# Contrasting coastal polynya and open ocean phytoplankton productivity with SOCCOM BGC-Argo Float data: A Ross Sea case study

**Eli Mally**
***Mentors: Guillaume Liniger and Nick Bock***
***Summer 2026***
**Keywords:** Phytoplankton productivity, Southern Ocean, BGC-Argo Floats

**ABSTRACT**
The Ross Sea Polynya (RSP) is a region of intense nutrient upwelling that supports a high abundance of life in the Southern Ocean. Although it is an important measure of ecosystem health and the magnitude of carbon sequestration, net primary productivity (NPP) in the RSP is still not yet well known, especially during austral winter months where sea ice cover and weather conditions limit sampling. This study uses satellite observations and data from nine SOCCOM BGC-Argo Floats to quantify and predict NPP in the RSP and the surrounding open ocean (OO). Additionally, we use two different machine learning models, a Random Forest model and a Neural Network, to predict NPP during winter months when sea ice cover prevents estimations. Here, we find clear hydrographic and biogeochemical differences between the RSP and the OO. Both models predicted known NPP well during summer months using float-derived temperature and salinity. However, both models struggle to predict NPP during winter months, with the Neural Network predicting especially unrealistic values with time and depth. These results show that year-round NPP cannot be predicted using summer data alone, and emphasize the importance of collecting year-round measurements in the RSP.

**INTRODUCTION**

Phytoplankton are microscopic organisms that form the base of the ocean food chain and produce half of the world's oxygen (Naselli-Flores et al., 2022), making them critical drivers of ocean carbon cycling. By converting inorganic carbon into organic carbon biomass, phytoplankton provide energy for higher trophic levels and influence the exchange of carbon between the atmosphere and the ocean. Net primary productivity (NPP), defined as the rate at which phytoplankton produce biomass after accounting for respiration, is therefore an important measure of ocean ecosystem health and carbon cycling.

The Southern Ocean is especially important to these processes, regulating global cycles of heat, carbon, and the climate of the earth and taking up 40% of the earth's carbon dioxide emissions (Sarmiento et al., 2023). Macronutrients that fertilize phytoplankton blooms, including nitrate and phosphate, are upwelled in this area and distributed to the rest of the world's oceans (Uchida et al., 2019). At the same time, the Southern Ocean experiences high seasonality with extreme fluctuations in sea ice cover, temperature, and wind intensity throughout the year. Antarctic coastal polynyas, which are areas of open water surrounded by sea ice near the coast, occur within this highly dynamic environment (Morales Maqueda et al., 2004). Polynyas are regions of intense nutrient upwelling that support a high abundance of life in the Southern Ocean (Arrigo et al., 2003). Amongst Antarctic coastal polynyas, the Ross Sea Polynya (RSP) has the largest area in the Southern Ocean (Nunes et al., 2025) and creates 28% of the phytoplankton productivity of the Southern Ocean, higher than any other polynya (Park et al., 2019). The RSP is located adjacent to the Ross Sea Ice Shelf, off the coast of Victoria Land, in the Pacific sector of the Southern Ocean. The RSP is a region of rapid sea ice formation that increases salinity and water density, contributing to global circulation (Nunes et al., 2025). Surface changes occur rapidly in the RSP, where rapid phytoplankton growth and accumulation usually occur over six weeks in the summer when sea ice cover decreases enough (Cao et al., 2025). Phytoplankton are mainly limited by light and nutrients—particularly iron—in the RSP (Hague et al., 2021). Although light is the main limitation of phytoplankton growth during ice-covered winter periods, previous findings have shown that phytoplankton blooms occur well before sea ice retreat (Hague et al., 2021; McClish et al., 2023). The RSP is additionally undergoing physical, chemical, and biological changes with the effects of a changing climate (Kaufman et al., 2017). It has had a large positive trend in sea ice cover since 1978 (Parkinson, 2019), but in recent years has experienced extensive sea ice loss (Jena et al., 2024). Although the RSP has been monitored through ship-based observations, these observations are are heavily seasonally biased, with available data limited during austral winter months when sea ice cover is extensive. Satellites can measure many physical and biological properties with high spatial and temporal resolution, but are limited by ice and cloud cover. Additionally, satellite measurements can only give information about the surface layer, creating large gaps in data as phytoplankton distributions and productivity can extend far below the surface (Behrenfeld et al., 2010; Stoer and Fennel, 2024).

To address these spatial and temporal gaps in the Southern Ocean, the Southern Ocean Carbon and Climate Observations and Modeling program (SOCCOM) was created to deploy a vast array of biogeochemical (BGC) Argo floats. BGC-Argo floats are autonomous platforms drifting in a quasi-Lagrangian way that measure core variables (temperature, salinity, density) and biogeochemical

ocean properties. Floats sample as deep as 2000 meters, offering a 3D perspective of the ocean interior and overcoming the challenges faced by traditional shipboard sampling strategies. Recent findings using autonomous profiling floats and gliders have characterized hydrographic variables and revealed insights about phytoplankton productivity in the RSP (Porter et al., 2019; Cao et al., 2025; Portela et al., 2025). However, much is still not yet known about NPP in the area, especially during the winter months.

This paper aims to quantify environmental conditions and model NPP in the RSP and surrounding area using SOCCOM BGC-Argo Float data. Considering the importance of this region and the current gaps, this works explores the following questions:

1. Using BGC-Argo Float data, what are the patterns of hydrographic conditions and phytoplankton productivity in the Ross Sea Polynya?
2. Can Ross Sea net primary productivity be predicted using float data and machine learning?

## MATERIALS AND METHODS

### 1. SOCCOM BGC-Argo Floats

Typical to autonomous profiling floats, BGC-Argo Floats are quasi-Lagrangian, drifting freely in the water. On average, floats sample every 10 days and as deep as 2000 meters. They measure, in addition to core variables (temperature, salinity, density), up to six biogeochemical ocean properties: pH, nitrate, dissolved oxygen, chlorophyll-a fluorescence, suspended particles (backscattering at 700nm), and downwelling irradiance. Other variables can also be derived from these measurements, such as particulate organic carbon (POC), total alkalinity (TALK), dissolved inorganic carbon (DIC), and surface partial pressure of $CO_2$ ($pCO_2$).

Parameters (oxygen, nitrate, pH, chlorophyll-a fluorescence, and backscattering at 700nm) from nine SOCCOM BGC-Argo Floats in the Ross Sea area was accessed in high resolution Empirical Seawater Property Estimation Routines (ESPER) format from the SOCCOM and GO-BGC Snapshot (Johnson et al., 2017; Johnson et al., 2026). The Snapshot data is pre-quality controlled and includes the derived variables POC, TALK, $pCO_2$, and DIC. Additional quality control was performed on the float data, removing physically impossible values. Four of these floats (7900823, 7900824, 7900825, and 7900827) were deployed off the Ross Sea continental shelf from the *R/V Aaron* between January 26 and February 7, 2023. A fifth float (7900826) was deployed at the same time, but measured only three profiles before discontinuing, so its data was not used for this paper. Five additional floats were deployed in the open ocean areas neighboring the Ross Sea: 5904860, 5904858, and 5905636 were

deployed from the *R/V Palmer* between December 2016 and April 2018; 5906556 and 5906557 were deployed from the *R/V Revelle* between April and June 2023.

Data was collected and made freely available by the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR-1425989, with extension NSF OPP-1936222), and by the Global Ocean Biogeochemistry Array (GO-BGC) Project funded by the National Science Foundation, Division of Ocean Sciences (NSF OCE-1946578), supplemented by NASA, and by the International Argo Program and the NOAA programs that contribute to it. The Argo Program is part of the Global Ocean Observing System (https://doi.org/10.17882/42182, https://www.ocean-ops.org/board?t=argo).

## 2. Satellite data products

Satellite sea ice concentration (SIC) data for 2014-2025 was obtained from the University of Bremen Institute of Environmental Physics AMSR2 ASI Sea Ice Concentration data archive. The Antarctic 3.125-km gridded product (s3125), version 5.4, was used. The data was generated using the ARTIST Sea Ice (ASI) algorithm applied to satellite AMSR2 observations (Spreen et al., 2008). Level 3 mapped SNPP-VIIRS photosynthetically active radiation (PAR) data for 2017-2026 were obtained from the NASA Ocean Biology Distributed Active Archive Center (NASA, 2025). Both satellite products were matched with BGC-Argo Floats through time and space using nearest-neighbor matching with the Python SciPy KDTree function (Virtanen et al., 2020).

## 3. NPP calculation

Net primary productivity (NPP) was estimated for the top 200 meters of the ocean using the Carbon Based Productivity Model (Westberry et al., 2008), following methods described in Yang et al. (2021):

$$NPP = C_{phyto} * \mu,$$

where $C_{phyto}$ was defined as phytoplankton carbon (mg C $m^{-3}$) and $\mu$ was defined as phytoplankton specific growth rate ($d^{-1}$). Based on Graff et al. (2015), $C_{phyto}$ was calculated as:

$$C_{phyto} = 12128 * b_{bp}(470) + 0.59,$$

where $b_{bp}(470)$ was converted from float-measured $b_{bp}(700)$ (optical backscattering at 700nm) as described in Morel and Maritorena (2001):

$$b_{bp}(470) = b_{bp}(700)(\frac{400}{700})^{-1}.$$

Phytoplankton growth rate μ was calculated following Westberry et al. (2008) using float-derived chlorophyll-a (Chl):

$$\mu = \frac{\frac{2*Chl}{Cphyto}(1 - e^{-5Iz})}{0.022 + (0.045 - 0.022) * e^{-3Iz}},$$

where $I_z$ (mol photon $m^{-2}$ $d^{-1}$) corresponds to PAR at each depth (z). As none of the floats were equipped with PAR sensors, PAR at depth was estimated using satellite measurements of surface PAR ($I_0$):

$$I_z = I_0 * e^{-kdPAR*z}.$$

$kd_{PAR}$ (the attenuation coefficient for PAR) was calculated based on Morel et al. (2007) as:

$$kd_{PAR} = 0.0655 + 0.874*kd_{490} - (0.00121*kd_{490}^{-1}),$$

where $kd_{490}$ was calculated as:

$$kd_{490} = 0.0166 + a*Chl^{b},$$

where values of 0.0426086 and 0.4403899 were used for constants *a* and *b,* respectively, based on findings in Cornec et al. (2021).

## 4. NPP derived from machine learning

### 4.1. Selecting and transforming predictors

To predict NPP on float profiles, a subset of predictors was selected from the available float variables: latitude, longitude, temperature, salinity, pressure, day of year (DOY), oxygen, apparent oxygen utilization (AOU), pH, nitrate, DIC, and POC. AOU was estimated using the Python GSW toolbox. Data was subset to include only points taken at 200 meters depth and above. Different combinations of predictor variables were tested in Random Forest to determine the best model. Longitude, pressure, and DOY variables were transformed following Sauzède et al., 2017, Carter et al., 2021, and Liniger et al. (2025):

$$DOY_{corr} = \frac{DOY * \pi}{182.645}$$

$$Pressure_{corr} = \frac{Pressure}{20000\ dbar} + \frac{1}{(1 + exp(-\frac{Pressure}{300\ dbar}))^{3}}$$

$$Longitude_{1,corr} = \cos(Longitude - 20)$$

$$Longitude_{2,corr} = \cos(Longitude - 110)$$

As DOY and longitude are circular, these variables were transformed to preserve the periodicity. Pressure was transformed to account for a wide range of values.

### 4.2. Random Forest

To predict NPP, Random Forest models were performed in Python using the 'RandomForestRegressor' function from the scikit-learn package (Pedregosa et al., 2011). A random subset of 80% of the data points with known NPP (n = 4968) was selected for training the model, with the remaining 20% of points being used for testing. Models were run to predict unknown NPP ten times, each time with a different random 80/20 training and testing split, and the predicted NPP values were averaged across all ten runs.

4.3. Neural Network

Neural Network models (multilayer perceptrons) were performed to predict net primary productivity using the MATLAB 'feedforwardnet' function with a Bayesian regularization (trainbr) to minimize overfitting. As for RF, , a random 80/20 split of data points was used to train and test the model. The Neural Network was built with two hidden layers, and was run five times with five different combinations of neurons in the first and second hidden layer ([10 and 10], [15 and 15], [10 and 20], [20 and 10], and [20 and 20]). Data was square transformed to prevent negative predicted NPP values from appearing. The predicted NPP output for each of the five runs was then averaged to result in one overall predicted NPP value for each sample.

All data analysis and visualization was performed in Python v3, except for the Neural Network (Methods Section 4.3), which was performed in MATLAB R2026a.

**RESULTS**

1. Study area and float positions

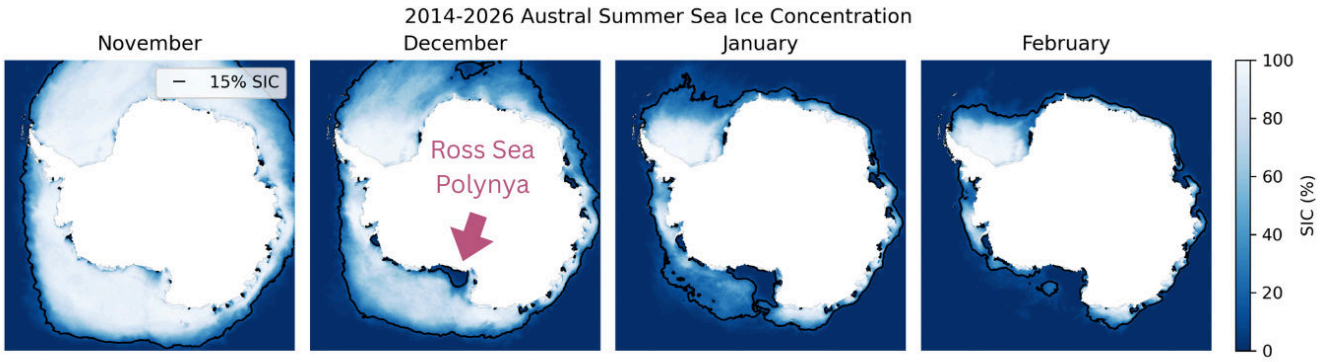



*Figure 1. Southern Ocean sea ice concentrations (SIC) for November-February, averaged from 2014-2026. White color represents high SIC; dark blue represents open water. The black contour line*

*represents a 15% SIC threshold commonly used in oceanography to define the boundaries of coastal polynyas.*

Southern Ocean SIC was highly seasonal, with extreme fluctuations between summer and winter months. On average, SIC was at its maximum extent during September and at its minimum during February. During austral summer months (around November-January), the Ross Sea Polynya (RSP) began to appear and widen, outlined by a 15% SIC contour line in Figure 1.

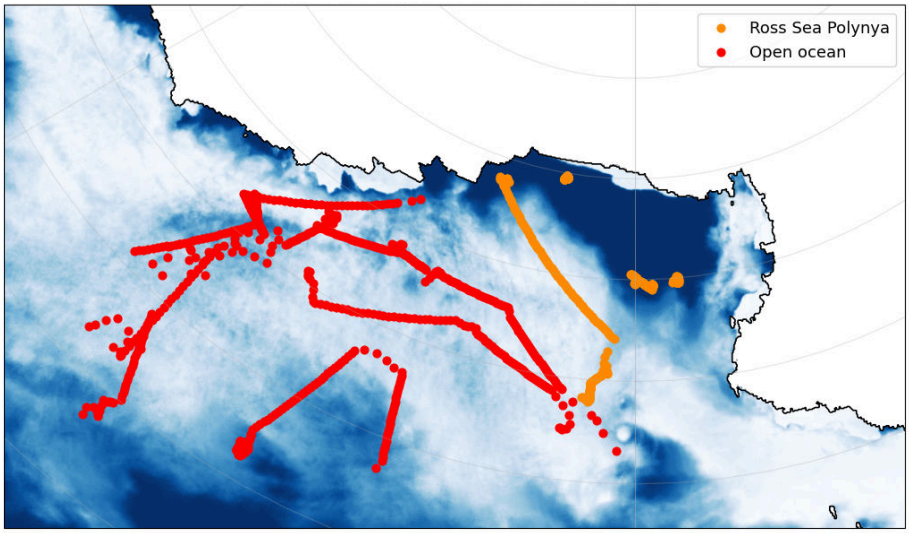



*Figure 2. Float trajectories for the nine studied SOCCOM BGC-Argo floats, where each point represents the location of a different float profile. Orange points represent profiles from floats in the "Ross Sea Polynya" group (n=441); red points represent profiles from floats in the "Open Ocean" group (n=659). The blue background image corresponds to the average SIC in December 2025, where white denotes a high concentration and dark blue denotes open water. The white area, outlined in black, is Antarctic continental ice.*

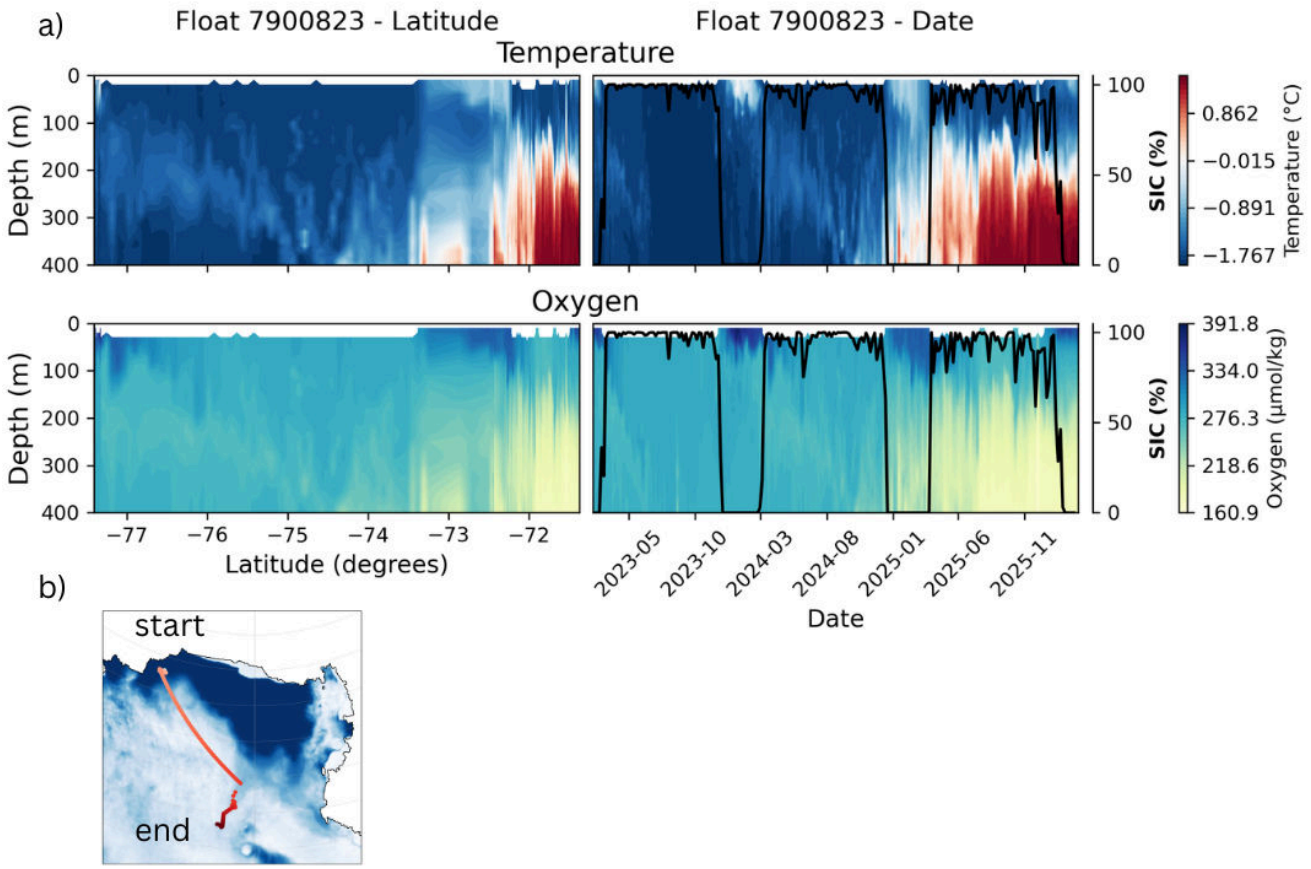



*Figure 3. a) Contour plots of temperature and oxygen with depth for Ross Sea float 7900823 as a function of latitude (left column) and time (right column).The bold black line overlaid on the right column corresponds to sea ice concentration. b) A map of the trajectory of Float 7900823, beginning near the edge of the Ross Ice Shelf and drifting further into the open ocean.*

As autonomous profiling floats are quasi-Lagrangian, the BGC-Argo Floats studied in this paper drifted freely through space and time and captured changes in their environment. All floats stayed in either the RSP region or the open ocean (OO) region (Figure 2), with the exception of Float 7900823, which drifted from the Ross Sea continental shelf into the open ocean (Figure 3). Hydrographic and biogeochemical variables shifted throughout the float's sampling period. The RSP area was cold and rich in oxygen, but when the float drifted further out, a warmer and deoxygenated water mass was observed.

## 2. Hydrographic variables

### 2.1. Contrast between the Ross Sea Polynya and the open ocean

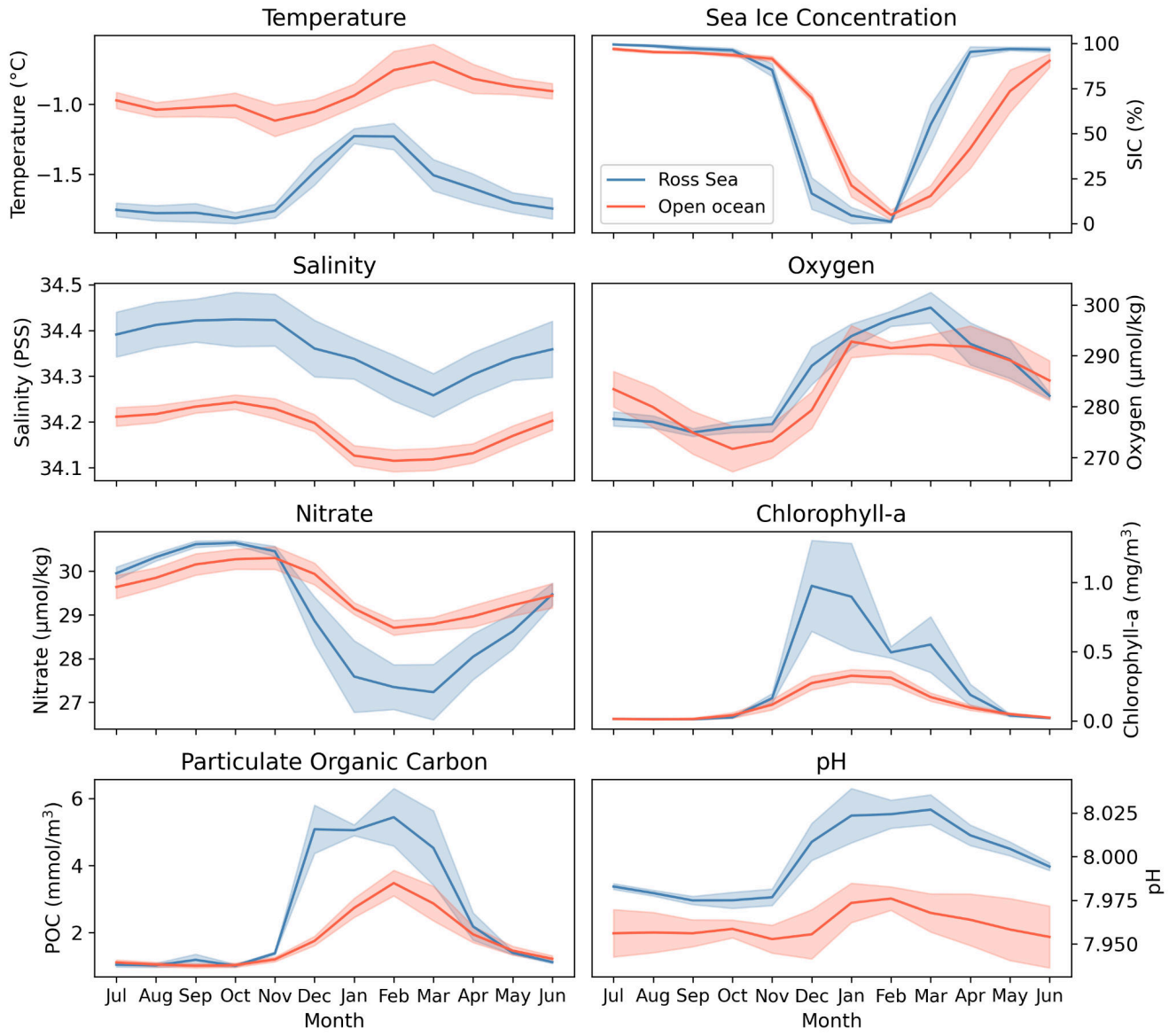



*Figure 4. Monthly climatologies of different hydrographic and biogeochemical variables (temperature, sea ice concentration, salinity, oxygen, nitrate, chlorophyll-a, particulate organic carbon, and pH) for the top 200m of the water column. Blue corresponds to floats in the Ross Sea group; red corresponds to floats in the Open Ocean group. Shaded regions indicate standard error across floats.*

Both the RSP and the OO showed seasonal fluctuations in the top 200m. Winter months (June-October) were characterized by ~100% SIC, low temperatures (RSP: between -1.81 and -1.75°C; OO: between -1.04 and -0.91°C), and near-zero chlorophyll-a biomass and POC. High nitrate concentrations (between 29 and 30 µmol/kg for both regions) were observed, while low oxygen concentrations (RSP: between 279 and 281 µmol/kg; OO: between 280 and 290 µmol/kg) and pH

(RSP: between 7.975 and 8.000; OO: between 7.940 and 7.975) were seen. During and after November (the beginning of the summer months), SIC dropped sharply to near-zero, temperatures rose (peaking in January-February in the RSP and March in the OO), and chlorophyll-a concentrations increased with high variability between floats. Additionally, nitrate concentrations decreased, while POC and pH became elevated.

The RSP and OO floats also exhibited clear regional differences in hydrographic and biogeochemical variables. On average, the RSP was colder and much more saline than the neighboring OO. It also had slightly higher dissolved oxygen and dissolved inorganic carbon concentrations. The RSP showed much greater chlorophyll-a concentrations than the OO during the summer months, exhibiting local maxima during December and March. SIC began to decrease earlier, with the sharpest decline occurring in December in the RSP but January in the OO. This earlier SIC decrease aligned with an earlier increase and peak in temperature in the RSP.

## 2.2. Under-ice phytoplankton activity

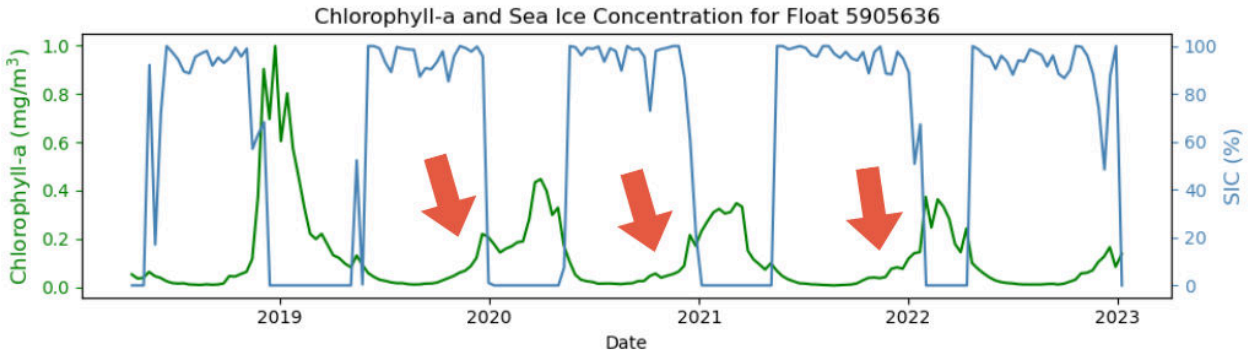



*Figure 5. SIC (blue) and chlorophyll-a concentrations averaged over the top 200 meters (green) over time for open ocean Float 5905636. Red arrows indicate phytoplankton biomass increase (high chlorophyll-a) that appear well before SIC drops.*

Chlorophyll-a concentrations were highly dependent on SIC. Concentrations were close to or at zero for most of the winter months (April-November) when SIC were at their peak. In the summer, SIC decreased sharply and chlorophyll-a concentrations consequently spiked sharply. However, phytoplankton biomass often increases well before sea ice began to retreat in the summer: Figure 5 displays chlorophyll-a and SIC over time for an OO float (5905636), where chlorophyll-a concentrations began to spike well before SIC had dropped dramatically.

3. Net primary production

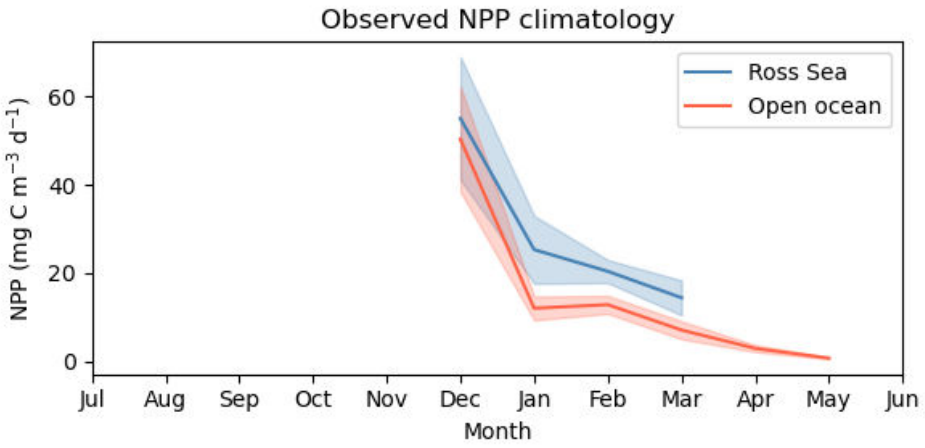



*Figure 6. Monthly climatology for observed net primary productivity (NNP) for both the Ross Sea Polynya (blue) and the open ocean (red). Lines represent monthly averages for all floats; shaded areas represent standard error. No data are available during winter months due to ice coverage.*

Net primary productivity (NPP) displayed clear seasonal patterns. For both the RSP and the OO, NPP was by far the highest and most variable in December, reaching values of 49 mg C $m^{-3}$ $d^{-1}$. Productivity dropped quickly in January and decreased thereafter (until near-zero in May for the OO and close to 20 mg C $m^{-3}$ $d^{-1}$ in March for the RSP), with the RSP exhibiting almost 10 mg C $m^{-3}$ $d^{-1}$ more productivity from January through March than the OO on average. However, large portions of the year contained no NPP estimations at all, as sea ice cover blocked the retrieval of satellite downwelling irradiance data required to calculate productivity. NPP could only be calculated for the summer months of December-March in the RSP, with no available data during the rest of the year. The OO had slightly more data, with NPP calculations from December-May, but still displayed no wintertime NPP. A contour plot of estimated NPP with depth for each float can be found in the Supplementary Appendix (Figure S1).

4. Machine learning

4.1. Random Forest model

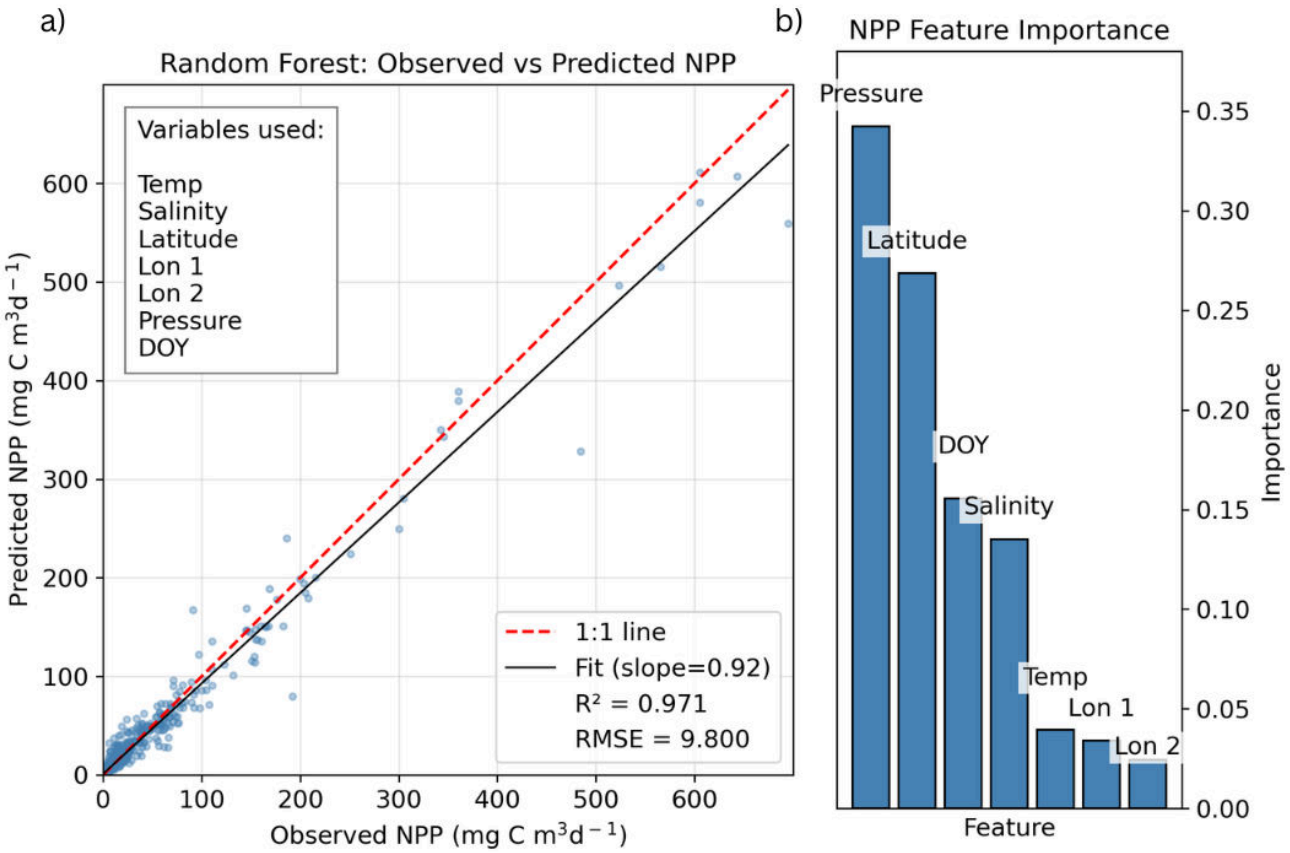



*Figure 7. a) Observed vs. predicted NPP from the test dataset using the best Random Forest predictor model (temperature, salinity, latitude, longitude, pressure, and day of the year (DOY)). Blue points represent individual data points (n = 1264). The black line represents the linear fit of the plot with a slope of 0.92. The red dashed line is a 1:1 line. b) Contribution of each feature used to predict NPP, where a higher value represents a larger contribution to the model.*

Random Forest (RF) models were performed on selections of BGC-Argo Float parameters in the upper 200 meters of the water column to attempt to predict NPP. An 80/20 split was done on the known NPP data points to create training (n = 5053) and testing (n = 1264) datasets. The best predictor model was found to use temperature, salinity, latitude, longitude, pressure (depth), and day of year (DOY) as predictor variables. Adding biological variables such as nitrate, pH, or oxygen made the model perform worse. Pressure, latitude, DOY, and salinity had the greatest influence on the model, with some influence by temperature and longitude (Figure 7b). When observed and predicted NPP for the test dataset were plotted against each other (Figure 7a), the model resulted in an $R^2$ value of 0.971 with a slope of 0.920 (slightly underpredicting NPP, especially when values were high) and an RMSE of 9.800 mg C $m^3$ $d^{-1}$.

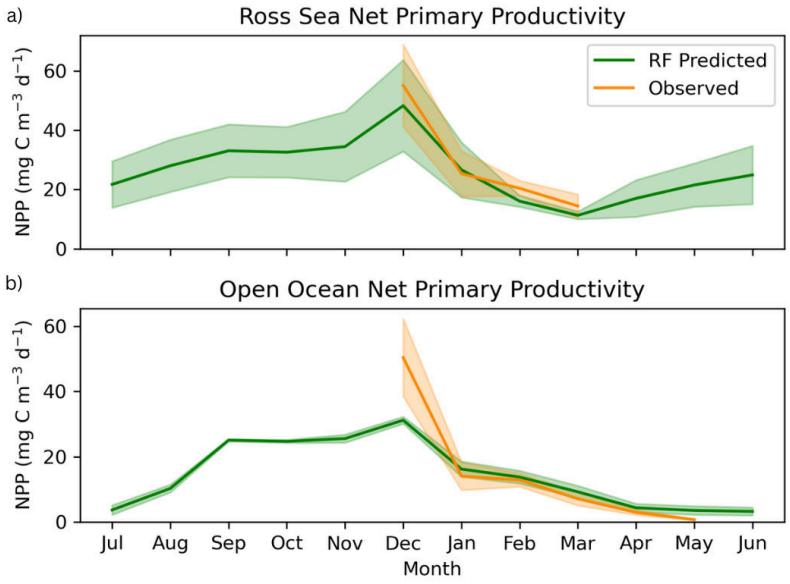



*Figure 8. Monthly climatologies of net primary productivity for the upper 200 meters of the water column in the Ross Sea (a) and the open ocean (b). Solid lines represent monthly averages across all data; shaded areas represent standard error across floats. Orange corresponds to observed NPP; green corresponds to NPP predicted by the Random Forest model.*

The RF model predicted monthly average NPP values within the range of what had been observed, with observed values ranging from 14.430 ± 3.942 to 55.052 +- 13.907 mg C $m^{-3}$ $d^{-1}$ in the RSP and 0.738 +- 0.209 to 50.333 +- 11.900 mg C $m^{-3}$ $d^{-1}$ in the OO. In the RSP (Figure 8a), average monthly observed values in the summer months (when open water allowed the retrieval of satellite PAR data) aligned well with what the model had predicted, decreasing from 49 mg C $m^{-3}$ $d^{-1}$ in December to 11 mg C $m^{-3}$ $d^{-1}$ in March. From April to November, the predicted NPP values increased steadily with high variability. In the OO (Figure 8b), predicted values decreased fastest from December to January and continued decreasing until May, similar to what was observed; However, the model underpredicted NPP in December by a factor of 2. From April to November in the OO, the RF model also predicted steadily increasing NPP values. Contour plots of predicted NPP with depth for each float can be found in the Supplementary Appendix (Figure S2).

## 4.2. Neural Network model

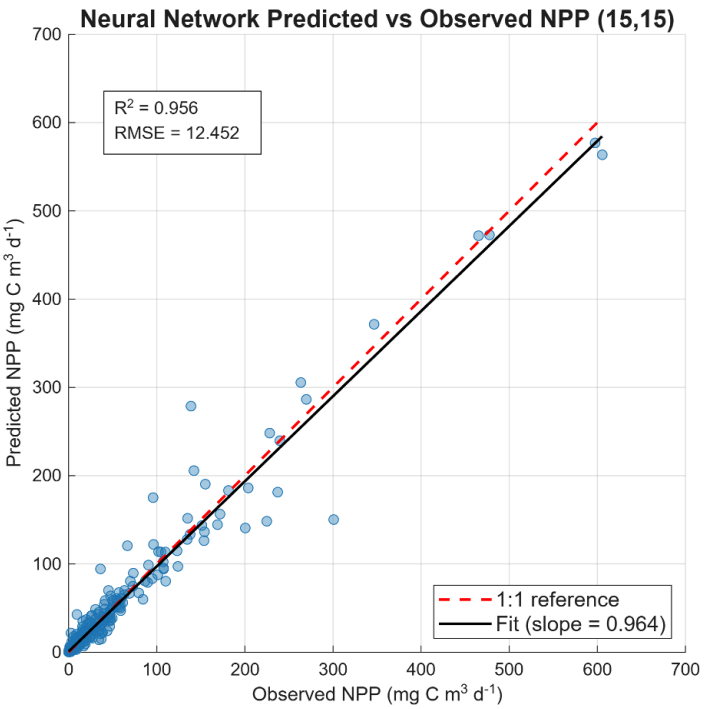



*Figure 9. Observed vs. predicted NPP using a Neural Network predictor model (with the same variables used in the Random Forest model) with two hidden layers of 15 neurons each, one of five models that was averaged to predict NPP. Blue points represent individual data points. The black line represents the linear fit of the plot with a slope of 0.988. The red dashed line is a 1:1 reference line.*

A neural network (NN) model was performed using the same variables as in the best Random Forest model (temperature, salinity, latitude, longitude, pressure, and DOY). Five NN models were run, each with two hidden layers and varying numbers of neurons inside each layer. When observed and predicted NPP were plotted against each other, the five models resulted in a range of $R^2$ values (0.943 to 0.971), linear fits (0.936 to 1.017), and RMSE values (10.002 to 14.064 mg C $m^3$ $d^{-1}$) (specifics in Supplementary Appendix, Table S1 and Figures S2-6). In comparison, the RF model had an $R^2$ in the same range (0.971) and a slightly lower fit (0.920) and RMSE (9.800 mg C $m^3$ $d^{-1}$), performing similarly to the NN model when predicting known NPP.

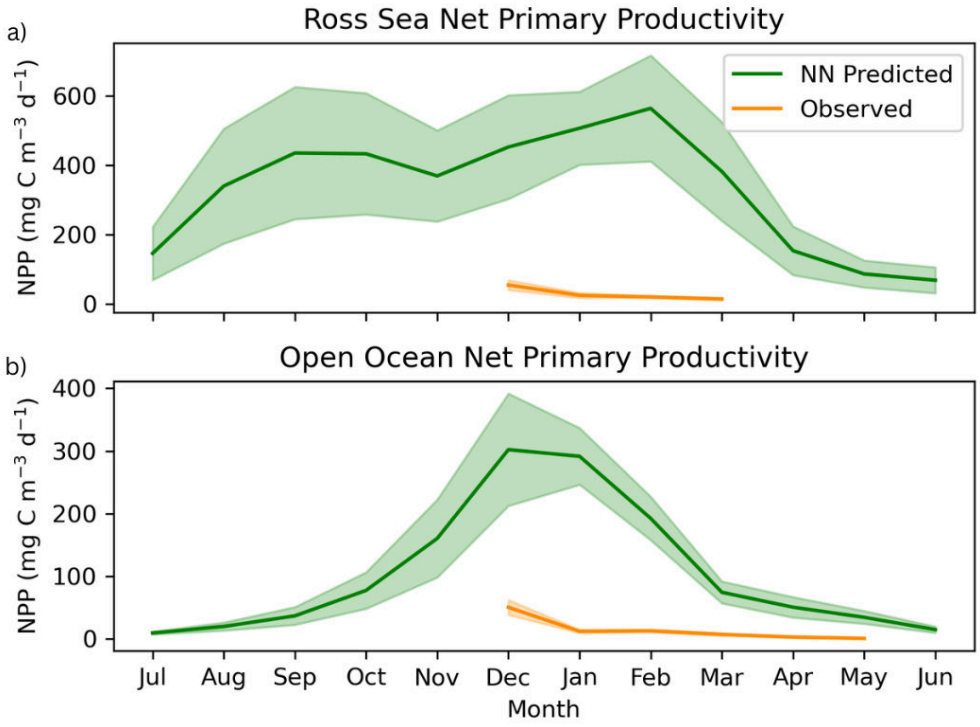



*Figure 10. Monthly climatologies of net primary productivity in the Ross Sea (a) and the open ocean (b). Solid lines represent monthly averages across all data; shaded regions represent standard error across floats. Orange corresponds to observed NPP; green corresponds to NPP predicted by the Neural Network model.*

When extrapolating to predict unknown NPP values, the NN model struggled to predict NPP year-round. Monthly average predicted values ranged from 68.689 ± 37.414 to 564.496 ± 153.050 mg C $m^3$ $d^{-1}$ in the RSP and from 9.394 ± 2.243 to 302.012 ± 89.703 mg C $m^3$ $d^{-1}$ in the OO, an order of magnitude above the highest monthly average observed NPP in either region (RSP: 55.052, OO: 50.333 mg C $m^3$ $d^{-1}$). In the RSP, predicted NPP displayed two maxima in February and September-October and decreased to its lowest rate in June, providing a stark contrast to the observed NPP's December peak. In the OO, NPP was highest in December and decreased for the rest of the year until July, at which point it steadily increased. Contour plots of predicted NPP with depth for each float can be found in the Supplementary Appendix (Figure S8).

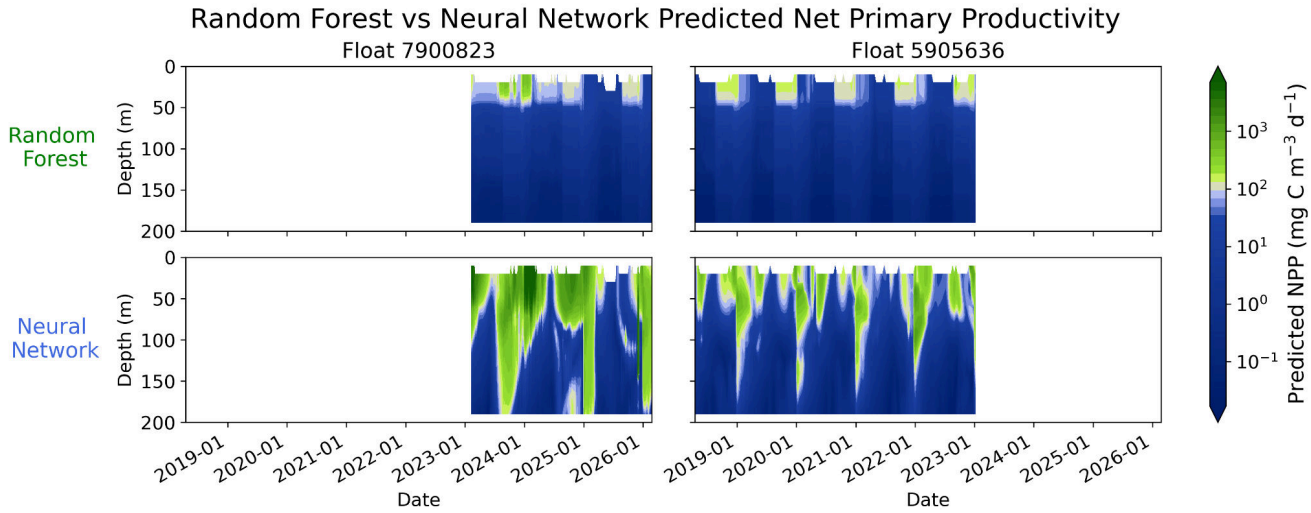



*Figure 11. Contour plots of predicted NPP for RSP float 7900823 (left column) and OO float 5905636 (right column) across time for the top 200m of the water column. The top row represents the Random Forest-predicted NPP, while the bottom row represents the Neural Network-predicted NPP. Dark blue colors correspond to zero or near-zero values of NPP, while dark green colors correspond to high NPP. White colors represent areas of no data.*

As well as predicting different orders of magnitude, the RF and NN models exhibited different distributions of NPP across time and depth. In the RF model, predicted NPP remained close to or at zero below the top 50 meters of the water column, and often (but not always) during the winter months. Although many floats had periods of low NPP during the winter, some did not have a clear seasonal trend. In the NN model, predicted NPP showed extreme variability across time and depth, with high rates extending from the surface up to 200 meters in many of the summer months. Predictions were also more inconsistent than in the RF model, with a lack of a clear low-NPP winter period.

## DISCUSSION

The Ross Sea Polynya and the open ocean exhibited different hydrographic and biogeochemical conditions. On average, floats in the RSP displayed lower temperatures year-round than floats in the OO, likely due to Antarctic katabatic winds that descend downslope from mountains and push the forming sea ice away, exposing the water to a colder atmosphere (Falco et al., 2024). The RSP also exhibited year-round higher salinity due to brine rejection (where the formation of sea ice pushes salt into the surrounding water) (Falco et al., 2024), and higher pH, potentially due to the more intense phytoplankton activity that occurred and removed $CO_2$ (Rivaro et al., 2014). As well as

displaying different average year-round conditions, the RSP and OO had different temporal patterns. In the RSP, sea ice concentrations decreased faster and earlier than in the OO, but autumn sea ice also returned faster. Temperature followed a similar trend, where RSP temperatures increased and peaked earlier than in the OO. One float (7900823), which had drifted between the RSP and the OO during its lifetime, observed a higher-temperature and deoxygenated water mass at depth after reaching the OO.

Both regions exhibited clear phytoplankton blooms during the austral summer months, evidenced by increases in chlorophyll-a concentrations. These peaks aligned with sharp decreases in nitrate from phytoplankton nutrient uptake, as well as increased POC from the increased mass of carbon in the water. Dissolved oxygen concentrations and pH were elevated and DIC decreased during this time due to photosynthetic activity by phytoplankton. The RSP experienced earlier and greater phytoplankton blooms and more nitrate depletion than the OO. Additionally, while the OO displayed a smooth curve in chlorophyll-a concentration, peaking in January, the RSP had two clear peaks in chlorophyll-a: once in December and once in March. The rate of NPP was consistently slightly higher in the RSP than in the OO. From what was estimated with the available PAR data, NPP was at its highest in December (the earliest month with available PAR for both regions), thereafter sharply decreasing. NPP values were not directly compared to previous findings from the Ross Sea (Chen et al., 2025). However, as phytoplankton blooms have been shown to occur under sea ice (Figure 5; Hague et al., 2021; McClish et al., 2023), there is likely primary productivity under sea ice that is not captured by the temporal extent of the NPP estimation.

The Random Forest and Neural Network models showed comparable results when predicting known NPP (when PAR was available in the summer), using the predictor variables temperature, salinity, DOY, pressure, latitude, and longitude. Random Forest predicted unknown summertime NPP well, while the neural network model struggled and highly overpredicted by an order of magnitude. However, both models struggled to predict unknown NPP in the winter months. Although missing data meant that wintertime NPP could not be calculated directly for comparison, the predicted wintertime NPP values were higher than what was biologically probable for both models—the Random Forest model predicted an almost linear NPP increase from March to December for the RSP, with similar increases from May to December for the OO. Likewise, the neural network model predicted increases in NPP throughout the winter. Float-derived chlorophyll-a and POC (Figure 4), which often (but not always, due to photoacclimation [Westberry et al., 2008]) mirror NPP, showed a Gaussian distribution with elevated values during summer months, but near-zero values from May to October. LiDAR-derived NPP calculations in the Ross Sea have also shown a Gaussian distribution (Chen et al., 2025), further supplying evidence that the model predictions were inaccurate. Both models likely

failed to accurately predict NPP during winter months because of a lack of wintertime data. The causes of the extreme overpredictions by the neural network are still not entirely known, but it is possible that the matrix-based model is too complex for a small NPP dataset.

Multiple caveats need to be considered in the context of these results. Foremost, limited quality-controlled float data limited the full range of hydrographic and biogeochemical observations, potentially causing the machine learning models to have difficulty predicting anything outside of the training data range. In addition, the NPP estimation is subject to error because nearest neighbor matching was used between float coordinates and satellite PAR. Additionally, although Float 7900823 was classified into the "RSP" category, it drifted into the OO with time and may have been better classified into the "OO" category in the middle of its trajectory. Although 200m was used as a depth threshold for NPP calculations, the euphotic zone depth may have been a better proxy for the productive layer. Additionally, due to time constraints, the only combination of variables tested on the neural network were the ones that had been found to produce the best Random Forest model. A different combination of variables could have potentially produced a more accurate neural network.

In conclusion, we found clear environmental differences between the RSP and OO regions. We additionally emphasize that it is currently not possible to calculate NPP where there is sea ice through BGC-Argo Float and satellite-based methods alone. Future work could involve expanding this analysis to include more BGC-Argo floats in the Southern Ocean, especially those equipped with PAR sensors to avoid reliance on satellite-derived PAR. In addition, the performance of other machine learning models (such as XGBoost) could be tested. Predicting NPP could also involve incorporating SIC (by calculating depth-integrated instead of depth-resolved NPP) or accessing light data from other sources such as numerical models. Despite the above caveats, we tested the efficacy of two different machine learning models in predicting NPP in the Ross Sea area, emphasizing the need to collect more year-round data to understand phytoplankton productivity.

**ACKNOWLEDGEMENTS**

I would like to thank my mentors, Guillaume Liniger and Nicholas Bock, for their constant guidance and support throughout this internship. I would also like to thank George Matsumoto, Megan Bassett, and Jessica Chapman for tirelessly working to coordinate internship logistics, as well as the Float Data team at MBARI. I am grateful towards all the other MBARI summer interns for their curiosity and encouragement. The MBARI Summer Internship Program is generously supported by the David and Lucile Packard Foundation. Additional funding is provided by Dalio Philanthropies and the Maxwell/Hanrahan Foundation.

**REFERENCES**


Arrigo, K. R., & van Dijken, G. L. (2003). Phytoplankton dynamics within 37 Antarctic coastal polynya systems. *Journal of Geophysical Research: Oceans*, *108*(C8), 2002JC001739. https://doi.org/10.1029/2002jc001739

Cao, R., Smith, W. O., Zhong, Y., Riser, S., Johnson, K. S., & Talley, L. (2025). The seasonal patterns of hydrographic and biogeochemical variables in the Ross Sea: A BGC-Argo Analysis. *Deep Sea Research Part II: Topical Studies in Oceanography*, *219*, 105436. https://doi.org/10.1016/j.dsr2.2024.105436

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y. Y., Álvarez, M., Wanninkhof, R., Feely, R. A., & Barbero, L. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, *19*(12), 785–809. https://doi.org/10.1002/lom3.10461

Chen, P., Zhang, Z., Bisson, K., Cédric Jamet, Girolamo, P. Di, Dionisi, D., Hu, Y., Zhang, S., & Pan, D. (2025). Underestimated accelerated Antarctic phytoplankton net primary production in winter over past decade from spaceborne LiDAR. *Nature Communications*, *16*(1), 11430–11430. https://doi.org/10.1038/s41467-025-66275-w

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., D'Ortenzio, F., Gentili, B., & Schmechtig, C. (2021). Deep Chlorophyll Maxima in the Global Ocean: Occurrences, Drivers and Characteristics. *Global Biogeochemical Cycles*, *35*(4). https://doi.org/10.1029/2020gb006759

Falco, P., Aulicino, G., Castagno, P., Capozzi, V., de Ruggiero, P., Garzia, A., Ferola, A. I., Cotroneo, Y., Colella, A., Fusco, G., Pierini, S., Budillon, G., Zambianchi, E., & Spezie, G. (2024). Ocean-atmosphere-ice processes in the Ross Sea: A review. *Deep Sea Research Part II: Topical Studies in Oceanography*, *218*, 105429. https://doi.org/10.1016/j.dsr2.2024.105429

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., Dongen-Vogels, V. van, Reifel, K. M., & Behrenfeld, M. J. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, *102*, 16–25. https://doi.org/10.1016/j.dsr.2015.04.006

Hague, M., & Vichi, M. (2021). Southern Ocean Biogeochemical Argo detect under-ice phytoplankton growth before sea ice retreat. *Biogeosciences*, *18*(1), 25–38. https://doi.org/10.5194/bg-18-25-2021

Jena, B., Kshitija, S., Bajish, C. C., Turner, J., Holmes, C., Wilkinson, J., Mohan, R., & Thamban, M. (2024). Evolution of Antarctic sea ice ahead of the record low annual maximum extent in september 2023. *Geophysical Research Letters*, *51*(7), e2023GL107561. https://doi.org/10.1029/2023gl107561

Johnson, Kenneth S.; Riser, Stephen C.; Boss, Emmanuel S.; Talley, Lynne D.; Sarmiento, Jorge L.; Swift, Dana D.; Plant, Josh N.; Maurer, Tanya L.; Key, Robert M.; Williams, Nancy L.; Wanninkhof, Richard H.; Dickson, Andrew G.; Feely, Richard A.; Russell, Joellen L. (2017). Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Float Data Archive. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0TX3C9X

Johnson, Kenneth S.; Fassbender, Andrea; Gray, Alison; Nicholson, David; Purkey, Sarah; Riser, Stephen C.; Takeshita, Yui; Talley, Lynne D.; Wijffels, Susan E.; Bartoloni, Sarah; Bock, Nicholas; Gilson, John; Grady, Logan A.; Guisewhite, Nicola J.; Maurer, Tanya L.; Parise, Katherine; Plant, Josh N.; Robbins, Pelle; Rupan, Rick A.; Swift, Dana D. (2026). SOCCOM and GO-BGC float data - Snapshot 2026-07-01. In Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) and Global Ocean Biogeochemistry (GO-BGC) Biogeochemical-Argo Float Data Archive. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0Z320M0

Kaufman, D. E., Friedrichs, M. A. M., Smith, W. O., Hofmann, E. E., Dinniman, M. S., & Hemmings, J. C. P. (2017). Climate change impacts on southern Ross Sea phytoplankton composition, productivity, and export. *Journal of Geophysical Research: Oceans*, *122*(3), 2339–2359. https://doi.org/10.1002/2016jc012514

Liniger, G., Sharp, J. D., Takeshita, Y., & Johnson, K. S. (2025). Two Decades of Increase in Southern Ocean Net Community Production Revealed by BGC‐Argo Floats. *Global Biogeochemical Cycles*, *39*(8). https://doi.org/10.1029/2024gb008371

McClish, S., & Bushinsky, S. M. (2023). Majority of Southern Ocean Seasonal Sea Ice Zone Bloom Net Community Production Precedes Total Ice Retreat. *Geophysical Research Letters*, *50*(20). https://doi.org/10.1029/2023gl103459

Morales Maqueda, M. A., Willmott, A. J., & Biggs, N. R. T. (2004). Polynya Dynamics: a Review of Observations and Modeling. *Reviews of Geophysics*, *42*(1). https://doi.org/10.1029/2002rg000116

Morel, A., & Maritorena, S. (2001). Bio‐optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research: Oceans*, *106*(C4), 7163–7180. https://doi.org/10.1029/2000jc000319

Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., & Tièche, F. (2007). Optical properties of the “clearest” natural waters. *Limnology and Oceanography*, *52*(1), 217–229. https://doi.org/10.4319/lo.2007.52.1.0217

NASA Ocean Biology Processing Group. (2025). *Suomi-NPP VIIRS Level-3 Global Mapped Photosynthetically Active Radiation (PAR) Data, version 2022.0* [Dataset]. NASA Ocean Biology Distributed Active Archive Center. https://doi.org/10.5067/SUOMI-NPP/VIIRS/L3M/PAR/2022.0

Naselli-Flores, L., & Padisák, J. (2022). Ecosystem services provided by marine and freshwater phytoplankton. *Hydrobiologia*, *850*, 2691–2706. https://doi.org/10.1007/s10750-022-04795-y

Nunes, G. S., Ferreira, A., & Brito, A. C. (2025). Long-term satellite data reveals complex phytoplankton dynamics in the Ross Sea, Antarctica. *Communications Earth & Environment*, *6*(1). https://doi.org/10.1038/s43247-025-02590-w

Park, J., Kim, J. H., Kim, H., Hwang, J., Jo, Y. H., & Lee, S. H. (2019). Environmental Forcings on the Remotely Sensed Phytoplankton Bloom Phenology in the Central Ross Sea Polynya. *Journal of Geophysical Research: Oceans*, *124*(8), 5400–5417. https://doi.org/10.1029/2019jc015222

Parkinson, C. L. (2019). A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic. *Proceedings of the National Academy of Sciences*, *116*(29), 14414–14423. https://doi.org/10.1073/pnas.1906556116

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Portela, E., Meyer, M. G., Heywood, K. J., & Smith, W. O. (2025). Unprecedented Summer Phytoplankton Bloom in the Ross Sea. *Geophysical Research Letters*, *52*(3). https://doi.org/10.1029/2024gl111264

Porter, D. F., Springer, S. R., Padman, L., Fricker, H. A., Tinto, K. J., Riser, S. C., & Bell, R. E. (2019). Evolution of the Seasonal Surface Mixed Layer of the Ross Sea, Antarctica, Observed With Autonomous Profiling Floats. *Journal of Geophysical Research: Oceans*, *124*(7), 4934–4953. https://doi.org/10.1029/2018jc014683

Rivaro, P., Messa, R., Ianni, C., Magi, E., & Budillon, G. (2014). Distribution of total alkalinity and pH in the Ross Sea (Antarctica) waters during austral summer 2008. *Polar Research*, *33*(1), 20403. https://doi.org/10.3402/polar.v33.20403

Sarmiento, J. L., Johnson, K. S., Arteaga, L. A., Bushinsky, S. M., Cullen, H. M., Gray, A. R., Hotinski, R. M., Maurer, T. L., Mazloff, M. R., Riser, S. C., Russell, J. L., Schofield, O. M., & Talley, L. D. (2023). The Southern Ocean carbon and climate observations and modeling (SOCCOM) project: A review. *Progress in Oceanography*, *219*, 103130–103130. https://doi.org/10.1016/j.pocean.2023.103130

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., & Johnson, K. S. (2017). Estimates of water-column nutrient concentrations and carbonate system parameters in the Global Ocean: A novel approach based on neural networks. *Frontiers in Marine Science*, *4*, 128. https://doi.org/10.3389/fmars.2017.00128

Spreen, G., Kaleschke, L., & Heygster, G. (2008). Sea ice remote sensing using AMSR-E 89 GHz channels. *Journal of Geophysical Research: Oceans, 113*(C2), C02S03. https://doi.org/10.1029/2005JC003384

Stoer, A. C., & Fennel, K. (2024). Carbon-centric dynamics of Earth's marine phytoplankton. *Proceedings of the National Academy of Sciences*, *121*(45). https://doi.org/10.1073/pnas.2405354121

Uchida, T., Balwada, D., Abernathey, R., Prend, C. J., Boss, E., & Gille, S. T. (2019). Southern Ocean Phytoplankton Blooms Observed by Biogeochemical Floats. *Journal of Geophysical Research: Oceans*, *124*(11), 7328–7343. https://doi.org/10.1029/2019jc015355

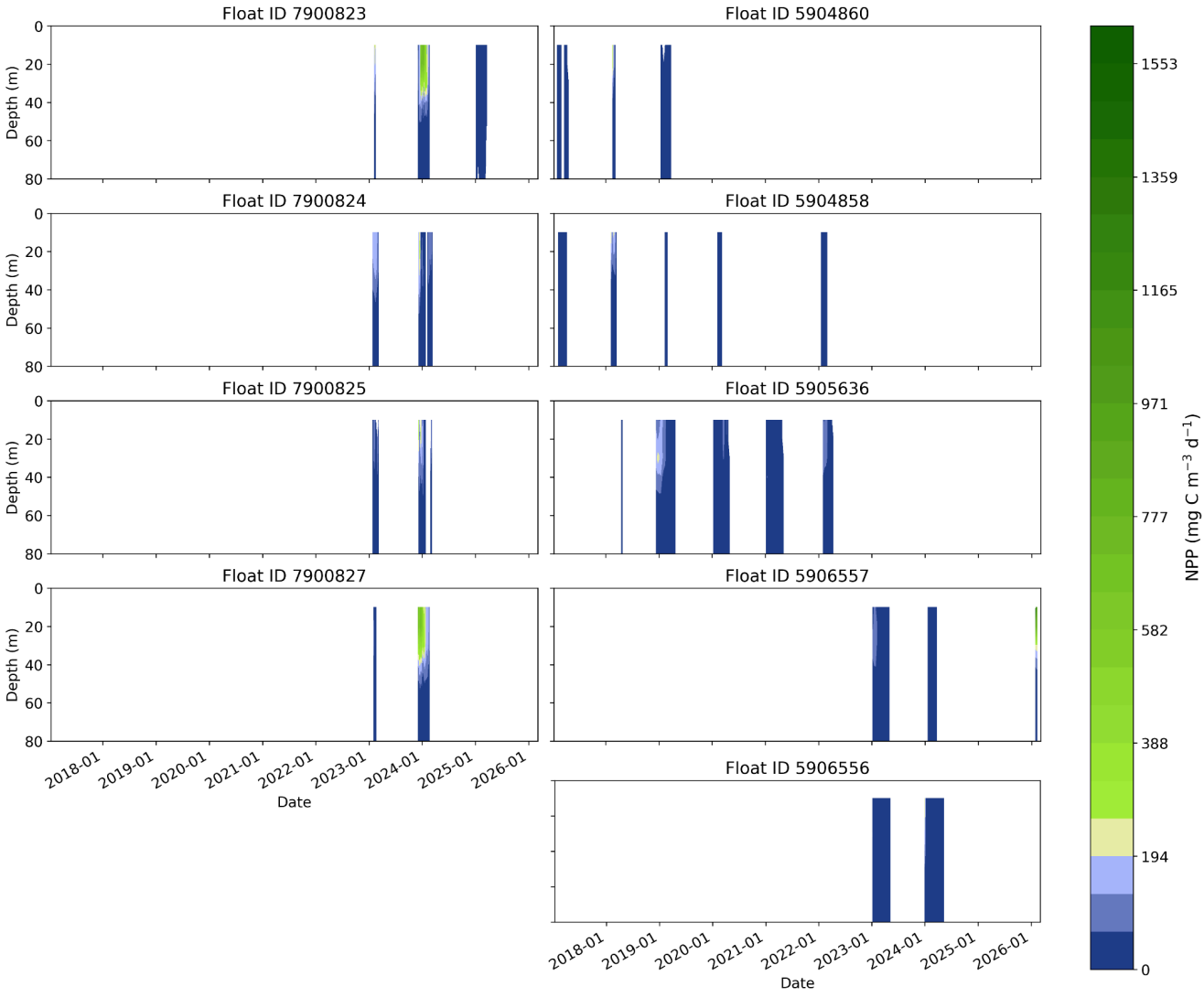
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://www.nature.com/articles/s41592-019-0686-2

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), n/a–n/a. https://doi.org/10.1029/2007gb003078

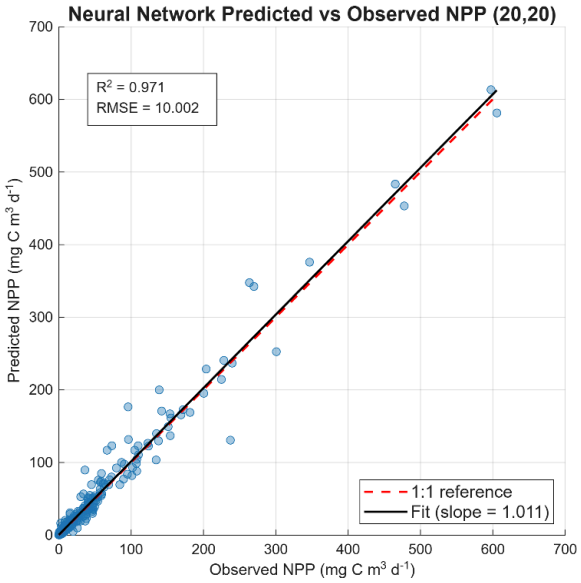
Yang, B., Fox, J., Behrenfeld, M. J., Boss, E. S., Haëntjens, N., Halsey, K. H., Emerson, S. R., & Doney, S. C. (2021). In Situ Estimates of Net Primary Production in the Western North Atlantic With Argo Profiling Floats. *Journal of Geophysical Research: Biogeosciences*, *126*(2). https://doi.org/10.1029/2020jg006116

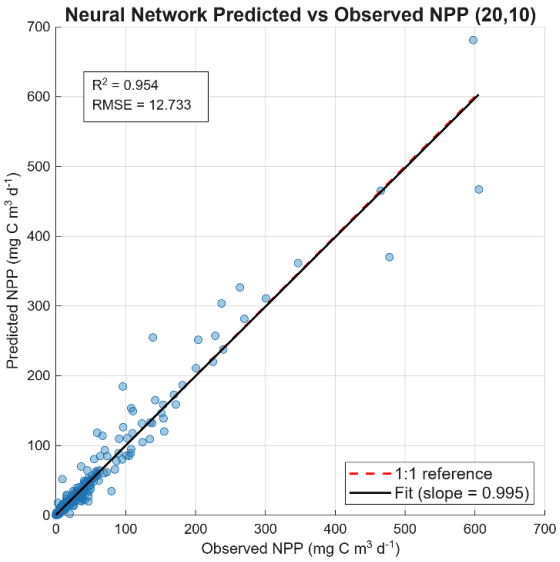
**SUPPLEMENTARY APPENDIX**

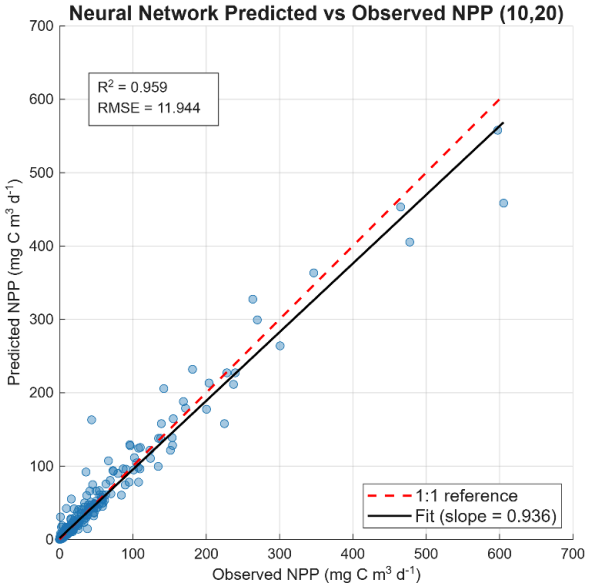
Figure 1: Observed net primary productivity for all floats

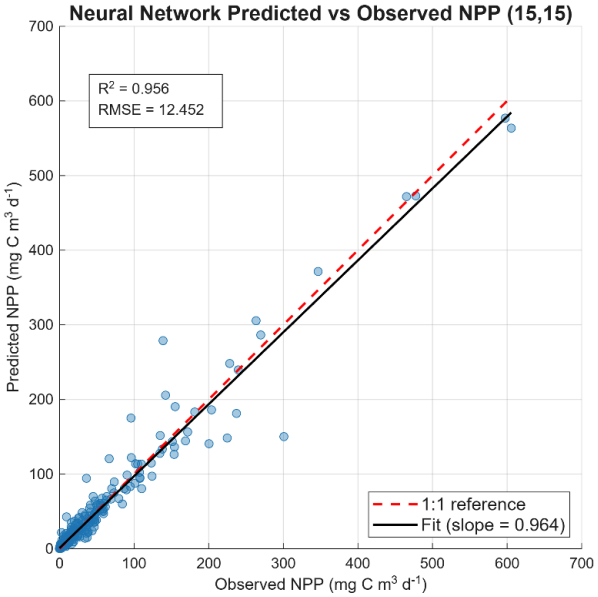



Figures 2-6: Observed vs predicted neural network NPP on the test dataset for all five runs


Neural Network Predicted vs Observed NPP (20,20)
R² = 0.971
RMSE = 10.002
1:1 reference
Fit (slope = 1.011)
Predicted NPP (mg C m³ d⁻¹)
Observed NPP (mg C m³ d⁻¹)


Neural Network Predicted vs Observed NPP (20,10)
R² = 0.954
RMSE = 12.733
1:1 reference
Fit (slope = 0.995)
Predicted NPP (mg C m³ d⁻¹)
Observed NPP (mg C m³ d⁻¹)


Neural Network Predicted vs Observed NPP (10,20)
R² = 0.959
RMSE = 11.944
1:1 reference
Fit (slope = 0.936)
Predicted NPP (mg C m³ d⁻¹)
Observed NPP (mg C m³ d⁻¹)


Neural Network Predicted vs Observed NPP (15,15)
R² = 0.956
RMSE = 12.452
1:1 reference
Fit (slope = 0.964)
Predicted NPP (mg C m³ d⁻¹)
Observed NPP (mg C m³ d⁻¹)

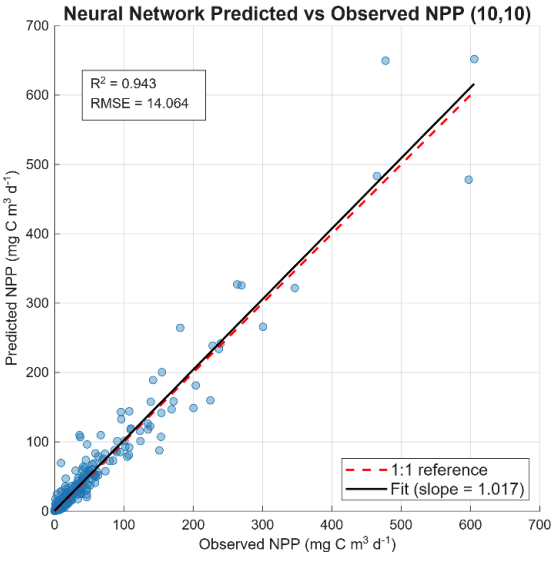

Neural Network Predicted vs Observed NPP (10,10)
R² = 0.943
RMSE = 14.064
1:1 reference
Fit (slope = 1.017)
Predicted NPP (mg C m³ d⁻¹)
Observed NPP (mg C m³ d⁻¹)

Table 1: Validation metrics for all five neural network runs

| Neurons | MSE | RMSE | $R^2$ | Slope |
|---|---|---|---|---|
| [10 10] | 197.7931 | 14.0639 | 0.9433 | 1.0168 |
| [15 15] | 155.0581 | 12.4522 | 0.9555 | 0.9636 |
| [10 20] | 142.6684 | 11.9444 | 0.9591 | 0.9360 |
| [20 10] | 162.1167 | 12.7325 | 0.9535 | 0.9950 |
| [20 20] | 100.0346 | 10.0017 | 0.9713 | 1.0108 |