Biogeosciences Discuss., 10, 8103–8157, 2013
www.biogeosciences-discuss.net/10/8103/2013/
doi:10.5194/bgd-10-8103-2013

Biogeosciences Discussions Open Access

Distribution of phytoplankton functional types in high-nitrate low-chlorophyll waters in a new diagnostic ecological indicator model

A. P. Palacz[1], M. A. St. John[1], R. J. W. Brewin[2], T. Hirata[3], and W. W. Gregg[4]

[1]National Institute of Aquatic Resources, Denmark Technical University, Jægersborg Allé 1, 2920 Charlottenlund, Denmark
[2]Plymouth Marine Laboratory, Plymouth, UK
[3]Faculty of Environmental Earth Science, Hokkaido University, N10W5, Sapporo, 060-0810 Hokkaido, Japan
[4]NASA Global Modeling and Assimilation Office, Greenbelt, MD, USA

Received: 24 April 2013 – Accepted: 2 May 2013 – Published: 15 May 2013

Correspondence to: A. P. Palacz (arpa@aqua.dtu.dk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Modeling and monitoring plankton functional types (PFTs) is challenged by insufficient amount of field measurements to ground-truth both plankton models and bio-optical algorithms. In this study, we combine remote sensing data and a dynamic plankton model to simulate an ecologically-sound spatial and temporal distribution of phyto-PFTs. We apply an innovative ecological indicator approach to modeling PFTs, and focus on resolving the question of diatom-coccolithophore co-existence in the subpolar high-nitrate and low-chlorophyll regions. We choose an artificial neural network as our modeling framework because it has the potential to interpret complex nonlinear interactions governing complex adaptive systems, of which marine ecosystems are a prime example. Using ecological indicators that fulfill the criteria of measurability, sensitivity and specificity, we demonstrate that our diagnostic model correctly interprets some basic ecological rules similar to ones emerging from dynamic models. Our time series highlight a dynamic phyto-PFT community composition in all high latitude areas, and indicate seasonal co-existence of diatoms and coccolithophores. This observation, though consistent with in situ and remote sensing measurements, was so far not captured by state-of-the-art dynamic models which struggle to resolve this "paradox of the plankton". We conclude that an ecological indicator approach is useful for ecological modeling of phytoplankton and potentially higher trophic levels. Finally, we speculate that it could serve as a powerful tool in advancing ecosystem-based management of marine resources.

1 Introduction

We are yet to obtain a consistent and complete view of the global biogeography of plankton functional types (PFTs), groups of organisms composed of many different species identified by a common biogeochemical function rather than a common phylogeny. We deal with large uncertainty due to insufficient amount of field measurements

to ground-truth both plankton models (Anderson, 2005) and bio-optical phytoplankton PFT algorithms (Brewin et al., 2011). Knowing how PFT distributions are changing is key to projecting biological responses to global climate change, both in the geological past (Wells et al., 1999), present and future (Beaugrand et al., 2012; Bopp et al., 2005; Boyd and Doney, 2002). A requirement is to investigate regions such as the Southern Ocean, the subarctic North Pacific and the equatorial Pacific, the so-called high-nitrate low-chlorophyll (HNLC) regions (Minas and Minas, 1992) where the ocean acts as a source of CO_2 to the atmosphere (Takahashi et al., 2002) despite being sufficiently productive. According to modeling studies, these regions may also constitute PFT diversity hotspots (Prowe et al., 2012) or, to the contrary, PFT diversity deserts (Barton et al., 2010).

A variety of recently-developed phytoplankton size class (phyto-PSC) and phytoplankton PFT (phyto-PFT) bio-optical algorithms may be used to monitor regional and global phytoplankton distributions at an intraseasonal resolution (Alvain et al., 2008; Bracher et al., 2009; Brewin et al., 2010; Hirata et al., 2011). Among the key phyto-PFTs are silicifiers (diatoms) whose biomass can be estimated from satellite (Sathyendranath et al., 2004; Alvain et al., 2008; Hirata et al., 2013), and calcifiers (coccolithophores) whose biomass estimates up until very recently (Sadeghi et al., 2012) were limited to bloom occurrence and area calculation (Alvain et al., 2008; Moore et al., 2012) and inference from particulate inorganic carbon (PIC) measurements (Balch et al., 2011). However, field measurements of biological processes, on the level of PFTs in particular, are not sufficiently resolved compared to chemical and physical properties of the ocean (Claustre et al., 2010). Thus, synoptic relationships between PFT biomass (derived from in situ phytoplankton cell and pigment abundance/biomass) and optical properties of the surface ocean (derived from remote sensing) cannot be constrained and evaluated consistently over all biogeochemical provinces. Borders between biogeochemical provinces shift dramatically on annual and longer scales (Boyd and Doney, 2002; Devred et al., 2007) challenging the integration of regional algorithms into global models (Cetinić et al., 2012).

Dynamic plankton models coupled to ocean circulation models take changing environmental conditions into account. Moreover, they derive PFT estimates from knowledge of the underlying mechanistic processes. However, whereas they can discern between several to tens of different groups (Follows et al., 2007), they struggle with the long debated paradox of the plankton (Hutchinson, 1961). To our knowledge, none of the models operating on the PFT level can simulate the observed coexistence of more than one dominant group under limiting resources when their biogeochemical functions are similar yet need to be parameterised separately. For instance, fast-growing diatoms in the subarctic North Pacific and the Southern Ocean outcompete coccolithophores preventing their biomass to build up (e.g. Gregg and Casey, 2007; Le Quéré et al., 2005; Sinha et al., 2010) in contrast to what remote sensing observations suggest (e.g. Alvain et al., 2008; Balch et al., 2011; Sadeghi et al., 2012). A PFT biogeochemical model is sensitive to the type of PFT classification, choice of zooplankton grazing formulation (Hashioka et al., 2012) and, especially for diatoms and coccolithophores, to the mixing formulation in the chosen physical model coupled to the plankton model (Sinha et al., 2010). The MARine Ecosystem Model Intercomparison Project (MAREMIP) (Hashioka et al., 2012) currently examines the ability of these models to identify key processes and evaluate the role of functional groups in the whole ecosystem. Complementary to MAREMIP, the Satellite phyto-PFT Intercomparison Project (Hirata et al., 2012) is investigating the performance of satellite phyto-PFT models in the global ocean.

In a two-dimensional phytoplankton niche space defined by turbulence and nutrient concentration ("Margalef's Mandala"), coccolithophores traditionally fall between diatoms which thrive in well-mixed, high nutrients regimes, and dinoflagellates which dominate the stratified, low-nutrient regimes (Margalef, 1978). Today we know that picoeukaryotic and prokaryotic autotrophs (e.g. *Prochlorococcus*, *Synecoccocus*) successfully compete with dinoflagellates for their niche. Light is another important niche descriptor. Balch (2004) suggested that daylength is as important as light intensity for the onset of coccolithophore (and likely other phyto-PFT) blooms, and should form

the third dimension of phytoplankton mandala. The ability to locate phyto-PFTs in their ecological niches allows us to derive ecological rules that can verify both plankton models and bio-optical algorithms. However, ecological rules can serve another purpose: they help identify key ecological indicators of change of PFTs – a task imposed by the rapidly changing climate.

Having acknowledged the challenges in monitoring and modeling biological processes explicitly, an ecological indicator approach is an alternative means of describing and managing marine ecosystems (e.g. Dale and Beyeler, 2001; Blanchard et al., 2010). In this study, we explore the possibility of applying an ecological indicator approach to simulate an ecologically consistent global distribution of phyto-PFTs, with particular focus on diatoms and coccolithophores in the HNLC regions. We choose an artificial neural network (ANN) as our modeling framework because this artificial intelligence tool has the potential to interpret complex nonlinear interactions governing complex adaptive systems (Holland, 1995) of which marine ecosystems are a prime example (Levin, 1998). In order to enable projection of past and future phyto-PFT states, and their potential application in ecosystem management of marine resources (Palacz, 2012), we select ecological indicators that fullfil the criteria of indicators of Good Environmental Status (GES) (Commission, 2008). These criteria, described by Link et al. (2010) include: (i) measurability – the availability of data to estimate the indicator, (ii) sensitivity – the ability to detect change in an ecosystem, and (iii) specificity – the ability to link the said change in an indicator as a response to a known intervention or pressure.

The idea of using an ecological indicator approach to phyto-PFT modeling is not a new one. For instance, Raitsos et al. (2006) attempted to explain variability in North Atlantic blooms of coccolithophores by identifying their ecological indicators of change obtained from a combination of in situ, satellite and model data. Application of ANNs in the ecological approach to phyto-PFT modeling was pioneered by Raitsos et al. (2008) who estimated probability of diatom occurrence in the North Atlantic from ecological (e.g. sea surface temperature, photosynthetically available radiation,

surface chlorophyll *a* concentration) and geographical (e.g. latitude, longitude) indicators. ANNs based on ecological indicators were also used to simulate the distribution of pCO_2 in the North Atlantic (Telszewski et al., 2009), and to compare patterns of biological production in eastern boundary upwelling regions (Lachkar and Gruber, 2012).

In contrast to these earlier studies, we attempt to use only ecological indicators to simultaneously model biomass distribution of four phyto-PFTs in key biogeochemical provinces, including the open ocean HNLC regions. We hypothesize that our phyto-PFT ecological indicator model (hereafter PhytoANN) will:

- interpret the complex nonlinear interactions between four phyto-PFTs and their ecological indicators in a variety of distinct biogeochemical conditions;
- improve the existing model estimates of monthly climatology and time series distribution of diatoms and coccolithophores in the HNLC regions.

2 Methods

2.1 Source of indicators

We selected the following ecological indicators as principal inputs into the PhytoANN model: (i) sea surface temperature (SST), (ii) wind speed (Wspd), (iii) photosynthetically available radiation (PAR), (iv) surface chlorophyll *a* concentration (Chl), and (v) mixed layer depth (MLD). Although we considered two additional indicators, modeled surface nitrate (NO_3) and surface iron (Fe) concentration, we did not include them in the final PhytoANN because they displayed significant bias with respect to observations in several biogeographic provinces, and they did not add significantly towards explaining patterns of phyto-PFT variability. We assume that changes in NO_3 distribution can largely be explained by associated changes in SST and Chl. This assumption has been used to derive a NO_3 index (Nelson et al., 2004) and NO_3 maps from satellite data alone (Silió-Calzada et al., 2008). We do not consider geographical and time

indicators because they do not fulfill the specificity criterion (latitude, longitude, month of the year) or can be replaced by a similar ecological indicator (daylength by PAR).

SST data came from NOAA's optimum interpolation version 2 product (NOAA-OI-SST-V2) provided by the NOAA/OAR/ESRL PSD in Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/). Wspd data was downloaded from NOAA Ocean Winds (http://www.ncdc.noaa.gov/oa/rsad/air-sea/seawinds.html). The wind speeds were generated by blending observations from multiple satellites. Zhang et al. (2006) describe the details of the Wspd algorithm. PAR and Chl data came from Sea-viewing Wide Field-of-view Sensor (SeaWiFS). We downloaded the processed and gridded (monthly, 9 km by 9 km) data from National Aeronautics and Space Administration (NASA) Ocean Productivity project (http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.par.seawifs.php and http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.chl.seawifs.php). Information about MLD came from the multiyear Simple Ocean Data Assimilation (SODA) model (Carton and Giese, 2008) that attempts to reconstruct the changing physical climate of the global ocean based on a sequential data assimilation approach. A forecast of MLD was derived from the forecast density fields based on a 0.125 potential density criterion (following e.g. Kara et al., 2003). We obtained the processed and gridded data (monthly, 9 km by 9 km) also from NASA Ocean Productivity (http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.mld.soda.php). Information about length of time series and spatial and temporal resolution used in the model analysis is included in Table 1.

2.2 Source of phyto-PFTs

We chose to model four phyto-PFTs: (i) diatoms, (ii) coccolithophores, (iii) cyanobacteria and (iv) chlorophytes. In order to train the PhytoANN model to associate phyto-PFT biomass with environmental conditions, we obtained biomass estimates from NASA Ocean Biogeochemical Model (NOBM) (Gregg et al., 2003). Processed and gridded (monthly, 1° by 1°) data were downloaded from NASA Giovanni (http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=ocean_model). NOBM is a coupled

three-dimensional general circulation, biogeochemical, and radiative model of the global oceans which assimilates SeaWiFS chlorophyll *a*. It was validated using data from in situ (Joint Global Ocean Flux Study (JGOFS)) and satellite data sources (SeaWiFS and MODIS Ocean Color instruments). Biogeochemical processes in the model were controlled by factors such as: circulation and turbulence dynamics, irradiance availability, and the interactions among four phytoplankton functional groups (diatoms, chlorophytes, cyanobacteria, and coccolithophores), four nutrients (nitrate, ammonium, silica, and dissolved iron) and one herbivore group (Gregg et al., 2003). The model architecture is fully described in Gregg (2000) and Gregg (2002). In this model, chlorophytes were intended to represent prasinophytes, pelagophytes, and other flagellates. Cyanobacteria were meant to encompass all pico-prokaryotes. NOBM phyto-PFT results have been thoroughly evaluated against in situ and satellite estimates in Gregg and Casey (2007).

The choice of NOBM model data used for training the PhytoANN was dictated by two factors. First, there is not enough in situ data at a sufficient spatial and temporal resolution that is sampled from the necessary range of biogeochemical provinces. This is despite of the fact that a global atlas of PFT biomass measurements has now been assembled by the MAREDAT project (Smith and Pesant, 2012). Unfortunately, its coverage of diatom (Leblanc et al., 2012) and coccolithophore (O'Brien et al., 2012) measurements is not enough to produce complete monthly climatology maps beyond the North East Atlantic. Second, only by training the PhytoANN on other model phyto-PFT results can we later directly compare two model outputs to independent phyto-PFT estimates within and outside of the training domain, and thus verify the second hypothesis of our study.

Additional independent phyto-PFT estimates come from in situ and remote sensing observations. We used the same set of JGOFS in situ data as Gregg and Casey (2007), assuming that differences in borders of their biogeographic provinces and our regions of interest are small enough to allow for such comparison. Where available,

we referred to new data, for example from the Equatorial Biocomplexity cruises in the eastern equatorial Pacific (Taylor et al., 2011).

Remote sensing phyto-PFT biomass estimates came from a bio-optical algorithm (Hirata et al., 2011, 2013). This algorithm established error-quantified, synoptic-scale relationships between SeaWiFS Chl and ten phytoplankton pigment groups at the sea surface to determine phyto-PSC and phyto-PFT estimates. Phyto-PFTs include diatoms, dinoflagellates, green algae, prymnesiophytes (haptophytes), pico-eukaryotes, prokaryotes and *Prochlorococcus* sp., while phyto-PSCs include micro-, nano- and picoplankton. An earlier version of the PSC component (Hirata et al., 2008) was recently evaluated through a global intercomparison of such algorithms (Brewin et al., 2011).

Remote sensing estimates of PIC concentration may be used to discern regional patterns of seasonal to interannual variability in coccolithophore biomass. PIC data were derived using a two-band algorithm based on normalized water-leaving radiance at 440 and 550 nm (Balch, 2005). Though not directly correlated to living coccolithophore biomass, PIC was established as a good proxy for monitoring changes in the distribution of this PFT (e.g. Balch et al., 2011; Moore et al., 2012). Data were downloaded from the NASA Giovanni portal. Information about length of time series and spatial and temporal resolution used in the model analysis is included in Table 1.

2.3 Spatial and temporal domains

Figure 1 shows the geographical extent of the areas used for training and evaluation (confirmatory analysis) and projection of new states (exploratory analysis). Four Atlantic regions, corresponding to the North East Atlantic Ocean (NEAtl), the Norwegian Sea (NorwSea), the western central Atlantic Ocean (WCAtl) and the Equatorial Atlantic Ocean (EqAtl), were included in both the confirmatory and exploratory part of the analysis. These areas were selected for training because: (i) together they provided a very wide geographical and seasonal range of input and target values sufficient to make the PhytoANN sensitive to most biogeochemical conditions, and (ii) their NOBM PFT biomass estimates were in general in very good agreement with observations available

from the Atlantic basin (Gregg and Casey, 2007). We used a total of 87 000 data points from these regions. Regions used exclusively for exploratory analysis included the eastern equatorial Pacific Ocean (EEP), the Central Subtropical Pacific Ocean (CPac), the Subarctic North East Pacific Ocean (NEPac) and the Antarctic Atlantic Ocean (AntAtl). Information about length of time series and spatial and temporal resolution used in the two parts of model analysis is included in Table 1.

2.4 PhytoANN training and evaulation

In this study we use a basic feedforward ANN which contains inputs, outputs, one hidden and one output layer fully connected via ANN's free parameters (weights and biases) as illustrated in Fig. 2. The feedforward structure means that there is a unidirectional flow of information without feedback from the output back to the input layer. We use supervised learning to train the ANN to interpret complex and nonlinear relationships between ecological indicators (inputs) and phyto-PFT biomass (targets) by iteratively introducing a number of input–output example sets. We carry out the training using a common backpropagation (BP) algorithm that involves the forward and backward phase. During the first phase, free parameters of the network are fixed, and the input signal is propagated through the network layer by layer (Fig. 2). The activation value of each processing unit, called a neuron, is determined by the sum of inputs multiplied by the connection weight. The forward phase finishes with the computation of an error term being the difference between the generated output and the known target. During the backward phase, the error term is propagated through the ANN in the backward direction. It is during this phase that adjustments are applied to weights and biases of the network so as to minimize the error term according to the mean-squared error (MSE) criterion. The adjustment of these free parameters is performed according to the gradient descent rule. The described procedure is common for many ecological applications of ANNs (Lek and Guegan, 1999). We use Mathworks MATLAB ANN toolbox to perform all calculations.

In order to avoid overfitting and assure the generality of our model, we use the early stopping procedure. In this technique, the available data is divided into three subsets: training, validation and testing. The training set (70 % of confirmatory analysis data) is used for computing the gradient and updating the network weights and biases. The validation set (15 % of confirmatory analysis data) is used to monitor the error during the training process because it initially decreases but later typically increases as the network begins to overfit the data. Hence, when the validation error increases for a specified number of iterations (six in this study), the training is stopped, and the weights and biases at the minimum of the validation error are returned. The error of calculated from the testing set (15 % of confirmatory analysis data) is not used during training explicitly. However, it is plotted during the training process to monitor whether the error in the testing set reaches a minimum at a significantly different iteration number than the validation set error. In such cases, it would indicate a poor division of the data set and a need for re-training.

It is not required to select only linearly independent indicators as inputs to the ANN because it performs dimensionality reduction on its own. Similarly, we do not remove any outliers because we assume the ANN is capable of distinguishing between signal and noise after analysing a sufficient number of training examples. We do however need to normalize the input data onto a common min-max scale to prevent indicators with highest absolute values from overpowering the neurons. Moreover, if the indicators or targets have a non-normal distribution, the ANN will produce results biased towards the more populated end of their range. In this study, we thus log-normalize MLD and Chl in the input layer and all four phyto-PFTs in the output layer. All values are then back-transformed to linear scales only after training and simulation are completed.

We also note that climatology and time series results presented in this study are in fact ensemble mean results from 10 PhytoANNs. No two ANNs are exactly the same because they are trained using random weight and bias vector initialization and a random distribution of datapoint indices assigned to training, evaluation or testing. Though individual simulations reveal variable absolute values of Phyto-PFT biomass, their time

series distribution patterns are very robust. Therefore, we consider ensemble means as optimal representations of PhytoANN results.

In Table 2 we provide the details of the PhytoANN architecture common to all ensemble members, chosen as optimal for this study. Results in Table 3 show how increasing the number of neurons in the hidden layer (ANN complexity) increases the fit to the entire confirmatory data (both log-transformed and linear). In theory, the optimum number of neurons depends on the degree of linear independence of the patterns in hidden layer space (Teoh et al., 2006). We note that a good fit is obtained even using the simplest architecture. However, an inspection of time series distributions reveals that not all expected patterns are well captured by a five-neuron ANN. On the other hand, the 10- and 15-neuron nets are likely overfitted to the training data because they do not increase the fit substantially (Table 3). We concluded that the eight-neuron ANN was well-fitted yet general enough to simulate phyto-PFTs, also in a relatively short time (average of 228 s per training).

Table 4 revealed that the Levenberg–Marquardt and the BFGS Quasi-Newton training algorithms provide equally good fits to the data and take little time to perform. Except for the fact that the Levenberg–Marquardt algorithm is more common in feed-forward type of ANNs, its choice here is arbitrary. We relied on vast literature accounts of the usage of transfer and performance functions and did not test the sensitivity to the choice of these functions here.

In summary, we describe the entire procedure with consecutive steps from data selection and processing to PhytoANN model development and its application:

1. Select ecological indicators that fulfill the criteria of Link et al. (2010).
2. Assemble a matrix of input and target data and place them on a unified spatial and temporal grid.
3. Divide all available data between confirmatory and exploratory datasets.
4. Inspect histograms of individual input and target data, and transform them onto a log-10 scale if their distribution is non-normal.

5. Normalize all processed input and target data onto a common minimum–maximum range (e.g. −1 to 1) in order to avoid bias towards high value inputs/outputs.
6. Divide the confirmatory dataset into training (70 %), testing (15 %) and evaluation (15 %) subsets, either randomly or systematically.
7. Set up a feedforward ANN with one input, one hidden layer and one output layer.
8. Select the type of transfer, performance and training function, and initial ANN parameters. Initialize connection weights and biases randomly within the network.
9. Train the network with early stopping.
10. Evaluate ANN performance within confirmatory regions by calculating error statistics.
11. Perform sensitivity analysis on key parameters listed above and retrain the ANN to maximize performance.
12. Apply a total of 10 trained nets to time series from exploratory regions.
13. Normalize phyto-PFT biomass with respect to total Chl to assure conservation of Chl biomass.
14. Compare the ensemble mean output with target as well as independent data.

3 Results and discussion

3.1 Ecological niches of PFTs

The first aim is to investigate how the PhytoANN interprets the interactions between PFTs and their environmental indicators. We break down the problem into four

questions here. Which interactions are linear and which nonlinear? How do the interpretations vary across PFTs? Does the model capture the same relationships that are desrcribed mathematically by the NOBM that was used to train it? How are the weights distributed among the interactions?

We answer the first three questions using information presented in Figs. 3 and 4. By plotting PhytoANN estimates phyto-PFT biomass against an observed range of one input at a time, we detect very nonlinear relationships in all cases, for all phyto-PFTs. However, some linear patterns emerge after we separate the results according to the spatial domain of origin. Consequently, we notice two distinct physical and/or biogeochemical regimes that can be separated into high and low latitude regions. All four phyto-PFTs exhibit similar responses to conditions in the NEAtl, NorwSea, NEPac and AntAtl (hereafter HighLat Regime), but very different type of responses to conditions in the EqAtl, EEP, WCAtl and CPac (hereafter LowLat Regime). In general, HighLat Regime is characterized by relatively lower SST and lower PAR but higher Wspd and higher Chl compared to the LowLat Regime.

Within a single regime, phyto-PFT biomass shows stronger relationships to individual inputs. For example, coccolithopore biomass is positively correlated with PAR in the HighLat Regime (Fig. 3d), and cyanobacteria biomass is highly correlated with PAR in the LowLat Regime (Fig. 4c). Compared to coccolithopores and cyanobacteria, diatoms and chlorophytes exhibit little if any visible relationship with individual indicators, except with Chl. We observe characteristic relationships between Chl and all phyto-PFTs contribution to total Chl [%], which are in agreement with patterns of co-variability between Chl and phyto-PFT contribution [%] derived for bio-optical algorithms (Hirata et al., 2011). Hirata et al. (2011) concluded that Chl is not only an index of phytoplankton biomass but also an index of phytoplankton community structure at synoptic scale. In case of diatoms, we note a mean exponential increase in percentage of total Chl as a function of log-10 of Chl. Our spectrum differs from Hirata et al. (2011) in that there is a larger scatter around the main trend towards the high Chl concentration end. This scatter is associated with lower diatom contributions found in the NEAtl, NorwSea

and partially also the NEPac (Fig. 3i). Comparing the contribution of PhytoANN coccolithophores with haptophytes from the bio-optical model, the relationship differs at the higher end of the Chl range. The PhytoANN estimates that coccolithophores make up a greater proportion of total Chl under bloom conditions observed in the subarctic regions (Fig. 3j). This comparison should be viewed with caution because coccolithopores constitute only a fraction of the larger haptophyte phyto-PFT group considered by Hirata et al. (2011).

When comparing % of total Chl in cyanobacteria in the PhytoANN (Fig. 4i) with *Prochlorococcus* in the bio-optical model (Hirata et al., 2011, their Fig. 2i), we see that maximum percentage contribution is associated with low total Chl in both models. Although the shapes of the two distributions along the Chl gradient are very similar, there are some differences in the magnitude of fractional contribution at lowest Chl values. However, this difference would be minimal if we assumed that NOBM's (and thus PhytoANN's) very broad cyanobacteria group overlaps not only with *Prochlorococcus* but partially also with prokaryotes classified separately in the bio-optical model (Hirata et al., 2011, their Fig. 2g).

Chlorophytes are perhaps most difficult to evaluate because they should to some extent functionally resemble more than one group in the bio-optical algorithm. In the PhytoANN, chlorophytes contribute the most to total Chl under moderate Chl levels (Fig. 4j). Such a pattern is in general close to that of green algae in the bio-optical algorithm (Hirata et al., 2011, their Fig. 2h). The only clear discrepancy occurs for the NorwSea region where the PhytoANN predicts very high percentage contribution of chlorophytes even for the highest Chl values. The comparisons with the bio-optical model therefore suggest that the PhytoANN overestimates the contribution to total Chl of both coccolithophores and chlorophytes in the North Atlantic basin.

It is interesting to compare PhytoANN relationships with those in the NOBM (Figs. A1 and A2). In general, we see that the PhytoANN was able to capture the same general relationships described mechanistically by the NOBM. Considering that NOBM's physical model is partially forced with similar sources of data as used for indicators

of change in the PhytoANN (e.g. SST, wind stress), this result need not be surprising and may suggest that our derived relationships are artefacts of the empirical or mechanistic relationships in the NOBM. Nevertheless, we observe important differences in how phyto-PFTs are distributed along environmental gradients in the two models, especially in areas modeled only during exploratory analysis. The most striking difference is that in the NOBM, NEPac and AntAtl coccolithophores seem to fall into a separate environmental regime, unlike in the PhytoANN. In the NOBM, coccolithophores are poorly correlated with SST, PAR, Wspd and MLD in those regions. Consequently, their large percentage contribution to total Chl is also not associated with high Chl concentration there. This analysis confirms what was previously described by Gregg and Casey (2007), namely that the NOBM does not predict coccolithophore blooms coincident with high diatom biomass concentrations anywhere outside of the nutrient-replete North Atlantic basin.

The results of this analysis also describe the ecological niches of individual phyto-PFTs which can be compared to the traditional phytoplankton mandala. In the PhytoANN, diatom blooms occur under SST between 5 and 15 °C, PAR between 20 and 45 $W m^{-2}$, Wspd between 5 and 10 $m s^{-1}$, and Chl above 0.25 $mg m^{-3}$. Coccolithophores are in general more abundant under low mixing (shallower MLD) regimes – consistent with what we know of their ecology (Balch, 2004). Still, they occupy a similar niche to diatoms. Their biogeographical extent is thus considerably greater than in the NOBM (Gregg and Casey, 2007) or another global dynamic PFT model – PlankTOM (Sinha et al., 2010). High latitude blooms predicted by PhytoANN both in the Atlantic and the Pacific are in agreement with in situ and remote sensing coccolithophore estimates in surface waters, as well as with geological records of coccoliths in bottom sediments (Balch, 2004, and references therein). Cyanobacteria and chlorophytes dominate the LowLat Regime which in general can be described by high SST, high PAR, low Wspd and shallow MLD conditions. Their ecological niche is consistent with Margalef's mandala and numerous field studies which concluded that the intensity of surface blooms of cyanobacteria is regulated by a combination of climatic factors, such

as water temperature, solar radiation, and wind speed (Stal et al., 2003; Whitton and Potts, 2000).

Results from Figs. 3 and 4 also indicate the role of interactions between two or more indicators. For example, under same levels of PAR, there is large build up of cyanobacteria biomass in the LowLat regime but no such buildup in the HighLat. This is clearly because of coinciding differences in SST but likely also variable Wspd and MLD conditions (Fig. 4). Of course these interactions are well known and appear also in the NOBM (Figs. A1 and A2). Yet, it is important to note that the PhytoANN is capable of interpreting these complex and often nonlinear interactions between phyto-PFTs and their ecological indicators because it enables us to verify the first hypothesis of this study.

In order to now say what is the distribution of weights assigned to these interactions, we use a Hinton diagram from one of ten PhytoANNs used to form the ensemble. In Fig. 5 we see that Chl, SST, PAR and Wspd are strongly correlated within a single neuron. MLD appears to be the least significant indicator on its own (but could be significant in combination with others). While it may often be closely associated with Wspd, it can also have an opposite sign assigned, as in the 5th and 7th neuron. We note that in this net, only the 6th and 8th neurons store very similar information about the interactions between inputs. In most other nets included in the ensemble, all eight neurons provide unique information. With respect to connections from hidden layer to the output later, sign and strength of correlations differ from one neuron to another. Note that any one neuron may store important information about one or two PFTs but at the same time provide insignificant information about the remaining phyto-PFTs. This indicates a rather unique phyto-PFT response to various combinations of environmental conditions interpreted by the PhytoANN.

3.2 Annual average phytoplankton community composition

The relationship between phyto-PFTs and ecological indicators also reveals important regional differences in phytoplankton community composition. In Fig. 6 we compare

NOBM and PhytoANN annual mean relative contribution to total Chl biomass, and relate it to any available field estimates. In diatoms, we note that PhytoANN's estimates are in line with those of NOBM in areas where NOBM was used for training. However, significant differences are noted elsewhere. In the NEPac and the AntAtl, PhytoANN predicts less than 20 % annual contribution of diatoms which is much closer to observations compared to NOBM (Fig. 6). In the EEP, although PhytoANN estimates a much smaller diatom contribution relative to NOBM (18 vs. 65 %), it is still twice as high as the less than 10 % observed contribution.

In NOBM, high diatom contribution is almost always at the expense of severely underestimating coccolithophores. This is in contrast to PhytoANN which predicts their 60 % contribution compared to almost 0 % in the NOBM (Fig. 6). It should be noted that NOBM appears to overestimate coccolithophore contribution in two out of four regions used to train the PhytoANN (with two others not evaluated against observations). Hence, it is expected that the PhytoANN overestimates coccolithophores in those and other regions as well (e.g. EEP, NEPac and AntAtl). Nevertheless, the fact that PhytoANN predicts a significant contribution of coccolithophores suggests the existence of suitable ecological niches for coccolithophores in the NEPac and AntAtl, as well as in the EEP.

Similarity of ecological niches of the NEAtl, NEPac and AntAtl are evident both in PhytoANN and NOBM. For example, both models rely on a similar total Chl which is not surprising considering that NOBM assimilates SeaWiFS Chl during its simulation. Therefore, the main difference between the model phyto-PFT distribution in these regions originates primarily from distinct partitioning of Chl between the phyto-PFTs. In NOBM it depends on nutrient uptake and light availability in a dynamical context, while in PhytoANN only on the favourability of environmental conditions.

As for cyanobacteria, the largest difference between the two models is observed in the EEP (Fig. 6). Here, higher biomass estimates from PhytoANN are closer but still significantly lower than observed. On the other hand, though similar to NOBM, PhytoANN biomass estimate is much too low in the EqAtl relative to observations. This

is likely the cause of its underestimate in the ecologically similar EEP. In the WCAtl box, both models predict a proportion of cyanobacteria biomass that is higher than reported from the Bermuda Atlantic Time Series (BATS) (Lomas and Bates, 2004).

In general, both NOBM and PhytoANN struggle to reflect the observed contribution of chlorophytes to total phytoplankton biomass (Fig. 6). We note that NOBM underestimates their contribution in the NEAtl and the WCAtl – regions used to train PhytoANN. We had no available data to compare chlorophytes specifically in the NorwSea and the EqAtl. As for the exploratory regions, PhytoANN projects chlorophyte biomass levels in better agreement with observations. Nevertheless, it underestimates their contribution in the EEP, the AntAtl and the NEPac. In the CPac, it is actually not capable of simulating any chlorophytes, contrary to NOBM. This is likely related to the fact that PhytoANN's training fit is lowest in the WCAtl, which is also ecologically most similar to the CPac.

3.3 Seasonal succession of phyto-PFTs

How do these ecologically driven predictions of phyto-PFT distributions affect what we know about seasonal succession of phytoplankton in these selected regions? In Fig. 7a–d we see that our model can reproduce the monthly climatologies of NOBM phyto-PFTs very well in regions used for training. In the two adjacent boxes of NEAtl and NorwSea, coccolithophores increase substantially in biomass during the spring bloom and become by far the most dominant group by summer time. The spring phytoplankton bloom also shows a large increase in diatom biomass, both in the NOBM and PhytoANN. Chlorophytes make up the most of the smaller fall bloom, while cyanobacteria constitute an insignificant part of the community. Coccolithophores reach their peak in June which is in agreement with field measurements of *Emiliania huxleyi* blooms (Fernandez et al., 1993; Raitsos et al., 2006) and remote-sensing PIC estimates (Balch, 2004) (Fig. 8a). The timing of the peak of the large spring (April) and smaller fall diatom bloom (October) match well with field studies (Edwards and Richardson, 2004; Barton et al., 2013) but also with estimates from the bio-optical model of

Hirata et al. (2011) (Fig. 8a). As expected, the fall bloom, centered around August and September, is mostly made up of chlorophytes. This is also in agreement with longterm observations at these latitudes (Edwards and Richardson, 2004; Barton et al., 2013). The seasonally changing ratio of diatoms to dinoflagellates is thus well represented in the PhytoANN, and is similar to NOBM results (Fig. 8a) and in situ data (Leterme et al., 2005).

PhytoANN predicts that coccolithophore biomass is higher than that of diatoms, even at the peak of the spring phytoplankton bloom. Even though coccolithophore blooms are very abundant in the North Atlantic, especially southwest off Iceland and in the Norwegian Sea, the model is likely to overestimate their contribution here, especially in the spring. In the NorwSea, the bio-optical model of Hirata et al. (2011) shows haptophyte biomass lower than that of diatoms during March–June period (Fig. 8a). On the other hand, the most recently improved PhytoDOAS bio-optical algorithm predicts spring and summer coccolithophore biomass as high as in NOBM and PhytoANN (Sadeghi et al., 2012, their Fig. 7). The summer peak in coccolithophore biomass is also seen in the PIC time series (Fig. 8a). However, the distribution is much narrower around the June peak compared to NOBM and PhytoANN. It also shows a more pronounced decline in biomass in late summer and early fall. Alvain et al. (2008) analyzed monthly climatology results from their bio-optical PHYSAT diagnostic model in the North Atlantic box (40–70° N, 60–20° W) that has a similar latitudinal extent to our NEAtl box. They concluded that diatoms contribute only at most 20 % of total Chl during the spring phytoplankton bloom (their Fig. 8). This does not mean however that the remaining 80 % from nanoeukaryotes explains the really high proportion of coccolithophores in the PhytoANN. One reason for coccolithophore overestimation in the NEAtl and NorwSea might be the fact that both NOBM (Fig. A1j) and PhytoANN (Fig. 3j) associated coccolithophore outbursts with highest Chl values. This is in contrast to the findings of Fernandez et al. (1993) who reported a huge coccolithophore bloom characterized by high PIC levels but relatively low Chl (less than $1\,mg\,m^{-3}$) and particulate organic

carbon concentrations. It is possible that modeling calcifier biomass in general should not be based on units of Chl.

It is noteworthy that the distribution of coccolithophores and diatoms even in a three-month average satellite image is extremely patchy (Sadeghi et al., 2012). Patchiness in phyto-PFT distributions suggests some spatial variability in ecological niches within the North Atlantic biogeochemical domain. Probability of diatom occurrence modeled by Raitsos et al. (2008) also varied greatly across the basin even during a 1 week snapshot image. The recent North Atlantic Bloom Experiment (NABE) in 2008 confirmed the extreme patchiness in patterns of biological productivity, phytoplankton species composition and associated carbon fluxes in this region (Alkire et al., 2012; Mahadevan et al., 2012). The relatively coarse resolution of NOBM data used in this study does not resolve the effects of this patchiness. We are currently testing the hypothesis that, given a higher spatial and temporal resolution of inputs, PhytoANN might generate a more complex image of phyto-PFT biogeography in response to a dynamic physical and biogeochemical environment. The purpose of this study however is to evaluate the performance of our novel approach only in terms of domain-averaged monthly climatologies of four phyto-PFTs.

In the absence of strong seasonality in the EqAtl, the most pronounced feature is the summer to early fall increase in diatoms and chlorophytes (Fig. 7c). Coccolithophores show little annual variability, and cyanobacteria remain at an almost constant biomass level. Here, PhytoANN results match those of NOBM very well. Minimum discrepancies in the timing of the maximum phyto-PFT biomass are insignificant considering the low (less than 0.15 $mg\,m^{-3}$) total Chl biomass levels.

In Fig. 7d PhytoANN exhibits a very different community composition pattern characteristic of the WCAtl. Here, cyanobacteria dominate from summer to winter, while coccolithophores temporarily dominate the community from March until May. Diatoms reach a peak in their relative contribution in March. These results, similar to NOBM, are consistent with observations at BATS, where haptophytes make up between 25 and 46 % of total phytoplankton biomass (Lomas and Bates, 2004). At the BATS

site, haptophytes are mostly composed of coccolithoporids such as *Emiliania huxleyi* (Haidar and Thierstein, 2001). PhytoANN and NOBM predict very little chlorophyte activity under these environmental conditions. However, according to Lomas and Bates (2004), pelagophytes and prasinophytes (but not dinoflagellates) can make up a substantial proportion of the winter/spring bloom population, on average around 15 %. We note that PhytoANN and NOBM do not capture the observed significant interannual variability in the ratio of chlorophytes to cyanobacteria during the bloom season (Lomas and Bates, 2004).

Very low phyto-PFT biomass levels are also estimated for the CPac region (Fig. 7f). PhytoANN captures patterns simulated by NOBM very well. Namely, cyanobacteria are most abundant year-round. Coccolithophores increase slightly in concentration during summer months. Diatoms and chlorophytes have extremely low biomass levels. Compared to the WCAtl box, they show no increase during the winter/spring bloom period.

We find large differences between PhytoANN and NOBM results in all HNLC regions used for exploratory analysis. In the EEP, both models show a similar weak seasonal variability with only a slight increase in early fall (Fig. 7e). The peak can be attributed to a coincident maximum tropical instability wave activity that enhances upwelling of cold, nutrient-rich waters and thus promotes higher new production (e.g. Evans et al., 2009; Strutton et al., 2001; Vichi et al., 2008). PhytoANN predicts chlorophytes to dominate the phytoplankton community, and diatoms and coccolithiphores to be much lower in biomass. This is in marked contrast to the climatology picture from NOBM where diatoms are by far the most dominant functional group, indicating a very different interpretation of their role in the EEP ecosystem. According to the PHYSAT diagnostic and PISCES dynamic models (Gorgues et al., 2010) diatoms rarely dominate in this region. In fact, their contribution to total phytoplankton biomass does not exceed 10 % on average. This long-term low diatom biomass level has also been confirmed by several field estimates conducted at distinct locations and in different times of the year (Blanchot et al., 2001; Dandonneau et al., 2004; Taylor et al., 2011). Average diatom biomass estimated by the bio-optical algorithm is even lower than in PhytoANN (Fig. 8b), and

is thus closest to field observations. Our PhytoANN model predicts no marked differences in the seasonal distribution of coccolithophores which are more abundant than diatoms on an annual average. This is consistent with bio-optical estimates of haptophytes (Fig. 8b). PIC climatology also indicates little seasonal variability and concentrations levels much lower than in the NEAtl (Fig. 8b compared to Fig. 8a).

The NEPac region reveals the largest inconsistencies between models and observations in diatom and coccolithophore monthly climatologies. According to NOBM, diatoms reach highest absolute biomass in NEPac, higher even than in the North Atlantic. They reach their spring peak in May and their fall peak in September (Fig. 7g). Chlorophytes maintain low biomass but flourish in the fall. Maximum chlorophyte biomass level is approximately half that of diatoms during their peak in September. Coccolithophores and cyanobacteria appear not affected by strong seasonality in the environmental forcing and contribute little if at all to total phytoplankton biomass. PhytoANN on the other hand predicts a very different phyto-PFT distribution in response to changing environmental conditions. Here, coccolithophores remain the dominant group throughout the year. Their biomass is highest in May but does not decline much until October. Diatoms peak in May but have drastically lower biomass levels from July to March. In contrast to NOBM, PhytoANN diatoms do not bloom in the fall. In PhytoANN, this late peak is attributed to coccolithophores and chlorophytess.

How does this correspond to what we know about climatology of these phyto-PFTs? Analyzing the seasonality of coccolithophore blooms in the Bering Sea, Iida et al. (2002) found significant interannual variability in the peak areal coverage between 1998 and 2001. The general pattern however suggested a spring peak between April and May and a fall peak between August and September. Up to two month shifts were recorded for exceptionally warm years (1998) and cold years (1999) caused by anomalous ENSO activity in the equatorial Pacific. According to the bio-optical model, haptophytes follow the diatom distribution closely. They exhibit a strong but smaller than diatom increase in biomass in May and a weaker but greater than diatom increase in biomass in November (Fig. 8c). This picture is thus most similar to NOBM. PIC

estimates suggest that coccolithophores are very abundant from March to October, and reach a maximum in September. It appears that both NOBM and the bio-optical algorithm attribute much of the summer and fall Chl to diatoms rather than coccolithophores, in contrast to in situ (Iida et al., 2002) and remote sensing observations (Fig. 8c). We know that the PhytoANN overestimates the proportion of coccolithophores on the annual basis (Fig. 6c), yet, its monthly climatology is closer to the one revealed by PIC and in situ data compared to NOBM and the bio-optical model. It is puzzling however that PIC does not reveal a maximum spring coccolithophore peak seen in Iida et al. (2002) and simulated by all the models. Regardless of the discrepancies between timing and magnitude of diatom and coccolithophore blooms in the NEPac, it is clear that PhytoANN correctly predicts that coccolithophores are at least as dominant as diatoms in the region. Therefore, it provides an improvement over its training model, NOBM, which does not allow coccolithophores to utilize a very favorouble ecological niche. We consider this result the most important evaluation of our PhytoANN ecological indicator model.

In the AntAtl, NOBM predicts an absolute dominance of diatoms that reaches its maximum levels between November and January (Fig. 7h). There is a small chlorophyte increase following the winter diatom bloom, but no seasonal response of coccolithophores. This is in marked contrast to both the bio-optical model and PIC estimates. The bio-optical model shows that haptophytes follow the seasonal distribution of diatoms and that their biomass is higher than diatoms in all seasons except for November–January when they are more or less equal. PIC climatology is similarly characterized by a winter maximum and a summer minimum. We also observe that AntAtl PIC maximum is lower than the boreal summer and early fall maximum in NEPac, and at the same time higher than the boreal winter PIC minimum in NEPac. This observation is consistent with the monthly global climatology analysis of Moore et al. (2012) whose coccolithophore bloom classifier describes eight optical water types from multiple satellite sensors.

PhytoANN's monthly coccolithophore climatology matches well with bio-optical algorithm results, also in relation to other basins. Furthermore, PhytoANN suggests that chlorophytes constitute a substantial portion of total Chl from April to September (close to 50 %). This is in close agreement with another bio-optical model result of Alvain et al. (2008, their Fig. 10) who based on 1998–2006 monthly climatology predict that diatoms contribute at most 52 % to total Chl and that nanoflaggelates (most equivalent to our chlorophytes) contribute over 80 % from April to September in the Southern Ocean (40–70° S, 180°–180° E). We conclude that PhytoANN is able to improve phyto-PFT monthly climatology picture from NOBM in the AntAtl using the knowledge of ecological rules inferred during training from the NOBM itself.

Quantifying coccolithophore biomass has been extremely difficult both in dynamic and diagnostic models. Except for the recently improved PhytoDOAS algorithm (Sadeghi et al., 2012), most bio-optical models show the spatial extent of bloom areas (Alvain et al., 2008; Moore et al., 2012). When analyzing the results of the Dynamic Green Ocean Model, Le Quéré et al. (2005) noted that model calcifiers grow between 40° N and 40° S but they are almost absent poleward of these latitudes. Satellite observations following the method of Brown and Yoder (1994) reveal highest coccolithophore bloom frequencies in the 40–70° latitude band of both hemispheres (Le Quéré et al., 2005, their Fig. 10). Le Quéré et al. (2005) suggests that this is because the traits defined for calcifiers and the zooplankton that graze on them do not give calcifiers a competitive advantage at high latitudes. Compared to other PFTs, they have a lower maximum growth rate, higher light affinity, and lower resistance to darkness. In the NOBM, coccolithophore are very abundant beyond 40° of latitude but only in the North Atlantic. In high latitudes of other basins they are severely outcompeted by diatoms (Gregg and Casey, 2007). Current dynamic model results indicate that we have an insufficient knowledge of either traits of calcifiers (e.g. vital rates), or their protective defenses against zooplankton grazing (Strom, 2002). Even though our PhytoANN considers neither nutrient competition nor zooplankton grazing responses, it provides a more

realistic and ecologically-consistent picture of coccolithophore distribution in the high latitude regions.

3.4 Dramatic shifts in phyto-PFT distribution

We also wanted to check how our diagnostic model performs when it comes to projecting changes under extreme input conditions. Here, we choose to focus on the EEP where interannual El Niño Southern Oscillation (ENSO) contributes the most to temporal variability (Wang and Fiedler, 2006, and references therein). ENSO cycles not only shift total biomass by almost two orders of magnitude but also alter the phytoplankton community composition significantly. In order to really test the PhytoANN response to the extreme 1997–1999 El Niño and La Niña events, we not only excluded the EEP but also the 10.1997–12.1999 period from all data used for confirmatory analysis.

In Fig. 9a we first note that the NOBM captures a sharp decrease in diatoms percentage contribution to total Chl during El Niño. The coincident decrease in chlorophytes contribution is not distinct from the monthly climatological pattern (Fig. 7e). Coccolithophores and cyanobacteria respond positively to the warmer but nutrient poor El Niño conditions. In the PhytoANN, both diatoms and chlorophytes contribute much less to total Chl from October 1997 to summer 1998 than on a long-term average (Fig. 9b). Also, contrary to the NOBM, coccolithophores do not increase their contribution to total Chl in response to El Niño. In turn, cyanobacteria dominate the phytoplankton population as they constitute up to 80 % of total Chl biomass around the peak of El Niño conditions.

In response to the subsequent La Niña event, diatoms in the NOBM restore their high contribution achieving more than 80 % of total Chl by fall 1998. However, we also note that the long-term mean diatom percentage contribution is grossly overestimated. In the NOBM, cocolithophores and cyanobacteria reach their 1997–2005 all time low by the beginning of 1999. In the PhytoANN, we see a different response of the phytoplankton community. All phyto-PFTs return to their long-term average contribution levels but do not reveal the expected equally dramatic changes in phytoplankton

community composition associated with La Niña high Chl blooms. The substantial increase in Chl is mostly attributed to chlorophytes and only secondly to diatoms. Compared to the NOBM results, the PhytoANN suggests more coexistence of phyto-PFTs even under the extreme La Niña conditions. We also note that along the entire time series we see evidence of coccolithophores becoming the dominant group at times (fall 1999 and 2000) in PhytoANN (Fig. 9b). NOBM simulates interannual variability in coccolithophore biomass as well but in different years (fall 2002 and 2003) and never beyond the diatom levels (Fig. 9a). This suggests that the two models also predict different ecological responses of phyto-PFTs to interannual changes in the ecology of the EEP.

In the bio-optical model (Fig. 9c), haptophytes, which include but are not limited to coccolithophores, dominate the biomass spectrum. consistently throughout the entire time series. The sum of *Prochlorococcus* and prokaryotes are on a similar level to cyanobacteria modeled by the PhytoANN, but markedly higher than in the NOBM. Diatoms remain on a very low (less than 10 %) level of contribution to total Chl, in close agreement with in situ observations (e.g. Taylor et al., 2011). The bio-optical model reveals a moderate increase in *Prochlorococcus* and prokaryotes contribution to total Chl during El Niño, and it also captures the large increase in diatom contribution during the subsequent La Niña event. During this time diatoms far exceed their long-term average levels. Such a response was also indicated by in situ (Chavez et al., 1999), some remote sensing (Masotti et al., 2011) and dynamic modeling studies (Gorgues et al., 2010), which report that diatoms first decreased and later increased significantly in response to El Niño and La Niña, respectively.

Based on these results, we can conclude that our diagnostic model appears to be moderately sensitive to extreme environmental perturbations and can detect only some significant temporal shifts in phytoplankton community composition. Neither the NOBM nor the PhytoANN fully capture the observed strong changes in diatom contribution during an extreme ENSO cycle. Both models perform better when simulating changes in response to El Niño rather than La Niña conditions. Higher spatial and temporal

resolution of model estimates would allow to test if our model is also sensitive to high-frequency submesoscale shifts in phyto-PFT distribution attributed to passing tropical instability waves (Palacz and Chai, 2012; Parker et al., 2011).

3.5 Assessment of model limitations

First, the ecological indicators used by PhytoANN do not include nutrients concentration and zooplankton grazing explicitly. Effects of changing nutrient concentrations are largely inferred from changes in SST, as SST and nitrate are often closely correlated, e.g. at the BATS site (Nelson et al., 2004). Including field measurements of Si and Fe as separate indicators is expected to improve our model predictive skill. However, using NOBM nutrient fields as inputs did not improve PhytoANN performance in either confirmatory or exploratory analysis in any significant manner. We speculate that some grazing effects are only implicitly included in our model through considering total Chl as an indicator. However, this does not resolve phyto-PFT-specific grazing pressures. Consequently, PhytoANN's interpretation of ecological rules distinguishing between phyto-PFTs is strongly biased towards bottom-up control processes. We are currently developing a prototype of a similar model that includes feedback from key zooplankton PFTs. It is to be implemented in the North Atlantic basin where basin-wide longterm coverage of Continuous Plankton Recorder data enables such an experiment (Raitsos et al., 2008).

Second, the source of several ecological indicators is common to the NOBM and the PhytoANN. This means that PhytoANN's interpretation of some the relationships between phyto-PFTs and the environment is not necessarily independent from NOBM's. However, this does not hinder the verification of our hypotheses, especially considering the large differences in phyto-PFT distributions in regions used only for exploratory analysis.

Third, we assume that the NOBM phyto-PFT distribution used in training most closely represents the in situ conditions in the Atlantic basin. This assumption is based on the fact that NOBM assimilates SeaWiFS Chl data to correct for total PFT biomass levels

and that it was generally positively evaluated against observations in the North Atlantic (Gregg and Casey, 2007) among other regions. Although in situ data would provide the ideal input, currently available in situ phyto-PFT biomass data, with strong seasonal and geographical bias, is insufficient to represent most of biogeochemical conditions in the open ocean.

Fourth, PhytoANN shows only a surface picture of phyto-PFT biogeography. This limitation will be difficult to overcome considering that all remote sensing indicators provide a surface view themselves, although, there are approaches that try to account for vertical changes in bio-optical algorithms (e.g. Uitz et al., 2006; Brewin et al., 2010).

4 Implications

The results of this study highlight the benefits of using advance statistical techniques to unravel complex and highly non-linear ecological interactions, with implications for biogeosciences and marine ecosystem management alike. We demonstrate that through an artificial neural network we can combine remote sensing and dynamic model results to generate new, ecologically-consistent estimates of phyto-PFT distribution in a wide range of biogeographic conditions. If ecological rules can be extracted from weights assigned to connections within PhytoANN, then we can provide biogeochemically-specific parameterizations of phyto-PFT growth functions which are currently too rigid to capture the global variability in phytoplankton vital rates. This approach should help model the global distribution of silicifiers and calcifiers correctly so that we can reduce the uncertainty on how much atmospheric carbon is being fixed into biomass and how much is being exported into the deep ocean (Francois et al., 2002; Rost and Riebesell, 2004; Sarmiento et al., 2002). In turn, this will improve our future projections of global carbon fluxes and climate mitigation plans.

Unlike other diagnostic PFT models, PhytoANN can be used to make future projections under scenarios of climate-induced changes to key environmental indicators. This is because it takes inputs that are also modeled by most coupled NPZD models

running in forecast mode. In this study, we verify the hypothesis that PhytoANN is able to interpret the complex and nonlinear interactions between phyto-PFTs and the environment, at least to the same extent as the original training model, yet in a fraction of time required to make a dynamic simulation. We speculate that our PhytoANN could be used to interpret similar relationships in an ensemble of coupled models and later applied to future time series of indicators of change. This would provide a novel framework for constructing such ensemble model projections to examine the differences and similarities between them, and eventually lead to better constrained future projections of PFT states.

Furthermore, PhytoANN type of models can further be developed to look at phytoplankton size classes and their dependency on changing environmental conditions, such as temperature. Shifts in community size spectrum in response to rising temperatures are suggested by some studies (e.g. Hilligsøe et al., 2011), yet in others they are shown to depend primarily on total biomass and productivity (e.g. Maranón et al., 2012). Hence, ANNs could prove useful in examining these interactions in the context of variability among biogeochemical provinces rather than global average trends.

Similarly, this approach can be potentially expanded to include higher trophic levels, from zooplankton functional groups to fish species. Depending on data availability among other things, this could first be tested within a single ecosystem. If successful, such a model could provide an alternative framework to the newly proposed General Ecosystem Models (GEMs) that aim at resolving complex and adaptive properties of ecosystems (Purves et al., 2013). ANNs are already heavily relied upon in system control and management applications in other disciplines such as electrical engineering, medical science or business and economics (Anandarajan et al., 2001; Jaeger and Haas, 2004; Khan et al., 2001). Here, we demonstrate that only very few measurable, specific and sensitive indicators of change of phyto-PFTs are sufficient to capture key seasonal to interannual patterns of their distribution. While it is far more difficult to come up with indicators of change of key zooplankton and fish species, such efforts are being currently undertaken under the auspices of the European Union 7th

Framework EURO-BASIN (**E**uropean **B**asin-scale **A**nalysis, **S**ynthesis and **In**tegration) project (http://www.euro-basin.eu/).

Acknowledgements. The authors would like to thank our many collaborators from the EURO-BASIN project for valuable suggestions and criticism at the early stages of this work. We also thank Patrizio Mariani, Thomas Kiorbøe, Ken Andersen, Fi Prowe and Ivo Grigorov from DTU-Aqua for comments and suggestions. The research was (co-)funded by European Union Seventh Framework Programme project EURO-BASIN, European Basin-scale Analysis, Synthesis and Integration in the North Atlantic (Grant Agreement no. 264933).

References

Alkire, M. B., D'Asaro, E., Lee, C., Jane Perry, M., Gray, A., Cetinić, I., Briggs, N., Rehm, E., Kallin, E., Kaiser, J., and Gonzalez-Posada, A.: Estimates of net community production and export using high-resolution, Lagrangian measurements of O_2, NO_3^-, and POC through the evolution of a spring diatom bloom in the North Atlantic, Deep-Sea Res. Pt. I, 64, 157, doi:10.1016/j.dsr.2012.01.012 2012. 8123

Alvain, S., Moulin, C., Dandonneau, Y., and Loisel, H.: Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: a satellite view, Global Biogeochem. Cy., 22, 1–15, doi:10.1029/2007GB003154, 2008. 8105, 8106, 8122, 8127

Anandarajan, M., Lee, P., and Anandarajan, A.: Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks, Intelligent Systems in Accounting, Finance and Management, 10, 69–81, doi:10.1002/isaf.199, 2001. 8132

Anderson, T. R.: Plankton functional type modelling: running before we can walk?, J. Plankton Res., 27, 1073–1081, doi:10.1093/plankt/fbi076, 2005. 8105

Balch, W., Drapeau, D., Bowler, B., Lyczkowski, E., Booth, E., and Alley, D.: The contribution of coccolithophores to the optical and inorganic carbon budgets during the Southern Ocean Gas Exchange Experiment: new evidence in support of the Great Calcite Belt hypothesis, J. Geophys. Res., 116, C00F06, doi:10.1029/2011JC006941, 2011. 8105, 8106, 8111

Balch, W. M.: Re-evaluation of the physiological ecology of coccolithophores, in: Coccolithophores. From Molecular Processes to Global Impact, edited by: Thierstein, H. R. and Young, J. R., Springer, Berlin, 165–190, 2004. 8106, 8118, 8121

Balch, W. M.: Calcium carbonate measurements in the surface global ocean based on Moderate-Resolution Imaging Spectroradiometer data, J. Geophys. Res., 110, C07001, doi:10.1029/2004JC002560, 2005. 8111

Barton, A., Finkel, Z., Ward, B., Johns, D., and Follows, M.: On the roles of cell size and trophic strategy in North Atlantic diatom and dinoflagellate communities, Limnol. Oceanogr, 58, 254–266, doi:10.4319/lo.2013.58.1.0254, 2013. 8121, 8122

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., and Follows, M. J.: Patterns of diversity in marine phytoplankton, Science, 327, 1509–1511, doi:10.1126/science.1184961, 2010. 8105

Beaugrand, G., McQuatters-Gollop, A., Edwards, M., and Goberville, E.: Long-term responses of North Atlantic calcifying plankton to climate change, Nat. Clim. Change, 3, 263–267, doi:10.1038/NCLIMATE1753, 2012. 8105

Blanchard, J. L., Coll, M., Trenkel, V. M., Vergnon, R., Yemane, D., Jouffre, D., Link, J. S., and Shin, Y.-J.: Trend analysis of indicators: a comparison of recent changes in the status of marine ecosystems around the world, ICES J. Mar. Sci., 67, 732–744, doi:10.1093/icesjms/fsp282, 2010. 8107

Blanchot, J., André, J.-M., Navarette, C., Neveux, J., and Radenac, M.-H.: Picophytoplankton in the equatorial Pacific: vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions, Deep-Sea Res. Pt. I, 48, 297–314, doi:10.1016/S0967-0637(00)00063-7, 2001. 8124

Bopp, L., Aumont, O., Cadule, P., Alvain, S., and Gehlen, M.: Response of diatoms distribution to global warming and potential implications: a global model study, Geophys. Res. Lett., 32, L19606, doi:10.1029/2005GL023653, 2005. 8105

Boyd, P. W. and Doney, S. C.: Modelling regional responses by marine pelagic ecosystems to global climate change, Geophys. Res. Lett., 29, 1–4, doi:10.1029/2001GL014130, 2002. 8105

Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., and Peeken, I.: Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data, Biogeosciences, 6, 751–764, doi:10.5194/bg-6-751-2009, 2009. 8105

Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M., and Hardman-Mountford, N. J.: A three-component model of phytoplankton size class for the Atlantic Ocean, Ecol. Model., 221, 1472–1483, doi:10.1016/j.ecolmodel.2010.02.014, 2010. 8105, 8131

Brewin, R. J. W., Hardman-Mountford, N. J., Lavender, S. J., Raitsos, D. E., Hirata, T., Uitz, J., Devred, E., Bricaud, A., Ciotti, A., and Gentili, B.: An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing, Remote Sens. Environ., 115, 325–339, doi:10.1016/j.rse.2010.09.004, 2011. 8105, 8111
Brown, C. and Yoder, J.: Coccolithophorid blooms in the global ocean, J. Geophys. Res., 99, 7467–7482, doi:10.1029/93JC02156, 1994. 8127
Carton, J. and Giese, B.: A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA), Mon. Weather Rev., 136, 2999–3017, doi:10.1175/2007MWR1978.1, 2008. 8109
Cetinić, I., Perry, M., Briggs, N., Kallin, E., D'Asaro, E., and Lee, C.: Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment, J. Geophys. Res., 117, C06028, doi:10.1029/2011JC007771, 2012. 8105
Chavez, F., Strutton, P., Friederich, G., Feely, R., Feldman, G., Foley, D., and McPhaden, M.: Response of the equatorial Pacific Ocean to the 1997–1998 El Nino, Science, 286, 2126–2131, 1999. 8129
Claustre, H., Bishop, J., Boss, E., Bernard, S., Berthon, J.-F., Coatanoan, C., Johnson, K., Lotiker, A., Ulloa, O., Perry, M. J., D'Ortenzio, F., Fanton D'andon, O. H., and Uitz, J.: Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies: potential synergies with ocean color remote sensing, in: Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society, vol. 2, edited by: Hall, J., Harrison, D. E., and Stammer, D., European Space Agency, Rome, ESA Publ. WPP-306, Venice, Italy, doi:10.5270/OceanObs09.cwp.17, 7 pp., 2010. 8105
Commission, E.: Directive 2008/56/EC of the European Parliament and of the Council establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive), Official Journal of the European Union, L164, 19–40, 2008. 8107
Dale, V. H. and Beyeler, S. C.: Challenges in the development and use of ecological indicators, Ecol. Indic., 1, 3–10, doi:10.1016/S1470-160X(01)00003-6, 2001. 8107
Dandonneau, Y., Deschamps, P.-Y., Nicolas, J.-M., Loisel, H., Blanchot, J., Montel, Y., Thieuleux, F., and Bécu, G.: Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific, Deep-Sea Res. Pt. II, 51, 303–318, doi:10.1016/j.dsr2.2003.07.018, 2004. 8124
Devred, E., Sathyendranath, S., and Platt, T.: Delineation of ecological provinces using ocean colour radiometry, Mar. Ecol.-Prog. Ser., 346, 1–13, doi:10.3354/meps07149, 2007. 8105

Edwards, M. and Richardson, A. J.: Impact of climate change on marine pelagic phenology and trophic mismatch, Nature, 430, 881–884, doi:10.1038/nature02808, 2004. 8121, 8122
Evans, W., Strutton, P. G., and Chavez, F. P.: Impact of tropical instability waves on nutrient and chlorophyll distributions in the equatorial Pacific, Deep-Sea Res. Pt. I, 56, 178–188, 2009. 8124
Fernandez, E., Boyd, P., Holligan, P., and Harbour, D.: Production of organic and inorganic carbon within a large-scale coccolithophore bloom in the northeast Atlantic Ocean, Mar. Ecol.-Prog. Ser., 97, 271–285, 1993. 8121, 8122
Follows, M. J., Dutkiewicz, S., Grant, S., and Chisholm, S. W.: Emergent biogeography of microbial communities in a model ocean, Science, 315, 1843–1846, doi:10.1126/science.1138544, 2007. 8106
Francois, R., Honjo, S., Krishfield, R., and Manganini, S.: Factors controlling the flux of organic carbon to the bathypelagic zone of the ocean, Global Biogeochem. Cy., 16, 1087, doi:10.1029/2001GB001722, 2002. 8131
Gorgues, T., Menkes, C., Slemons, L., Aumont, O., Dandonneau, Y., Radenac, M. H., Alvain, S., and Moulin, C.: Revisiting the La Niña 1998 phytoplankton blooms in the equatorial Pacific, Deep-Sea Res. Pt. I, 57, 567–576, 2010. 8124, 8129
Gregg, W.: A coupled ocean general circulation, biogeochemical, and radiative model of the global oceans: seasonal distributions of ocean chlorophyll and nutrients, NASA Technical Memorandum 2000-209965, 33 pp., 2000. 8110
Gregg, W. and Casey, N.: Modeling coccolithophores in the global oceans, Deep-Sea Res. Pt. II, 54, 447–477, doi:10.1016/j.dsr2.2006.12.007, 2007. 8106, 8110, 8112, 8118, 8127, 8131
Gregg, W. W.: Tracking the SeaWiFS record with a coupled physical/biogeochemical/radiative model of the global oceans, Deep-Sea Res. Pt. II, 49, 81–105, doi:10.1016/S0967-0645(01)00095-9, 2002. 8110
Gregg, W. W., Ginoux, P., Schopf, P. S., and Casey, N. W.: Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model, Deep-Sea Res. Pt. II, 50, 3143–3169, doi:10.1016/j.dsr2.2003.07.013, 2003. 8109, 8110, 8152
Haidar, A. T. and Thierstein, H. R.: Coccolithophore dynamics off Bermuda (N. Atlantic), Deep-Sea Res. Pt. II, 48, 1925–1956, doi:10.1016/S0967-0645(00)00169-7, 2001. 8124
Hashioka, T., Vogt, M., Yamanaka, Y., Le Quéré, C., Buitenhuis, E. T., Aita, M. N., Alvain, S., Bopp, L., Hirata, T., Lima, I., Sailley, S., and Doney, S. C.: Phytoplankton competition dur-

ing the spring bloom in four Plankton Functional Type Models, Biogeosciences Discuss., 9, 18083–18129, doi:10.5194/bgd-9-18083-2012, 2012. 8106

Hilligsøe, K. M., Richardson, K., Bendtsen, J., Sørensen, L.-L., Nielsen, T. G., and Lyngsgaard, M. M.: Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO_2 flux, Deep-Sea Res. Pt. I, 58, 826–838, doi:10.1016/j.dsr.2011.06.004, 2011. 8132

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T., and Barlow, R.: An absorption model to determine phytoplankton size classes from satellite ocean colour, Remote Sens. Environ., 112, 3153–3159, doi:10.1016/j.rse.2008.03.011, 2008. 8111

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., and Yamanaka, Y.: Synoptic relationships between surface Chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types, Biogeosciences, 8, 311–327, doi:10.5194/bg-8-311-2011, 2011. 8105, 8111, 8116, 8117, 8122

Hirata, T., Hardman-Mountford, N., and Brewin, R. J. W.: Comparing satellite-based phytoplankton classification methods, Eos Trans. AGU, 93, 59–60, doi:10.1029/2012EO060008, 2012. 8106

Hirata, T., Saux-Picart, S., Hashioka, T., Aita-Noguchi, M., Sumata, H., Shigemitsu, M., Allen, J. I., and Yamanaka, Y.: A comparison between phytoplankton community structures derived from a global 3D ecosystem model and satellite observation, J. Marine Syst., 109–110, 129–137, doi:10.1016/j.jmarsys.2012.01.009, 2013. 8105, 8111, 8155

Holland, J. H.: Hidden Order: How Adaptation Builds Complexity, Helix Books, Addison-Wesley Publishing Company, 1995. 8107

Hutchinson, G. E.: The Paradox of the Plankton, Am. Nat., 95, 137–145, 1961. 8106

Iida, T., Saitoh, S., Miyamura, T., Toratani, M., Fukushima, H., and Shiga, N.: Temporal and spatial variability of coccolithophore blooms in the eastern Bering Sea, 1998–2001, Prog. Oceanogr., 55, 165–175, doi:10.1016/S0079-6611(02)00076-9, 2002. 8125, 8126

Jaeger, H. and Haas, H.: Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, Science, 304, 78–80, doi:10.1126/science.1091277, 2004. 8132

Kara, A., Rochford, P., and Hurlburt, H.: Mixed layer depth variability over the global ocean, J. Geophys. Res., 108, 3079, doi:10.1029/2000JC000736, 2003. 8109

Khan, J., Wei, J. S., Ringnér, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S.: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nat. Med., 7, 673–679, doi:10.1038/89044, 2001. 8132

Lachkar, Z. and Gruber, N.: A comparative study of biological production in eastern boundary upwelling systems using an artificial neural network, Biogeosciences, 9, 293–308, doi:10.5194/bg-9-293-2012, 2012. 8108

Le Quéré, C., Harrison, S., Colin Prentice, I., Buitenhuis, E., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Glob. Change Biol., 11, 2016–2040, 2005. 8106, 8127

Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E., Marshall, H., Peloquin, J., Piontkovski, S., Poulton, A. J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., van Leeuwe, M. A., Varela, M., Widdicombe, C., and Yallop, M.: A global diatom database – abundance, biovolume and biomass in the world ocean, Earth Syst. Sci. Data, 4, 149–165, doi:10.5194/essd-4-149-2012, 2012. 8110

Lek, S. and Guegan, J.-F.: Artificial neural networks as a tool in ecological modelling, an introduction, Ecol. Model., 120, 65–73, 1999. 8112

Leterme, S. C., Edwards, M., Seuront, L., Attrill, M. J., Reid, P. C., and John, A. W. G.: Decadal basin-scale changes in diatoms, dinoflagellates, and phytoplankton color across the North Atlantic, Limnol. Oceanogr., 50, 1244–1253, doi:10.4319/lo.2005.50.4.1244, 2005. 8122

Levin, S. A.: Ecosystems and the biosphere as complex adaptive systems, Ecosystems, 1, 431–436, 1998. 8107

Link, J. S., Yemane, D., Shannon, L. J., Coll, M., Shin, Y.-J., Hill, L., and Borges, M. D. F.: Relating marine ecosystem indicators to fishing and environmental drivers: an elucidation of contrasting responses, ICES J. Mar. Sci., 67, 787–795, doi:10.1093/icesjms/fsp258, 2010. 8107, 8114

Lomas, M. and Bates, N.: Potential controls on interannual partitioning of organic carbon during the winter/spring phytoplankton bloom at the Bermuda Atlantic time-series study (BATS) site, Deep-Sea Res. Pt. I, 51, 1619–1636, doi:10.1016/j.dsr.2004.06.007, 2004. 8121, 8123, 8124

Mahadevan, A., D'Asaro, E., Lee, C., and Perry, M. J.: Eddy-driven stratification initiates North Atlantic spring phytoplankton blooms, Science, 337, 54–58, doi:10.1126/science.1218740, 2012. 8123

Maranón, E., Cermeno, P., Latasa, M., and Tadonléké, R.: Temperature, resources, and phytoplankton size structure in the ocean, Limnol. Oceanogr., 57, 1266–1278, doi:10.4319/lo.2012.57.5.1266, 2012. 8132

Margalef, R.: Life-forms of phytoplankton as survival alternatives in an unstable environment, Oceanol. Acta, 1, 493–509, 1978. 8106

Masotti, I., Moulin, C., Alvain, S., Bopp, L., Tagliabue, A., and Antoine, D.: Large-scale shifts in phytoplankton groups in the Equatorial Pacific during ENSO cycles, Biogeosciences, 8, 539–550, doi:10.5194/bg-8-539-2011, 2011. 8129

Minas, H. J. and Minas, M.: Net community production in high nutrient-low chlorophyll waters of the tropical and antarctic oceans – grazing vs. iron hypothesis, Oceanol. Acta, 15, 145–162, 1992. 8105

Moore, T. S., Dowell, M. D., and Franz, B. A.: Detection of coccolithophore blooms in ocean color satellite imagery: a generalized approach for use with multiple sensors, Remote Sens. Environ., 117, 249–263, doi:10.1016/j.rse.2011.10.001, 2012. 8105, 8111, 8126, 8127

Nelson, N., Siegel, D., and Yoder, J.: The spring bloom in the northwestern Sargasso Sea: spatial extent and relationship with winter mixing, Deep-Sea Res. Pt. II, 51, 987–1000, doi:10.1016/j.dsr2.2004.02.001, 2004. 8108, 8130

O'Brien, C. J., Peloquin, J. A., Vogt, M., Heinle, M., Gruber, N., Ajani, P., Andruleit, H., Arístegui, J., Beaufort, L., Estrada, M., Karentz, D., Kopczyńska, E., Lee, R., Pritchard, T., and Widdicombe, C.: Global marine plankton functional type biomass distributions: coccolithophores, Earth Syst. Sci. Data Discuss., 5, 491–520, doi:10.5194/essdd-5-491-2012, 2012. 8110

Palacz, A. P.: Advancing Ocean Management in the North Atlantic with Integrated Ecosystem Assessments and Artificial Neural Networks: What is the Potential of Using Coupled NPZD model and Satellite-Derived GES Time Series?, Technical University of Denmark (unpublished), 2012. 8107

Palacz, A. P. and Chai, F.: Spatial and temporal variability in nutrients and carbon uptake during 2004 and 2005 in the eastern equatorial Pacific Ocean, Biogeosciences, 9, 4369–4383, doi:10.5194/bg-9-4369-2012, 2012. 8130

Parker, A. E., Wilkerson, F. P., Dugdale, R. C., Marchi, A. M., Hogue, V. E., Landry, M. R., and Taylor, A. G.: Spatial patterns of nitrogen uptake and phytoplankton in the equatorial upwelling zone (110° W–140° W) during 2004 and 2005, Deep-Sea Res. Pt. II, 58, 417–433, doi:10.1016/j.dsr2.2010.08.013, 2011. 8130

Prowe, A. F., Pahlow, M., Dutkiewicz, S., Follows, M., and Oschlies, A.: Top-down control of marine phytoplankton diversity in a global ecosystem model, Prog. Oceanogr., 101, 1–13, doi:10.1016/j.pocean.2011.11.016, 2012. 8105

Purves, D., Scharlemann, J. P. W., Harfoot, M., Newbold, T., Tittensor, D. P., Hutton, J., and Emmott, S.: Ecosystems: time to model all life on Earth, Nature, 493, 295–297, doi:10.1038/493295a, 2013. 8132

Raitsos, D., Lavender, S., Maravelias, C., Haralabous, J., Richardson, A., and Reid, P.: Identifying four puhytoplankton functional types from space: an ecological approach, Limnol. Oceanogr., 53, 605–613, 2008. 8107, 8123, 8130

Raitsos, D. E., Lavender, S. J., Pradhan, Y., Tyrrell, T., Reid, P. C., and Edwards, M.: Coccolithophore bloom size variation in response to the regional environment of the Subarctic North Atlantic, Limnol. Oceanogr., 51, 2122–2130, 2006. 8107, 8121

Rost, B. and Riebesell, U.: Coccolithophores and the biological pump: responses to environmental changes, in: Coccolithophores – From Molecular Processes to Global Impact, edited by: Thierstein, H. R. and Young, J. R., Springer, Berlin, 76–99, 2004. 8131

Sadeghi, A., Dinter, T., Vountas, M., Taylor, B. B., Altenburg-Soppa, M., Peeken, I., and Bracher, A.: Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data, Ocean Sci., 8, 1055–1070, doi:10.5194/os-8-1055-2012, 2012. 8105, 8106, 8122, 8123, 8127

Sarmiento, J., Dunne, J., Gnanadesikan, A., Key, R., Matsumoto, K., and Slater, R.: A new estimate of the $CaCO_3$ to organic carbon export ratio, Global Biogeochem. Cy., 16, 1107, doi:10.1029/2002GB001919, 2002. 8131

Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C., and Maass, H.: Discrimination of diatoms from other phytoplankton using ocean-colour data, Mar. Ecol.-Prog. Ser., 272, 59–68, 2004. 8105

Silió-Calzada, A., Bricaud, A., and Gentili, B.: Estimates of sea surface nitrate concentrations from sea surface temperature and chlorophyll concentration in upwelling areas: a case study for the Benguela system, Remote Sens. Environ., 112, 3173–3180, doi:10.1016/j.rse.2008.03.014, 2008. 8108

Sinha, B., Buitenhuis, E. T., Quéré, C. L., and Anderson, T. R.: Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models, Prog. Oceanogr., 84, 204–224, doi:10.1016/j.pocean.2009.10.003, 2010. 8106, 8118

Smith, W. and Pesant, S. (Eds.): MAREDAT – towards a world atlas of marine plankton functional types, Special Issue, Earth Syst. Sci. Data, 2012. 8110

Stal, L. J., Albertano, P., Bergman, B., von Bröckel, K., Gallon, J. R., Hayes, P. K., Sivonen, K., and Walsby, A. E.: BASIC: Baltic Sea cyanobacteria, an investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea – responses to a changing environment, Cont. Shelf Res., 23, 1695–1714, doi:10.1016/j.csr.2003.06.001, 2003. 8119

Strom, S.: Novel interactions between phytoplankton and microzooplankton: their influence on the coupling between growth and grazing rates in the sea, Hydrobiologia, 480, 41–54, doi:10.1023/A:1021224832646, 2002. 8127

Strutton, P. G., Ryan, J. P., and Chavez, F. P.: Enhanced chlorophyll associated with tropical instability waves in the equatorial Pacific, Geophys. Res. Lett., 28, 2005–2008, 2001. 8124

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, Deep-Sea Res. Pt. II, 49, 1601–1622, 2002. 8105

Taylor, A. G., Landry, M. R., Selph, K. E., and Yang, E. J.: Biomass, size structure and depth distributions of the microbial community in the eastern equatorial Pacific, Deep-Sea Res. Pt. II, 58, 342–357, doi:10.1016/j.dsr2.2010.08.017, 2011. 8111, 8124, 8129, 8152

Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D. W. R., and Wanninkhof, R.: Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405–1421, doi:10.5194/bg-6-1405-2009, 2009. 8108

Teoh, E., Tan, K., and Xiang, C.: Estimating the number of hidden neurons in a feedforward network using the singular value decomposition, IEEE T. Neural Networ., 17, 1623–1629, doi:10.1109/TNN.2006.880582, 2006. 8114

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll, J. Geophys. Res.-Oceans, 111, C08005, doi:10.1029/2005JC003207, 2006. 8131

Vichi, M., Masina, S., and Nencioli, F.: A process-oriented model study of equatorial Pacific phytoplankton: the role of iron supply and tropical instability waves, Prog. Oceanogr., 78, 147–162, doi:10.1016/j.pocean.2008.04.003, 2008. 8124

Wang, C. Z. and Fiedler, P. C.: ENSO variability and the eastern tropical Pacific: a review, Prog. Oceanogr., 69, 239–266, 2006. 8128

Wells, M. L., Vallis, G. K., and Silver, E. A.: Tectonic processes in Papua New Guinea and past productivity in the eastern equatorial Pacific Ocean, Nature, 398, 601–604, 1999. 8105

Whitton, B. and Potts, M.: The Ecology of Cyanobacteria: Their Diversity in Time and Space, Kluwer Academic Publishers, 2000. 8119

Zhang, H. M., Bates, J. J., and Reynolds, R. W.: Assessment of composite global sampling: sea surface wind speed, Geophys. Res. Lett., 33, L17714, doi:10.1029/2006GL027086, 2006. 8109

Table 1. Spatial and temporal charasteristics of the input and target data used in confirmatory and exploratory analysis.

	confirmatory analysis	exploratory analysis
spatial resolution	1°	box-average
temporal resolution	monthly	monthly
time series length	Jan 2000–Dec 2004	Oct 1997–Dec 2004

Table 2. Default PhytoANN parameters chosen for this study.

Parameter	Value
ANN type	Feedforward
#hidden layers	1
#neurons in hidden layer	8
#inputs	5
#outputs	4
1st layer transfer function	tangential sigmoidal
2nd layer transfer function	linear
Learning rate parameter	dynamic
Training algorithm	Levenberg–Marquardt
Data division mode	random, every sample
training : validation : testing [%]	70 : 15 : 15
Generalization scheme	early stopping
Performance function	mean squared error

Table 3. Results of the sensitivity analysis on PhytoANN's number of neurons in the hidden layer. Metrics are: number of training iterations (epochs), training time, correlation coefficient and normalized mean-squared error on log-transformed and normal data. Presented statistics are ensemble means from 5 runs under same configuration but different sample division and weight and bias initialization.

# N	epochs	time [s]	$R_{\log 10}$	R_{lin}	$\text{NMSE}_{\log 10}$ [%]	NMSE_{lin} [%]
5	144	182	0.88 (0.82-0.84-0.84-0.93)	0.51 (0.45-0.64-0.65-0.52)	28 (50-41-40-15)	96 (96-96-99-210)
8	**125**	**227**	**0.89 (0.83-0.88-0.85-0.94)**	**0.54 (0.61-0.83-0.54-0.45)**	**26 (46-29-38-13)**	**81 (84-58-74-110)**
10	150	312	0.90 (0.83-0.89-0.86-0.94)	0.79 (0.61-0.87-0.57-0.73)	25 (45-25-36-13)	47 (110-34-72-76)
15	179	489	0.90 (0.84-0.90-0.87-0.95)	0.81 (0.68-0.88-0.62-0.66)	23 (41-24-33-12)	45 (78-31-65-82)

Table 4. Results of the sensitivity analysis on PhytoANN's type of training algorithm. Metrics are: correlation coefficient and normalized mean-squared error on log-transformed data, number of training iterations (epochs), and training time. Presented statistics are ensemble means from 5 runs under same configuration but different sample division and weight and bias initialization.

Training algorithm	Training fit (R-coeff)	NMSE [%]	epochs	time [s]
Levenberg–Marquardt	**0.89 (0.83-0.88-0.85-0.94)**	**26 (46-29-38-13)**	**125**	**227**
Variable learning rate	0.75(0.47-0.44-0.81-0.44)	76(290-209-54-454)	98	21
Bayesian regularization	0.89 (0.83-0.89-0.85-0.94)	25 (46-25-38-13)	362	644
BFGS Quasi-Newton	0.89 (0.83-0.87-0.85-0.93)	27 (46-31-40-15)	218	90

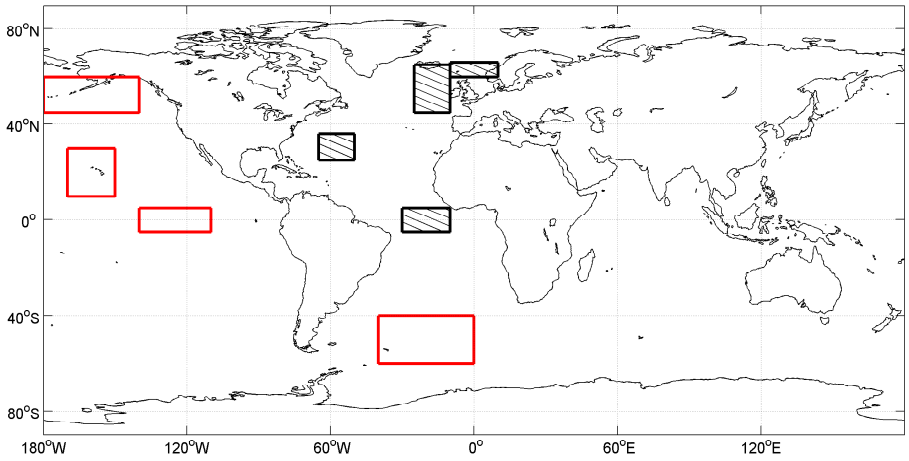

Fig. 1. Location of the areas used for both confirmatory and exploratory analysis (black) and only exploratory analysis (red). Black boxes correspond to the NorwSea, NEAtl, WCAtl and EqAtl when looking from north to south. Red boxes correspond to the NEPac, CPac, EEP and AntAtl when looking from north to south.

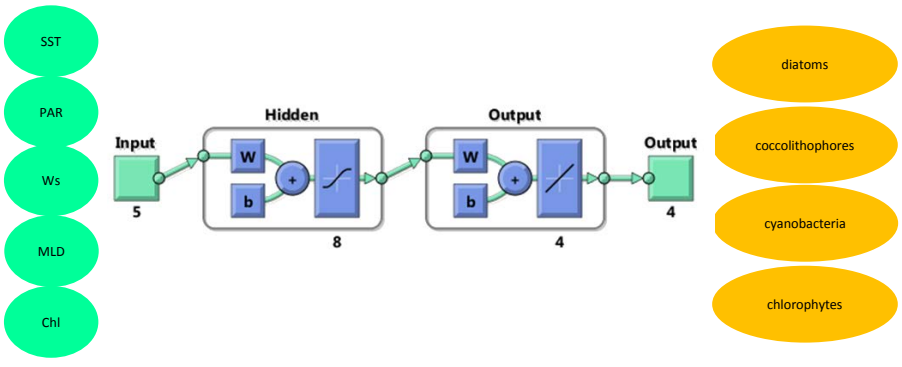

Fig. 2. A conceptual model illustrating how the ANN interprets the relationship between input variables, and between input and target variables. Each hidden neuron in layer 1 computes the sum of input values multiplied by connection weights, and calculates the activation value of each neuron (being a nonlinear sum of all inputs). Each output neuron in layer 2 is a linear sum of the activation values from all layer 1 neurons multiplied by respective connection weights.

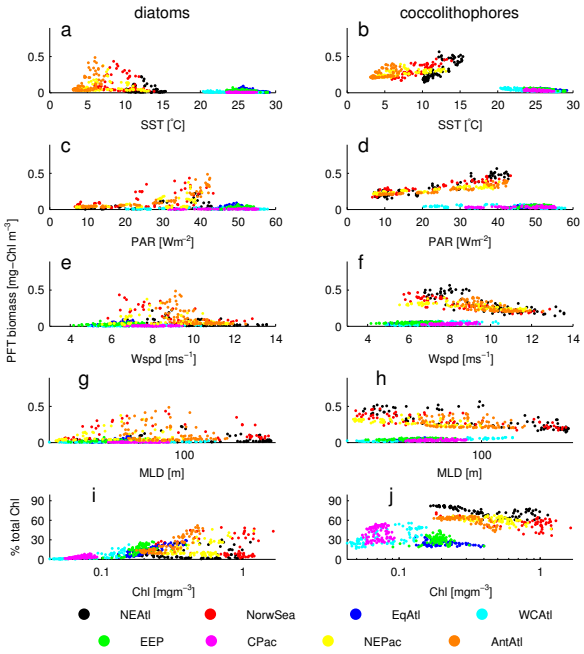

Fig. 3. Scatter plots showing the emerging relationship between PhytoANN diatom (left column) and coccolithophore (right column) biomass [mg Chl m^{-3}] **(a–h)** or % of total Chl **(i, j)** vs. individual ecological indicators color coded according to domain of origin. On the x-axis from top to bottom: SST, PAR, Wspd, MLD, Chl. All data points come from exploratory analysis only, thus they are box-average and monthly average estimates from October 1997 to December 2004.

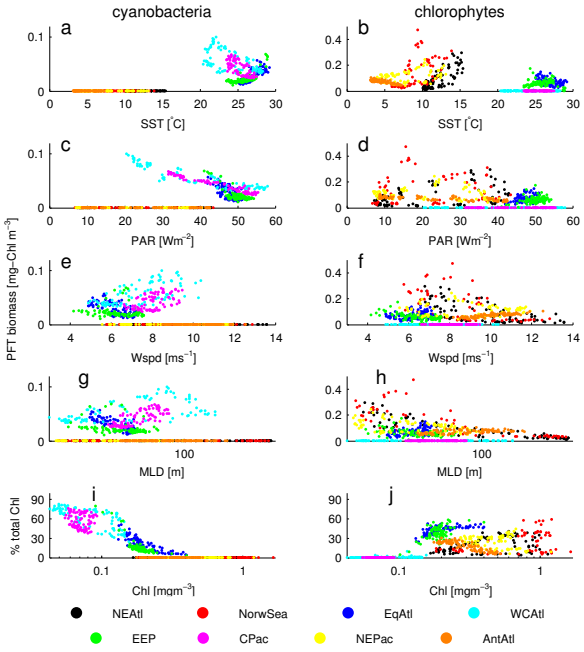

Fig. 4. Scatter plots showing the emerging relationship between PhytoANN cyanobacteria (left column) and chlorophyte (right column) biomass [mg Chl m^{-3}] **(a–h)** or % of total Chl **(i, j)** vs. individual ecological indicators color coded according to domain of origin. On the x-axis from top to bottom: SST, PAR, Wspd, MLD, Chl. All data points come from exploratory analysis only, thus they are box-average and monthly average estimates from October 1997 to December 2004.

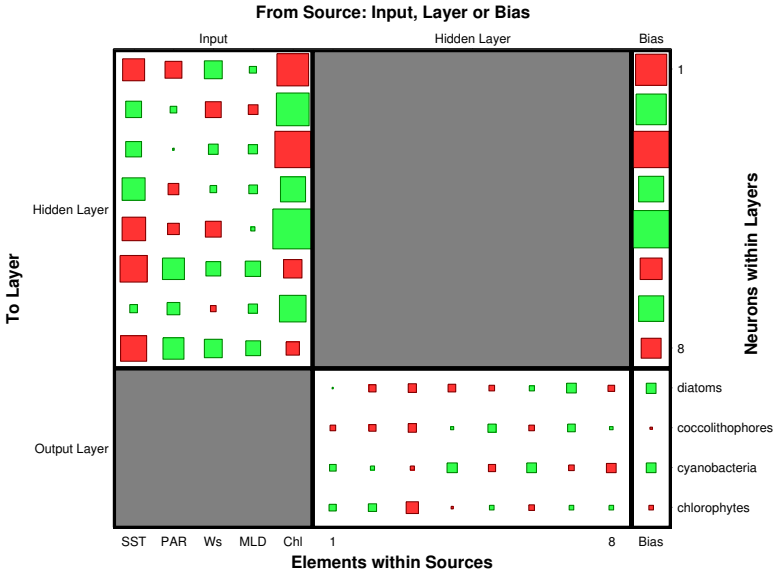

Fig. 5. Hinton-weight diagram depicting relative weights assigned to all connections within one of ten PhytoANNs used to create the ensemble. Green color corresponds to positive correlation, red color negative correlation. Area of the square is proportional to the strength of correlation. Left hand size illustrates the connection weights between the input and the hidden layer. The right hand side illustrates the connection weights between the hidden layer and the output layer.

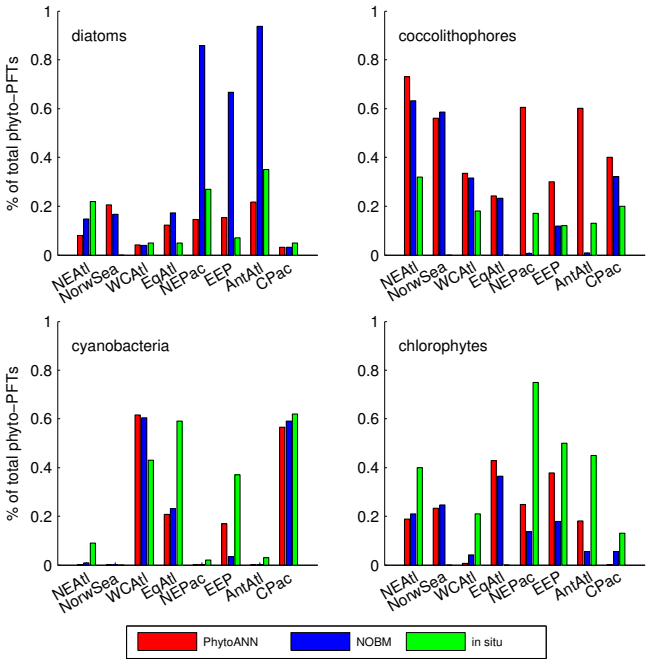

Fig. 6. Approximate box or basin-average comparison between longterm (1997–2004) annual mean percentage contribution to total Chl biomass of diatoms, coccolithophores, cyanobacteria and chlorophytes according to NOBM (blue), PhytoANN (red) and in situ (green) estimates. In situ estimates were derived from both C- and Chl-based measurements. Here, estimates were derived from a combination of data collected by Gregg et al. (2003) (http://gmao.gsfc.nasa.gov/research/oceanbiology/data.php) and EEP data from Taylor et al. (2011, their Table 3).

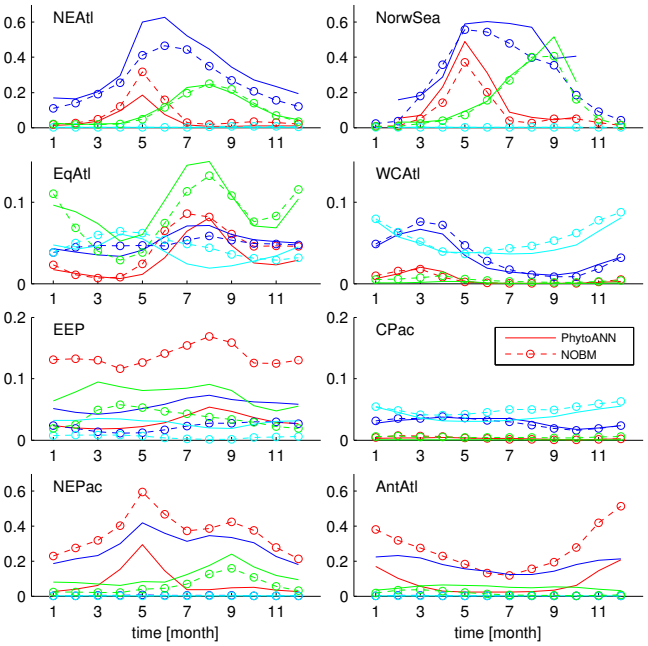

Fig. 7. 1997–2004 longterm monthly climatology of NOBM (dashed lines with open circles) and PhytoANN (solid lines) phyto-PFT biomass estimates from confirmatory (**a** NEAtl, **b** NorwSea, **c** EqAtl, **d** WCAtl) and exploratory areas (**e** EEP, **f** CPac, **g** NEPac, **h** AntAtl). Each plot includes four PFTs: diatoms (red), coccolithophores (blue), chlorophytes (green) and cyanobacteria (cyan). All phytoPFT biomass in units of mg Chl m^{-3}.

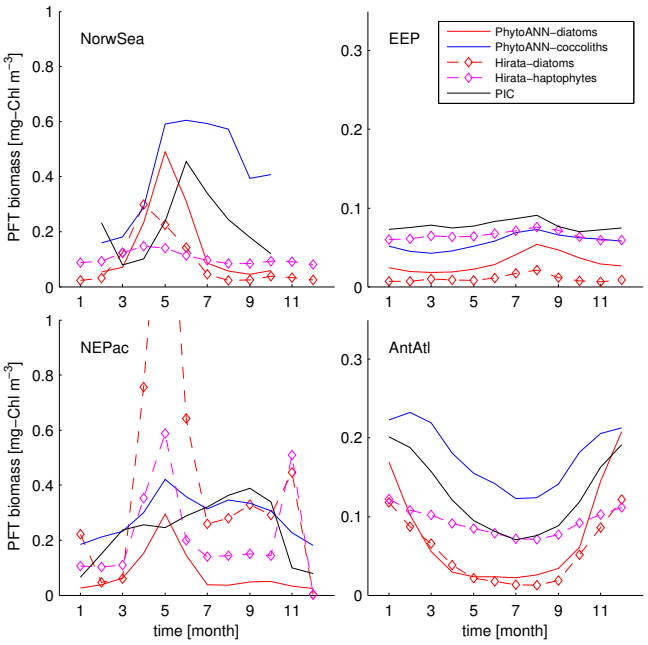

Fig. 8. Longterm monthly climatology of diatom (red) and coccolithopore (blue) biomass from PhytoANN compared with diatoms and haptophytes (magenta) from the bio-optical algorithm as well as PIC from space (black) in four areas: **(a)** NorwSea, **(b)** EEP, **(c)** NEPac and **(d)** AntAtl. PhytoANN and PIC results are marked with a solid line, while bio-optical algorithm with dashed line. PIC values are multiplied by an arbitrary factor of 300 to match scales.

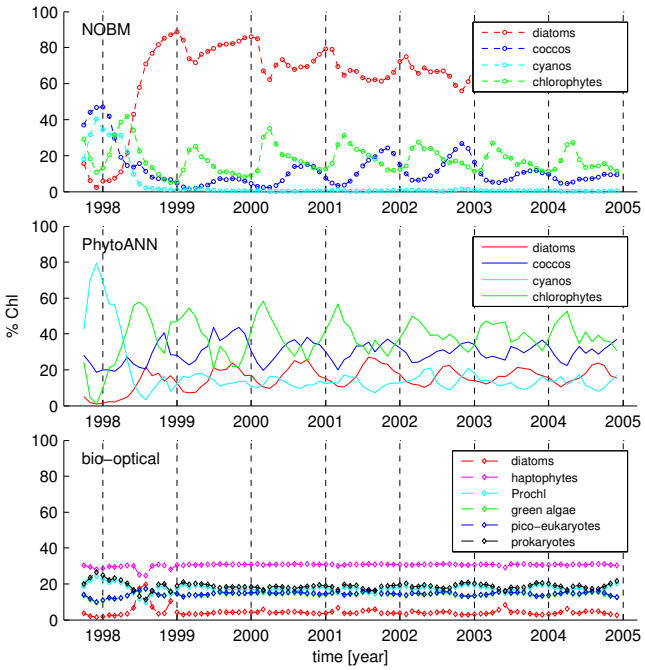

Fig. 9. October 1997–December 2004 time series of the percentage contribution to total Chl of four phyto-PFTs in the EEP from **(a)** NOBM, **(b)** PhytoANN, and **(c)** of six phyto-PFTs from the bio-optical algorithm (Hirata et al., 2013).

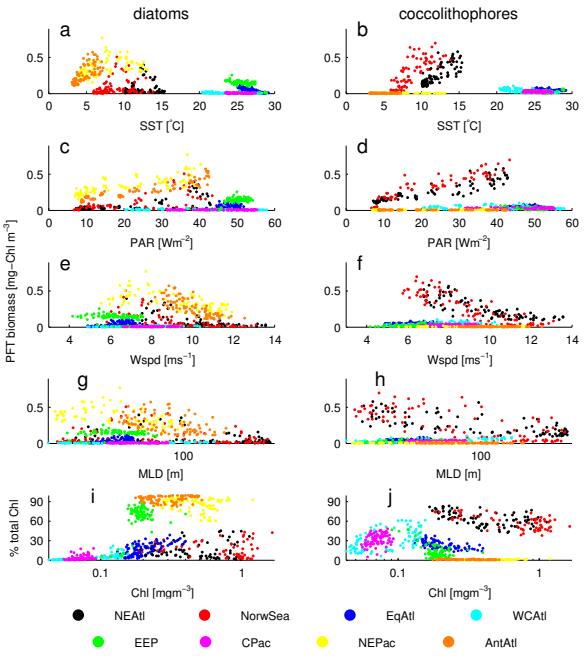

Fig. A1. Scatter plots showing the emerging relationship between NOBM diatom (left column) and coccolithophore (right column) biomass [mg Chl m^{-3}] **(a–h)** or % of total Chl **(i, j)** vs. individual ecological indicators color coded according to domain of origin. On the x-axis from top to bottom: SST, PAR, Wspd, MLD, Chl. All data points come from exploratory analysis only, thus they are box-average and monthly average estimates from October 1997 to December 2004.

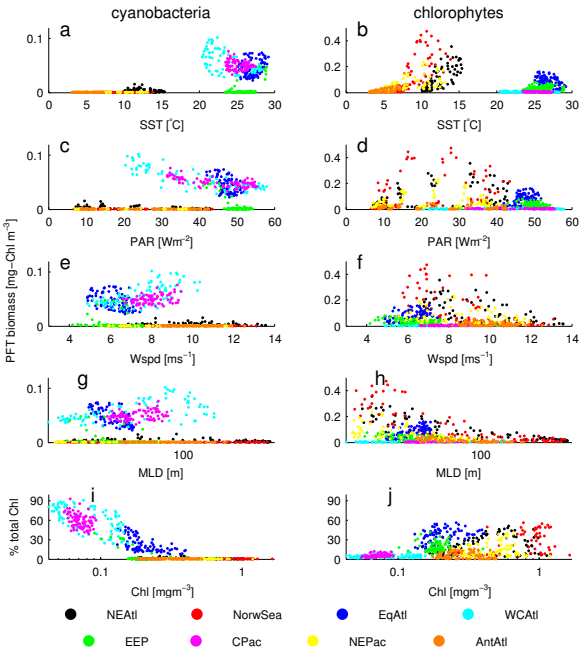

Fig. A2. Scatter plots showing the emerging relationship between NOBM cyanobacteria (left column) and chlorophyte (right column) biomass [mg Chl m^{-3}] **(a–h)** or % of total Chl **(i, j)** vs. individual ecological indicators color coded according to domain of origin. On the x-axis from top to bottom: SST, PAR, Wspd, MLD, Chl. All data points come from exploratory analysis only, thus they are box-average and monthly average estimates from October 1997 to December 2004.