Machine learning fidelity verification exposes management-relevant tradeoffs in satellite-based ocean eco-province inference

Makayla McDevitt[1*], Maike Sonnewald[1,2,3], Stephanie Dutkiewicz[4]

[1] University of California Davis, Davis, 95616, USA.

[2] University of Washington, Seattle, 98195, USA.

[3]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, 08540, USA.

[4]Massachusetts Institute of Technology, Cambridge 02139, USA.

[*]Corresponding author. Email: mrmcdevitt@ucdavis.edu

Marine ecosystems are increasingly impacted by climate change, necessitating tools to identify and predict spatial habitat information for applications, including fisheries management and carbon sequestration efforts. To build such tools, ecological marine provinces, "eco-provinces," ecologically meaningful regions in the global ocean can be used. We use unsupervised machine learning (ML) to identify eco-provinces with corresponding uncertainty measures based on output of a global simulation of phytoplankton functional types. We then develop a hierarchy of explainable dense ensemble networks to infer how well the eco-provinces can be detected from remotely observable fields. Key results include that the delineated eco-provinces are both ecologically meaningful and can be inferred with high skill. However, no straightforward relationship was found where adding more input data to the network consistently improves inference skill, and there is an intricate tradeoff between inputs and prediction fidelity. Our work is a case for optimism and a cautionary tale of needing careful validation of

prediction fidelity and uncertainty quantification.

Teaser

Ocean biomes, or phytoplankton ”eco-provinces”, are identified and predicted using machine learning workflows that allow for fidelity validation and uncertainty quantification, revealing a nuanced relationship between input data and prediction accuracy.

Introduction

As the impacts of climate change become more severe, there is an increasing need for a comprehensive characterization to represent the ecology of the global ocean. Due to the shortage of global ocean in situ data, eco-provinces will need to be objectively discovered, then inferable, or predictable, from remote data. We use an unsupervised machine learning (ML) workflow to determine eco-provinces based on modeled phytoplankton functional type (PFT) concentrations, and a supervised dense network ensemble (DNE) approach to infer eco-provinces based on remotely observable input data. Both delineation and inference of the eco-provinces utilize novel ML methodologies that allow for trustworthiness through quantification of uncertainty, validation of fidelity, and explainability of predictions.

Our work presents PFT eco-provinces that offer a comprehensive representation of the global ocean and can be inferred based on remotely observable data. When designing inference tools, a question that is pertinent with new remote sensing products is: What input data is needed to make accurate predictions? Using the fidelity verification methodology for inference, we demonstrate, using a hierarchy of ML models, what remotely observable inputs led to increases or decreases in skill. Our inference framework is a useful and trustworthy tool because it allows for prediction of ecological meaning based on remotely inferable input, combined with a prediction uncertainty measure and explanation of what inputs influenced correct and incorrect predictions.

In order to explore how our eco-province identification and prediction workflow could be potentially useful for operational efforts, we consider hypothetical fisheries management and carbon sequestration examples as case studies. Previous work demonstrated the importance of ocean

province characterizations for categorization of fishery biomass trends of exploited fish populations (*1*), and that satellite-derived phytoplankton phenology can help elucidate commercial fish survival rates (*2–5*). We consider two case studies, one considering upwelling and high-diatom eco-provinces for fisheries management, and one taking into account high-cocolithophore eco-provinces for carbon sequestration insights. We also consider two case studies of incorrect predictions, one with high uncertainty and one with low uncertainty. These case studies, discussed futher below, provide two examples for evaluating the fidelity and uncertainty of our workflow.

Our work offers a much-needed methodological advancement in identification and prediction of ecologically meaningful ocean provinces, but future work is needed before using it in real world applications.

Ocean Province Characterizations

Previous ocean province characterizations have been based on examination of near-surface chlorophyll fields and expert knowledge (*6*), taxonomic configurations and evolutionary patterns (*7*), taxonomic similarities and oceanographic processes (*8*), and marine plant and animal distributions (*9*). Due to the large and highly complicated input data required for global ocean characterizations, more recent approaches have used ML in order to parse the input in a more data-driven manner. For example, (*10*) used k-means clustering on species distribution, temperature, salinity, and nutrient data. (*11*) used a self organizing map (SOM) and hierarchical agglomerative clustering to identify "seascape" provinces in the global ocean based on satellite-derived ocean color products, sea surface temperature, and chlorophyll a. (*11*) seascapes are incorporated into the NOAA Coastwatch platform, illustrating the utility of such predictions for fisheries management. Our work stands in complement to the (*11*) seascapes, but explores identification based on PFTs.

A previous characterization that utilizes phytoplankton abundance as an input is (*12*) who identify eco-provinces based on detailed ecological model output, probabilistic projection, density based clustering and graphs. Other phytoplankton-based global ocean characterizations include (*13*) who used a SOM and hierarchical clustering to characterize the global ocean based on phytoplankton species data, (*14*) who used a co-occurance ecological network of plankton taxa to identify six phytoplankton community types based on in situ rDNA metabarcoding, then predicted them using

a support vector machine (SVM) based on satellite ocean color and temperature products. Most recently, (*15*) used a SOM and agglomerative clustering to identify spatiotemporally varying phytoplankton biomes based on satellite ocean color, chlorophyll-a, and an in situ gene dataset that reveals phytoplankton relative cell abundance. These ocean biomes reveal more detail in the sparse regions where the in situ gene dataset is located, so much of the open ocean is characterized as the same biome. Our work compliments (*15*), in that it determines global PFT-based eco-provinces with complexity in pelagic regions.

Validation in the general field of geosciences, and in particular global ocean ecological characterizations using ML, is extremely challenging. This difficulty to validate is due to the sparcity of observational data and highly complex covariance structures within both modeled and available observational data. Our approach, the Native Emergent Manifold Interrogation (NEMI) Method (*16*), is designed for intuitive validation and is an unsupervised clustering workflow for identifying meaningful structures in highly complicated and non-linear data. Described fully in the methods section, NEMI carefully accounts for nonlinear interactions in the input, quantifies uncertainty of eco-province identification using entropy, and allows for extensive external validation. Validation in the context of geospatial unsupervised learning is extremtly important, and different internal validation metrics can be misleading (*17*). NEMI allows for rigorous validation, using external metrics, and facilitates parameter selection by allowing the user to compare embeddings with different parameters in the 3D space. Our workflow allows us to choose any number of eco-provinces based on the desired application. For this paper, we consider the example of 10 eco-provinces as a starting point before considering higher complexity cases (i.e. a higher number of eco-provinces).

Prediction of Eco-Provinces Based on Remotely Sensed Input Data

Our work aims to establish an initial proof of concept for predicting eco-provinces based on remotely observable data. In our work, we use modeled input data to identify what data, and how much, is useful for eco-province inference. This work paves the way for using remotely sensed satellite data such as NASA's PACE satellite for example, which includes ocean color, chlorophyll, and phytoplankton products.

A goal of this paper is to lay foundations and provides a proof of concept for remotely inferring

eco-provinces based on real world remotely sensed data, such as from ocean color satellites. As a test case, we use phytoplankton functional group biomass from a computer simulation (see methods) as input to characterize eco-provinces, and modeled ocean color data to explore how well remote products can predict those provinces.

Inferring eco-systems remotely remains a formidable challenge. We demonstrate that in our case this remote inference can be successful accomplished, and provide a detailed analysis of the various tradeoffs associated with the number of inputs needed. In order to predict eco-provinces based on remotely sensed fields, we explore the use of deep learning first in terms of what machinery is required for optimal network skill, a robust uncertainty quantification, and transparency in what inputs influenced a particular prediction to ensure fidelity of predictions, following the framework in (*18*). We optimize the predictive skill and uncertainty quantification by creating an ensemble of neural networks all dedicated to the same task (see methods), then average their results to increase robustness following (*19, 20*). To prioritize transparency in network decisions, we use eXplainable Artificial Intelligence (XAI, see methods), described in (*21, 22*). XAI allows for transparency in network decisions, addressing the concern of neural network inference methodology acting as a "black box." This additional step supporting trustworthy decision making is necessary when using neural networks to make high-stakes management decisions. In addition, we tested what type and how much information is needed for our dense network ensemble (DNE) to function by considering which inputs influenced correct or incorrect predictions (see methods). We consider different versions of the DNE with varying inputs with the rationale of starting simple with fewer inputs, then assessing any improvements from adding input fields.

Results

Our ML inputs are from the Darwin Biogeochemistry Model (*23*) (see methods) which is a marine biogeochemical-ecosystem numerical model that includes an explicit radiative transfer module, and hence able to provide reflected irradiances (similar to what a satellite might sense), nutrients, chlorophyll, and over 50 types of phytoplankton which we group into 5 PFTs, pico-prokayrotes, pico-eukaryotes, coccolithophers, diatoms and mixotrophic dinoflagellates (see methods). We use PFTs for eco-province identification (Fig. 1a) and input similar to what satellite sensors estimate (blue

reflectance, green reflectance, chlorophyll, and phytoplankton carbon biomass) for eco-province prediction (Fig. 1b). Eco-provice identification uses the NEMI clustering algorithm, based on boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF) phytoplankton functional type biomass estimates averaged across 1993-2015.

Seasonal Eco-Province Identification

Fig. 2a illustrates the NEMI eco-provinces. Each point corresponds to a spatial grid cell (latitude and longitude) on the JJA or DJF plot, and the color represents the eco-province label. The eco-province assignment for each grid cell corresponds to an uncertainty percentage quantified by entropy, presented in Fig. 2b. The final eco-provinces were chosen based on the ensemble member that resulted in the least uncertainty. As the color bar indicates, dark blue regions represent eco-provinces with lower uncertainty compared to lighter blue and green regions. Here, we chose the number of eco-provinces to be 10, but any number could be chosen based on the desired application. Note that the number of points presented in Fig. 2 is randomly subsampled for visualization.

As each point in Fig. 2a represents one location, we can project these from the embedding space back onto geographical space. Fig. 2c and Fig. 2d present the eco-provinces plotted geospatially, for averaged JJA (Fig. 2c) and boreal winter DJF (Fig. 2d). The same eco-provinces are identified in both JJA and DJF highlighting a seasonal shift in eco-province location, especially in higher latitudes. Although expected, the nuances in eco-province grouping can not be pre-determined without the unsupervised NEMI methodology. For example, just viewing the phytoplankton carbon or chlorophyll biomass in Fig. 1b would not indicate the nuances in the Arctic in eco-provinces 8 (dark purple) and 9 (light purple), or the split of eco-province 2 between northern and southern latitudes in JJA.

While eco-provines in the lower latitudes remain more consistent across the seasons, clusters in high latitudes are more variable. Since eco-province 3 (light orange) exists in low nutrient regions, we will refer to it as the oligotrophic eco-province. Similarly, since eco-province 1 (light blue) serves as a thin transition between two larger eco-provinces, we will refer to it as the border eco-province. Eco-provinces 8 (dark purple) and 9 (light purple) are located in the Arctic, so we will refer to them as the Arctic provinces. These eco-provinces (oligotrophic, border, and Arctic) do

not change very noticeably depending on the season. In contrast, eco-provinces 2 (dark orange), 4 (dark green), 5 (light green), 6 (red), and 7 (salmon) shift from high southern to northern (or vise versa) depending on the season. Eco-province 2 characterizes high latitudes, while eco-province 5 encompasses regions with high diatom biomass. Note that eco-provinces 0 (encompassing coastal and upwelling regions, dark blue), 5 (light green) and 6 (red), for the most part tend towards northern latitudes in JJA and southern latitudes in DJF. In contrast, eco-provinces 2, 4, and 7 exhibit the opposite pattern, tending towards southern latitudes in JJA.

Fig. 2e and Fig. 2f show the corresponding JJA and DJF entropy percentages plotted spatially. Note that lighter blue and green regions have higher uncertainty than dark blue regions. For example, eco-provinces 1 and 6 have more uncertainty than eco-province 3. In general, the uncertainty is higher between clusters, which is expected. The core regions of each cluster are less variable, but the borders vary slightly on each run of the algorithm. This uncertainty quantification is important since it allows us to map an uncertainty percentage to each point in the ocean to its specific eco-province. Note that since the cluster assignments and corresponding uncertainty quantifications are for each grid cell individually, the certainty can vary across one cluster. In general, high uncertainty regions are those where the DNE struggles to make predictions.

We compare the clustering results with the original modelled PFT biomass (Fig. 2g) in order to gauge the ecological fidelity of the results. There are more diatoms, mixotrophic dinoflagellates, picoeukaryotes, and pico-prokaryotes in cluster 5 and more coccolithophores and pico-prokaryotes in cluster 0. The functional types most abundant in cluster 5 have a strong seasonal component where they exist in the mid to high northern latitudes in JJA, then mid to high southern latitudes in DJF. The functional types most abundant in cluster 0 exist surrounding continents and oligotrophic gyres, with a slight seasonal polarity that emphasizes a strong presence in the North Atlantic and Pacific during JJA.

The transition zones between eco-provinces have higher uncertainty compared to larger, core eco-provinces. Some general eco-province structures can be explained by physics-based knowledge. For example, the Southern Ocean is broken up into two provinces during both JJA and DJF, due to ACC (Figs. 2 a,b). The oligotrophic gyres, which are strongly influenced by global wind patterns, are characterized by eco-province 3 in both JJA and DJF. General province structures, most noticeably the horizontal v shape surrounding the Eastern Tropical Pacific, align with large-scale “planetary

waves" caused by Earth's rotation. However, the eco-provinces can not be expected from physical knowledge alone. For example, the location of the low nutrient oligatrphic eco-province does not align with expected gyre structure in the North Atlantic.

Seasonal Eco-Province Inference

Having defined the eco-provinces from PFT biomass input, we now consider how well remotely sensed products could be used to infer the provinces. Our work considers three versions (V1, V2, and V3) of the DNE, with different input features as specified in Fig. 1b. V1 includes blue and green irradiance reflectance, a measure of ocean color, as input. V2 also includes total chlorophyll as an input. V3 additionally incudes of total phytoplankton carbon biomass. These inputs (reflectance , Chl, carbon biomass) are typical satellite products. Note that Chl and Carbon biomass are determined from the reflectance, so these fields are correlated (*24*). Fig. 3 outlines the general inference and explainability workflow, which includes the process of DNE prediction with uncertainty and explaibability for each grid cell.

Fig. 4 displays the V1 predicted 2005 eco-provinces (after training on 1993-2004 input data) with corresponding true false values and entropy. The true false values indicate correct and incorrect network predictions, respectively. Prediction entropy is a measure of prediction certainty, and is calculated differently than eco-province identification entropy (see methods). In Fig. 4c and Fig. 4d, the correct DNE predictions are displayed in green compared to the incorrect predictions in purple. Note the incorrect predictions in the Arctic, Western Pacific, and in between eco-provinces. Fig. 4e and Fig. 4f display the uncertainty of the prediction quantified as entropy, where darker regions have less uncertainty. We find interesting deviations from the correspondence of skill and prediction uncertainty, where the DNE was certain the prediction was correct, but it was wrong. The incorrect prediction of the boreal summer (JJA) Western Pacific has fairly low uncertainty, which is unexpected for an incorrect prediction. We discuss the low uncertainty incorrect prediction of the Western Pacific later as a case study of a cautionary tale that emphasizes the need for prediction explainability and uncertainty quantification.

Figs. 5 and 6 are analogous to Fig. 4, but for network versions 2 and 3. When we add chlorophyll as an input, the prediction improves in the JJA North Atlantic, but degrades in JJA near the top of

the Southern Ocean and Western Pacific (Fig. 4). Explanations for why the prediction degraded in these locations are explored through the use of XAI below. The incorrect prediction near the top of the Southern Ocean has high uncertainty, which is expected of an incorrect prediction. Adding phytoplankton carbon biomass as an input further improves the prediction in the JJA Arctic, but the incorrect predictions of the JJA top of the Southern Ocean and Western Pacific remain.

The DNE predicts core ocean provinces (Table S1) such as the oligotrophic gyres, Southern Ocean, and upwelling provinces with more accuracy (96%, 97%, 87% respectively) compared to border provinces such as those surrounding the oligotrophic gyres and in between the Arctic and North Atlantic (78% and 63% respectively). Better predictability of core regions aligns with expectations, as they do not change as frequently in space and time, and were initially identified with higher certainty.

Comparing all three versions of the DNE in terms of the overall accuracy, more input variables lead to better performance, but with interesting deviations. The addition of the extra inputs (going from V1 to V2 to V3) increases the maximum training accuracy from 0.78 to 0.85 and 0.88, respectively. The most improvement is observed in the prediction of the JJA North Atlantic and JJA Arctic with increasing versions. Across the three versions, the maximum validation accuracy increases from 0.67 to 0.81 and 0.85. Furthermore, the respective minimum training and validation losses decrease from 0.53 and 0.82 for V1, to 0.37 and 0.49 for V2, to 0.30 and 0.37 for V3. Overall, there are more occurrences of false predictions on the boundaries compared to the cores of each province. Accordingly, the prediction entropy is higher between provinces. The entropy distributions are lower for correctly classified grid cells compared to incorrectly classified grid cells. From a DNE performance perspective, V3 is desirable compared to the others. The DNE's prediction accuracy increases with the addition of chlorophyll and phytoplankton carbon bioamss as inputs because additional inputs have an increased variety of maxima and minima values to distinguish distinct provinces.

In our evaluation, we match skill with uncertainty, where largely a poor prediction has high uncertainty, but this is not always the case. The DNE predicts core provinces such as the oligotrophic gyres and areas in the Southern Ocean with much more accuracy compared to border provinces such as cluster 1 where it surrounds the oligotrophic gyres, across all versions.

Our results strongly indicate that the best version of input data depends on the user's region

of interest. Fig. S3 displays the 3 compositions of DNE inputs in each eco-province, so that we can determine whether high or low concentrations of each input contributed to correct or incorrect predictions. In general, either high or low input concentrations are more informative for DNE predictions compared to moderate concentrations.

Inference Fidelity Assessment Across Three Versions of DNE

SHapley Additive exPlanations (SHAP) is an XAI technique that uses a game theoretic approach to computing the contribution of each input to a prediction (*25*). SHAP Values reveal the importance of input features (Figs. 7-10, feature importance is color coded and separated based on whether the DNE's prediction was correct or not). In the correctly predicted eco-provinces, blue represents low SHAP values where the input decreases the likelihood of the correct prediction, compared to red which indicates high SHAP values which increase the likelihood of the correct prediction. Similarly, the incorrectly predicted regions have low SHAP values in green, going through orange as neutral, and reaching purple as high SHAP values. The correctly predicted eco-provinces are outlined in black, while the incorrectly predicted regions are outlined in light yellow. We obtained the SHAP values for all eco-provinces and inputs, but for conciseness are highlighting eco-provinces 5, 0, and 7 (Figs. 7,8) as examples that may have potential implications for fisheries management and carbon sequestration. We also consider eco-provinces 3 and 6 (Figs. 9, 10) as case studies of incorrect predictions with low and high uncertainty, respectively.

Fig. 7 displays SHAP values for blue and green reflectance, including correct and incorrect predictions plotted together, for V1, V2, and V3, for eco-provinces 0, 5, and 7. Fig. 7 reveals how the importance of blue and green reflectance decreases across versions with the addition of chlorophyll and phytoplankton carbon biomass.

Comparing the ocean color inputs in Fig. 7, blue reflectance is more important for prediction on the boundary of eco-province 0, compared to green reflectance which is more influential in the center of the eco-province. Considering the addition of chlorophyll and phytoplankton carbon biomass inputs, chlorophyll is most important for predicting eco-province 0, as revealed by the SHAP values displayed in Fig. 8. Also note the high prediction accuracy of eco-province 0 corresponds to high levels of all V3 inputs (Fig. S3). Note that chlorophyll is more important in some cases for

predicting the inside region of eco-provine 0, compared to phytoplankton carbon biomass which is more important on the outside of the eco-province.

Considering all of the inputs, both blue and green reflectance are most influential for predicting eco-province 5 in V1 (Fig. 7) and chlorophyll is most influential in V's 2 and 3. The high prediction accuracy of eco-province 5 corresponds in Fig. S3 to high levels of all inputs except for blue reflectance. Green reflectance is most influential for predicting eco-province 7 in V1 (Fig. 7), while chlorophyll is most influential in V's 2 and 3. Fig. S3 reveals that eco-province 7 has medium levels of blue reflectance and fairly high levels of the other inputs. Fig. 9 displays the JJA and DJF SHAP values for Eco-Province 3 for versions 2 and 3, including blue and green reflectance inputs. Comparing all inputs, both blue and green reflectance, but especially blue reflectance, are influential for the incorrect prediction of the JJA Western Pacific eco-province 3, with decreasing importance across versions with the addition of chlorophyll and phytoplankton carbon biomass as inputs. (Fig. 9). Figs. 9 and 10 demonstrate the high importance of chlorophyll in the V2 and phytoplankton carbon biomass in the V3 incorrect prediction of the Southern Ocean. Fig. S3 demonstrates medium to high blue reflectance, and medium green reflectance feeding into the JJA eco-province 6 prediction.

Overall SHAP results demonstrate that the difference in prediction accuracy and uncertainty in core vs. border eco-provinces may relate to these eco-provinces containing either very high or very low input concentrations, which is more meaningful for the DNE to distinguish meaningful eco-provinces compared to less extreme input values. Overall SHAP results reveal that high blue reflectance, low green reflectance, low plankton, and low chlorophyll increase the likelihood of low plankton eco-provinces, such as the oligotrophic eco-province, being predicted. Similarly, low blue reflectance, high green reflectance, high plankton, and high chlorophyll help predict high plankton eco-provinces such as coastal upwelling provinces, the Eastern Tropical Pacific, DJF Southern Ocean, and JJA north pacific. Furthermore, these border regions had more uncertainty during the identification phase, so the DNE having more uncertainty predicting these regions confirms expectations.

Discussion

We identify eco-provinces on a global 1/2 degree spatial grid for boreal summer (JJA) and boreal winter (DJF) averages spanning 1993-2015 from model output of phytoplankton biomass. We determine how well satellite products may be able to predict these eco-provinces by inferring them based on remotely observable inputs. We predict the eco-provinces using a series of DNEs, assessing how prediction accuracy and learning strategies are shaped by the inputs, training on 1993-2004 input data and then predicting 2005 global eco-provinces. To delve into the root of correct and incorrect predictions across the three versions, we use XAI to reveal input influence for the case study examples with potential implications for fisheries management, carbon sequestration, and incorrect predictions. This framework, demonstrated in Fig. 3, aligns with the general workflow demonstrated in (*18*) as a quantitative pathway to demonstrate fidelity. As we increase the number of inputs, overall accuracy increases, but the accuracy in particular regions decreases. This result aligns with expectations, because increasing input information the DNE uses has the potential benefit of increasing the context necessary for predictions, but it can also increase input data uncertainty. We challenge and test the assumption that "more input data is better," identifying spatially which predictions improve as a result of adding more inputs. Insight from asking such questions guides our recommendations on how the DNE should be used based on different goals and in different geographical provinces.

Identified Eco-Provinces

The identified eco-provinces exhibit a detailed and ecologically meaningful characterization of the global ocean. As presented in Fig. 2, our eco-provinces provide a comprehensive characterization of the open ocean, with very high certainty in larger and core provinces such as the upwelling eco-province 0, the oligotrophic eco-province 3, and the provinces that encompass the Southern Ocean such as eco-provinces 5 and 6. Our eco-provinces offer a comprehensive delineation of global ocean plankton ecology, and thus can be used to compliment past characterizations such as (*11*) Seascapes and the (*15*) Ocean Biomes, that reveal more detail in coastal rather than pelagic regions. Furthermore, since our eco-provinces are based on phytoplankton functional type estimates on a 1/2 by 1/2 degree global ocean grid, their visible structure reveals the nuanced behavior representing

the fluid boundaries and ever-changing nature of the global ocean. Our curved and intricate eco-provinces are very similar to those presented by (*12*), but on a higher spatial resolution from 1 by 1 degree, and including a seasonal rather than annual average. Similarly, our eco-provinces are higher in resolution and complexity compared to the eight 1 by 1 degree biomes presented in (*13*) and the six phytoplankton communities in (*14*). Since our eco-provinces reveal a high level of ecologically informed detail in both coastal and pelagic regions while also encompassing seasonal variability, they provide potentially useful information for informing fisheries management and carbon sequestration efforts.

Eco-Province Identification Methodology

Our identification and prediction methodology is distinct from previous approaches such as (*6, 11–14*). Our use of PFT input offers a simplified approach to (*12*) who also considered zooplankton and nutrient flux data. The current work, (*12*), and (*13*) prioritize the ecological information gained from phytoplankton inputs, compared to (*11*), who used satellite sea surface temperature, ocean color, and chlorophyll a for an approach that is directly compatible with satellite data. (*6*) also utilized satellite chlorophyll data, combined with decades of observational experience, in order to define global ocean provinces. We used the ML NEMI workflow to identify eco-provinces, which reduces the dimensionality of the seasonal phytoplankton functional type inputs using UMAP (*26*). UMAP is preferred to the probabilistic approach t-SNE used in (*12*), since the categorical cross entropy used to create the embedding in the low dimensional representation preserves both local and global, rather than just local, structures in the data. In addition, UMAP has been shown to have better computational performance compared to t-SNE (*17, 26*). (*11, 13*) utilized a SOM for dimension reduction, which is a two layer neural network that uses competitive learning to map the input to a discrete grid (*27*). While SOMs are useful for mapping onto a fixed grid while prioritizing local topology, UMAP is preferred for capturing global structures with more efficiency when working with large datasets.

For clustering, the NEMI workflow, in addition to (*11, 13*), uses agglomerative clustering where the user sets the desired number of clusters. The use of agglomerative clustering is desired compared to density-based clustering in (*12*), which does not allow the user to set the number of

clusters. As a result, the approach presented in (*12*) requires additional aggregation using the Bray-Curtis dissimilarity metric (*28*). In addition, a comprehensive analysis of clustering algorithms has demonstrated that hierarchical clustering can better grapple with outlier sensitivity compared to DBSCAN (*29*).

(*14, 15*) used machine learning algorithms involving in situ genetic data to identify ocean biomes, alongside satellite data, for global ocean biome characterization. Our work is similar in that it includes a phytoplankton component, but from modeled PFT data, which allows us to characterize the global ocean, including pelagic regions, with higher detail and complexity.

After dimension reduction and clustering in our eco-province identification methodology, the NEMI workflow increases robustness of province identification by quantifying the uncertainty using entropy. Entropy quantifies the uncertainty of each cluster assignment for each grid cell through ensembling, which allows us to choose the final eco-provinces based on the ensemble with the parameter combinations offering the least uncertainty. (*12*) also addressed stochasticity using ensembling, but our use of entropy allows a more nuanced interpretation proportional to the number of eco-regions. Furthermore, our approach identifies the functional type composition within each eco-province (Fig. 2g) which is an advancement compared to previous characterizations that do not consider PFTs. Overall, our eco-province identification approach utilizes advanced ML methodology and detailed ecological input in order to create a unique global ocean characterization, complimenting past work.

Eco-Province Inference

After the eco-province identification step, we predict the eco-provinces based on input data that can be remotely sensed. (*30*)'s comprehensive review on ensemble deep learning demonstrates the advantage of ensemble and deep learning compared to traditional algorithms (*30*). In addition, (*31*) discussed using ensemble methods to improve robustness of deep learning for image classification in marine environments, finding emsembles optimal in terms of robust uncertainty quantification and concluding that ensembles should be the standard for using deep learning for benthic image automation (*31*). To add another level of interpretability to our results, we utilize the XAI technique SHAP, presented in (*25*), in order to understand what inputs were important for correct and

incorrect predictions of eco-provinces of interest. Using XAI for domain specific validation has been demonstrated to be insightful in (*19*). Overall, our inference methodology combining ensembled predictions, uncertainty quantification, and explainability, poses as a useful framework to build off of for eventual predictions in the real ocean.

Our approach of predicting PFT-based eco-provinces based on remotely sensed input data balances the trade off between satellite compatibility and identifying the complex ecological structures present throughout the open ocean. We take inspiration from (*11, 14, 15*)'s compatibility with satellite data, but add more complexity to the characterization of the open ocean at all latitudes.

Inference Method Fidelity Comparison

Remotely sensed phytoplankton carbon biomass has more uncertainty than chlorophyll biomass, which has more uncertainty than blue and green reflectance (*32*). The uncertainty in chlorophyll and phytoplankton carbon biomass are a result of parameter estimates in their calculation; chlorophyll being calculated using a blue green light reflectance ratio and phytoplankton carbon biomass being derived from an empirical relationship to the particle backscattering properties of water (*33*). As we added inputs across the three DNE versions, overall accuracy increased. Most notable is the improved prediction of the North Atlantic and Arctic. In particular, going from V2 to V3 with the addition of phytoplankton carbon biomass, the prediction of the Arctic greatly improves. However, some regional locations decreased in accuracy with the addition of inputs. For example, across the three versions the prediction of the JJA Western Pacific and JJA Southern Ocean deteriorates. These examples are discussed below.

Case Studies

A primary application of our work is to characterize the open ocean for potential future use in fisheries management. Previous work has demonstrated the benefit of global ocean characterizations for fisheries management efforts. For example, (*1*) utilized (*7*)'s Marine Ecoregions to categorize fishery biomass trends of exploited fish (*1*). In addition, (*2–5*) have demonstrated that satellite-derived phytoplankton phenology can help elucidate commercial fish survival rates. Our eco-provinces are uniquely useful in that they offer a visually and ecologically detailed characterization

of the global ocean, with uncertainty measures associated with identification and prediction. In our work, eco-provinces 0 and 5 are potential eco-provinces of interest for fisheries management stakeholders, since eco-province 0 encompasses coastal upwelling regions, the Eastern Tropical Pacific, and the North Atlantic. Many of the world's fisheries are located in coastal regions, and do well in regions with high fish abundance. The explicit connection between upwelling regions and fish has been demonstrated in (*34, 35*). Eco-province 5 has the largest proportion of diatoms, which are important for food web and energy transport (*36, 37*). From a fisheries management perspecitive it is useful to know how how well and with how much certainty these provinces can be predicted. Figs. 6 and 7 compare the SHAP values across all three versions and inputs, where blue and green reflectance decrease in importance with the addition of chlorophyll and phytoplankton carbon biomass. Overall, chlorophyll is most influential for the prediction of these eco-provinces. The prediction accuracies of eco-provinces 0 and 5 increase going from V2 to V3 with the addition of phytoplankton carbon biomass. Since both versions contain chlorophyll, V2 could be used to prioritize lower input uncertainty while V3 could be used to prioritize prediction accuracy.

Previous work has demonstrated the need for marine conservation efforts to mitigate the effects of climate change. (*38*) emphasized the importance of marine ecosystems and their associated biodiversity for sustaining life on Earth. In addition to maintaining carbon cycles, marine ecosystem services are important for maintaining global oxygen cycles, producing food and energy, and sustaining human wellbeing (*38*). Here, we focus on how eco-provinces could be useful for management efforts supporting, or better understanding, carbon sequestration, since carbon sequestration mitigates climate change (*39*). The importance of coccolithophores for carbon sequestration is demonstrated by (*40, 41*), as coccolithophores affect the carbon chemistry of the ocean through the production and sinking of their calcium carbonate shells (*40*). Figs. 6 and 7 can be analyzed from the lens of identifying eco-provinces with high carbon sequestration potential, since there are the highest concentrations of coccolithophores in eco-provinces 0 and 7. Management efforts could focus on protecting such eco-provinces in order to support carbon sequestration, while choosing the DNE that prioritizes important features while taking into account input uncertainty. Similarly to the fisheries management case, Versions 2 and 3 would both be useful as they include chlorophyll, where V2 has less input uncertainty and V3 has more output accuracy.

Incorrect Predictions

Figs. 8 and 9 display the importance of medium input values to the incorrect prediction of the JJA eco-province 6 in the Southern Ocean. Blue and green reflectance are influential for this prediction, decreasing in importance across versions with chlorophyll being the most important in V2 and phytoplankton carbon biomass being the most important in V3. This incorrect prediction may relate to the input values being more mid range and close in value to those in the Arctic which prevents the DNE from distinguishing the regions. This incorrect prediction corresponds to high prediction entropy, which is expected for an incorrect prediction. For applications geared toward making predictions in the JJA Southern Ocean, V1 of the DNE would be most useful.

A cautionary tale is in the incorrect prediction of eco-province 1 as eco-province 3 in the JJA Western Pacific. Figs. 9 and 10 reveal the role of all inputs to this incorrect prediction, most notably chlorophyll. It is important to note that in a real world situation, there is more uncertainty in satellite-derived chlorophyll and phytoplankton carbon biomass since they are estimated based on ocean color. These increased uncertainties may complicate real-world results. In the case of our prediction using modeled data, the incorrect prediction may relate to the fact that the NEMI algorithm labeled that region as eco-province 3 in DJF but as eco-province 1 in JJA. The low uncertainty in the prediction is a reminder to keep some skepticism when using mathematical tools such as artificial intelligence, and emphasizes the importance of transparency methods such as XAI in order to identify any potential issues. V1 is best suited for applications in the JJA western Pacific.

Shortcomings of numerical models of phytoplankton

Since we used modeled PFT concentrations in order to identify the eco-provinces, our work serves as a conceptual framework that can be extended to be applicable in the real ocean. Although the Darwin Model is validated on in situ data, there still large assumptions made throughout the model equations and uncertainties in the parameters, as would be the case when using any numerical model. The use of modeled data provides high spatial and temporal resolution of how PFTs vary throughout the global ocean, which would be impossible to obtain from sparse and disjoint in situ datasets. Futhermore the model does not perform as well near the coasts as it does not account for finer resolution coastal dynamics. The model only captures a subset of the full range of physical

and chemical forcings of the in situ ocean, and only a portion of the diversity of phytoplankton types.

Despite the caveats, our approach is necessary and valuable as a proof of concept. Future work could begin to access what amount of in situ data would be needed for such classification in the real ocean, and could address the level of uncertainty from newer satellite algorithms for estimating phytoplankton types from space from hypersopectral sensors (e.g. NASA PACE mission) was needed to use such inputs. Furthermore, our ability to infer these PFT-based eco-provinces based on remotely sensed data is especially substantial in the effort to capture ecological information from satellite data alone. We have demonstrated success predicting eco-provinces based on blue and green reflectance, and will have greater potential when using higher spectral resolution. Additional optical data will be especially useful for inferring border eco-provinces, and as we increase the number of eco-provinces identified by NEMI.

Conclusion and future work

Here, we have established a proof of concept for identifying global eco-provinces using NEMI, then predicting them based on modeled ocean color input data. Here we have specified 10 provinces, but the number can be changed according to the desired application. XAI reveals the importance of each remotely observable input for the correct and incorrect predictions of each eco-province. In this paper, we identify the importance of blue reflectance, green reflectance, chlorophyll, and phytoplankton carbon biomass for predicting eco-provinces that could be of interest for stakeholders such as fishery management and carbon sequestration. In addition, we delve into possible explanations for an incorrect prediction with low, compared to high, uncertainty. Future work will focus on increasing temporal resolution, in order to identify and predict eco-provinces on finer time and space scales. The temporal resolution will be increased to monthly averages in order to be more applicable to management efforts that change frequently according to the season. For increasing spatial resolution, we will consider the updated Darwin Biogeochemistry Model of Ocean color, which includes 13 ocean color wavebands at a 1/6th degree resolution rather than just blue and green reflectance at a 1/2 degree resolution. These additional inputs at a higher resolution will offer improved predictive capability. A possible expanded DNE framework could include inputs from

PACE satellite data which encompasses ocean color hyperspectrally, along with chlorophyll, phytoplankton carbon biomass products, and potentially estimates of phytoplankton functional types. Our study provides a framework for eco-province identification and prediction that will be useful in developing such products for real ocean management.

Methodology

Darwin Model

This paper presents eco-province identification based on PFT concentrations estimated from the Darwin Biogeochemistry model (*23*) which is based on differential equations that follow the movement of matter through inorganic nutrient, living and dead organic matter and remineralization of detritus. The model also captures dynamically evolving Chl distributions, as well as the absorption and scattering of three streams of light on optically important constituents in the ocean. The model outputs irradiance reflectance which is a measure of ocean color, similar to what is captured by satellite sensors, and is divided into wavebands across the visible light spectrum. The model output spans 1993-2011, and in this study we consider phytoplankton biomass, chlorophyll, and irradiance reflectance values for the 425 nm (blue reflectance) and 550 nm (green reflectance) wavebands.

Phytoplankton growth is parameterized as dependent on temperature, irradiance, and nutrients. The model captures 50 types of plankton that differ in terms of trophic strategy, functional roles and size (ranging from 0.6 to over 1000 micrometers in equivalent spherical diameter). Here we focus on groups of phytoplankton known as "phytoplankton functional types" (PFTs): diatoms (that require silca), coccolithophores (that calficy), mixotrophic dinoflagellates (which graze as well as photosynthesize), picoeukaryotes, and pico-prokaryotes (which are well adapted to low nutrient conditions).

All model output used here consists of 3-day averaged values on a 1/2 by 1/2 degree (about 50km) spatial grid. Here we only use the surface (0-10m) values, though the model includes 50 levels ranging in resolution of 10m at the surface to 100m at depth. To analyze general seasonal changes, we considered JJA and DJF averaged datasets across the time period of the dataset.

The model captures the higher concentration of phytoplankton carbon biomass in northern lati-

tudes JJA, compared to southern latitudes in DJF due to warmer weather and upwelling supporting phytoplankton blooms. Fig. 1 presents a) JJA and b) DJF diatoms thrive in the North Pacific in JJA and Southern Ocean in DJF, but are in lower abundance in the ologtrophic regions due to their higher nutrient and silica requirements. Coccolithophores are most abundant in the North Atlantic in JJA (Fig. 1c) and mid- to high southern latitudes (the calcite belt, Balch et al REF) in DJF (Fig. 1d). Mixotrophic dinoflagellates are abundant in mid- to high northern latitudes in JJA (Fig. 1e) and mid- to high southern latitudes in DJF (Fig. 1f). Picoeukaryotes and picoprokaryotes follow similar seasonal patterns, with greater persistence in oligotrophic waters due to their small size and higher nutrient affinity (Figs. 1g-j).

Eco-Province identification using NEMI

Eco-provinces are identified using the "Native Emergent Manifold Interrogation" (NEMI) method, developed by (*16*) to determine regions of interest in large or highly complicated and nonlinear input data. The first step of NEMI is UMAP, "Uniform Manifold Approximation and Projection for Dimension Reduction," (*26*) a general non-linear dimensionality reduction technique that constructs a simplified, three-dimensional representation of the five dimensional input called an "embedding." Each of the original dimensions of the input dataset corresponds to each of the five considered phytoplankton functional types. The embedding is important, because it offers a three dimensional representation, which can be easily visualized. The UMAP dimension reduction technique is sophisticated in its minimization of categorical cross entropy between the high and low dimensional spaces, resulting in a locally and globally balanced embedding (*26*). Furthermore, each point in the embedding represents one point in geographical space, so no information is lost in the process of dimension reduction. On this simplified representation, NEMI uses hierarchical agglomerative clustering to identify meaningful areas of interest in the low-dimensional representation. Step two of NEMI addresses the sensitivity of clustering results to noise and uncertainty through the use of an ensemble methodology. Each iteration, which we define as one singlular run of the NEMI dimension reduction and clustering workflow with the same parameters, is repeated several times in order to create a group or "ensemble" of results. Each ensemble member has its own cluster assignments for each original data point. Entropy quantifies the uncertainty of whether each

original input point belongs to the assigned cluster. In information theory, the entropy of a random variable measures the uncertainty of the variable's potential states. For each sample i, the entropy $H_i = -\sum_{j=1}^{N_I} p_{ij} log(p_{ij})$, where N_I is the number of possible outcomes and p_{ij} is the probability that outcome j occurs for each sample (*42*). In the case of NEMI, we are determining entropy for each datapoint i by comparing the cluster assignments across each member of the ensemble. Therefore p_{ij} is the proportion of times point i is assigned to cluster j across the ensemble. The final eco-provinces are chosen based on the ensemble member that has the least entropy. Thus each point in the global ocean is mapped to a specific eco-province, alongside a quantitative measure of uncertainty.

Unlike commonly used dimension reduction and clustering techniques such as Principle Component Analysis and k-Means, NEMI does not have strict assumptions about the linearity or spherical clusterability of the underlying input data. NEMI performs well with large and highly elaborate and nonlinear data, and is less parametric compared to other methods. Compared to (*12*), NEMI has fewer parameters to tune and therefore less risk of noise interference. Due to the straightforward nature of the algorithm, NEMI is intuitive to use for global applications (*16*). Since NEMI allows the user to set the number of clusters, a desirable level of complexity can be achieved for the application of interest.

DNE for Eco-Province Inference Based on Ocean Color Data

After eco-province identification, we test how well we can predict eco-provinces from remotely-sensed products. We using a dense ensemble multilayer perceptron (MLP) on Darwin Model output that is simular to products from satellite sensors (reflectance irradiance, total chlorophyll and total phytoplankton carbon biomass). There is some stochasticity involved in training a neural network due to random weight initialization, random mini-batch (subset) sampling, and Adam optimization which is an extension of stochastic gradient descent. In addition, there is a risk of overfitting or becoming stuck in local minima of the loss function. To address this stochasticity, we use a group, "ensemble," of 10 neural networks dedicated to the same prediction task. Each neural network uses categorical cross entropy for the loss function. The final prediction is a mean of the ensemble's results, which is much more robust than a prediction from a single neural network because it

averages out randomness, reduces overfitting, and allows uncertainty estimation. If all ensemble members agree on the prediction, the final result is more reliable. Average ensemble entropy quantifies the uncertainty of the ensemble's results, which is represented in Fig. 3. Our code for ensemble prediction and uncertainty quantification follows that presented in (*43*). We further assess the impact on ensembling using XAI, as described below. The combination of using a network ensemble for inference and XAI for interpretation increases the fidelity of our workflow (*18*).

We utilize three versions of the DNE in order to consider varying levels of uncertainty in our input fields. V1, which has the least input uncertainty, only considers blue (Figs. 1k and 1l) and green (Figs. 1m and 1n) irradiance reflectance (ocean color) inputs. Figs. 1k and 1l indicates that there is more blue reflectance in the oligotrophic gyres due to the low abundance of phytoplankton carbon biomass in that region. In contrast, there is more green reflectance in high phytoplankton biomass regions such as upwelling regions mid to high northern latitudes in JJA (Fig. 1m), and Southern Ocean in DJF (Fig. 1n). V2 considers ocean color and total chlorophyll (Figs. 1o and 1p) measured in $\frac{\text{mg Chl}}{m^3}$. Chlorophyll follows the seasonal pattern noted previously. V3 includes the addition of total phytoplankton biomass (Figs. 1q and 1r) measured in $\frac{\text{mmol Carbon}}{m^3}$. Although there are some differences, phytoplankton carbon biomass follows similar seasonal patterns (Figs. 1g and 9r).

In order to train the DNE, we first split the input data into training, test, and validation sets by splitting up longitudinal regions on the globe (Fig. S2). For an additional level of validation of the DNE, we predict the 2005 eco-provinces for the entire globe. Each version of the model is associated with a quantified uncertainty, computed across the ensemble.

SHapley Additive exPlanations (SHAP)

Finally, SHAP, (*25*), was employed to identify the influence of particular features on where the DNE correctly and incorrectly predicted 2005 eco-provinces, based on the labels for the 1993-2004 dataset. SHAP takes a game theoretic approach, testing every combination of inputs in order to assess the contribution of each individual input. High SHAP values indicate that the input variable increased the probability of the prediction, while low SHAP values reveal the feature decreased the probability of the prediction. In our analysis, we distinguished SHAP values in locations where the

neural network made correct vs. incorrect predictions, in order to more accurately pinpoint input influence.

DNE & XAI Workflow Overview

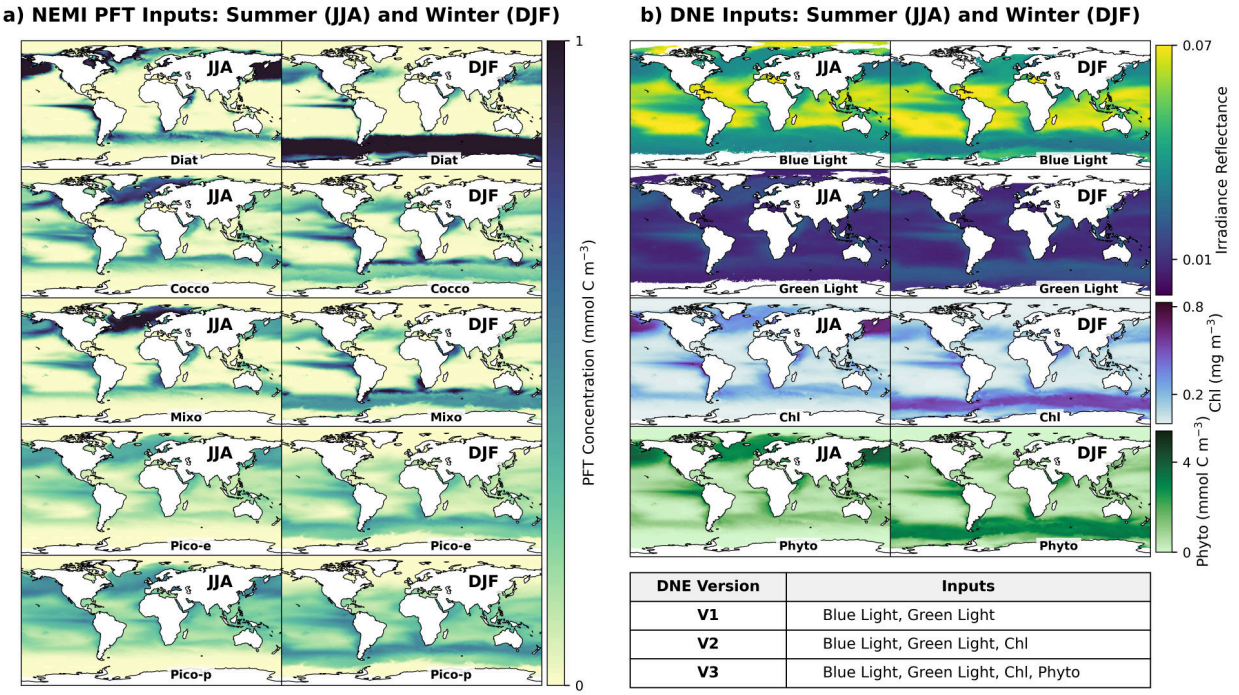
Fig. 3 presents the neural network construction for all three versions predicting a single year along with an uncertainty quantification, and the workflow for using SHAP to assess the contribution of each input to correct and incorrect predictions.

Advantages of Our Methodology

Our methodology considers optimal processes throughout both the eco-provonce identification and prediction stage which is important and necessary based on the intricate nature and large size of our input data. In the identification phase, NEMI uses UMAP which is a topological dimension reduction technique that involves the use of categorical cross entropy to optimize representation of both high and low dimensional space in the embedding which preserves structure well for large and highly complicated datasets, as demonstrated for marine geospatial data in (*17*). The agglomerative clustering approach allows the user to select the number of clusters, which makes our approach more customizable to the desired application. In addition, our NEMI eco-provinces have a mathematical uncertainty measure for each grid cell. Our DNE construction is much more robust compared to just a single neural network prediction. The ensemble prediction includes prediction entropy averaged across the ensemble, which allows for a more transparent and interpretable inference tool. Furthermore, three versions of our DNE allow us to identify the importance of the number of inputs for prediction accuracy in different locations across the globe. SHAP adds another layer of transparency to the predictions, assigning a measure importance to each input towards correct or incorrect predictions of each class. (*18*) emphasizes the importance of such a workflow for verifying the fidelity of learned physical dependencies to known dynamics. Not only is our approach statistically sound, it is also useful as a theoretical foundation for informing operational efforts. The ability to predict verified ocean community structure based on remotely sensed input data, alongside an uncertainty measure, is very important and timely as a baseline to monitor climate change.

Data availability

All input data is publicly available through the Simons Collaborative Marine Atlas Project. All code, eco-province and entropy data, DNEs, and SHAP results will be publicly available on GitHub.



DNE Version	Inputs
V1	Blue Light, Green Light
V2	Blue Light, Green Light, Chl
V3	Blue Light, Green Light, Chl, Phyto

Figure 1: **NEMI PFT Inputs and DNE Inputs**. Panel a) displays in the left column boreal summer (June, July, August: JJA) and in the right column boreal winter (December, January, February: DJF) NEMI PFT concentrations. The PFTs are Diatoms (Diat), Coccolithophores (Cocco), Mixotrophic Dinoflagellates (Mixo), Picoeukaryotes (Pico-e), and Picoprokaryotes (Pico-p). Panel b) presents boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF) DNE inputs. Included is irradiance reflectance for wavebands 425 nm (blue reflectance), 550nm (green reflectance), chlorophyll (chl), and phytoplankton carbon biomass (phyto). The bottom right table indicates which inputs are included in each version of the DNE.

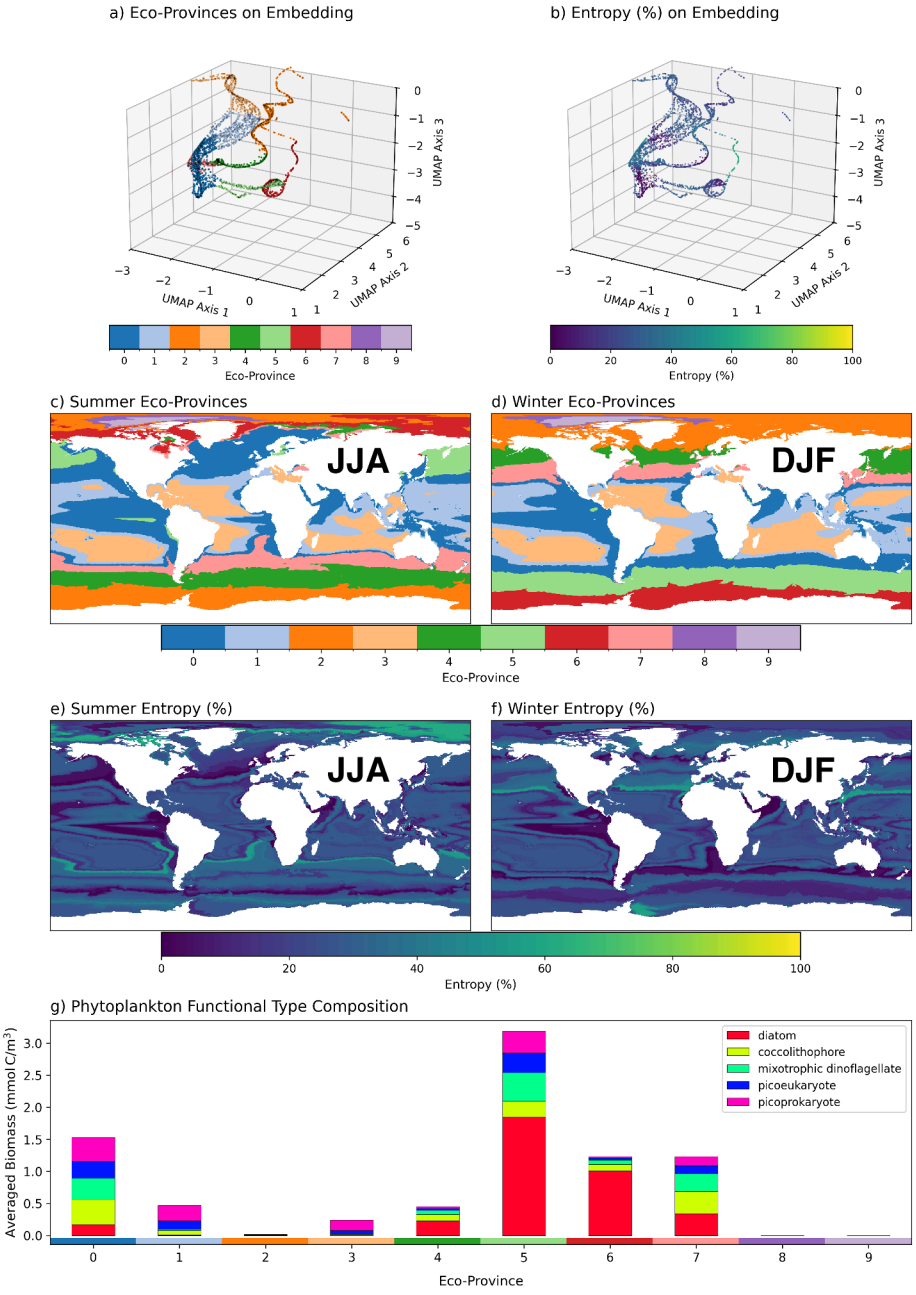


Figure 2: **Embedded and spatial Eco-Provinces With Corresponding Uncertainty, and PFT Input Composition**. Section a) presents the 10 June, July, August (JJA) boreal summer and December, January, February (DJF) boreal winter eco-provinces plotted on the embedding, with the corresponding entropy (uncertainty) in b). Sections c) and d) display the 10 JJA and DJF eco-provinces plotted geographically, with the corresponding uncertainty measured with entropy in e) and f). g) displays the phytoplankton functional type composition for each eco-province.

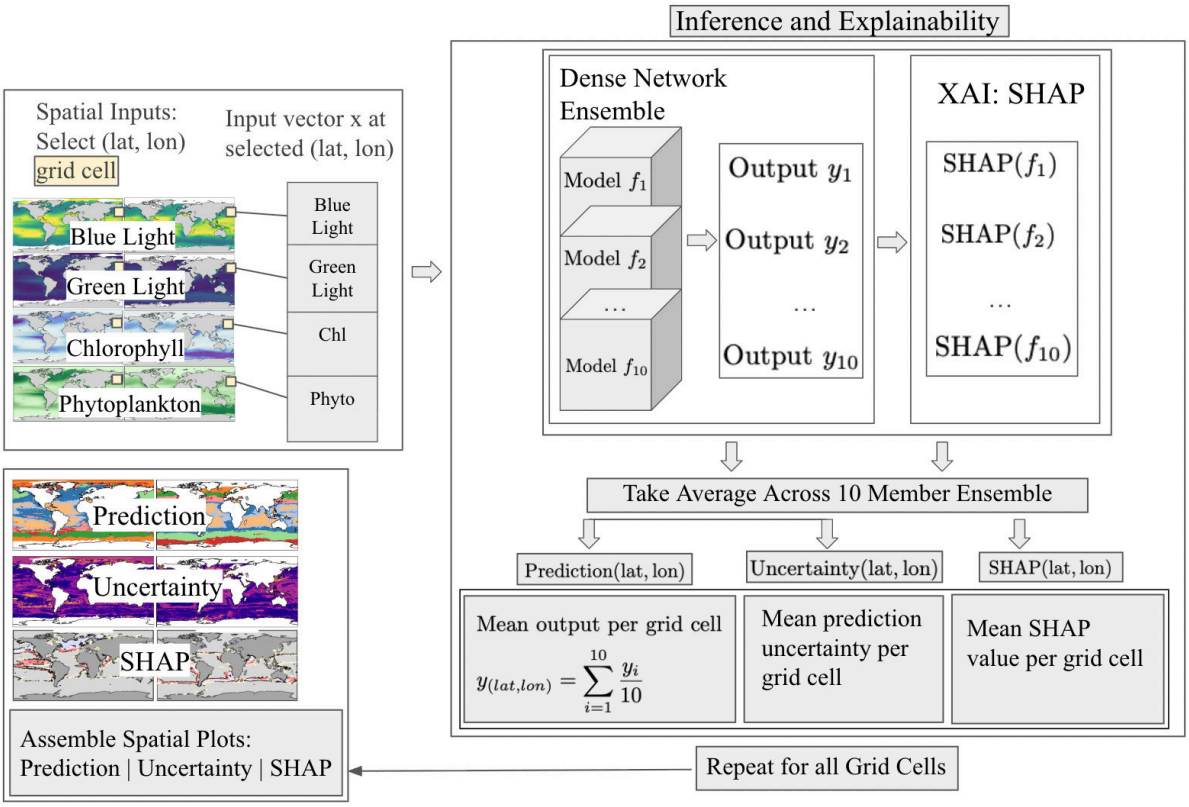


Figure 3: **DNE Construction**. The construction of the DNE, including the modeled input data that can be remotely sensed, a grid cell by grid cell input into the robust DNE prediction, repetition for all grid cells, average prediction and entropy across the ensemble, and SHAP to determine the importance of each input for correct and incorrect grid cell predictions is displayed.

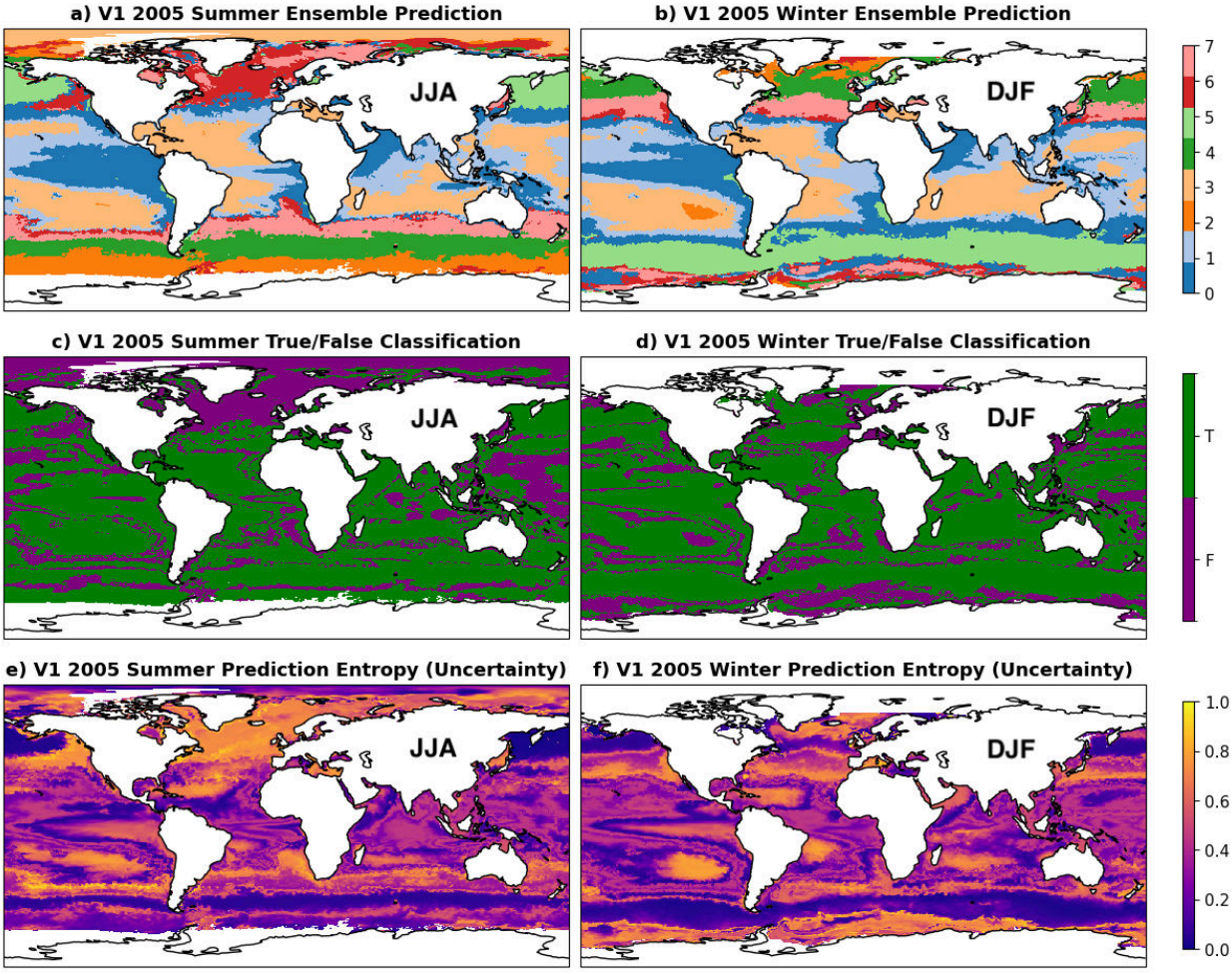


Figure 4: **V1 predicted eco-provinces, with corresponding true-false and entropy plots**. Sections a) and b) illustrate the V1 averaged prediction of 2005 boreal summer (June, July August: JJA) and boreal winter (December, January, February: DJF) eco-provinces. V1 includes blue and green reflectance as DNE inputs. Sections c) and d) present where the prediction was correct in green and incorrect in purple, compared to the actual NEMI eco-province labels. Note the incorrect predictions in the Arctic, Western Pacific, and in between eco-provinces. Sections e) and f) display the averaged entropy (uncertainty) of the prediction. The incorrect prediction of the boreal summer (JJA) Western Pacific has fairly low uncertainty.

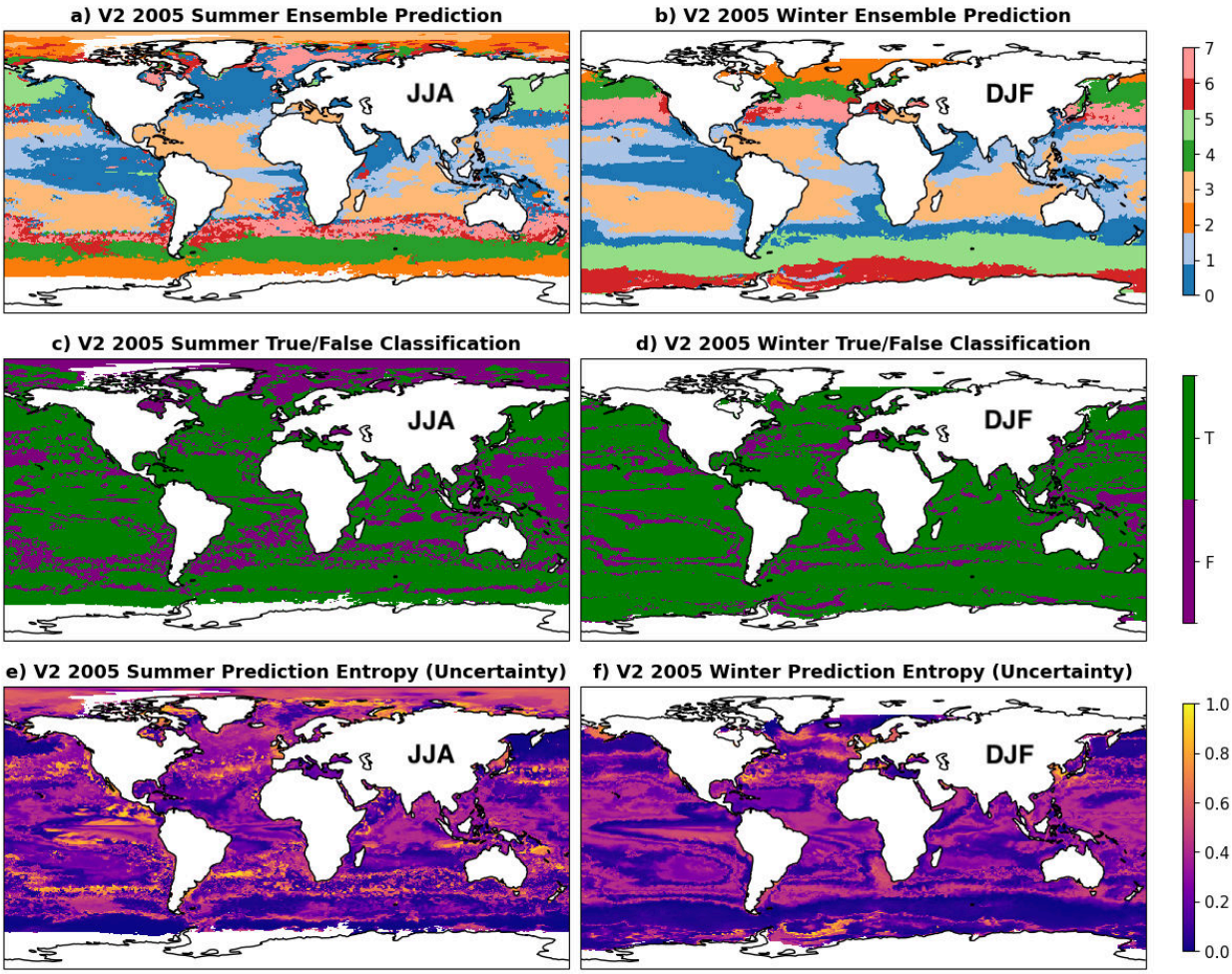


Figure 5: **V2 predicted eco-provinces, with corresponding true-false and entropy plots**. Fig. 5 has the same format as Fig. 4, but for V2 which has chlorophyll as a DNE input, in addition to the blue and green reflectance in V1. Boreal summer corresponds to June, July, August (JJA) and boreal winter corresponds to December, January, February (DJF). Compared to V1, the prediction improves in the North Atlantic, but gets worse in the Western Pacific and the Southern Ocean. Note the high entropy in the incorrect Southern Ocean prediction, but low entropy associated with the Western Pacific incorrect prediction.

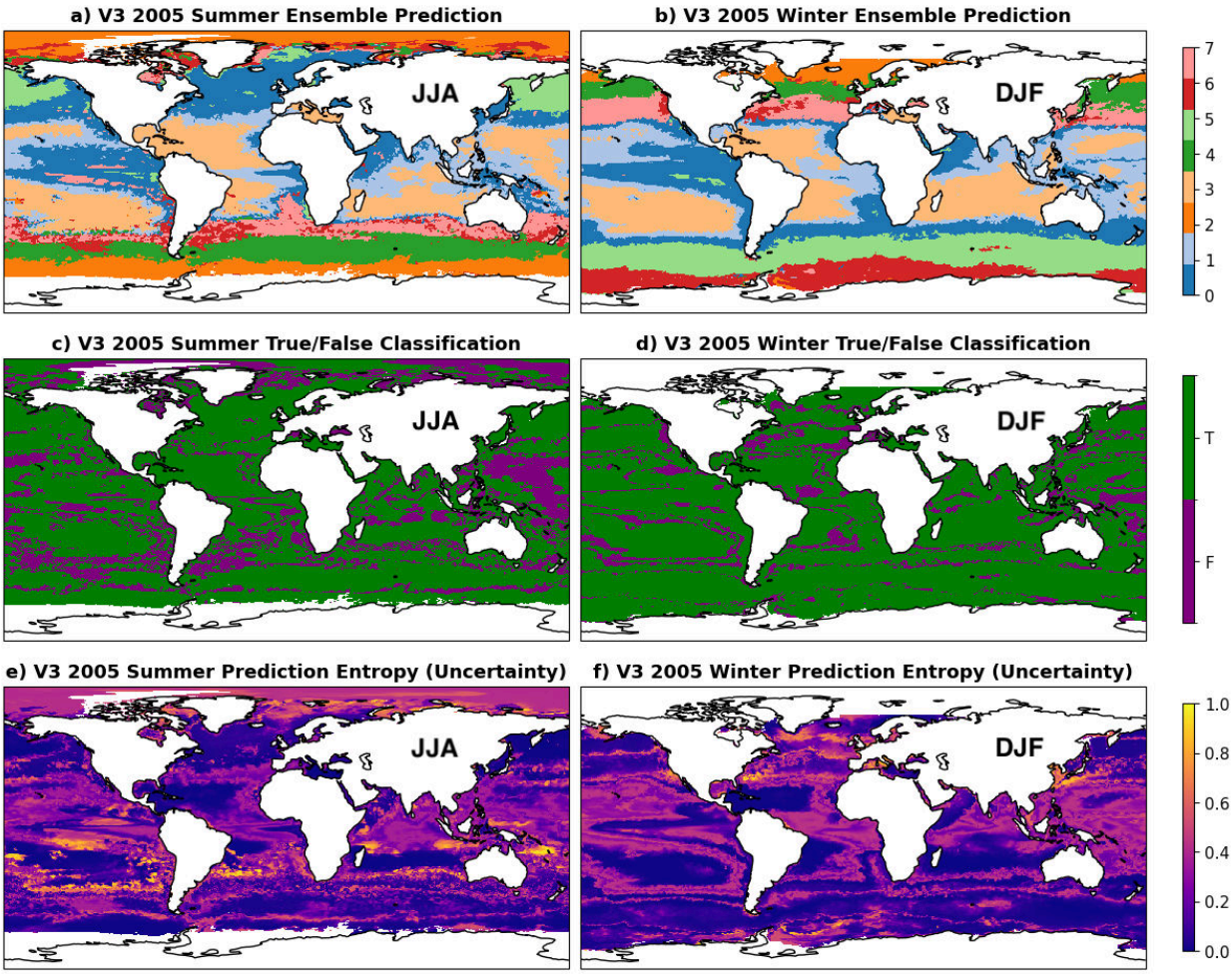


Figure 6: **V3 predicted eco-provinces, with corresponding true-false and entropy plots**. Here are the boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF) results of V3, which includes phytoplankton carbon biomass as an input in addition to the inputs in V2. Note the continued improvement in the Arctic, alongside sustained difficulty in the summer (JJA) Southern Ocean and Western Pacific. As previously observed, the uncertainty of the incorrect Southern Ocean prediction is high compared to the uncertainty of the incorrect Western Pacific prediction.

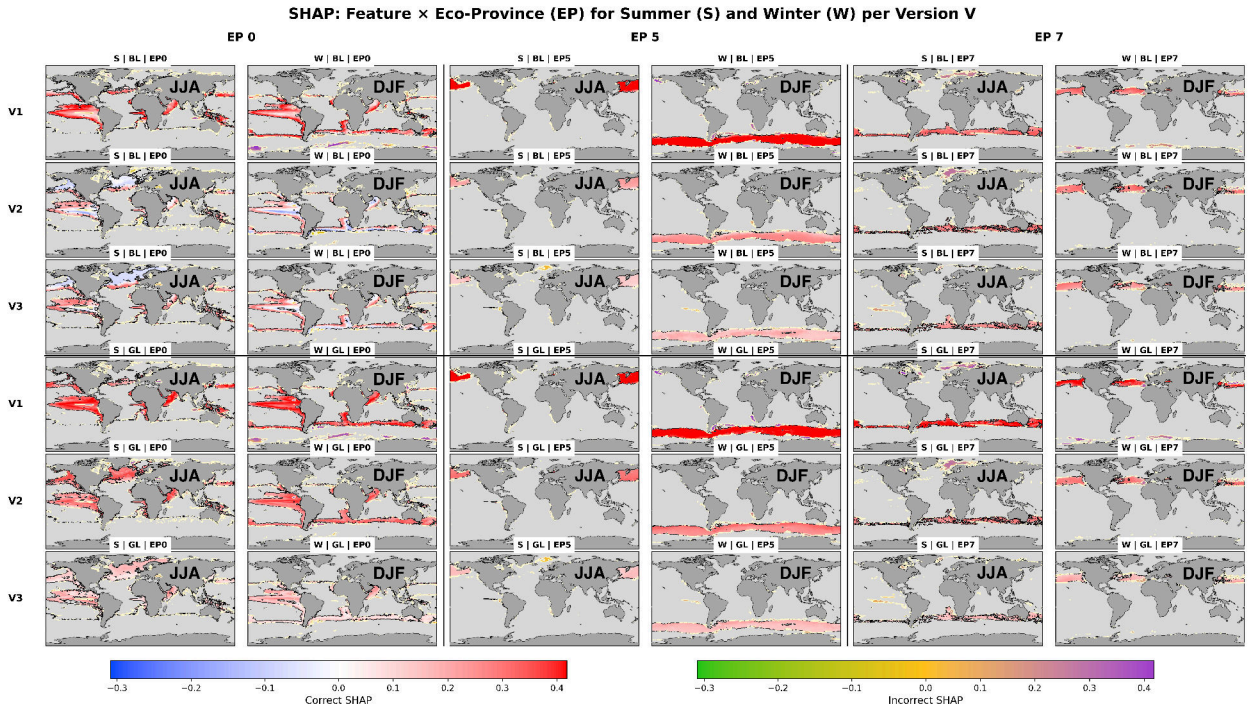


Figure 7: **SHAP results for Eco-Provinces (EPs) 0, 5, and 7 for all three versions of the DNE, including blue and green reflectance inputs.** For boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF) the importance of blue and green reflectance decreases across versions with the addition of chlorophyll and phytoplankton carbon biomass. Blue reflectance is more important for prediction on the boundary of eco-province 0, compared to green reflectance which is more influential in the center of the eco-province. Also note the high accuracy of predictions for these core eco-provinces 0, 5, and 7, which in S3 correspond to fairly high DNE inputs. More specifically, eco-province 0 has high levels of all inputs, eco-province 5 has high levels of all inputs except for blue reflectance, and eco-province 7 has medium levels of blue reflectance and fairly high levels of the other inputs.

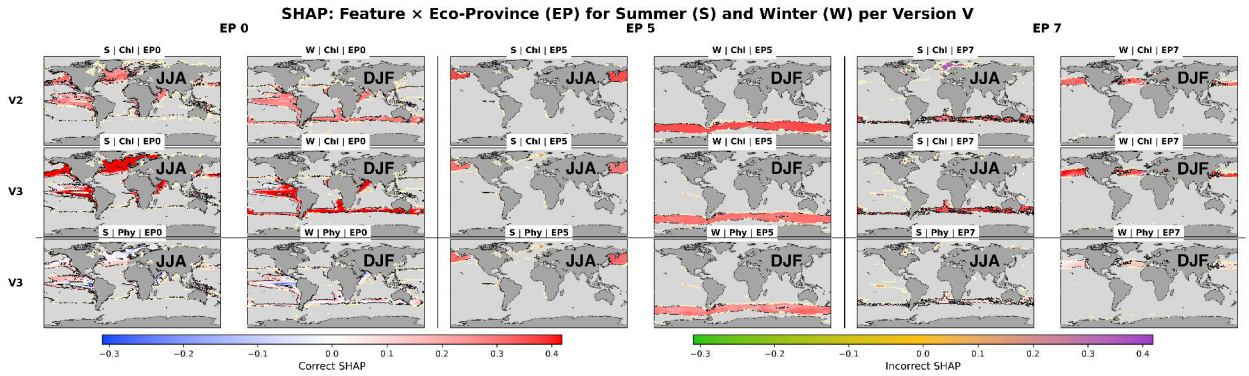


Figure 8: **SHAP results for Eco-Provinces (EPs) 0, 5, and 7 for all three versions of the DNE, including chlorophyll and phytoplankton carbon biomass inputs.** Displayed are the boreal summer (June, July, August: JJA) and boreal winter (December, Janurary, February: DJF) SHAP values for eco-provinces 0, 5 and 7 for chlorophyll and phytoplankton carbon biomass inputs across all three versions. Overall, chlorophyll is most influential in these predictions. Note that chlorophyll is more important in some cases for predicting the inside region of eco-provine 0, compared to phytoplankton which is more important on the outside of the eco-province.

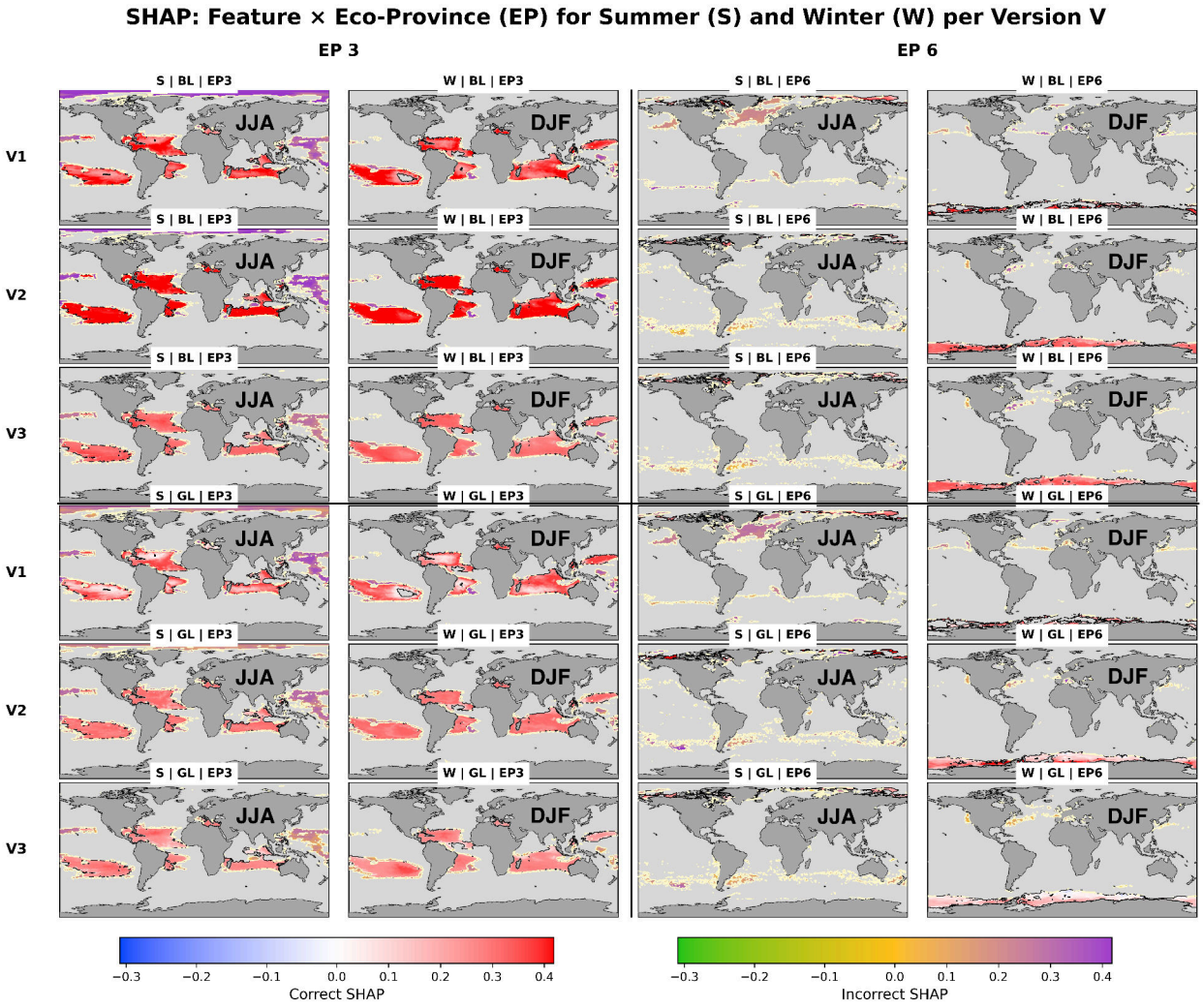


Figure 9: **SHAP values for Eco-Provinces (EPs) 3 and 6 for versions 2 and 3, including blue and green reflectance inputs**. The importance of blue and green reflectance across all three versions for the boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF) prediction of eco-provinces 3 and 6 is demonstrated. Both inputs contribute to the incorrect predictions of these eco-provinces. Note that the incorrect prediction of eco-province 3 in the boreal summer (JJA) Western Pacific is much more surprising compared to the incorrect prediction of eco-province 6 in the boreal summer (JJA) Southern Ocean, due to the low uncertainty of the eco-province 3 incorrect prediction.

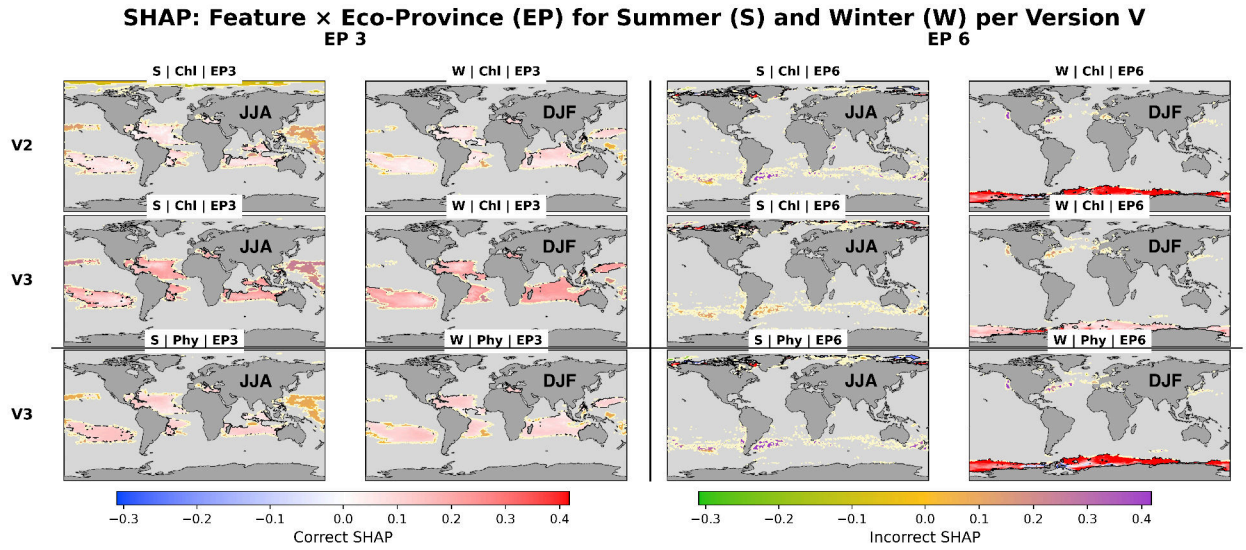


Figure 10: **SHAP results for eco-provinces (EPs) 3 and 6, for chlorophyll and phytoplankton carbon biomass as inputs, across versions 2 and 3**. The importance of chlorophyll to the incorrect prediction of boreal summer (June, July, August: JJA) eco-province 3, especially in V3 which includes all four inputs, is emphasized. For eco-province 6, chlorophyll is most influential to the incorrect prediction in V2, while phytoplankton carbon biomass is most influential for V3.

References and Notes

1. M. L. D. Palomares, *et al.*, Fishery biomass trends of exploited fish populations in marine ecoregions, climatic zones and ocean basins. *Estuarine* pp. Coastal and Shelf Science, 243, 106896 (2020), doi:10.1016/j.ecss.2020.106896.

2. T. Platt, C. Fuentes-Yaco, K. T. Frank, Spring algal bloom and larval fish survival. *Nature* **423** (6938), 398–399 (2003), doi:10.1038/423398b.

3. P. Koeller, *et al.*, Basin-scale coherence in phenology of shrimps and phytoplankton in the North Atlantic Ocean. *Science (New York* pp. N.Y.), 324(5928), 791–793 (2009), doi:10.1126/science.1170987.

4. M.-F. R. Jean-Baptiste Kassi, B. A. Mobio, T. Platt, R. Shubha Sathyendranath, K. A. D. E., Remotely Sensing the Biophysical Drivers of Sardinella aurita Variability in Ivorian Waters. *Remote Sensing* **10** (5), 785–785 (2018), doi:10.3390/rs10050785.

5. N. Menon, *et al.*, Satellite chlorophyll concentration as an aid to understanding the dynamics of Indian oil sardine in the southeastern Arabian Sea. *Marine Ecology Progress Series* **617**, –618, 137–147 (2019), doi:10.3354/meps12806.

6. A. Longhurst, S. Sathyendranath, T. Platt, C. Caverhill, An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research* **17** (6), 1245–1271 (1995), doi:10.1093/plankt/17.6.1245.

7. M. D. Spalding, *et al.*, Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience* **57** (7), 573–583 (2007), doi:10.1038/s41561-017-0028-x.

8. M. D. Spalding, V. N. Agostini, J. Rice, S. M. Grant, Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management* **60**, 19–30 (2012), doi:10.1016/j.ocecoaman.2011.12.016.

9. M. J. Costello, *et al.*, Marine biogeographic realms and species endemicity. *Nature Communications* **8** (1) (2017), doi:10.1038/s41467-017-01121-2.

10. R. Sayre, *et al.*, A Three-Dimensional Mapping of the Ocean Based on Environmental Data. *Oceanography* **30** (1), 90–103. Sonnewald. (2023). A hierarchical ensemble manifold methodology for new knowledge on spatial data: An application to ocean physics. ESS Open Archive (2017), doi:10.5670/oceanog.2017.116.

11. M. T. Kavanaugh, *et al.*, Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. *Progress in Oceanography* **120**, 291–304 (2014), doi:10.1016/j.pocean.2013.10.013.

12. M. Sonnewald, S. Dutkiewicz, C. Hill, G. Forget, Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces. *Science Advances* **6** (22) (2020), doi:10.1126/sciadv.aay4740.

13. U. Hofmann Elizondo, D. Righetti, F. Benedetti, M. Vogt, Biome partitioning of the global ocean based on phytoplankton biogeography. *Progress in Oceanography* **194**, 102530 (2021), doi:10.1016/j.pocean.2021.102530.

14. H. Kaneko, *et al.*, Predicting global distributions of eukaryotic plankton communities from satellite data. *ISME Communications* **3** (1) (2023), doi:10.1038/s43705-023-00308-7.

15. R. El Hourany, *et al.*, Linking satellites to genes with machine learning to estimate phytoplankton community structure from space. *Ocean Science* **20** (1), 217–239 (2024), doi:10.5194/os-20-217-2024.

16. M. Sonnewald, A hierarchical ensemble manifold methodology for new knowledge on spatial data: An application to ocean physics. *ESS Open Archive.* (2023), doi:https://doi.org/10.22541/essoar.168056792.25480169/v1.

17. Y. Jenniges, M. Sonnewald, S. Maneth, A. Olsen, B. P. Koch, Unveiling 3D ocean biogeochemical provinces in the North Atlantic: A systematic comparison and validation of clustering methods. *Ecological Informatics* **91**, 103390 (2025), doi:10.1016/j.ecoinf.2025.103390.

18. S. Suri, M. Sonnewald, Trusting machine learning with physics: A fidelity verification framework for complex systems. *ESS Open Archive* (2026), doi:10.22541/essoar.176894678.89831689/v1.

19. M. Clare, R. L. Maike Sonnewald, J. Deshayes, V. Balaji, Explainable Artificial Intelligence for Bayesian Neural Networks: Toward Trustworthy Predictions of Ocean Dynamics. *Journal of Advances in Modeling Earth Systems* **14** (11) (2022), doi:10.1029/2022ms003162.

20. M. Sonnewald, R. Lguensat, Revealing the Impact of Global Heating on North Atlantic Circulation Using Transparent Machine Learning. *Journal of Advances in Modeling Earth Systems* **13** (8) (2021), doi:10.1029/2021ms002496.

21. A. B. Arrieta, *et al.*, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* **58** (1), 82–115 (2020), doi:10.1016/j.inffus.2019.12.012.

22. M. Mersha, K. Lam, J. Wood, A. AlShami, J. Kalita, Explainable Artificial Intelligence: A Survey of Needs, Techniques, Applications, and Future Direction. *ArXiv.org* (2024), `https://arxiv.org/abs/2409.00265`.

23. S. Dutkiewicz, P. W. Boyd, U. Riebesell, Exploring biogeochemical and ecological redundancy in phytoplankton communities in the global ocean. *Global Change Biology* **27** (6), 1196–1213 (2021), doi:10.1111/gcb.15493.

24. J. Graff, *et al.*, Analytical phytoplankton carbon measurements spanning diverse ecossytems. *Deep Sea Research Part I: Oceanographic Research Papers* (2015), `https://doi.org/10.1016/j.dsr.2015.04.006`.

25. S. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions. *ArXiv:1705.07874 [Cs* p. Stat] (2017), `https://arxiv.org/abs/1705.07874`.

26. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv.org* (2018), `https://arxiv.org/abs/1802.03426`.

27. T. Kohonen, Essentials of the self-organizing map. *Neural Networks* **37**, 52–65 (2013), doi:10.1016/j.neunet.2012.09.018.

28. J. R. Bray, J. T. Curtis, An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27** (4), 325–349 (1957), doi:10.2307/1942268.

29. A. A. Wani, Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science* **10**, e2286–e2286 (2024), doi:10.7717/peerj-cs.2286.

30. A. Mohammed, R. Kora, A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges. *Journal of King Saud University - Computer and Information Sciences* **35** (2) (2023), https://www.sciencedirect.com/science/article/pii/S1319157823000228.

31. M. Wyatt, B. Radford, N. Callow, M. Bennamoun, S. Hickey, Using ensemble methods to improve the robustness of deep learning for image classification in marine environments. *Methods in Ecology and Evolution* (2022), doi:10.1111/2041-210x.13841.

32. IOCCG, *Uncertainties in Ocean Colour Remote Sensing*, Tech. Rep. 18, International Ocean Colour Coordinating Group, Dartmouth, Canada (2019), https://ioccg.org/wp-content/uploads/2020/01/ioccg-report-18-uncertainties-rr.pdf.

33. R. Vandermeulen, *et al.*, *Plankton, Aerosol, Cloud and Ocean Ecosystem (PACE) Satellite Data for Aquaculture and Fisheries Management*, Tech. Rep. NOAA Technical Memorandum NMFS-F/SPO-255, United States National Marine Fisheries Service, Office of Science and Technology (2025), doi:10.25923/5mtj-9h57, https://doi.org/10.25923/5mtj-9h57.

34. D. C. Reese, R. T. O'Malley, R. D. Brodeur, J. H. Churnside, Epipelagic fish distributions in relation to thermal fronts in a coastal upwelling system using high-resolution remote-sensing techniques. *ICES Journal of Marine Science* **68** (9), 1865–1874 (2011), doi:10.1093/icesjms/fsr107.

35. C. Anguita, S. Gelcich, M. Aldana, J. Pulgar, Exploring the influence of upwelling on the total allowed catch and harvests of a benthic gastropod managed under a territorial user rights for fisheries regime along the Chilean coast. *Ocean & Coastal Management* **195**, 105256 (2020), doi:10.1016/j.ocecoaman.2020.105256.

36. V. B-Béres, *et al.*, Ecosystem services provided by freshwater and marine diatoms. *Hydrobiologia* **850** (2022), doi:10.1007/s10750-022-04984-9.

37. P. Tréguer, *et al.*, Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience* **11** (1), 27—-37 (2008), doi:10.1038/s41561-017-0028-x.

38. D. Ward, *et al.*, Safeguarding marine life: conservation of biodiversity and ecosystems. *Reviews in Fish Biology and Fisheries* **32** (1), 1–36 (2022), doi:10.1007/s11160-022-09700-3.

39. U.S. Geological Survey, *Carbon Sequestration to Mitigate Climate Change*, Tech. Rep. Fact Sheet 2008–3097, U.S. Geological Survey (2008), https://pubs.usgs.gov/fs/2008/3097/pdf/CarbonFS.pdf.

40. S. Li, *et al.*, Multifaceted contribution of coccolithophores to ocean carbon export. *Ocean-Land-Atmosphere Research* (2024), doi:10.34133/olar.0049.

41. A. S. Rigual Hernández, *et al.*, Coccolithophore biodiversity controls carbonate export in the Southern Ocean. *Biogeosciences* **17** (1), 245–263 (2020), doi:10.5194/bg-17-245-2020.

42. I. Goodfellow, Y. Bengio, A. Courville, Deep Learning. *Deeplearningbook.org; MIT Press* (2016), https://www.deeplearningbook.org/.

43. W. Yik, M. Sonnewald, C. Mariana, R. Lguensat, Southern Ocean Dynamics Under Climate Change: New Knowledge Through Physics-Guided Machine Learning. *ArXiv.org* (2023), https://arxiv.org/abs/2310.13916.

Acknowledgments

Funding: This work was supported by startup funds from the University of California Davis and the UC Davis Global Affairs Grants for Advancing the United Nations Sustainable Development Goals (SDGs).

Author contributions: MM: Writing and running experiments. MS: Conceptualization, writing (reivew and editing), supervision, and funding aquisition. SD: Biogeochemical numerical model development for datasets, review and editing.

Competing interests: There are no competing interests to declare.

Supplementary Materials for

Machine learning fidelity verification exposes management-relevant tradeoffs in satellite-based ocean eco-province inference

Makayla McDevitt*, Maike Sonnewald, Stephanie Dutkiewicz

*Corresponding author. Email: mrmcdevitt@ucdavis.edu

This PDF file includes:

Supplementary Materials and Methods

Figures S1 to S3

Tables S1

Supplementary Materials and Methods

NEMI Input

To calculate summer (June, July, August: JJA) and winter (December, January, February: DJF) averaged datasets, we calculated monthly averages for each year based on the initial three day averages across the entire time period. After obtaining the monthly averages for each year, we found the monthly averages across the entire dataset by averaging each month across the 23 years. Finally, we found the seasonal averages by averaging the June, July, August total averages to create boreal summer datasets, and averaged December, January, February averages to find boreal winter datasets.

Eco-Province Identification

NEMI uses UMAP (*26*) for dimension reduction, a topological data reduction method that uses categorical cross entropy for optimization, which balances representation of both the high and low dimensional structures. To determine which clusters should be combined, NEMI uses Ward's linkage which minimizes the sum of squared differences within all of the clusters. (*17*) demonstrated how UMAP strengthens data associations, improving clustering outcomes. For clustering, NEMI uses hierarchical agglomeration, where each observation starts as its own cluster, then clusters are successively merged together to build nested clusters (*16*). We repeated the embedding, clustering, entropy, and functional type composition results for 15, 20, and 25 clusters. Similar patterns were observed with more complexity. A 20 cluster example is provided here in Fig. S1, for demonstration.

Using NEMI and assessing different numbers of eco-provinces, we were able to ensure that each cluster was cohearent and that as we increased the number of clusters, the clusters became more complex as progressions of eachother. This pattern of higher complexity clusters being progressions of lower complexity clusters was further validated by plotting in geographical space and relating back to PFT input concentrations. The locations and shapes of each eco-province, and the way in which the complexity progresses as we increase the number of clusters, matches expectations based on PFT input concentrations. For example, we observe distinct eco-provinces for low nutrtient oligatrophic regions, high nutrient coastal upwelling regions, and high-diatom regions in the JJA North Pacific and DJF Southern Ocean.

Fig. S2 demonstrates the neural network train, validation, and test split with testing on the western Atlantic and testing on the eastern Atlantic.

In each of the versions, the inputs are passed into four hidden layers, each with a hyperbolic tangent activation function. The output layer uses a softmax activation function to predict eight resulting eco-provinces. The neural network uses categorical cross entropy for the loss function and trains for 100 epochs.

Table S1 compares the accuracy of each version for boreal summer (June, July, August: JJA) and boreal winter (December, January, February: DJF), in addition to the average accuracy.

Fig. S4 compares the ensemble losses and accuracies across the three versions.

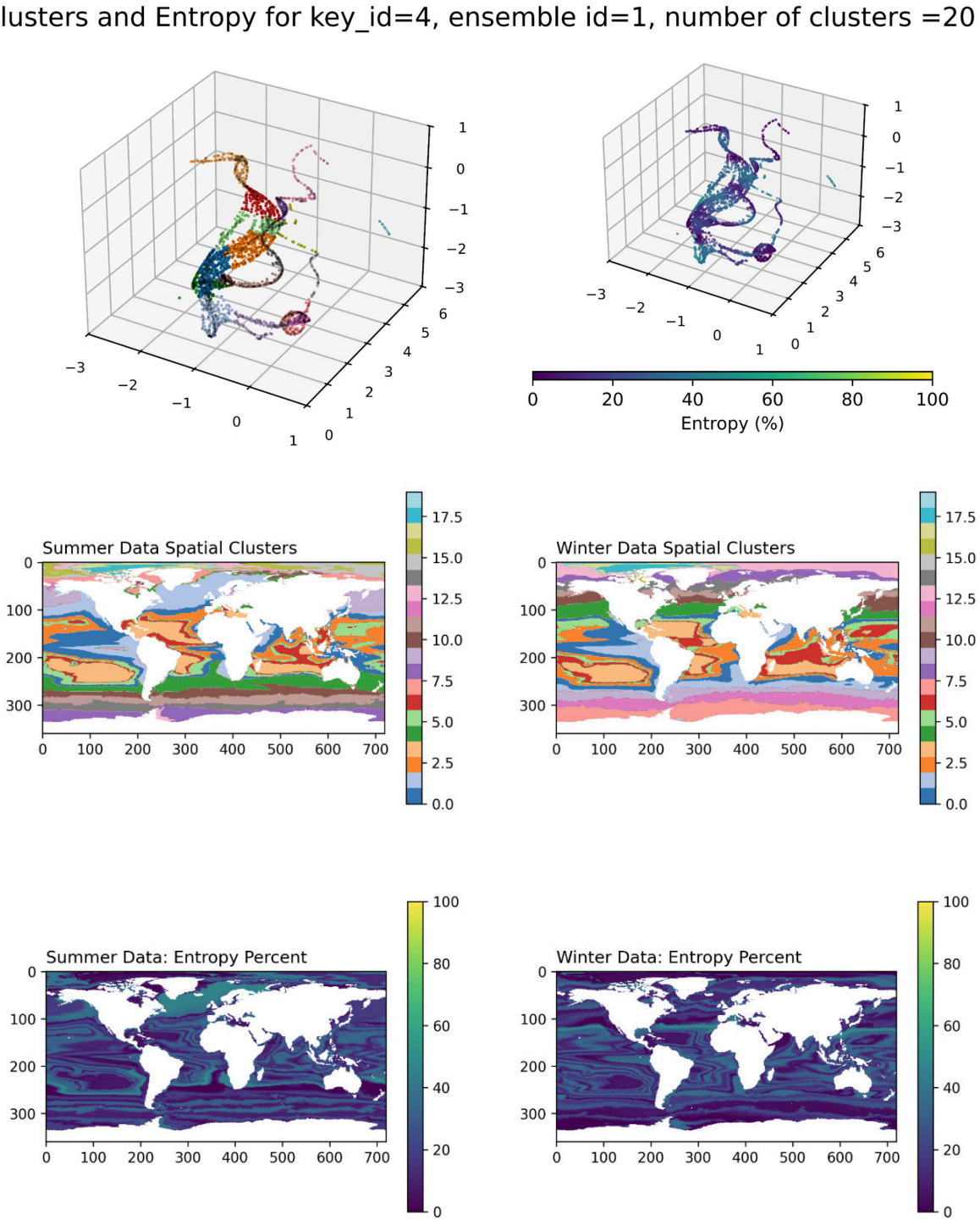


Figure S1: **20-Cluster Example.** Example of an eco-province characterization with 20 eco-provinces (clusters) and the corresponding entropy percentage.

Table S1: **Summer (JJA), Winter (DJF), and Average Accuracy Across Versions.** Table S1 Demonstrates the variety in accuracy among the three versions, revealing the increase in overall performance with the addition of inputs with some nuance in performance when considering each cluster individually.

Cluster	V1 Summer Accuracy	V1 Winter Accuracy	V1 Avg Accuracy
0	76.0%	69.4%	72.3%
1	60.2%	64.3%	62.4%
2	64.4%	45.7%	61.3%
3	46.8%	76.5%	59.1%
4	79.6%	67.4%	75.6%
5	77.7%	81.9%	81.0%
6	12.6%	27.8%	18.0%
7	68.0%	52.5%	61.7%
Overall	**60.7%**	**60.7%**	**61.4%**

Cluster	V2 Summer Accuracy	V2 Winter Accuracy	V2 Avg Accuracy
0	61.2%	75.8%	68.2%
1	52.3%	65.9%	59.8%
2	55.7%	63.0%	57.1%
3	49.6%	78.1%	61.8%
4	65.6%	74.3%	68.1%
5	80.2%	88.6%	86.9%
6	7.9%	76.6%	46.2%
7	54.6%	70.2%	61.1%
Overall	**53.4%**	**74.1%**	**63.6%**

Cluster	V3 Summer Accuracy	V3 Winter Accuracy	V3 Avg Accuracy
0	75.2%	79.0%	77.0%
1	57.4%	73.1%	65.3%
2	95.5%	96.0%	95.6%
3	89.5%	92.9%	91.3%
4	76.8%	77.2%	76.9%
5	86.5%	94.6%	92.9%
6	33.7%	88.3%	66.4%
7	57.5%	80.4%	65.8%
8	0.0%	NA	0.0%
Overall	**72.0%**	**84.3%**	**77.9%**

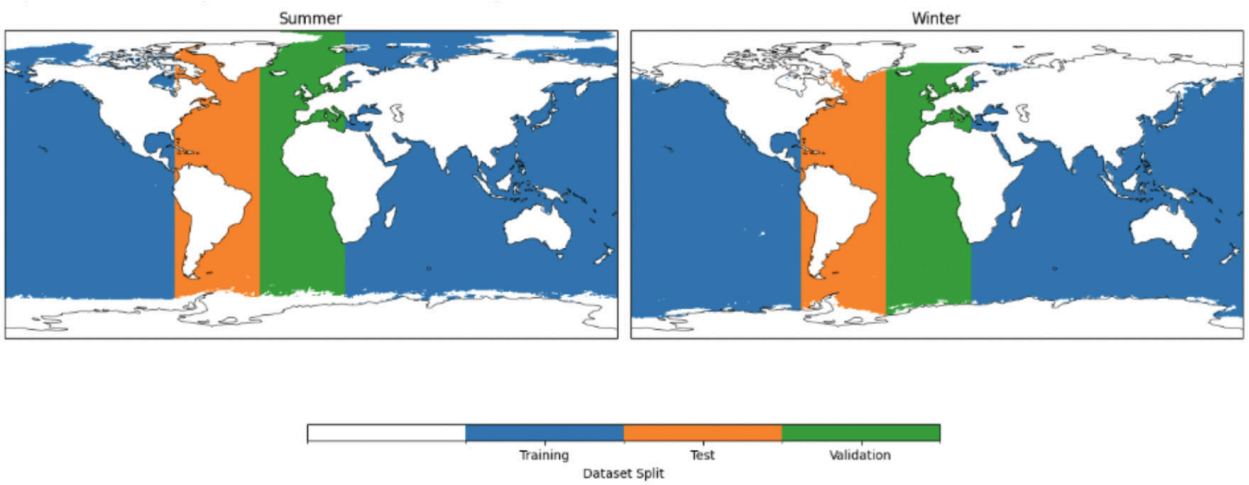


Figure S2: **Training, validation, and test split.** Fig. S2 demonstrates the vertical split used to train, validate, and test the neural network with input data averaging across 1993-2004. For an extra level of validation, we then tested the network on the entire globe for 2005.

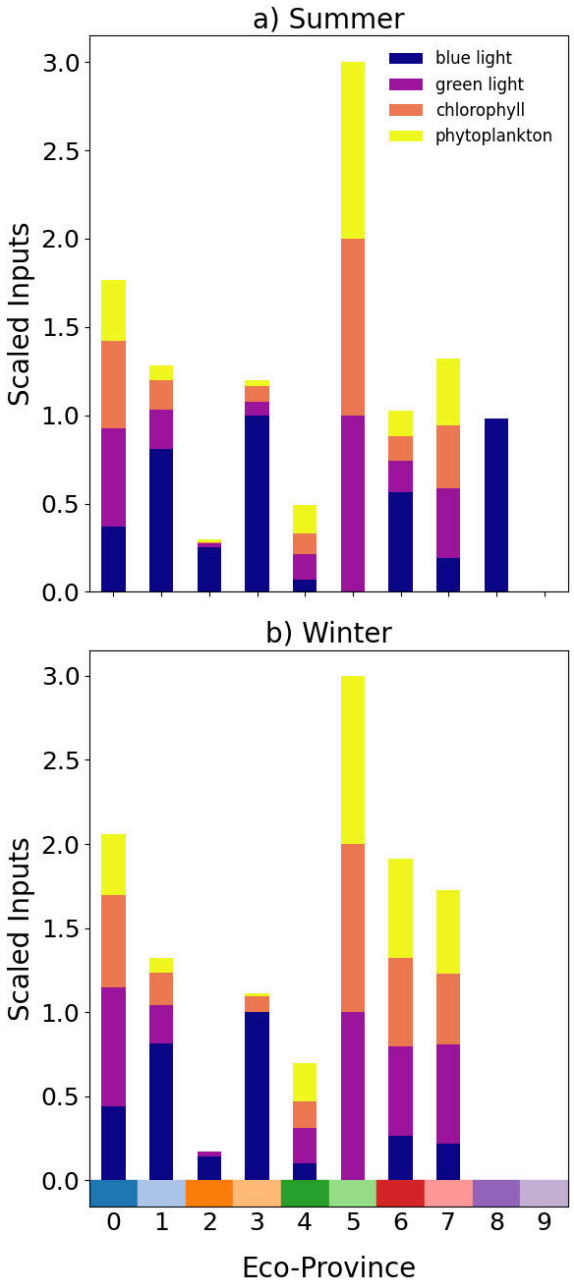


Figure S3: **DNE Input Composition**. Fig. S3 presents the scaled composition of blue reflectance, green reflectance, chlorophyll, and phytoplankton carbon biomass as inputs to the V3 DNE, broken up according to each eco-province. This composition allows us to identify the input composition for each different spatial region, in order to understand which inputs contributed to specific results. In general, either high or low input concentrations are more informative for DNE predictions compared to moderate concentrations.

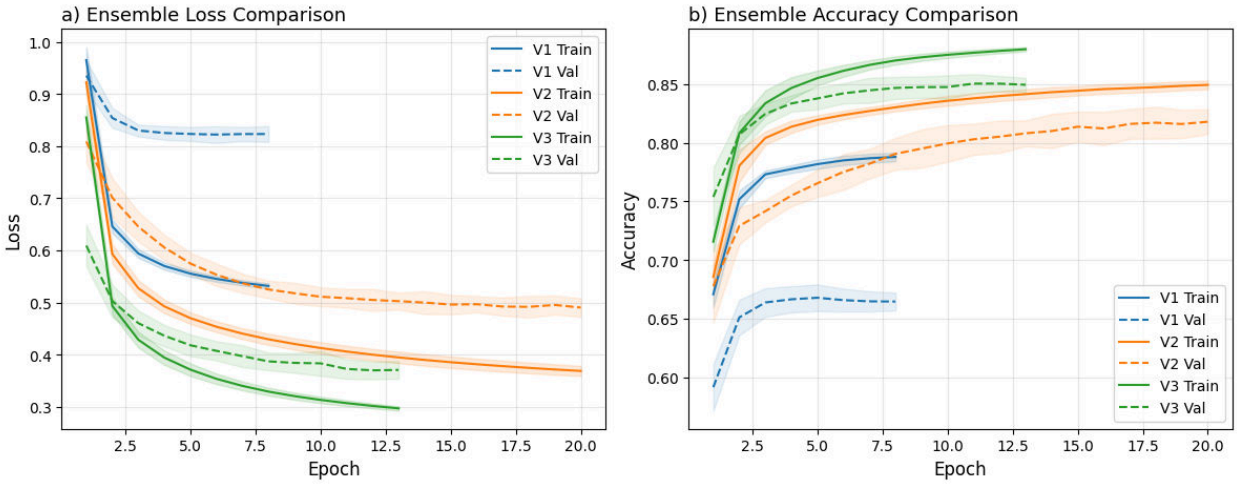


Figure S4: **Ensemble loss and accuracy comparison across three versions.**

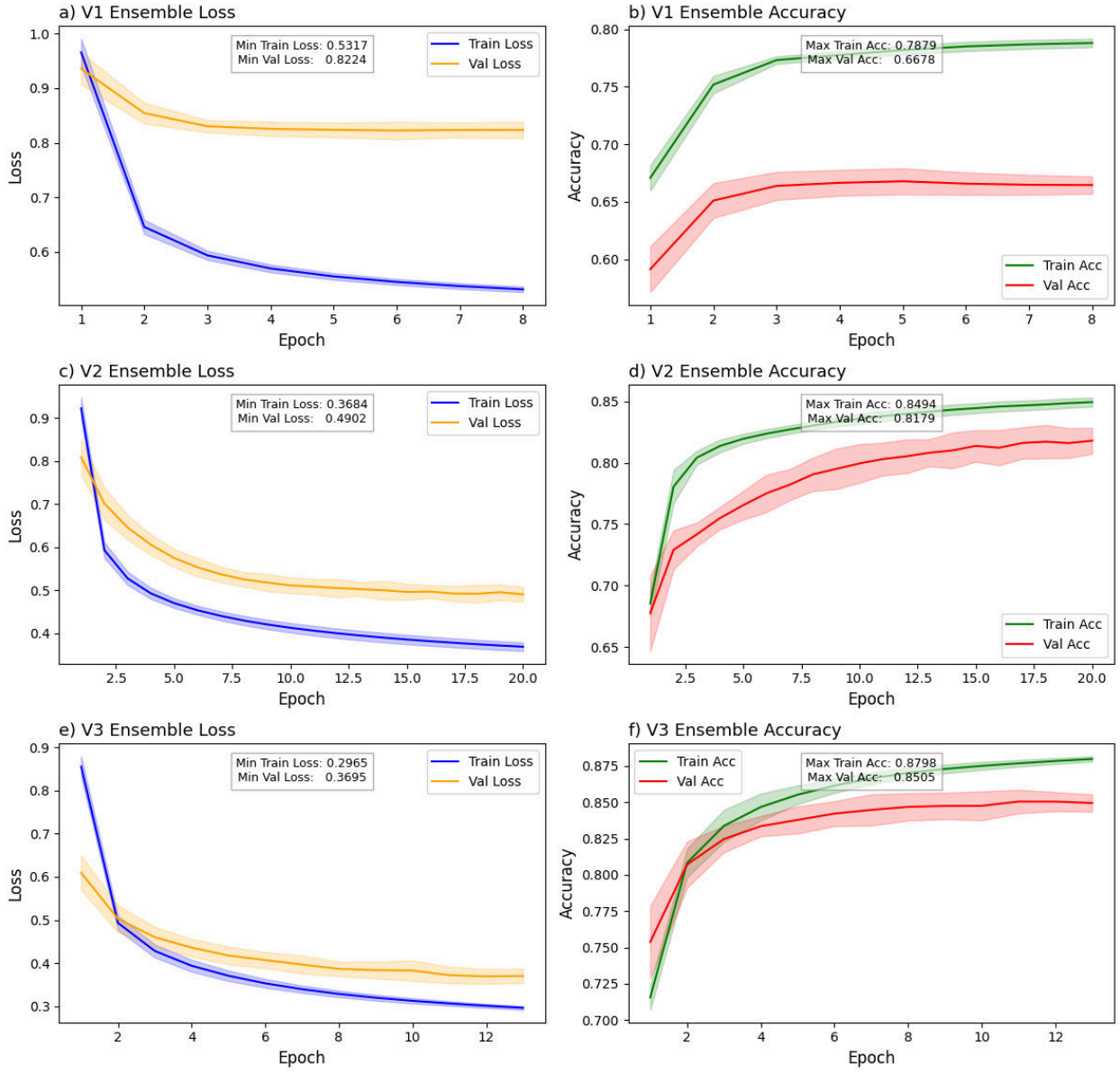


Figure S5: **Ensemble loss and accuracy comparison across three versions, with minimum loss and maximum accuracy values.** Fig. S5 also compares the ensemble losses and accuracies across the three versions, with minimum loss values and maximum accuracies.