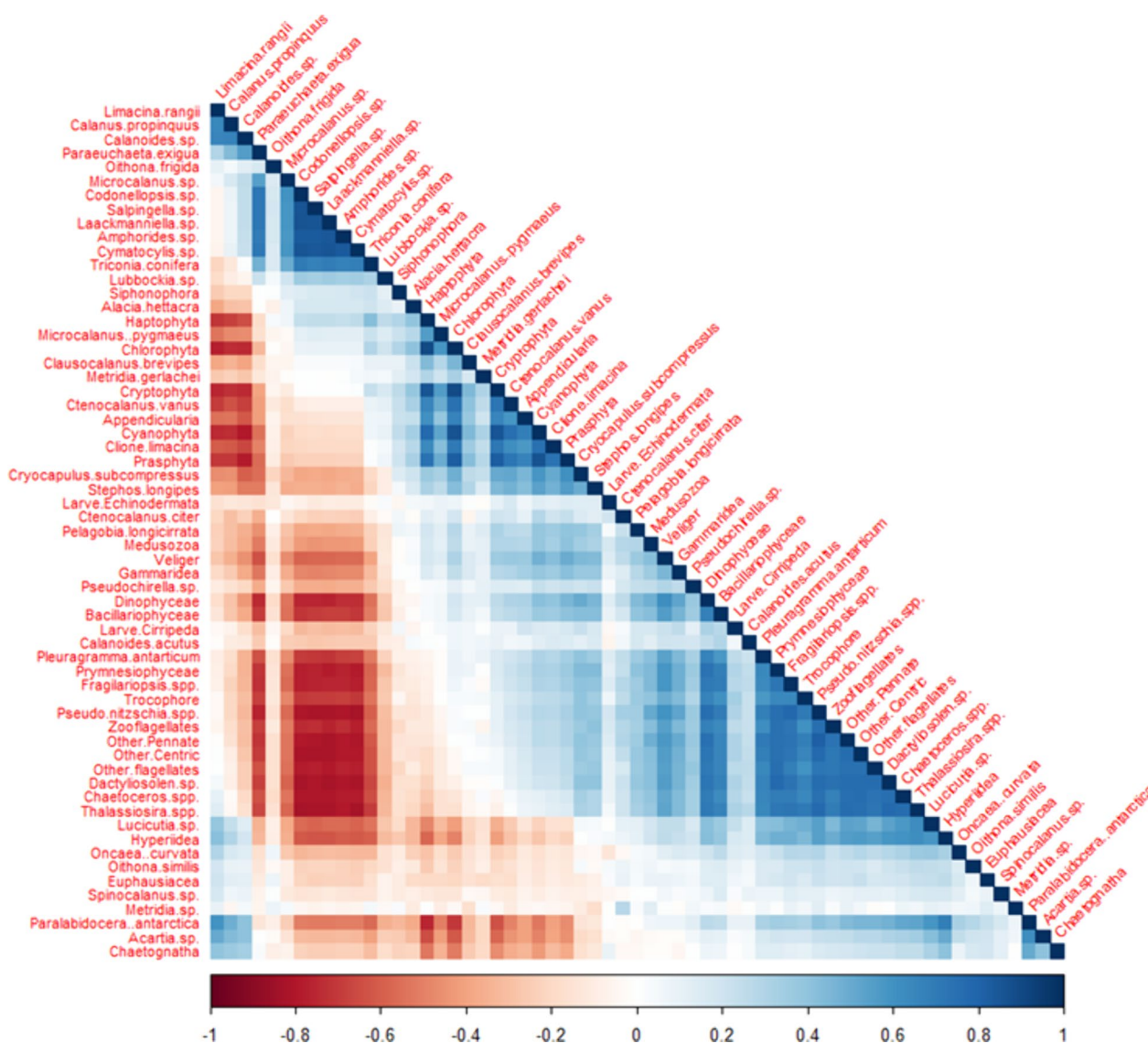AphiaID from the world register of marine species (WORMS, https://www.marinespecies.org)

Fig. 2 Correlation matrix in plankton community of Ross Sea sector estimated under co-occurrence model with marine environmental descriptors

models perform well with moderate accuracy, but can be improved also. In particular, the highest ROCs occurred in the copepods *Paralabidocera antarctica* (86%—CART; 82% TreeNet) and *Paraeuchaeta exigua* (81%—TreeNet), while, in the phytoplankton assemblages they were Cryptophyta (77%—CART) and Cyanophyta (76%—CART and 74%—TreeNet). Meanwhile, the lowest ROC values, in copepods, occur in *Calanoides acutus*, *Ctenocalanus citer*, *Oithona similis*, and *Metridia gerlachei* species (27%—RandomForest). In the phytoplankton assemblages the lowest value was obtained in the Chlorophyta (54%—RandomForest). The latter findings await more study. Further details in Appendix D.

Analysis of the importance of variables for each taxa and model, these scores represent the relative importance of the characteristics and thus help to rank the variables. The classifying of variable importance shows how significantly each variable contributes as a predictor variable in the models considered. Here we describe the variables influencing Bacillariophyceae and *P. antarctica* (Fig. 3). Detailed analyses for each of the other taxa considered can be found in Appendix E.

Concerning *Paralabidocera antarctica* in the CA model, significant variables include PRESSURE with the highest score of 100%, followed by SEAICE (69.80%), DECIMALLATITUDE (68.25%), and PHOSPHATE_50_M_ (55.98%). In the RF model, we

Table 3 Species list with sample size and ROC accuracy for both models

Taxa	CART ROC [%]	RANDOMFOREST ROC [%]	TREENET ROC [%]
Bacillariophycaea	62	62	63
Chlorophyta	47	54	63
Cryptophyta	70	63	71
Cyanophyta	72	63	74
Dinophyceae	62	62	66
Prymnesiophyceae	65	65	68
Calanoides acutus	49	27	45
Ctenocalanus citer	49	27	45
Oithona similis	49	27	45
Paralabidocera antarctica	83	79	82
Metridia gerlachei	49	27	45
Paraeuchaeta exigua	73	69	76

Red (very poor), orange (poor), yellow (good), and green (very good). ROC classes after Swets (1988) and Pearce and Ferrie (2000)

find a broader list of relevant variables, including DECI-MALLONGITUDE (13.58%), TEMPERATURE (12.41%), POTENTIAL (10.83%), and CONDUCTIVITY (10.19%). Lastly, in the TN model, there is a more extensive list of relevant variables, including CONDUCTIVITY (56.63%), SEAICE (51.83%), DECIMALLONGITUDE (100.00%), DECIMALLATITUDE (91.04%), and OXYGEN_MG_L (36.17%).

Concerning Bacillariophyceae, in both the CH and TN models, we find several significant environmental descriptors. For example, in CH, we have NOX (100.00%), POX (93.43%), NOX_200 (93.43%), NOX_SURFACE (93.43%), SILICATE_200 (93.43%), SILICATE_SURFACE (93.43%) and PHOSPHATE_SURFACE (92.90%). In TN, we find SEAICE (100.00%), DECIMALLONGITUDE (84.63%), SLOPE (76.01%), SALINITY (73.39%), FLUO-RESCENCE (72.96%), PRESSURE (70.68%) and OXYGEN_MG_L (69.19%).

Based on the predictions, generalization for the study area, Fig. 3 shows more types of trends between species of copepods and assemblages of phytoplankton, supporting that this organism prefers different Ross Sea zones. Based on the training data and ML with GIS layers, the use of an RIO index enabled us to generalize and categorize the different Ross Sea areas according to their occupation by copepod species.

In Appendix F (Fig. 1), the different correlation values between the RIOs of phytoplankton and copepods are showed. The assemblage of Primnesiophyceae show high values of correlation between Dinophyceae (0.99) and bacillariophyceae (0.91), followed by Dinophyceae showing high values of correlation with bacillariophyceae (0.91) and Chlorophyta (0.78). The lowest values are recorded in the Cyanophyta assemblage, with

Cryptophyta showing a value of −0.43, while bacillariophyceae show a slightly higher value of −0.41.

For copepods, the correlation values between the predicted distributions were between *O. similis*, *C. citer*, *M. gerlachei* and *C. acutus* (1.00). *P. antarctica* shows negative correlations with almost all species, except with *P. exigua*, which shows a value of 0.25. *P. exigua* correlates with a value of 0.01. Looking at the correlation values between the two planktonic groups we see different scenarios. The highest values occur between the Cyanophyta and *O. similis*, *C. citer*, *P. antarctica* and *C. acutus* (0.62), whereas the lowest value is among the same copepod species just mentioned with the assemblage Bacillariophyceae (−0.079).

The hierarchical clustering of the predicted distributions in the Ross Sea area for the 0–200 m depth class (Fig. 4) shows that the plankton groups have different correlation values. It can be noted that *Calanoides acutus*, *Metridia gerlachei*, *Oithona similis* and *Ctenocalanus citer* group together. Other groups with good correlation are Bacillariophyceae and *Paraeuchaeta exigua*, Dinophyceae and Prymnesiophyceae, Chlorophyta and Cryptophyta. This means they are found as an ecological community in the study areas, whereas *Paralabidocera antarctica* and Cyanophyta are less correlated.

Below are just two predictive maps with the corresponding RIO values for the Bacillariophyceae and the herbivorous copepod *Paralabidocera antarctica*. These two taxa were selected because they are fundamental to the Antarctic marine ecosystem and are considered key species. Consequently, the results of the models with high ROC performance are shown.

A total of 48 predictive distribution maps were generated for various species and phytoplankton assemblages,
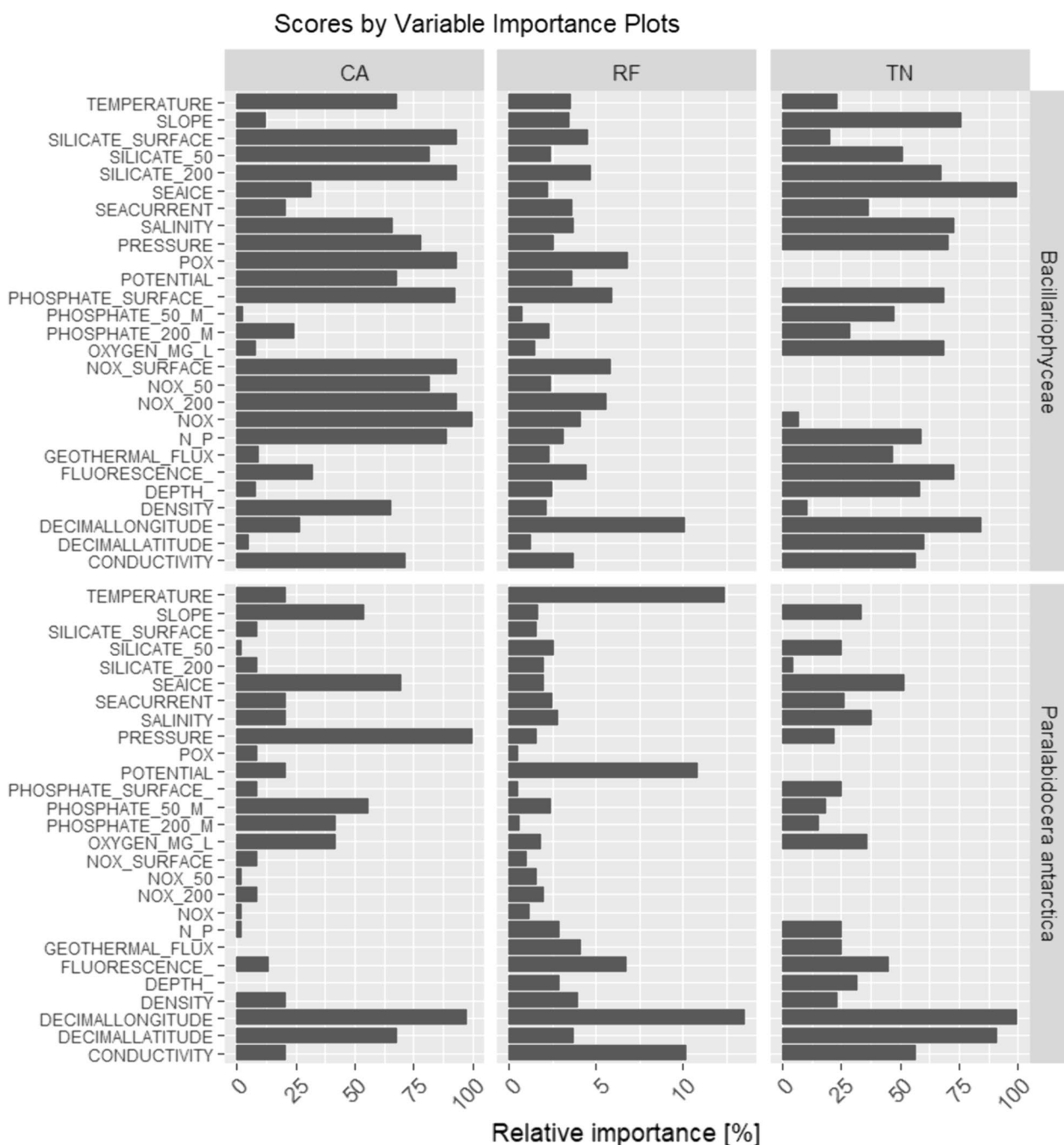
Grillo *et al. Ecological Processes* (2024) 13:56

Page 9 of 18

Fig. 3 Variable importance plots for predictive environmental descriptors from CA, RF and TN to predict the presence of Bacillariophyceae assemblages and *C. acutus*

with details on copepods provided in Appendix G. All RIO values are reported in Appendix H.

In Fig. 5, the presence of Bacillariophyceae is noteworthy that this significant assemblage of phytoplankton is observed in nearly all sampled stations. Concerning habitat suitability, the RIO index displays high values (0.6) across the entire study area. The lowest RIO values (0.49)

are identified in the north-eastern part of the sub-Antarctic belt and a few coastal areas. An additional noteworthy observation is found in marine protected areas, where Bacillariophyceae exhibit an increase in RIO values.

Other assemblages like Dinophyceae and Prymnesiophyceae (Appendix G: Figs. S17–S20 and S21–S24) present high values of presence in the whole area of the
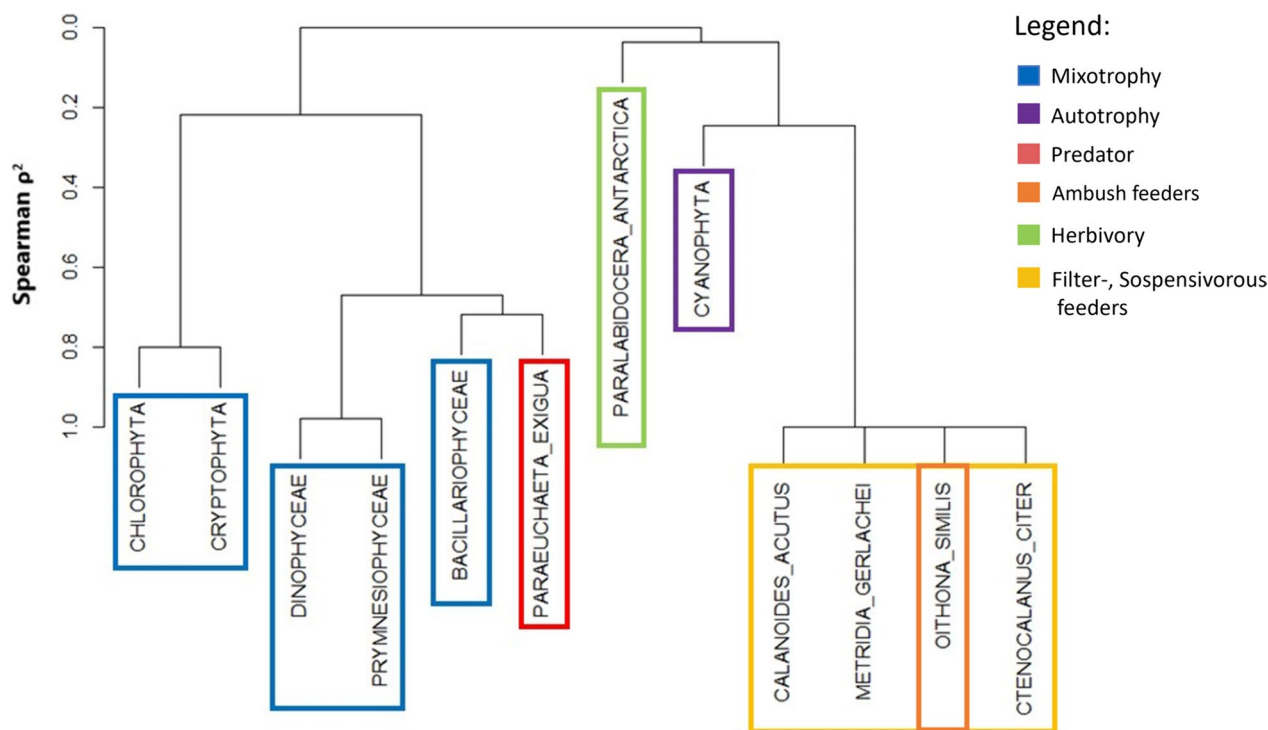
Fig. 4 Correlations of copepods and assemblages of phytoplankton based of RIO values (Algorithm Ensemble)

Ross Sea but show average low values (0.30) in the neritic zone. In Marine Protected Areas they show high presence values. Chlorophyta and Cryptophyta (Appendix G. Figs. S5–S8 and S9–S12) have average high RIO values (0.30) throughout the Ross Sea area. In general, Chlorophyta show high values in the sub-Antarctic pelagic zone, while Cryptophyta has high values in the sub-Antarctic belt (Appendix G. Figs. SS9–S12). Cyanophyta (Appendix G. Figs. S13 and S16) show high predicted values in the coastal and central areas of the Ross Sea, and low RIO values in the sub-Antarctic area.

Figure 6 shows the distribution of *Paralabidocera antarctica*, a herbivorous copepod. The original distribution is concentrated in four coastal stations. The predicted distribution shows the medium–high RIO values (0.37) throughout the Ross Sea area. The lowest values (0.33) can be found in the middle of the Ross Sea area. Furthermore, in all marine protected areas, the RIO index has medium values (0.35). *M. gerlachei, C. citer, O. similis and C. acutus* (Appendix G. Figs. S25, S29, S37, S41) show the medium–high RIO values (0.5–0.6) throughout the Ross Sea area. The lowest values (0.49) can be found, spot-like, in the northwest of the study area. Furthermore, in all marine protected areas, the RIO index has medium values (0.54). *Paraeuchaeta exigua* (Appendix G, Fig. S42)

has high RIO values over the entire Ross Sea area with low values in the marine protected area RS-GPZi.

The boxplots for RIOs (Fig. 7) showed that RA for the phytoplankton assemblages and TN for copepod performed better than the between models analysed.

In phytoplankton assemblages, the model with the lowest RIO values was CA (min=0, max=0.16, mean=0.04), followed by EN (min=0.12, max=0.38, mean=0.27) and TN (min=0.02, max=0.58, mean=0.27), while the RF model showed higher values than the other three models (min=0.32, max=0.64, mean=0.51).

Regarding copepods, the model with the lowest RIO values was once again CA (min=0, max=0.24, mean=0.19), followed by EN (min=0.12, max=0.50, mean=0.39) and RF (min=0.32, max=0.66, mean=0.52). However, the TN (min=0.04, max=0.71, mean=0.46) model showed higher values than the other three but also lower values, as predicted by the CA model.

Discussion
In Antarctic pelagic and coastal ecosystems, seasonal trends in species abundances and diversity are primarily influenced by ice cover duration, seawater temperature, dissolved nutrients, and atmospheric conditions (Pane et al. 2004; Ballard et al. 2012; Cecchetto et al. 2021;
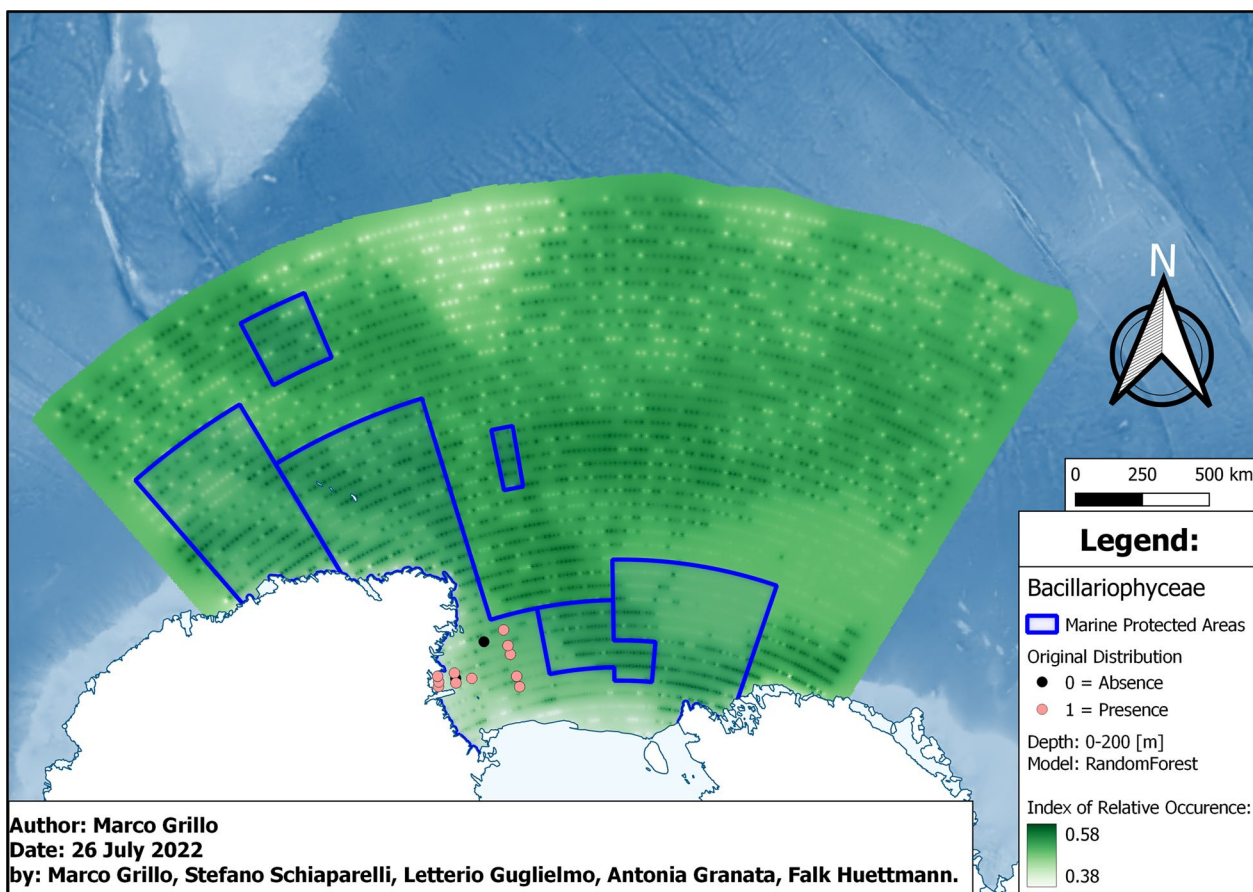
Fig. 5 Presence/absence points of raw survey location showing a predicted lattice grid summer distribution using the RandomForest algorithm for the depth class 0–200 m of the Bacillariophyceae. For details, see legend

Zwerschke et al. 2022). These factors serve as optimal environmental descriptors for species distribution modelling and therefore have been chosen as predictors.

Despite the availability of distributional data for the Antarctic plankton in the whole water column, modelling was so far restricted only to surface communities (e.g., Pinkerton et al. 2010b; Alvarez and Orgeira 2022; Lin et al. 2022) often sampled with the Continuous Plankton Recorder (CPR). This sampling method provided most of the quantitative data available and enabled covering large spatial portions of the Southern Ocean, unfortunately, the output is restricted to the surface layer of the water column.

To our knowledge, no prior attempts have really been made to model and generalize the three-dimensional distribution of plankton throughout the water column. In this study, we have leveraged phytoplankton and copepod distribution data down to a depth of 200 m, along with corresponding Open Access environmental data, to predict the distributions of these groups across the Ross Sea during the Austral summer months.

Before modelling, we have taken into account, here for the first time, the biological interactions that may exist within the Antarctic phyto and micro- and mesozooplankton communities together with the environmental descriptors of the surrounding environment, in order to model something closer to the "real world" data and not just an artificially sliced, predefined subset of species, interactions and variables.

Finally, we made predictions on likely distributions for selected key Antarctic taxa belonging to different trophic groups.

Correlation matrix in plankton community
The first step was to gain an understanding of the existing correlations among different planktonic taxa. Our analyses (Fig. 2) showed three types of interactions: positive (coexistence with other species and/or possible facilitating environmental conditions), neutral (i.e. unaffected by the presence of another species and environmental variables) and negative (i.e. antagonistic with other species and/or possible detrimental environmental conditions)
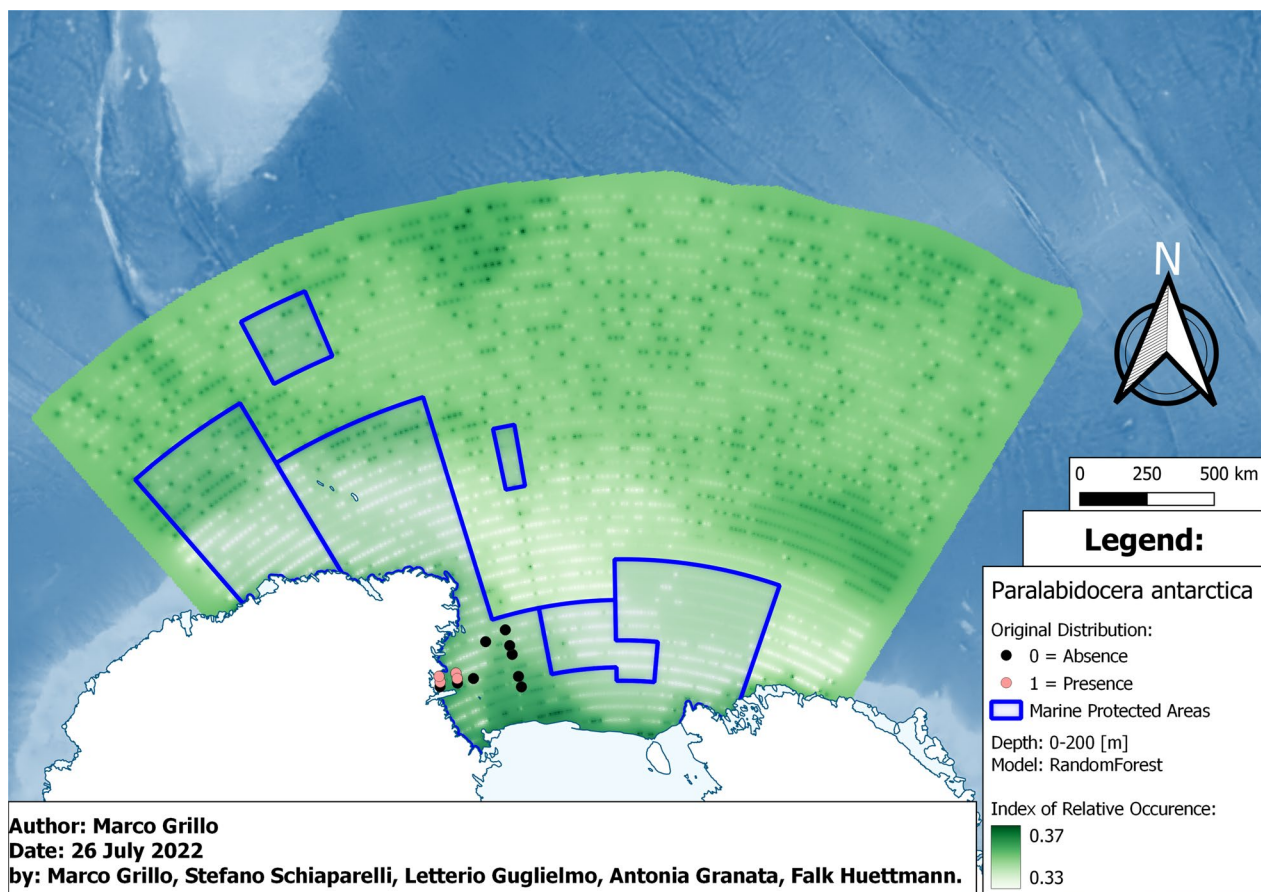
Fig. 6 Presence/absence points of raw survey location showing over a predicted lattice grid summer distribution using the CART algorithm for the depth class 0–200 m of the copepod *Paralabidocera antarctica*. For details, see legend

(Morales-Castilla et al. 2015). In our case, negative correlation values are observed for all zooplanktonic taxa which develop their trophic niche under similar environmental characteristics, regardless of their trophic ecology, thus encompassing predators, ambush feeders, suspension feeders, and filter feeders. Conversely, some taxa exhibit positive correlations, indicating the possibility of coexistence situations, as seen in the case of appendicularians interacting (i.e. coexisting) with other filter-feeding organisms. Appendicularians, functioning as specialized microfiltrators (Conley et al. 2018), "by-pass" food competition with other filter-feeding species thanks to their unique mucous feeding apparatus (Katija et al. 2017). Instead, in planktonic taxa occupying the lowest trophic levels, such as phytoplankton and microplankton, negative correlations emerge. This analysis offers a first but crucial glimpse of the potential biological correlations that might exist between planktonic taxa. Through these results, it is then possible to estimate how these taxa are potentially distributed, net of environmental descriptors and biological interactions, thus giving a glimpse of the realistic distributions of those species that sustain the ecological processes in a given sector.

VIPs and hierarchical clustering analysis

The analysis conducted by hierarchical clustering using Spearman's correlation coefficient (Fig. 4) provided important insights into the relationship between the different ecological niches occupied by the phytoplankton assemblages and copepod species analysed. Chlorophyta, Cryptophyta, Dinophyceae, Prymnesiophyceae, Bacillariophyceae and Cyanophyta were selected due to their contribution in abundance to the Antarctic phytoplankton communities (Biggs et al. 2019) and because they exert a significant role on food web dynamics, biogeochemical cycling and trophic carbon transfer in the marine environment (Finkel et al. 2010; Marañón 2015). The phytoplankton clustering showed the presence of four distinct groups.

In the first group, the co-presence of Chlorophyta and Cryptophyta, suggests significant similarities in the ecological niches occupied. These are mainly influenced by
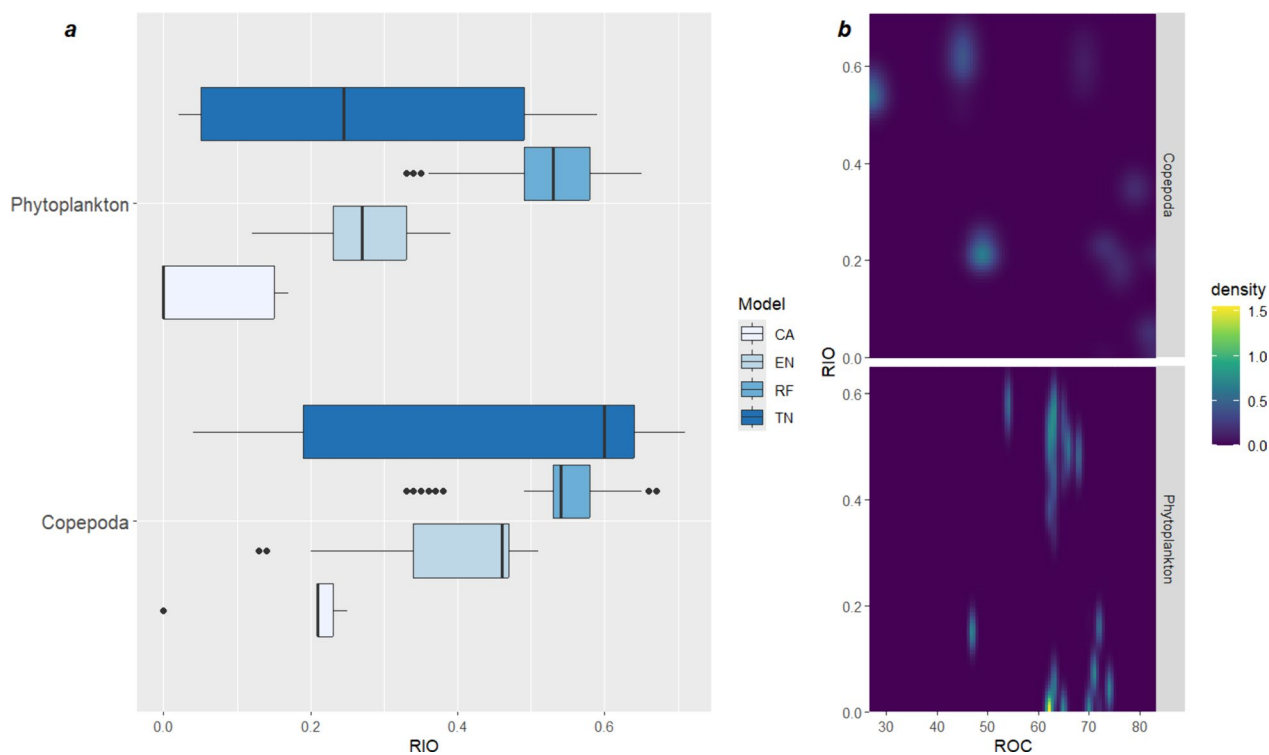
Fig. 7 Quality assessment of the models. **a** Boxplots of relative occurrence index (RIO) values for three models and the Ensemble model for Antarctic phytoplankton assemblages and the analysed copepod community; **b** evaluation of the quality of RIO values based on ROC

various oceanographic processes (see Fig. 3 and Appendix E), including the stabilisation of the upper mixed layer, sea ice melt and the development of frontal systems (Rodriguez et al. 2002). Similarly, the second group, comprising Dinophyta and Prymnesiophyceae, also shows similarities. Particlarly, these two assemblages are of particular ecological importance as the sea ice variations directly influences their cycles as the melting triggers their relative summer blooms (Selz et al. 2018; Anderson et al. 2018; Stoecker and Lavrentyev 2018). In addition, Prymnesiophyceae play a key role in the trophodynamics of the marine food web and the biogeochemical cycles of sulphur and carbon (Vancoppenolle et al. 2013). Despite the ecological niche of the Bacillariophyceae is alike with those of the Prymnesiophyceae and Dinophyceae, there are important distinctions, placing this assemblage in a third distinct group. In fact, Bacillariophyceae are also influenced by depth, light exposure and the availability of dissolved nutrients, while Prymnesiophyceae are influenced mostly by the upper mixed layer stabilization through ice (Rodriguez et al. 2002). It is relevant to emphasise that the Prymnesiophyceae and Bacillariophyceae show significant abundance peaks during the Austral summer in function of depth (Nuccio et al. 2000). Bacillariophyceae tend to be more coastal, while Prymnesiophyceae are more pelagic (DiTullio et al. 2000; Arrigo

et al. 2000). The Cyanophyta stands widely apart from the three other phytoplanktonic groups because of main differences in niche. This class generally inhabits shallow marine environments (Kosiba and Krztoń 2022) and their abundances depend on dissolved nutrients, water temperature and sea province (Karl and Bird 1993; Acosta Pomar et al. 2000).

A similar analysis can be applied to the copepod fauna (Fig. 4). The filter feeders (*C. acutus*, *M. gerlachei* and *C. citer*) (Michels and Schnack-Schiel 2005) and as well as the ambush feeders (*O. similis*) (Boxshall and Halsey 2004) are grouped into a single cluster. The species compete to feed on the same kind of prey and on suspended organic matter. They are co-existing in the same ranges of environmental variables analysed for their ecological niche (see Fig. 3 and Appendix E). *Paralabidocera antarctica*, instead, stands apart from the other trophic groups because, in addition to having an epipelagic ecological niche, it feeds mainly on sympagyc algae (Hoshiai et al. 1987; Swadling and Gibson 2000). These trophic-ecological characteristics lead to separate this species from other omnivorous copepod species. *P. exigua* is in a standing-alone group because of its trophic guild. It belongs to the group of carnivorous predators (Henschke et al. 2015). It is a eurybathic species, but other factors such as temperature, salinity and oxygen availability may

also influence its distribution and adaptation in the environment (Hagen et al. 1993; Carli et al. 2000; Hunt and Swadling 2021). In general, the genus *Paraeuchaeta* is an important link in the Antarctic food chain because, due to the large size of the individuals, it is also preyed upon by secondary consumers such as birds (Bocher 2002).

Species distribution models (SDMs) with ML
All the above-described situations are captured by the distribution maps. In general, planktonic organisms show different distributions according to their trophic niche (Chase and Leibold 2009). In our case, we analysed a "hypervolume with n-dimensions" where environmental descriptors determine presence-absence of an organism in that particular habitat (Hutchinson 1957). In this way, it is possible to quantitatively frame the trophic niches occupied through computational analyses (Elith et al. 2006; Drew et al. 2010). This has been previously done for some common copepod species (i.e. *Calanoides acutus, Ctenocalanus citer, Metridia gerlachei, Oithona similis, Paraeuchaeta exigua* and *Paralabidocera antartica*) by Grillo et al. (2022) but in this paper we bring the analysis one step further by analysing the phytoplankton assemblages (Chlorophyta, Cryptophyta, Dinophyceae, Prymnesiophyceae, Bacillariophyceae and Cyanophyta) as well. Furthermore, the sampling was undertaken in more recent expeditions (2016–2017), while in Grillo et al. (2022) data were from the austral summers 1987–1988, 1989–1990 and 1994–1995.

Our results indicate that also in the case of phytoplankton groups, it is possible to find distinct distributions that are picked up by the SDMs. Bacillariophyceae (diatoms) are found widely throughout the Ross Sea (Fig. 5), with significantly high RIO values in inshore areas and areas of maximum glacial cover. The Prymnesiophyceae are also found throughout the Ross Sea, however with more pronounced RIO values in the offshore areas. These observations confirm the ecology of the two phytoplanktonic assemblages, the documentation and description of which are already well-established (Nuccio et al. 2000).

When comparing the distribution of *M. gerlachei*, one of the most abundant calanoid copepod in the pelagic ecosystems of the Ross Sea (Voronina et al. 2001), with the modelled distribution of the same copepod in Grillo et al. (2022), variations in the predicted distribution emerge. During the 1980s and 1990s, *M. gerlachei* showed a widespread presence at all coastal stations sampled. As shown in Figures S54–S57 in Supplementary File S1 in Grillo et al. (2022), *M. gerlachei* was widespread in the Ross Sea area, in marine protected areas and along the entire water column analysed (0–750 m). This species showed high levels of occurrence in the neritic provinces, coastal habitats, sub-Antarctic areas and the Pennel Bank

area, with low values recorded in the marine area opposite Victoria Land. In 2016–2017, the distribution of this species remained about the same, but the values of its occurrence, including RSRMPAs, decreased by around 20%. This remarks the importance of establishing baselines for MPAs to compare with future situations.

Effectiveness of the models
Among the models analysed, the RF algorithm produced a very high range of RIO values for both phytoplankton assemblages and copepod community (Fig. 7). Specifically, for copepods, the TN model generated more extreme RIO values, both higher and lower. This confirms that the Random Forest model produces maps that are closer to reality (Mi et al. 2017) compared to the other algorithms used in this study. The RIO values obtained from the CA, in addition to being very low in general, do not seem to accurately represent the actual distribution of the studied species. Ensemble models seem to buffer those issues best (e.g., Hardy et al. 2011).

Our results, based on ROC analysis and boxplot graphs for RIO, suggest that there are differences in the generalization performance of the various modelling techniques (Fig. 7). Generally, the ROC values that generated high RIO values depend on the community analysed. For phytoplankton assemblages, medium–high ROC values predict high RIO values, while for the copepod community, medium–low ROC values predict high RIO values. Ideally, additional evidences for model predictions are to be used beyond ROC-types.

Future perspectives and improvements
The current analyses are constrained by a limited number of sampling points, which were then extrapolated to a larger spatial scale. The authors recognize the potential shortcomings of this design, and it is evident that further refinements are necessary. This could involve incorporating observations from additional years and, if feasible, expanding the sampling stations to enhance the robustness of the inferences drawn. Additionally, utilizing data from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/) and the Ocean Biogeographic Information System (OBIS, https://obis.org/) could significantly enrich the dataset, providing a more comprehensive understanding and enabling more accurate conclusions. Here we provide first baseline information for wider assessment and improvement.

Nevertheless, the findings remain highly relevant as they provide a first quantitative foundation for subsequent analyses of real-world planktonic communities. Machine Learning emerges as a powerful tool for

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 15 of 18

analysing such datasets characterized by complex non-linear patterns and a high volume of data.

Through this methodology, it becomes feasible to quantitatively estimate the potential 3D distribution of planktonic organisms (Drew et al. 2010; Steiner et al. 2023; Huettmann 2024; Guzzi et al. 2024). Given their pivotal ecological role, supporting the entire marine food web and ecological processes within the Ross Sea, these organisms are of paramount importance (Turner 2004). They serve as the link between dissolved organic and inorganic substances and primary and secondary consumers.

Another notable discovery is that even within the boundaries of the Ross Sea Region Marine Protected Area (RSRMPA), hotspots for phytoplankton species are already present. This underscores the global significance of the Ross Sea and its RSRMPA in contributing to global processes.

Conclusion

The findings of this study emphasize the pivotal role that key copepod species and phytoplankton assemblages play within the Ross Sea and the Ross Sea Region Marine Protected Area (RSRMPA) in maintaining the health of Antarctic marine ecosystems. The demonstrated effectiveness of marine protected areas in preserving primary trophic levels and associated predator populations underscores the global ecological significance of the Antarctic region. This research highlights the need for science-based management practices and provides robust Machine Learning and Open Access based quantitative models that can guide policymakers and conservation biologists. To ensure the continued effectiveness of conservation efforts and to mitigate the impacts of climate change on polar regions, ongoing monitoring and such dedicated research are imperative.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13717-024-00532-6.

> Supplementary Material 1. Appendix A. Absence and presence matrix with related environmental descriptors. Appendix B. Lattice grid of Ross Sea with environmental descriptors with rasters utilized. Appendix C. Correlation matrix values filed data and figure. Appendix D. Details models. Appendix E. VIP analysis values. Appendix F. Figure correlation plot with RIO values. Appendix G. Predictive maps for each copepod species and for each phytoplankton assemblages analysed. Appendix H. Predictor layers for each copepod and phytoplankton assemblage with the RIO-prediction for TreeNet, RandomForest, CART Decision Tree and Ensemble models. Appendix I. ISO-compliant metadata.

Author contributions

Conceptualization, M.G., F.H., and S.S.; methodology, M.G. and F.H.; software, F.H.; formal analysis, M.G. and F.H.; resources, M.G.; data acquisition L.G., A.G.; data curation, M.G., T.D. and S.S.; writing—original draft preparation, M.G., F.H., and S.S.; writing—review and editing, M.G., S.S., D.T., L.G., A.G., and F.H.; funding acquisition, F.H. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

The data and materials used in this work are available at the following link https://zenodo.org/records/11175588.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors approved the manuscript for publication in ecological processes.

Competing interests
No competing interest exists in this manuscript.

Author details
[1]Department of Physical Sciences, Earth and Environment (DSFTA), University of Siena, Siena, Italy. [2]Italian National Antarctic Museum (MNA, Section of Genoa), University of Genoa, Genoa, Italy. [3]Department of Earth, Environmental and Life Sciences (DISTAV), University of Genoa, Genoa, Italy. [4]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway. [5]Integrative Marine Ecology Department, Stazione Zoologica Anton Dohrn, Naples, Italy. [6]Institute of Polar Sciences, National Research Council (CNR), Messina, Italy. [7]Department of Biological, Chemical, Pharmaceutical, and Environmental Sciences (ChiBioFarAm), University of Messina, Messina, Italy. [8]-EWHALE Lab-, Biology & Wildlife Department, Institute of Arctic Biology, University of Alaska Fairbanks, Alaska, USA.

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 16 of 18

References

Acosta Pomar MLC, Maugeri TL, Bruni V (2000) Picoplankton abundance and biomass at Terra Nova Bay (Ross Sea, Antarctica) during the 1989–1990 austral summer. In: Faranda FM, Guglielmo L, Ianora A (eds) Ross Sea ecology. Springer, Berlin, pp 195–203

Alexander V, Niebauer HJ (1981) Oceanography of the eastern Bering Sea ice-edge zone in spring. Limnol Oceanogr 26:1111–1125

Alvarez F, Orgeira JL (2022) Krill finder: spatial distribution of sympatric fin (*Balaenoptera physalus*) and humpback (*Megaptera novaeangliae*) whales in the Southern Ocean. Polar Biol 45:1427–1440. https://doi.org/10.1007/s00300-022-03080-x

Anderson R, Charvet S, Hansen PJ (2018) Mixotrophy in chlorophytes and haptophytes—effect of irradiance, macronutrient, micronutrient and vitamin limitation. Front Microbiol 9:1704. https://doi.org/10.3389/fmicb.2018.01704

Arrigo KR, DiTullio GR, Dunbar RB et al (2000) Phytoplankton taxonomic variability in nutrient utilization and primary production in the Ross Sea. J Geophys Res Oceans 105:8827–8846. https://doi.org/10.1029/1998JC000289

Arrigo KR, van Dijken GL, Ainley DG, Fahnestock MA, Markus T (2002) Ecological impact of a large Antarctic iceberg. Geophys Res Lett 29:1104. https://doi.org/10.1029/2001GL014160

Arrigo KR, van Dijken GL, Bushinsky S (2008) Primary production in the Southern Ocean, 1997–2006. J Geophys Res Oceans 113:C08004. https://doi.org/10.1029/2007JC004578

Atkinson A (1998) Life cycle strategies of epipelagic copepods in the Southern Ocean. J Mar Syst 15:289–311

Atkinson A, Siegel V, Pakhomov E, Rothery P (2004) Long-term decline in krill stock and increase in salps within the Southern Ocean. Nature 432:100–103. https://doi.org/10.1038/nature02996

Ballard G, Jongsomjit D, Veloz SD, Ainley DG (2012) Coexistence of mesopredators in an intact polar ocean ecosystem: the basis for defining a Ross Sea marine protected area. Biol Conserv 156:72–82. https://doi.org/10.1016/j.biocon.2011.11.017

Biggs TEG, Alvarez-Fernandez S, Evans C et al (2019) Antarctic phytoplankton community composition and size structure: importance of ice type and temperature as regulatory factors. Polar Biol 42:1997–2015. https://doi.org/10.1007/s00300-019-02576-3

Bocher P (2002) Importance of the large copepod *Paraeuchaeta antarctica* (Giesbrecht, 1902) in coastal waters and the diet of seabirds at Kerguelen, Southern Ocean. J Plankton Res 24:1317–1333. https://doi.org/10.1093/plankt/24.12.1317

Bolinesi F, Saggiomo M, Ardini F et al (2020) Spatial-related community structure and dynamics in phytoplankton of the Ross Sea, Antarctica. Front Mar Sci 7:574963. https://doi.org/10.3389/fmars.2020.574963

Borghini F, Colacevich A, Bargagli R (2010) A study of autotrophic communities in two Victoria Land lakes (Continental Antarctica) using photosynthetic pigments. J Limnol 69:333–340. https://doi.org/10.3274/JL10-69-2-14

Boxshall GA, Halsey SH (2004) An introduction to copepod diversity. Ray Society, Andover

Boyd PW, Doney SC, Strzepek R et al (2008) Climate-mediated changes to mixed-layer properties in the Southern Ocean: assessing the phytoplankton response. Biogeosciences 5:847–864. https://doi.org/10.5194/bg-5-847-2008

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Brooks CM, Crowder LB, Österblom H, Strong AL (2020) Reaching consensus for conserving the global commons: the case of the Ross Sea, Antarctica. Conserv Lett 13:e12676. https://doi.org/10.1111/conl.12676

Carli A, Pane L, Stocchino C (2000) Planktonic copepods in Terra Nova Bay (Ross Sea): distribution and relationship with environmental factors. Ross Sea ecology. Springer, Berlin, pp 309–321

Cau A, Ennas C, Moccia D et al (2021) Particulate organic matter release below melting sea ice (Terra Nova Bay, Ross Sea, Antarctica): possible relationships with zooplankton. J Mar Syst 217:103510. https://doi.org/10.1016/j.jmarsys.2021.103510

Cecchetto M, Di Cesare A, Eckert E et al (2021) Antarctic coastal nanoplankton dynamics revealed by metabarcoding of desalination plant filters: detection of short-term events and implications for routine monitoring. Sci Total Environ 757:143809. https://doi.org/10.1016/j.scitotenv.2020.143809

Chase JM, Leibold MA (2009) Ecological niches: linking classical and contemporary approaches. University of Chicago Press, Chicago

Chown SL, Lee JE, Hughes KA et al (2012) Challenges to the future conservation of the Antarctic. Science 337:158–159. https://doi.org/10.1126/science.122282

Clarke A (1988) Seasonality in the Antarctic marine environment. Comp Biochem Physiol Part B Comp Biochem 90:461–473. https://doi.org/10.1016/0305-0491(88)90285-4

Clarke A, Harris CM (2003) Polar marine ecosystems: major threats and future change. Environ Conserv 30:1–25. https://doi.org/10.1017/S0376892903000018

Cole DN, Landres PB (1996) Threats to wilderness ecosystems: impacts and research needs. Ecol Appl 6:168–184. https://doi.org/10.2307/2269562

Colwell RK, Rangel TF (2009) Hutchinson's duality: the once and future niche. Proc Natl Acad Sci USA 106:19651–19658. https://doi.org/10.1073/pnas.0901650106

Conley KR, Lombard F, Sutherland KR (2018) Mammoth grazers on the ocean's minuteness: a review of selective feeding using mucous meshes. Proc R Soc B 285:20180056. https://doi.org/10.1098/rspb.2018.0056

Constable AJ, Melbourne-Thomas J, Muelbert MMC et al (2023) Marine ecosystem assessment for the Southern Ocean: summary for policymakers

Cordone A, D'Errico G, Magliulo M et al (2022) Bacterioplankton diversity and distribution in relation to phytoplankton community structure in the Ross Sea surface waters. Front Microbiol 13:722900. https://doi.org/10.3389/fmicb.2022.722900

Dayton PK, Watson D, Palmisano A et al (1986) Distribution patterns of benthic microalgal standing stock at McMurdo Sound, Antarctica. Polar Biol 6:207–213. https://doi.org/10.1007/BF00443397

Di Marco M, Ferrier S, Harwood TD et al (2019) Wilderness areas halve the extinction risk of terrestrial biodiversity. Nature 573:582–585. https://doi.org/10.1038/s41586-019-1567-7

DiTullio G, Grebmeier J, Arrigo K et al (2000) Rapid and early export of *Phaeocystis antarctica* blooms in the Ross Sea, Antarctica. Nature 404:595–598. https://doi.org/10.1038/35007061

Drew CA, Wiersma YF, Huettmann F (2010) Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer Science & Business Media, New York

El-Gabbas A, Van Opzeeland I, Burkhardt E, Boebel O (2021) Dynamic species distribution models in the marine realm: predicting year-round habitat suitability of baleen whales in the Southern Ocean. Front Mar Sci 8:802276. https://doi.org/10.3389/fmars.2021.802276

Elith J, Graham CH, Anderson RP et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550. https://doi.org/10.1038/nrmicro2832

Finkel ZV, Beardall J, Flynn KJ et al (2010) Phytoplankton in a changing world: cell size and elemental stoichiometry. J Plankton Res 32:119–137. https://doi.org/10.1093/plankt/fbp098

Fraser AD, Wongpan P, Langhorne PJ et al (2023) Antarctic Landfast Sea ice: a review of its physics, biogeochemistry and ecology. Rev Geophys 61:e2022RG000770. https://doi.org/10.1029/2022RG000770

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38:367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Garrison DL, Sullivan CW, Ackley SF (1986) Sea ice microbial communities in Antarctica. Bioscience 36:243–250. https://doi.org/10.2307/1310214

Gell FR, Roberts CM (2003) Benefits beyond boundaries: the fishery effects of marine reserves. Trends Ecol Evol 18:448–455. https://doi.org/10.1016/S0169-5347(03)00189-7

González-Herrero S, Navarro F, Pertierra LR et al (2023) Southward migration of the zero-degree isotherm latitude over the Southern Ocean and the Antarctic Peninsula: cryospheric, biotic and societal implications. Sci Total Environ 912:168473. https://doi.org/10.1016/j.scitotenv.2023.168473

Granata A, Weldrick CK, Bergamasco A et al (2022) Diversity in zooplankton and sympagic biota during a period of rapid sea ice change in Terra Nova Bay, Ross Sea, Antarctica. Diversity 14:425. https://doi.org/10.3390/d14060425

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 17 of 18

Grillo M, Huettmann F, Guglielmo L, Schiaparelli S (2022) Three-dimensional quantification of copepods predictive distributions in the Ross Sea: first data based on a machine learning model approach and open access (FAIR) data. Diversity 14:355. https://doi.org/10.3390/d14050355

Guglielmo L, Carrada GC, Catalano G et al (2000) Structural and functional properties of sympagic communities in the annual sea ice at Terra Nova Bay (Ross Sea, Antarctica). Polar Biol 23:137–146. https://doi.org/10.1007/s003000050019

Guillaumot C, Martin A, Fabri-Ruiz S et al (2016) Echinoids of the Kerguelen Plateau—occurrence data and environmental setting for past, present, and future species distribution modelling. ZooKeys 630:1–17. https://doi.org/10.3897/zookeys.630.9856

Guthery FS (2008) Statistical ritual versus knowledge accrual in wildlife science. J Wildl Manag 72:1872–1875

Guzzi A, Schiaparelli S, Balan M, Grillo M (2024) A beacon in the dark: grey literature data mining and machine learning enlightening historical plankton seasonality dynamics in the Ligurian Sea. Diversity 16:189. https://doi.org/10.3390/d16030189

Hagen W, Kattner G, Graeve M (1993) *Calanoides acutus* and *Calanus propinquus*, Antarctic copepods with different lipid storage modes via wax esters or triacylglycerols. Mar Ecol Prog Ser 97:135–142

Hardy SM, Lindgren M, Konakanchi H, Huettmann F (2011) Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaskan waters: a first open-access ensemble model. Integr Comp Biol 51(4):608–622

Hegel TM, Cushman SA, Evans J, Huettmann F (2010) Current state of the art for statistical modelling of species distributions. In: Cushman SA, Huettmann F (eds) Spatial complexity, informatics, and wildlife conservation. Springer, Tokyo, pp 273–311

Henschke N, Everett JD, Suthers IM et al (2015) Zooplankton trophic niches respond to different water types of the western Tasman Sea: a stable isotope analysis. Deep Sea Res Part I 104:1–8. https://doi.org/10.1016/j.dsr.2015.06.010

Hixon MA, Johnson DW, Sogard SM (2014) BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. ICES J Mar Sci 71:2171–2185. https://doi.org/10.1093/icesjms/fst200

Hoshiai T, Tanimura A, Watanabe K (1987) Ice algae as food of an Antarctic ice-associated copepod, *Paralabidocera antarctica* (IC Thompson). In: Proc NIPR Symp Polar Biol. Citeseer, p lll

Huettmann F (2024) A super SDM (species distribution model) 'in the cloud' for better habitat-association inference with a 'big data' application of the Great Gray Owl for Alaska. Sci Rep 14(1):7213

Huettmann F, Schmid M (2014) 9.1. Climate change and predictions of pelagic biodiversity components. Biogeographic Atlas of the Southern Ocean Scientific Committee on Antarctic Research, Cambridge, pp 470–475

Huettmann F, Kövér L, Robold R et al (2023) Model-based prediction of a vacant summer niche in a subarctic urbanscape: a multi-year open access data analysis of a 'niche swap' by short-billed Gulls. Ecol Inform 78:102364. https://doi.org/10.1016/j.ecoinf.2023.102364

Humphries GR, Magness DR, Huettmann F (2019) Machine learning for ecology and sustainable natural resource management. Springer, Cham

Hunt BPV, Swadling KM (2021) Macrozooplankton and micronekton community structure and diel vertical migration in the Heard Island Region, Central Kerguelen Plateau. J Mar Syst 221:103575. https://doi.org/10.1016/j.jmarsys.2021.103575

Hutchinson GE (1957) Concluding remarks. Cold Spring Harb Symp Quant Biol 22:415–427

Jiao S, Huettmann F, Guo Y et al (2016) Advanced long-term bird banding and climate data mining in spring confirm passerine population declines for the Northeast Chinese-Russian flyway. Global Planet Change 144:17–33. https://doi.org/10.1016/j.gloplacha.2016.06.015

Karl DM, Bird DF (1993) Bacterial-agal interactions in Antarctic coastal ecosystems. In: Trends in microbial ecology, sixth international symposium on microbial ecology. pp 37–40

Katija K, Sherlock RE, Sherman AD, Robison BH (2017) New technology reveals the role of giant larvaceans in oceanic carbon cycling. Sci Adv 3:e1602374. https://doi.org/10.1126/sciadv.1602374

Kiørboe T, Andersen A, Langlois VJ et al (2009) Mechanisms and feasibility of prey capture in ambush-feeding zooplankton. Proc Natl Acad Sci USA 106:12394–12399. https://doi.org/10.1073/pnas.0903350106

Kosiba J, Krztoń W (2022) Insight into the role of cyanobacterial bloom in the trophic link between ciliates and predatory copepods. Hydrobiologia 849:1195–1206. https://doi.org/10.1007/s10750-021-04780-x

Koubbi P, Broyer CD, Griffiths H et al (2014) Conclusions: present and future of Southern Ocean biogeography. In: Biogeographic Atlas of the Southern Ocean. pp 469–476

Leihy RI, Coetzee BWT, Morgan F et al (2020) Antarctica's wilderness fails to capture continent's biodiversity. Nature 583:567–571. https://doi.org/10.1038/s41586-020-2506-3

Lele SR, Dennis B, Lutscher F (2007) Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol Lett 10:551–563. https://doi.org/10.1111/j.1461-0248.2007.01047.x

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18–22

Lin S, Zhao L, Feng J (2022) Predicted changes in the distribution of Antarctic krill in the Cosmonaut Sea under future climate change scenarios. Ecol Ind 142:109234. https://doi.org/10.1016/j.ecolind.2022.109234

Marañón E (2015) Cell size as a key determinant of phytoplankton metabolism and community structure. Annu Rev Mar Sci 7:241–264. https://doi.org/10.1146/annurev-marine-010814-015955

Mariottini GL, Feletti M, Romano P et al (2000) An ultrastructural study of the Antarctic calanoid copepod *Metridia gerlachei* Giesbrecht, 1902. J Biol Res 76:73–80. https://doi.org/10.4081/jbr.2000.10799

Matsuoka K, Skoglund A, Roth G et al (2021) Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands. Environ Model Softw 140:105015. https://doi.org/10.1016/j.envsoft.2021.105015

McKie-Krisberg ZM, Gast RJ, Sanders RW (2015) Physiological responses of three species of Antarctic mixotrophic phytoflagellates to changes in light and dissolved nutrients. Microb Ecol 70:21–29. https://doi.org/10.1007/s00248-014-0543-x

Meiners KM, Vancoppenolle M, Thanassekos S et al (2012) Chlorophyll a in Antarctic sea ice from historical ice core data. Geophys Res Lett 39:L21602. https://doi.org/10.1029/2012GL053478

Meißner K, Fiorentino D, Schnurr S et al (2013) Distribution of benthic marine invertebrates at northern latitudes—an evaluation applying multi-algorithm species distribution models. J Sea Res 85:241–254

Melchiori V (2017) Rapporto sulla Campagna Antartica Estate Australe 2016–2017. ENEA—Programma Nazionale di Ricerca in Antartide (PNRA)

Mi C, Huettmann F, Guo Y et al (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ 5:e2849. https://doi.org/10.7717/peerj.2849

Michels J, Schnack-Schiel SB (2005) Feeding in dominant Antarctic copepods—does the morphology of the mandibular gnathobases relate to diet? Mar Biol 146:483–495. https://doi.org/10.1007/s00227-004-1452-1

Mittermeier RA, Mittermeier CG, Brooks TM et al (2003) Wilderness and biodiversity conservation. Proc Natl Acad Sci USA 100:10309–10313

Monti-Birkenmeier M, Diociaiuti T, Castagno P et al (2022) Pluridecadal temporal patterns of Tintinnids (Ciliophora, Spirotrichea) in Terra Nova Bay (Ross Sea, Antarctica). Diversity 14:604. https://doi.org/10.3390/d14080604

Moorthi S, Caron D, Gast R, Sanders R (2009) Mixotrophy: a widespread and important ecological strategy for planktonic and sea-ice nanoflagellates in the Ross Sea, Antarctica. Aquat Microb Ecol 54:269–277. https://doi.org/10.3354/ame01276

Morales-Castilla I, Matias MG, Gravel D, Araújo MB (2015) Inferring biotic interactions from proxies. Trends Ecol Evol 30:347–356. https://doi.org/10.1016/j.tree.2015.03.014

Norkko A, Thrush SF, Cummings VJ et al (2007) Trophic structure of coastal Antarctic food webs associated with changes in sea ice and food supply. Ecology 88:2810–2820

Nuccio C, Innamorati M, Lazzara L et al (2000) Spatial and temporal distribution of phytoplankton assemblages in the Ross Sea. In: Faranda FM, Guglielmo L, Ianora A (eds) Ross Sea ecology. Springer, Berlin, pp 231–245

O'Driscoll RL, Macaulay GJ, Gauthier S et al (2011) Distribution, abundance and acoustic properties of Antarctic silverfish (*Pleuragramma antarcticum*) in the Ross Sea. Deep Sea Res Part II 58:181–195

Grillo *et al. Ecological Processes*   (2024) 13:56

Page 18 of 18

Pane L, Feletti M, Francomacaro B, Mariottini GL (2004) Summer coastal zooplankton biomass and copepod community structure near the Italian Terra Nova Base (Terra Nova Bay, Ross Sea, Antarctica). J Plankton Res 26:1479–1488. https://doi.org/10.1093/plankt/fbh135

Park T (1994) Geographic distribution of the bathypelagic genus Paraeuchaeta (Copepoda, Calanoida). Hydrobiologia 292:317–332

Parker SJ, Sundby S, Stevens D et al (2021) Buoyancy of post-fertilised *Dissostichus mawsoni* eggs and implications for early life history. Fish Oceanogr 30:697–706

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133:225–245

Peck LS (2018) Antarctic marine biodiversity: adaptations, environments and responses to change. Oceanogr Mar Biol 56:105–236

Peel SL, Hill NA, Foster SD et al (2019) Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types. Methods Ecol Evol 10:1002–1014. https://doi.org/10.1111/2041-210X.13196

Pinkerton MH, Bradford-Grieve JM, Hanchet SM (2010a) A balanced model of the food web of the Ross Sea, Antarctica. CCAMLR Sci 17:1–31

Pinkerton MH, Smith AN, Raymond B et al (2010b) Spatial and seasonal distribution of adult *Oithona similis* in the Southern Ocean: predictions using boosted regression trees. Deep Sea Res Part I 57:469–485. https://doi.org/10.1016/j.dsr.2009.12.010

Reguera B, García-Portela M, Velasco-Senovilla E et al (2024) Dinophysis, a highly specialized mixoplanktonic protist. Front Protistol 1:1328026. https://doi.org/10.3389/frpro.2023.1328026

Rodriguez F, Varela M, Zapata M (2002) Phytoplankton assemblages in the Gerlache and Bransfield Straits (Antarctic Peninsula) determined by light microscopy and CHEMTAX analysis of HPLC pigment data. Deep Sea Res Part II Top Stud Oceanogr 49:723–747. https://doi.org/10.1016/S0967-0645(01)00121-7

Schiaparelli S, Alvaro MC, Cecchetto M et al (2019) Dalla rilevanza nazionale a quella internazionale: le strategie adottate dal Museo Nazionale dell'Antartide (MNA, Sede di Genova). Museologia Scientifica Memorie 19:40–44

Schnack-Schiel SB, Hagen W (1995) Life-cycle strategies of *Calanoides acutus*, *Calanus propinquus*, and *Metridia gerlachei* (Copepoda: Calanoida) in the eastern Weddell Sea, Antarctica. ICES J Mar Sci 52:541–548

Selz V, Lowry K, Lewis K et al (2018) Distribution of *Phaeocystis antarctica*-dominated sea ice algal communities and their potential to seed phytoplankton across the western Antarctic Peninsula in spring. Mar Ecol Prog Ser 586:91–112. https://doi.org/10.3354/meps12367

Smetacek V, Nicol S (2005) Polar ocean ecosystems in a changing world. Nature 437:362–368

Smith W Jr, Nelson DM (1990) Phytoplankton growth and new production in the Weddell Sea marginal ice zone in the austral spring and autumn. Limnol Oceanogr 35:809–821

Steiner M, Huettmann F, Bryans N, Barker B (2023) With super SDMs (machine learning, open access big data, and the cloud) towards more holistic global squirrel hotspots and coldspots. Sci Rep 14:5204. https://doi.org/10.21203/rs.3.rs-2883362/v1

Stoecker DK, Lavrentyev PJ (2018) Mixotrophic plankton in the polar seas: a pan-arctic review. Front Mar Sci 5:292. https://doi.org/10.3389/fmars.2018.00292

Swadling KM, Gibson JAE (2000) Grazing rates of a calanoid copepod (*Paralabidocera antarctica*) in a continental Antarctic lake. Polar Biol 23:301–308. https://doi.org/10.1007/s003000050449

Swadling KM, Constable AJ, Fraser AD et al (2023) Biological responses to change in Antarctic sea ice habitats. Front Ecol Evol 10:1073823. https://doi.org/10.3389/fevo.2022.1073823

Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293

Turner JT (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs. Zool Stud 43:255–266

Turner J, Barrand NE, Bracegirdle TJ et al (2014) Antarctic climate change and the environment: an update. Polar Rec 50:237–259

Van Goethem MW, Cowan DA (2019) Role of cyanobacteria in the ecology of polar environments. In: Castro-Sowinski S (ed) The ecological role of micro-organisms in the Antarctic environment. Springer International Publishing, Cham, pp 3–23

Vancoppenolle M, Meiners KM, Michel C et al (2013) Role of sea ice in global biogeochemical cycles: emerging views and challenges. Quatern Sci Rev 79:207–230. https://doi.org/10.1016/j.quascirev.2013.04.011

Villanova V, Spetea C (2021) Mixotrophy in diatoms: molecular mechanism and industrial potential. Physiol Plant 173:603–611. https://doi.org/10.1111/ppl.13471

Voronina N, Kolosova E, Melnikov I (2001) Zooplankton life under the perennial Antarctic Sea ice. Polar Biol 24:401–407. https://doi.org/10.1007/s003000100224

Warton DI, Blanchet FG, O'Hara RB et al (2015) So many variables: joint modeling in community ecology. Trends Ecol Evol 30:766–779. https://doi.org/10.1016/j.tree.2015.09.007

Yates P, Ziegler P, Welsford D et al (2019) Distribution of Antarctic toothfish *Dissostichus mawsoni* along East Antarctica: environmental drivers and management implications. Fish Res 219:105338. https://doi.org/10.1016/j.fishres.2019.105338

Zhang Q-C, Song J-J, Yu R-C et al (2013) Roles of mixotrophy in blooms of different dinoflagellates: implications from the growth experiment. Harmful Algae 30:10–26. https://doi.org/10.1016/j.hal.2013.08.003

Zwerschke N, Sands CJ, Roman-Gonzalez A et al (2022) Quantification of blue carbon pathways contributing to negative feedback on climate change following glacier retreat in West Antarctic fjords. Glob Change Biol 28:8–20. https://doi.org/10.1111/gcb.15898

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.