ORIGINAL ARTICLE

Using species distribution modelling to predict future distributions of phytoplankton: Case study using species important for the biological pump

Lasse Ørsted Jensen | Erik Askov Mousing | Katherine Richardson

Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

Correspondence
Lasse Ørsted Jensen, Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
Email: lasseoersted@gmail.com

Abstract

Some phytoplankton species have been predicted to contribute more to the biological pump than others. In this study, we examine the potential of species distribution modelling (SDM) for describing current and predicting future global distributions of two phytoplankton species: the diatom *Chaetoceros diadema* and the coccolithophore *Emiliania huxleyi*. Species distribution models (SDMs) were constructed using species data from the Ocean Biogeographic Information System and environmental layers from the Coupled Model Intercomparison Project Phase 5. The resulting distributions were evaluated by comparing predicted distributions with those found via a literature survey. The developed SDMs were then applied to predict future changes in the distributions of these species using environmental conditions based on the Intergovernmental Panel on Climate Change's Representative Concentration Pathways scenario 8.5 climate scenario, predicted for the year 2100. The model predicts that the total ocean area covered by *C. diadema* and *E. huxleyi* will decline under the examined climate scenario by 8% and 16%, respectively. Furthermore, the future ranges above depths >1,000 m are predicted to decline by 28% for *C. diadema*. As the biological pump is most active in deep ocean waters, this projected decrease in range in the deeper regions of the ocean may have implications for carbon sequestration, particularly for the diatom species. We conclude that SDM appears to be a robust tool for examining phytoplankton distributions. If the predicted changes in ranges of these two species under future ocean conditions are realised, this may result in a reduced contribution of these two phytoplankton species to carbon sequestration via the biological pump.

KEYWORDS
biological pump, *Chaetoceros diadema*, climate change, *Emiliania huxleyi*, phytoplankton, species distribution modelling

1 | INTRODUCTION

Historically, the potential contribution of phytoplankton to ocean carbon cycling has been assessed on the basis of bulk community characteristics such as community size structure, the presence of "ballasted" phytoplankton groups, biomass, etc. (Longhurst & Glen Harrison, 1989). It is, however, becoming increasingly clear that some species contribute disproportionally to carbon transport in the ocean (e.g.,

De La Rocha & Passow, 2007; Rynearson et al., 2013). This being the case, it becomes important to identify tools through which an understanding of the geographical distributions of individual phytoplankton species under different environmental conditions can be developed.

Species distribution modelling (SDM) is often employed to examine the distribution of terrestrial species (Robinson et al., 2011). SDM examines known species presences and absences in an environmental space to determine the niche size of the species in that particular

space. SDM has advanced from the application of simple linear models to machine learning methods and boosted regression trees. Studies have shown that SDM can provide reliable predictions through time and space (Araújo & Rahbek, 2006; Elith & Leathwick, 2009; Martínez-Meyer, Peterson, & Hargrove, 2004). Only recently has this approach been applied to marine species (Robinson et al., 2011) with marine applications of SDM now having been made for several different groups, i.e., fish (Robinson et al., 2011), bacteria (Flombaum, 2013) and, most recently, phytoplankton (Barton, Irwin, Finkel, & Stock, 2016).

Here, we wanted to assess the potential for using SDM for modelling the current and future distributions of two phytoplankton species that have been implicated as contributing disproportionally to vertical carbon transport: the diatom, *Chaetoceros diadema* (Rynearson et al., 2013) and the coccolithophorid, *Emiliania huxleyi* (Buitenhuis, Wal, & Baar, 2001). The aim of the study was, thus, to examine the ability of SDM to describe current distributions of these two species and to examine how the distribution of these may be expected to change under changing environmental conditions.

Owing to their heavy ballasting with silicate and calcium carbonate, respectively, both diatoms and coccolithophores (see Ragueneau, Schultes, Bidle, Claquin, & Moriceau, 2006) have been identified as being important for the transport of carbon from surface to deep ocean waters via the "biological pump" (Longhurst & Glen Harrison, 1989; Volk & Hoffert, 1985). Through this "pump," CO_2, converted to particulate organic carbon through photosynthesis, sinks to below the permanent thermocline and is sequestered. Thus, the biological pump is an important contributor to the global ocean's capacity to act as a carbon sink (Takahashi & Sutherland, 2009) in the global carbon cycle.

Diatoms as a group have been implicated as being important contributors to the biological pump because of their heavy silicate frustules. This phytoplankton group is abundant especially at high latitudes and is a common component of sediment cores (Alvain, Moulin, Dandonneau, & Loisel, 2008; Williams, 1986; Zielinski & Gersonde, 1997). Although positive net export has traditionally been associated with diatoms (Longhurst & Glen Harrison, 1989), not all diatoms appear to contribute equally to this export. Rynearson et al. (2013) followed the development of the phytoplankton population during the onset of a bloom and found a spore-forming diatom identified as *C. diadema* (Ehrenberg) Gran, 1897 to contribute disproportionally to vertical carbon flux in waters southwest of Iceland during the 2008 North Atlantic Bloom Experiment (NABE). Although this species was never found to be dominant in surface waters, its contribution to the diatom carbon present increased with depth and it dominated sediment traps during part of the study period. We chose, therefore, this species as one of the study organisms here.

As the other study organism, we chose the cosmopolitan coccolithophore, *E. huxleyi* (Lohmann) W.W. Hay & H.P. Mohler. The calcium carbonate forming coccolithophores have through geological time also been identified as being important vehicles of vertical carbon transfer in the world's oceans (Buitenhuis et al., 2001); Rost and Riebesell, 2004). As *E. huxleyi* is the dominant coccolithophore in the contemporary ocean (Rost and Riebesell, 2004), it was an obvious candidate for this study.

2 | MATERIAL AND METHODS

The product of the applied modelling procedure is a single ensemble model that can be used to predict the current and future distributions of the two species. The modelling procedure is outlined in Figure 1. All modelling was performed in the free and open source statistical software R (R Core Team, 2014) using the package "biomod2" (Thuiller, Georges, & Engler, 2014). Model input includes species presence points and environmental data (future and present), the details of which are outlined below.

2.1 | Species presence points

Data describing the current geographical distributions of the two species were obtained from the Ocean Biogeographic Information System (OBIS) and originate from various data sources (Table 1; OBIS, 2014). Only recent (since 1990) observations were used in this study. In addition to the OBIS data, our own data from the NABE 2008 (Rynearson et al., 2013) and from a research cruise off Eastern Greenland 2012 were included. In total, 718 data points for *C. diadema* and 9,430 for *E. huxleyi* were found and used as model input.

Step	1	2	3	4
Input	Calibration dataset	Full dataset	Full models	Environmental data
Models	Calibration models	Full models	Ensemble models	Ensemble models
Projections				Current and future projections

FIGURE 1 Modelling procedure. Step 1: the calibration data set consists of species presences, pseudo-absences and current environmental layers, split into two subsets, the evaluation (25%) and the calibration (75%) set. Calibration models are used to define which modelling techniques can be used in the future modelling steps. Step 2: the full models are generated by including the entire data set without the data split. Step 3: full models are combined into one ensemble model. Step 4: the ensemble model uses current and future environmental layers to predict the distribution of the species

TABLE 1 Literature review divided into geographical regions for *Chaetoceros diadema* and *Emiliania huxleyi*

Species	Region	Subregion	Reference
Chaetoceros diadema	Atlantic	Iceland	Rynearson et al. (2013)
		North	Knappertsbusch and Brummer (1995)
		South/Argentina	Popovich and Marcovecchio (2008)
		US East Coast	Durbin, Krawiec, and Smayda (1975)
		US East Coast	Leising, Pierson, Halsband-Lenk, Horner, and Postel (2005)
		US East Coast	Hargraves (1972)
		Hudson Bay/Hudson Strait	Harvey (1997)
		North Sea	Škaloud, Řezáčová, and Ellegaard (2006)
		North Sea	Hoppenrath (2004)
		South Africa	Pitcher (1990)
	Polar Seas	Kara Sea	Sukhanova, Flint, Sergeeva, and Kremenetskiy (2011)
		Bering Sea	Flint, Sukhanova, Kopylov, Poyarkov, and Whitledge (2002)
		North Pole	Katsuki et al. (2009)
		Coastal Canada	Mather, MacIntosh, and Kaczmarska (2010)
		Baltic sea (Kiel Bight)	Wasmund, Göbel, and Bodungen (2008)
	Others	Black Sea	Uysal and Sur (1995)
		Mediterranean Sea	Bouza and Aboal (2008)
		Cosmopolitan, coastal	Baars (1979)
		East China Sea	Jiang et al. (2014)
		Peru	Brodie and Kemp (1994)
Emiliania huxleyi	Atlantic	English Channel	Nielsen (1995)
		Atlantic	Nanninga and Tyrrell (1996)
		French Coast	Lampert, Quéguiner, Labasque, Pichon, and Lebreton (2002)
		North Sea	Richardson and Jackson (2007)
		North Sea	Škaloud et al. (2006)
	Low latitude regions	Mediterranean Sea	Bouza and Aboal (2008)
		The Arabian sea	Mergulhao and Mohan (2006)
	Others	Southern Ocean	Cubillos et al. (2007)
		South Pacific	Danbara and Shiraiwa (1999)
		High latitude regions	Alvain et al. (2008)

2.2 | Environmental data

The environmental data used were taken from the Coupled Model Intercomparison Project Phase 5 (CMIP5), which uses the Intergovernmental Panel on Climate Change's Representative Concentration Pathways scenario 8.5 climate scenario (Pachauri et al., 2014) to estimate and predict a large range of environmental variables from 1850 to 2100. The layers were generated by the Max-Planck-Institute-Earth System Model (MPI-ESM) used in the CMIP5 studies. The MPI-ESM is a global climate model that includes a component describing marine biogeochemical dynamics and simulates the oceanic cycles of 17 variables in the water column, including dissolved inorganic carbon, alkalinity and various nutrients (see Ilyina et al. (2013) for details). The MPI-ESM uses pre-industrial atmospheric greenhouse gas concentrations in the year 1850 and predicts monthly averages for a large variety of physical and chemical variables in the marine environment as a response of increasing CO_2 concentrations from 1850 up until 2100 (see Giorgetta et al. (2013) for details).

Six different environmental variables were used: sea surface temperature (SST), sea surface salinity (SSS), nitrate concentration (NO_3), mixed layer depth (MLD), silicate concentration (Si) and iron concentration (Fe). Each environmental variable was represented by monthly averages from the MPI-ESM predictions for the period 2006–2096, i.e., 1,080 layers in total. The layers were averaged for each 10-years period with "current" variables defined as being the average from 2006 to 2016 and "future" as the average of 2086–2096. All environmental layers have a cell size of 1° × 1° and employ a Mollweide projection.

Owing to the relatively low geographical resolution of the environmental data, it appeared that some of the species presence points were recorded as being on land (especially north of Norway and in the East China Sea). To retain the data in these regions where data were sparse, these points were arbitrarily moved to the nearest cell

containing environmental model estimates. If points on land were in a cluster, the whole cluster was moved by the same distance and direction to keep the relationship between the points intact. No points or clusters were moved more than 1°.

2.3 | Input data

The species distribution model uses one species presence point per environmental cell and, as some cells contained more than one species presence point, this meant that the number of presences used in the model was 122 for *C. diadema* and 452 for *E. huxleyi*.

The model requires both presences and absences but, as true absences of phytoplankton cannot be confidently extracted from the literature, we instead generated 10 pseudo-absence data sets, each with 500 pseudo-absence points, using a random selection conditioned by the absences not being allowed within 1° of a species presence (Barbet-Massin, Jiguet, Albert, & Thuiller, 2012). The models for *C. diadema* used all six environmental layers whereas the model for *E. huxleyi* was run without silicate.

2.4 | Model execution

Using an ensemble modelling approach has proven superior compared to single SDM (Araújo & New, 2007; Garcia, Burgess, Cabeza, Rahbek, & Araújo, 2012; Grenouillet, Buisson, Casajus, & Lek, 2011). Here we use 10 different algorithms to model both species: artificial neural network (ANN), classification tree analysis (CTA), flexible discriminant analysis (FDA), generalized additive modelling (GAM), boosted regression trees (GBM), generalized linear modelling (GLM), multiple adaptive regression splines (MARS), maximum entropy (MaxEnt), random forest (RF) and surface range envelope/bioclimatic variables (SRE/BIOCLIM).

The input data for the modelling execution consisted of a data set comprised of presences/pseudo-absences of the species and the environmental data sets with estimates for current and future oceanic conditions. All of the models were generated using a calibration subset of 75% of the input data set and an evaluation subset of 25%.

The models were initially evaluated by their "true skill statistic" (TSS; Allouche, Tsoar, & Kadmon, 2006). TSS measures the model performance based on sensitivity and specificity and includes commission errors (see Allouche et al., 2006 for details). Figure 2 shows the TSS score for all the initial model runs. Only modelling algorithms with an average TSS score above 0.8 were included in the subsequent modelling procedure. This ensures that the individual models used to build the ensemble models were the best suited and most accurately predicted occurrences. The model techniques passing the TSS test were used in the final modelling steps using the full data set. The resulting models are referred to as "full models" as they use all the presence and absence points and not just the 75% as the evaluation models. TSS was only calculated for the evaluation models, as the full models do not contain a data-split and, therefore, cannot be evaluated by TSS. Instead, the full models were evaluated by their response plots and environmental variables were excluded from the ensemble if they did not generate an effect response.
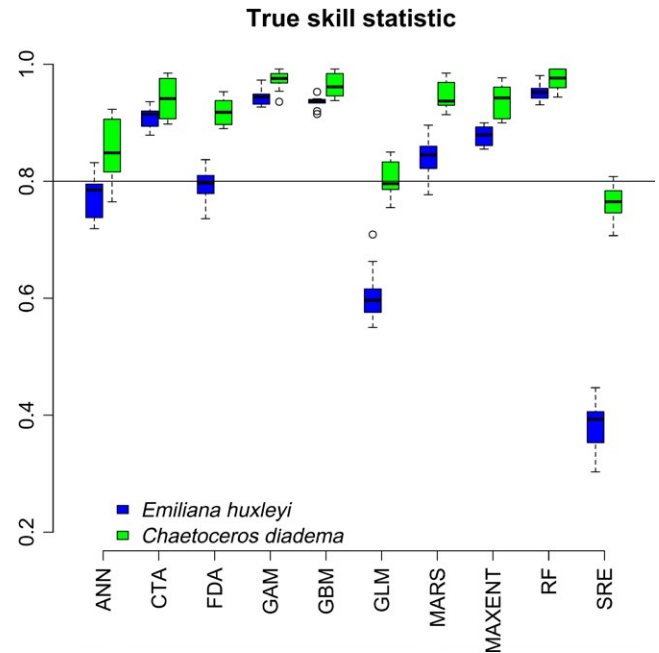

FIGURE 2 True skill statistic (TSS) scores for all model runs (not full models). ANN, artificial neural network; CTA, classification tree analysis; FDA, flexible discriminant analysis; GAM, generalized additive modelling; GBM, boosted regression trees; GLM, generalized linear modelling; MARS, multiple adaptive regression splines; MAXENT, maximum entropy; RF, random forest; SRE, surface range envelope/bioclimatic variables (BIOCLIM). Only models with a TSS score above 0.8 were included in the full/ensemble models

Models passing the above criteria were used to construct an ensemble model that can be used to predict the distribution of *C. diadema* and *E. huxleyi* for any 10-years period between 2006 and 2106. This ensemble model represents the consensus amongst the full models and is presented using a weighted mean of the full models, giving more weight to models with high TSS scores. The model output is a raster layer consisting of cells of 1° × 1°, spanning the globe. The cells contain values between 1 and 100, indicating the probability of finding the species in each cell.

In this study, we were interested in locating areas where *C. diadema* and *E. huxleyi* could be considered common. In order to locate these areas, we defined "common" as cells with a probability of occurrence higher than 20%. This cut-off value was chosen based on density kernel plots showing the number of cells at each probability level and represents a trade-off between the fluctuations in occurrence at low probabilities for *C. diadema* while retaining the resilience of the results (Figure S1). Thus, when choosing values lower than 20%, small differences in this value could significantly affect results and conclusions owing to large variation at this probability level. However, choosing values higher than 20% for both species had no significant influence on the conclusions of this study.

2.5 | Literature review

To investigate the validity of the current model predictions, a literature review was conducted. Using Google Scholar, a search for articles where

C. diadema or *E. huxleyi* were mentioned was performed. Only articles using data sets including the species in question and collected after 1990 were considered. The data sets used in the articles were cross-referenced to the OBIS database from which the model input was collected in order to ensure that the models were not being evaluated using data sets that had been used in model development and calibration.

3 | RESULTS

3.1 | Distribution of *Chaetoceros diadema* and *Emiliania huxleyi* in the contemporary ocean

The current projected distribution of *C. diadema* is centred around Arctic and sub-Arctic regions with some occurrences in upwelling areas, i.e., the Peruvian coast (Figure 3). Table 2 summarizes the distribution based on the literature review. Although some discrepancies between the modelled and actual distributions are noted, there was a high degree of compliance between the modelled and the actual reported distribution of the species.

The model predicts *E. huxleyi* to currently be associated with coastal zones with the highest occurrences being around major up-welling zones, i.e., Peru, the Benguela Current and the American West Coast (Figure 4). The literature review, by contrast, suggests that *E. huxleyi* has a cosmopolitan distribution (Table 1). We note, however, that the model-predicted distribution of *E. huxleyi* is largely in agreement with the reported global distribution of coccolithophorid blooms noted by remote sensing (Alvain et al., 2008). As *E. huxleyi* is the most

dominant coccolithophorid in the global ocean (Hagino et al., 2011), it seems fair to assume that most if not all of the blooms recorded by Alvain et al. (2008) represented mass occurrences of *E. huxleyi*. If we make this assumption, then the model seems reasonably good at identifying regions in which blooms of this species may occur.

3.2 | Distribution of *Chaetoceros diadema* and *Emiliania huxleyi* in the future ocean

The model scenario for the future distribution of *C. diadema* shows a trend of northward migration in the distribution of this species with an increasing probability of finding it in the Arctic Sea and declining probability at lower latitudes (Figure 3). The area of the range (number of cells with a predicted chance of occurrence above 20%) of *C. diadema* declines by 8% under the predicted future ocean conditions. The predicted future distribution of *E. huxleyi* also exhibits a migration towards colder regions and a decline near the equator (Figure 4). The distribution range of *E. huxleyi* is much wider than that of *C. diadema* and the model predicts that it will, in the future, be distributed in all of the world's oceans and in almost all temperature regimes. The model predicts *E. huxleyi* to be associated with coastal areas and especially concentrated around the major upwelling zones (Peru, the Benguela Current, American West Coast). The range of *E. huxleyi* declines in total by 17%.

The probability of finding *C. diadema* above deep ocean areas is also predicted to diminish in the future. For this species, the number of cells with a probability above 20% of being located in areas with water
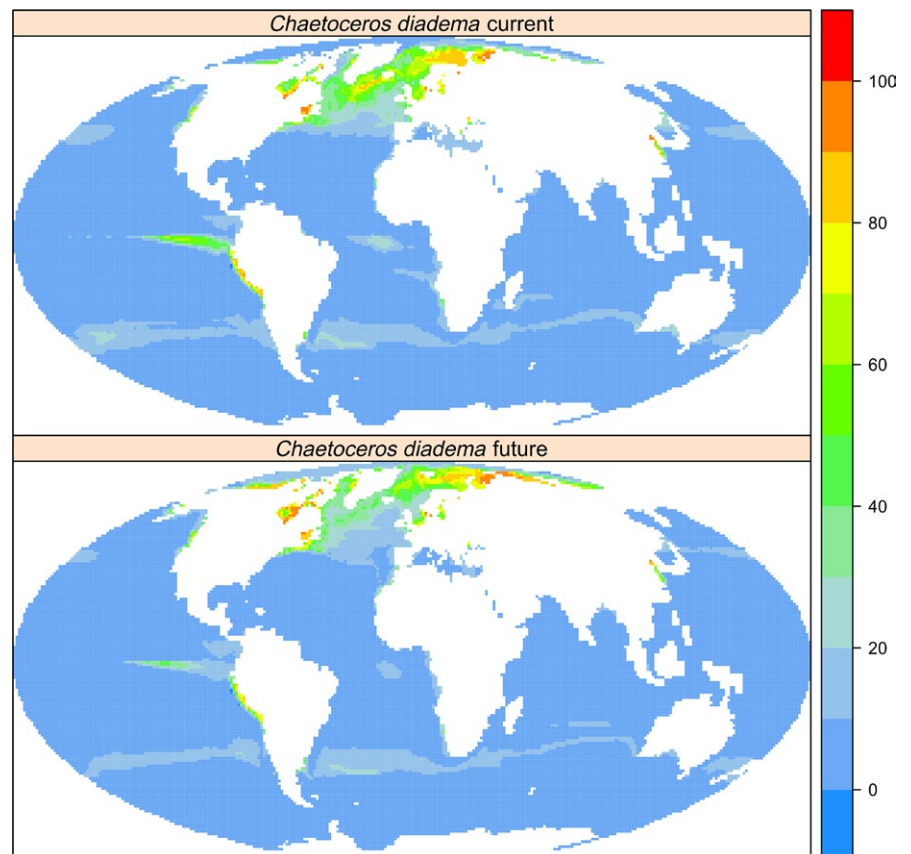

FIGURE 3 Current and future distribution of *Chaetoceros diadema*

TABLE 2 Species distribution modelling (SDM) data sources

Species	Data set	Geographical region
Chaetoceros diadema	MADAGASCAR National Oceanographic Data Centre – Diatoms	Madagascar
	Phytoplankton from the White Sea, Barents Sea, Norwegian Sea and Arctic Basin 1993–2003	White Sea, Barents Sea, Norwegian Sea and Arctic Basin
	Biogeographical data from BODC – British Oceanographic Data Centre North Atlantic	Atlantic and Polar Oceans
	Bay of Fundy Species List	North Atlantic/Bay of Fundy
	MICROBIS database	Global
	Phytoplankton Monitoring Network	Atlantic Ocean/French coast
	ICES biological community data set	North-East Atlantic, Greenland Sea, Baltic Sea and Norwegian Sea
	KOBIS database	Korea
	World Ocean Database 2009	Global
	Atlantic Zone Monitoring Program (AZMP) DFO Quebec region plankton data set	Canadian coasts
	JODC data set	Japan
	Phytoplankton of the Ukrainian Black Sea shelf (1985–2005)	Black Sea
	Phytoplankton Universidad Arturo Prat	Eastern Pacific
	Taxonomic Information System for the Belgian coastal area	North Sea
	REVIZEE South Score/Pelagic and Demersal Fish Database	South-West Atlantic
Emiliania huxleyi	Electron Micrograph Database	Antarctic
	Phytoplankton from the White Sea, Barents Sea, Norwegian Sea and Arctic Basin 1993–2003	Arctic Oceans
	Biogeographical data from BODC – British Oceanographic Data Centre Atlantic and Polar Oceans – various data from PANGEA	Global
	MICROBIS database	Global
	ICES biological community data set North-East Atlantic, Greenland Sea, Baltic Sea and Norwegian Sea	North-East Atlantic, Greenland Sea, Baltic Sea and Norwegian Sea
	World Ocean Database 2009	Global
	Australian National Algae Culture Collection	Global
	Atlantic Zone Monitoring Program (AZMP) DFO Quebec region plankton data set	Atlantic Ocean
	Phytoplankton of the Ukrainian Black Sea shelf (1985–2005)	Black Sea
	L4 Plankton Monitoring Programme	English Channel
	SNU-FF Phytoplankton and zooplankton (1999–2010)	Black Sea
	Continuous Plankton Recorder (Phytoplankton)	North Atlantic and North Sea
	Taxonomic Information System for the Belgian coastal area	North Sea
	Phytoplankton of the northern part of the Black Sea 1992–1993	Black Sea
	Nivå Bay species list	Baltic Sea/Øresund

DFO, The Departement of Fisheries and Oceans Canada; ICES, International Council for the Exploration of the Sea; JODC, Japan Oceanographic Data Center; KOBIS, Korea Ocen Beogeographic Information System; REVIZEE, Programa Nacional de Avaliação do Potencial Sustentável de Recursos Vivos da Zona Econômica Exclusiva; SNU-FF, Sinop University Fisheries Faculty.

depths >1,000 m declines by 28%, whereas there was no significant change in the occurrence of *E. huxleyi* above deep waters (Figure 5).

3.3 | Importance of different variables

For both species, the variation in model variable importance was high for the environmental variables and generally less pronounced for the nutrients nitrate, silicate and iron. SST and MLD are identified as being important for the model predictions of the species' distributions

whereas iron and SSS appear to be more important to the model describing the distribution of *E. huxleyi* than that describing the distribution of *C. diadema* (see Figure 6).

4 | DISCUSSION

This study demonstrates that species distribution modelling is a viable approach for describing current distributions of the two
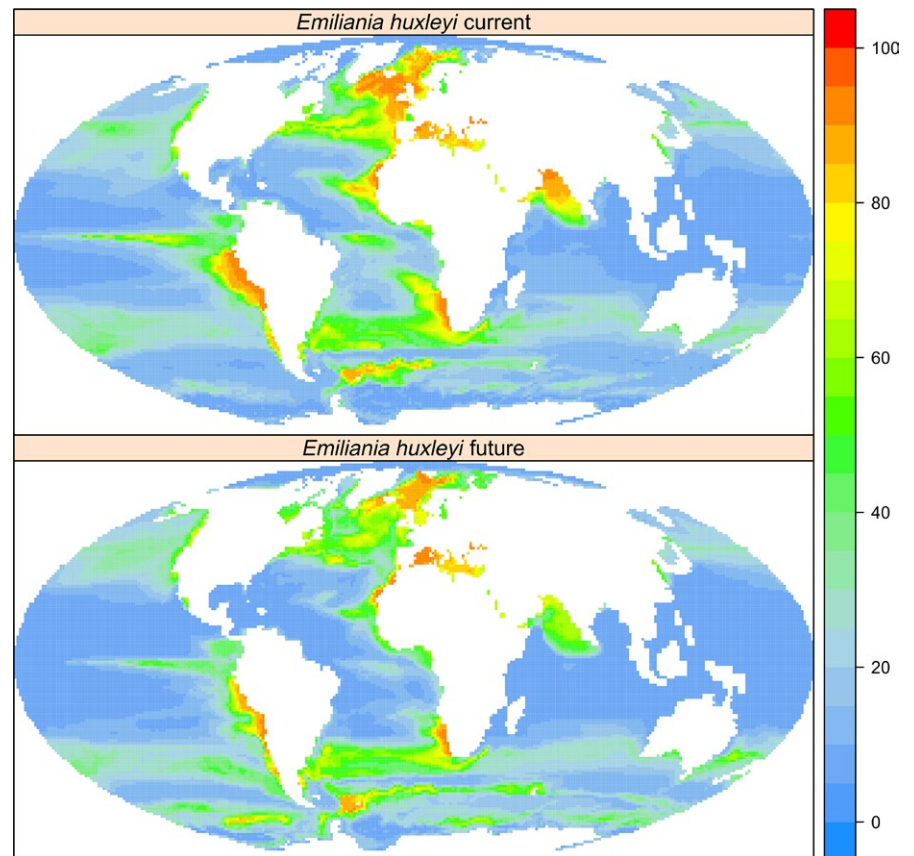
FIGURE 4 Current and future distribution of *Emiliania huxleyi*
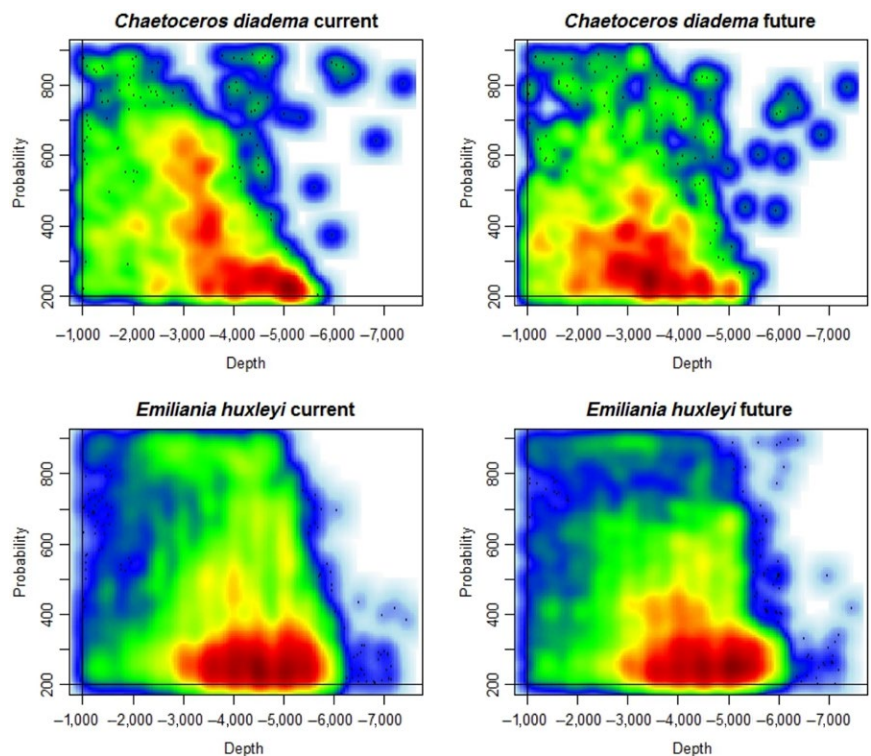

FIGURE 5 Heat map showing the current and future density of grid cells for the probability of occurrence of *Chaetoceros diadema* and *Emiliania huxleyi* above different depths

phytoplankton species studied here. The known distributions of these two species are very different, with *C. diadema* being mainly found in northern-to-Arctic latitudes and *E. huxleyi* exhibiting a cosmopolitan distribution. When we had established that the developed models could competently describe the current distributions of these two species, we then applied predicted future
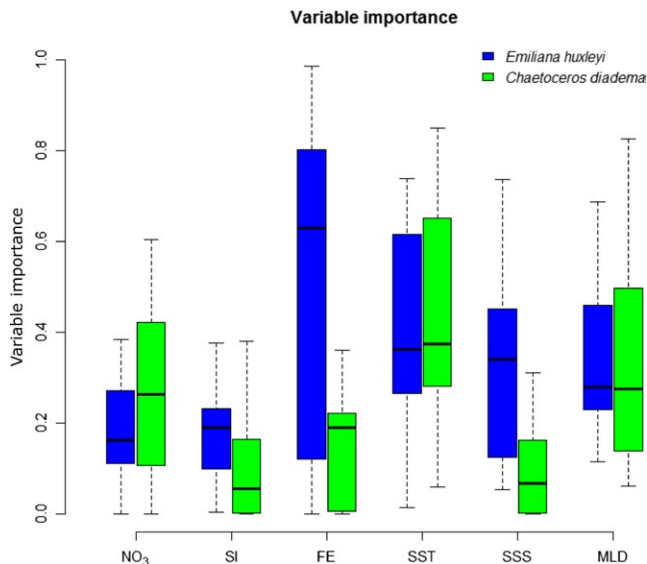
FIGURE 6 Variable importance in the full model for *Chaetoceros diadema* and *Emiliania huxleyi*. Variables are NO$_3$, nitrate; SI, silicate; FE, iron; SST, sea surface temperature; SSS, sea surface salinity; and MLD, mixed layer depth

environmental conditions and re-ran the models. These model runs predict a decline in the range of both study species as well as a reduction of habitat for *C. diadema* in ocean areas deeper than 1,000 m as a result of climate change. Such changes in the distribution of these species could, potentially, lead to an overall reduction in the vertical carbon transport mediated by these two species and, ultimately, potentially influence the global ocean's ability to sequester carbon.

Although literature reports of the distribution of these two species generally support the model predictions for the current distributions of these species (Figures 3 and 4), some discrepancies were noted. Most striking amongst these was the model prediction of a high presence of *C. diadema* in the Baltic Sea, where actual observations do not support this conclusion. We note, however, that the steep salinity gradient found in the Baltic Sea (ranging from 0 to 15) would decrease the likelihood of finding the species here (Brand, 1984). Brand (1984) tested 46 different marine phytoplankton for salinity tolerances, including four *Chaetoceros* sp. and *E. huxleyi*, and found the *Chaetoceros* sp. and *E. huxleyi* would not be able to survive in waters with salinity below 15. Likewise, McQuoid (2005) found that resting spores of *C. diadema* did not germinate at salinities below 15. The model predicts *C. diadema* to have a high chance of occurrence in the Baltic Sea, where the salinity can reach as low as 0 in the northernmost reaches.

This discrepancy between modelled and actual data where the model suggests *C. diadema* to occur in the Baltic may be an artefact caused by the temporal variability in the salinity variable used as model input: the vast majority of *C. diadema* presences used to build the model derive from the Barents and Kara Seas, on the northern coast of Russia. This area is subject to high river runoff in the early spring, resulting in significantly lower salinity in the surface water (Midttun, 1985) at this time than during the rest of the year. As the

salinity variable used to build the model is a 10-years average, the cells close to the Russian shore would have lower average salinities than the actual salinity at the time of sampling. The model interprets this as a high tolerance towards low salinities, resulting in a predicted distribution with low regard to the salinity variable (low variable importance). Future modelling attempts could be conducted with a higher temporal resolution (e.g., summer months only) to overcome this shortcoming.

Although there is a reasonably good agreement between the actual and the modelled distributions of these species, it is important to remember that, although the environmental variables used in the SDM are all known to have physiological implications for phytoplankton in general (e.g., Brand, 1984; Kiørboe, 1993; Mousing, Ellegaard, & Richardson, 2014), they were chosen on the basis of their being readily available, and not necessarily on the basis of an a priori expectation that all of them would have particular physiological importance for the species chosen in this study. In the models, the three most important variables for describing the distribution of *E. huxleyi* were iron concentration, surface temperature and surface salinity. For *C. diadema*, the three most important were surface temperature, mixed layer depth and nitrate concentration (Figure 6). This suggests that the two species respond differently to environmental conditions. However, species distribution models (SDMs) are purely correlative and we can, therefore, not be certain that the organisms are responding to the variables that we put into the model as it is just as likely that they are responding to something that correlates with these variables.

It is, for example, unlikely that mixed layer depth, in itself, has any physiological importance for the phytoplankton. Therefore, MLD must serve in the model as a proxy for another parameter, i.e., light availability or turbulence. Thus, we can conclude that the model results are consistent with the known fact that these two species are distributed differently with respect to temperature and salinity. Beyond that, the results suggest that water column stability (MLD) is important for both species and that these species are distributed differently in relation to ambient inorganic nutrient concentrations. We cannot, however, use these results to deduce the physiological requirements of the different species and we do not know that we have identified the most important environmental variables for describing the distributions of these species.

4.1 | *Chaetoceros diadema* and *Emiliania huxleyi* in a future ocean

The model predicts that there will be a high probability in the future of finding *C. diadema* in the Beaufort Sea, north of Canada and Alaska. Currently, sea ice covers this area during spring but the Representative Concentration Pathways scenario 8.5 climate scenario predicts total Arctic sea ice in September to disappear around 2060 and March sea ice around 2200 (both mean dates) (Hezel, Fichefet, & Massonnet, 2014). Considering the changes in sea-ice cover, it does seem plausible that *C. diadema* could have a realized niche in these areas by 2100.

The changing irradiance conditions in the polar seas could also have an effect on the proliferation of the phytoplankton here. As the

predicted phytoplankton distribution moves north it will experience periods of constant light in the summer, which might affect the efficiency of photosynthesis. Humphrey (1979) found a small increase in growth rate for *Chaetoceros didymium* under constant light, whereas others have observed no significant differences between *Chaetoceros* cells grown in constant light compared to cells in normal 12 hr light: 12 hr dark cycles (Toro, 1989). Katsuki, Takahashi, Onodera, Jordan, and Suto (2009) found high abundances of *Chaetoceros* spp. in close vicinity to the North Pole, including *C. diadema*. These observations suggest that *C. diadema* will probably be able to proliferate in the very high latitudes as suggested by the model here.

An important aspect of any SDM is whether the modelled species is able to move into the predicted niche, the so-called "accessible area" (Barve et al., 2011). Although there is currently much scientific debate concerning if and when phytoplankton are dispersal limited (e.g., Mousing, Richardson, Bendtsen, Cetinic, & Perry, 2016; Robinson et al., 2011), given the long time scale of this study, we do not consider it likely that dispersal barriers would prevent the species studied here moving into the new areas identified by the models. Based on the Atlantic Meridional Transect programme, Chust, Irigoien, Chave, and Harris (2013) showed that the environmental conditions were more important when selecting for community assembly than distance displaced for major groups of phytoplankton in the Atlantic. In addition, we note that *C. diadema* has high dispersal potential, as its resting spores can survive in cold and dark conditions over many months (French & Hargraves, 1980; Ragueneau et al., 2006; Smetacek, 1985). This means that spores plausibly could be transported north by the Norwegian current and be dispersed once they reach the Arctic Ocean.

The future pattern in the predicted distribution of *E. huxleyi* is not as radically different from the current distribution compared to the current versus future patterns of *C. diadema* and the projected changes in distribution for *E. huxleyi* relate primarily to a decline in some of the areas where it is found today. This might be expected as *E. huxleyi* has been demonstrated to have a much greater temperature tolerance than *C. diadema* and to be relatively warm-loving. Fisher and Susomo (1988/89), reported a temperature tolerance of 8–27°C for *E. huxleyi*, whereas a field study (Karentz & Smayda, 1984) have reported *C. diadema* to occur at temperatures between 0.9 and 10.2°C. Laboratory studies indicate a somewhat wider range of temperature tolerance for *C. diadema* (1.5–>17°C; Baars, 1979, 1981) but still suggest that this species is less likely to thrive at high temperatures than *E. huxleyi*.

4.2 | Directions for future research

Species distribution modelling is demonstrated here to have some promise for projecting the future distributions of phytoplankton species and its potential should be further developed both through refinement of the modelling approach and through its application to other species. For *E. huxleyi*, it would, for example, be obvious to include pH as a variable as the ability of this species to calcify is highly dependent on ocean acidity (Beaufort et al., 2011). However, pH has not been a standard environmental parameter measured when phytoplankton samples have been taken and the physiological responses of this species to changes in pH are not well enough constrained to be used in a model such as ours.

It should also be noted that SDM is sensitive to variations in sampling effort (Anderson, 2012). In the OBIS database, there is a bias towards higher sampling intensity in the Northern Hemisphere and, especially, in the coastal waters of Europe and North America. This might lead to a disproportionate probability of occurrence with high
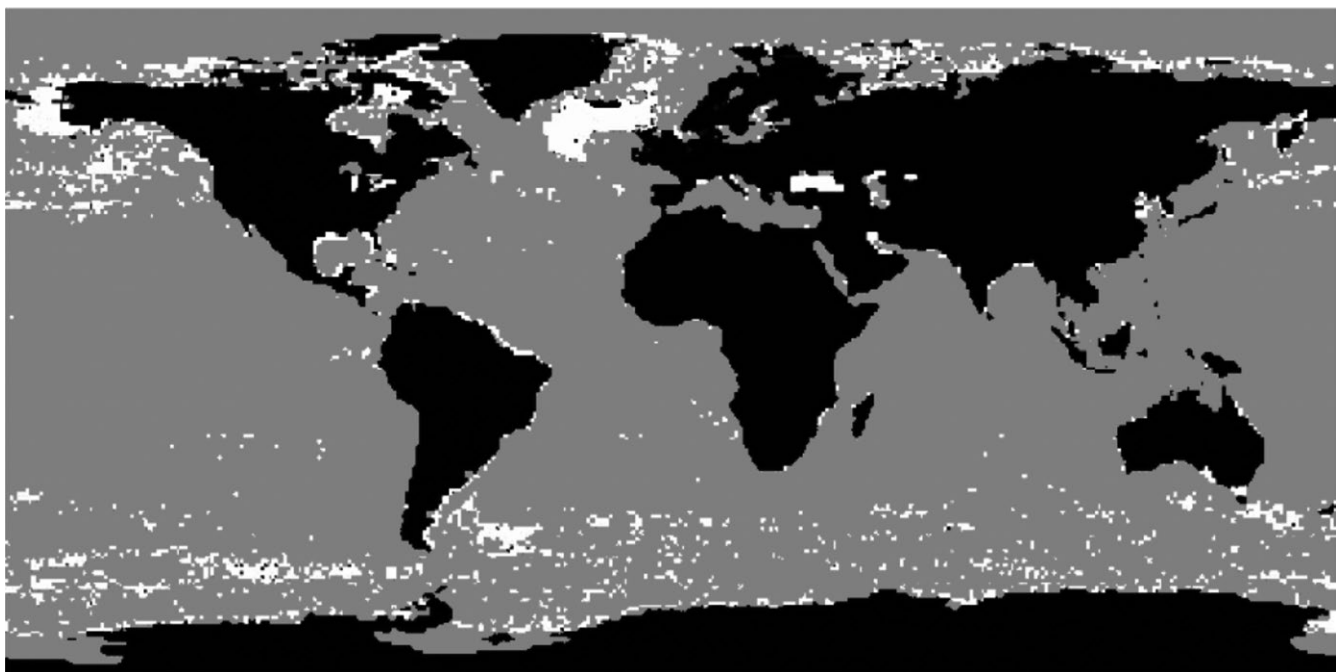

FIGURE 7 Global distribution of coccolithophorid blooms from Sea-viewing wide field-of-view sensor (SeaWiFS) (from Alvain et al., 2008)

accuracy of the actual distribution in the Northern Hemisphere and low and inaccurate probabilities in the Southern (OBIS, 2014). The uneven distribution of sampling effort might lead to a bias towards the more intensely sampled areas and we believe that this may have contributed to our finding that the predicted current distribution of regions with a high probability for the occurrences of *E. huxleyi* does not completely correspond to the satellite image shown in Figure 7. There is a clear lack of presences recorded from observations reported in the literature and in the modelled predictions in the Bering Strait, where the satellite images show frequent blooms (Alvain et al., 2008).

5 | CONCLUSIONS

This and other studies (Barton et al., 2016) indicate that SDM can be an effective tool for predicting distributions of individual phytoplankton species. Furthermore, this study suggests that changing environmental conditions (temperature, salinity and nutrients) may alter the location of the centres of distribution of two phytoplankton species that are potentially important contributors to the biological pump, i.e., for the vertical transport of carbon in the ocean. Specifically, the study suggests that changes in environmental conditions in coming decades may lead to a northward trajectory in the location of population centres for the distribution of the diatom, *C. diadema*. Decreases both in the absolute area in which substantial accumulations of this species are predicted to occur and in its occurrence over deep water (>1,000 m) are predicted. Thus, the predicted changes in future distributions of this species could, potentially, reduce this particular species' contribution to the biological pump. The study also suggests that the range of *E. huxleyi* may decrease in the future ocean. However, the models used here do not predict a major change in the occurrence of this species over depths of >1,000 m under future ocean conditions.

REFERENCES

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223–1232.

Alvain, S., Moulin, C., Dandonneau, Y., & Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochemical Cycles*, 22, GB3001.

Anderson, R. P. (2012). Harnessing the world's biodiversity data: Promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, 1260, 66–80.

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology and evolution*, 22, 42–47.

Araújo, M. B., & Rahbek, C. (2006). Ecology. How does climate change affect biodiversity? *Science (New York, N.Y.)*, 313, 1396–1397.

Baars, J. W. R. (1979). Autecological investigations on marine diatoms. I. Experimental results in biogeographical studies. *Hydrobiological Bulletin*, 13, 123–137.

Baars, J. W. R. (1981). Autecological investigations on marine diatoms. 2. Generations times of 50 species. *Hydrobiological Bulletin*, 15, 137–151.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3, 327–338.

Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11), 2964–2969.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., ... Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810–1819.

Beaufort, L., Probert, I., de Garidel-Thoron, T., Bendif, E. M., Ruiz-Pino, D., Metzl, N., ... de Vargas, C. (2011). Sensitivity of coccolithophores to carbonate chemistry and ocean acidification. *Nature*, 476, 80–83.

Bouza, N., & Aboal, M. (2008). Checklist of phytoplankton on the south coast of Murcia (SE Spain, SW Mediterranean Sea). In V. Evangelista, L. Barsanti, A. M. Frassanito, V. Passarelli & P. Gualtieri. (Eds.), *Algal toxins: Nature, occurrence, effect and detection*. NATO Science for Peace and Security Series A: Chemistry and Biology. Dordrecht: Springer.

Brand, L. E. (1984). The salinity phytoplankton tolerance isolates of forty-six marine phytoplankton species. *Estuarine, Coastal and Shelf Science*, 18, 543–556.

Brodie, I., & Kemp, E. S. A. (1994). Variation in biogenic and detrital fluxes and formation of laminae in late Quaternary sediments from the Peruvian coastal upwelling zone. *Marine Geology*, 116, 385–398.

Buitenhuis, E., Wal, P., & Baar, H. (2001). Blooms of *Emiliania huxleyi* are sinks of atmospheric carbon dioxide: A field and mesocosm study derived simulation. *Global Biogeochemical Cycles*, 15, 577–587.

Chust, G., Irigoien, X., Chave, J., & Harris, R. P. (2013). Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography*, 22, 531–543.

Cubillos, J., Wright, S., Nash, G., de Salas, M., Griffiths, B., Tilbrook, B., ... Hallegraeff, G. (2007). Calcification morphotypes of the coccolithophorid *Emiliania huxleyi* in the Southern Ocean: Changes in 2001 to 2006 compared to historical data. *Marine Ecology Progress Series*, 348, 47–54.

Danbara, A., & Shiraiwa, Y. (1999). The requirement of selenium for the growth of marine coccolithophorids, *Emiliania huxleyi*, *Gephyrocapsa oceanica* and *Helladosphaera* sp. (Prymnesiophyceae). *Plant and Cell Physiology*, 40, 762–766.

De La Rocha, C. L., & Passow, U. (2007). Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54, 639–658.

Durbin, E. G., Krawiec, R. W., & Smayda, T. J. (1975). Seasonal studies on the relative importance of different size fractions of phytoplankton in Narragansett Bay (USA). *Marine Biology*, 32, 271–287.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.

Flint, M. V., Sukhanova, I. N., Kopylov, A. I., Poyarkov, S. G., & Whitledge, T. E. (2002). Plankton distribution associated with frontal zones in the vicinity of the Pribilof Islands. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49, 6069–6093.

Flombaum, P. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences of the United States of America*, 11, 9824–9829.

French, F., & Hargraves, P. (1980). Physiological characteristics of plankton diatom resting stages. *Marine Biology Letters*, 1, 185–195.

Garcia, R. A., Burgess, N. D., Cabeza, M., Rahbek, C., & Araújo, M. B. (2012). Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, 18, 1253–1269.

Grenouillet, G., Buisson, L., Casajus, N., & Lek, S. (2011). Ensemble modelling of species distribution: The effects of geographical and environmental ranges. *Ecography*, 34, 9–17.

Hagino, K., Bendif, E. M., Young, J. R., Kogame, K., Probert, I., Takano, Y., ... Okada, H. (2011). New evidence for morphological and genetic variation in the cosmopolitan coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) from the COX1b-ATP4 genes. *Journal of Phycology*, 47, 1164–1176.

Hargraves, P. (1972). Studies on marine plankton diatoms. I. Chaetoceros diadema (Ehr.) Gran: Life cycle, structural morphology, and regional distribution. *Phycologia*, *11*, 247–257.

Harvey, M. (1997). Late-summer distribution of phytoplankton in relation to water mass characteristics in Hudson Bay and Hudson Strait (Canada). *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 1937–1952.

Hezel, P. J., Fichefet, T., & Massonnet, F. (2014). Modeled Arctic sea ice evolution through 2300 in CMIP5 extended RCPs. *The Cryosphere Discussions*, *8*, 1383–1406.

Hoppenrath, M. (2004). A revised checklist of planktonic diatoms and dinoflagellates from Helgoland (North Sea, German Bight). *Helgoland Marine Research*, *58*, 243–251.

Humphrey, G. (1979). Photosynthetic characteristics of algae grown under constant illumination and light-dark regimes. *Journal of Experimental Marine Biology and Ecology*, *40*, 63–70.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*, 287–315.

Jiang, Z., Liu, J., Chen, J., Chen, Q., Yan, X., Xuan, J., & Zeng, J. (2014). Responses of summer phytoplankton community to drastic environmental changes in the Changjiang (Yangtze River) estuary during the past 50 years. *Water research*, *54*, 1–11.

Karentz, D., & Smayda, T. J. (1984). Temperature and seasonal occurrence of 30 dominant phytoplankton species in Narragansett Bay over a 22-year period. *Marine Ecology Progress Series*, *18*, 277–293.

Katsuki, K., Takahashi, K., Onodera, J., Jordan, R., & Suto, I. (2009). Living diatoms in the vicinity of the North Pole, summer 2004. *Micropaleontology*, *55*, 137–170.

Kiørboe, T. (1993). Turbulence, phytoplankton cell size, and the structure of pelagic food webs. *Advances in Marine Biology*, *29*, 2–72.

Knappertsbusch, M., & Brummer, G.-J. A. (1995). A sediment trap investigation of sinking coccolithophorids in the North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*, 1083–1109.

Lampert, L., Quéguiner, B., Labasque, T., Pichon, A., & Lebreton, N. (2002). Spatial variability of phytoplankton composition and biomass on the eastern continental shelf of the Bay of Biscay (north-east Atlantic Ocean). Evidence for a bloom of *Emiliania huxleyi* (Prymnesiophyceae) in spring 1998. *Continental Shelf Research*, *22*, 1225–1247.

Leising, A. W., Pierson, J. J., Halsband-Lenk, C., Horner, R., & Postel, J. (2005). Copepod grazing during spring blooms: Does *Calanus pacificus* avoid harmful diatoms? *Progress in Oceanography*, *67*, 384–405.

Longhurst, A. R., & Glen Harrison, W. (1989). The biological pump: Profiles of plankton production and consumption in the upper ocean. *Progress in Oceanography*, *22*, 47–123.

Martínez-Meyer, E., Peterson, A. T., & Hargrove, W. W. (2004). Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography*, *13*, 305–314.

Mather, L., MacIntosh, K., & Kaczmarska, G. (2010) A checklist of diatom species reported (and presumed native) from Canadian coastal waters. *Canadian Technical Report of Fisheries and Aquatic Sciences*, *2881*.

McQuoid, M. R. (2005). Influence of salinity on seasonal germination of resting stages and composition of microplankton on the Swedish west coast. *Marine Ecology Progress Series*, *289*, 151–163.

Mergulhao, L., & Mohan, R. (2006). Coccolithophores from the central Arabian Sea: Sediment trap results. *Journal of earth system science*, *115*, 415–428.

Midttun, L. (1985). Formation of dense bottom water in the Barents Sea. *Deep Sea Research Part A. Oceanographic Research Papers*, *32*, 1233–1241.

Mousing, E. A., Ellegaard, M., & Richardson, K. (2014). Global patterns in phytoplankton community size structure – evidence for a direct temperature effect. *Marine Ecology Progress Series*, *497*(25), 38.

Mousing, E. A., Richardson, K., Bendtsen, J., Cetinic, I., & Perry, M. J. (2016). Evidence of small-scale spatial structuring of phytoplankton alpha- and beta-diversity in the open ocean. *Journal of Ecology*, doi:10.1111/1365-2745.12634

Nanninga, H., & Tyrrell, T. (1996). Importance of light for the formation of algal blooms by *Emiliania huxleyi*. *Marine ecology progress series*, *136*, 195–203.

Nielsen, M. (1995). Photosynthetic characteristics of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae) exposed to elevated concentrations of dissolved inorganic carbon. *Journal of Phycology*, *31*, 715–719.

OBIS (2014) *Ocean Biogeographic Information System. Intergovernmental Oceanographic Commission of UNESCO*. Retrieved from www.iobis.org. (Accessed: 2014-01-10)

Pachauri, R. K, Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., … Ypersele, J. (2014). *Climate change 2014: Synthesis Report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.

Pitcher, G. C. (1990). Phytoplankton seed populations of the Cape Peninsula upwelling plume, with particular reference to resting spores of *Chaetoceros* (Bacillariopyceae) and their role in seeding upwelling waters. *Estuarine, Coastal and Shelf Science*, *31*, 283–301.

Popovich, C. A., & Marcovecchio, J. E. (2008). Spatial and temporal variability of phytoplankton and environmental factors in a temperate estuary of South America (Atlantic coast, Argentina). *Continental Shelf Research*, *28*, 236–244.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ragueneau, O., Schultes, S., Bidle, K., Claquin, P., & Moriceau, B. (2006). Si and C interactions in the world ocean: Importance of ecological processes and implications for the role of diatoms in the biological pump. *Global Biogeochemical Cycles*, *20*, GB4S02.

Richardson, T. L., & Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science (New York, N.Y.)*, *315*, 838–840.

Robinson, L. M., Elith, J., Hobday, A. J., Pearson, R. G., Kendall, B. E., Possingham, H. P., & Richardson, A. J. (2011). Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, *20*, 789–802.

Rost, B., & Riebesell, U. (2004). Coccolithophores and the biological pump: responses to environmental changes. In H. R. Thierstein, & J. R. Young (Eds.), *Coccolithophores: From molecular processes to global impact* (pp. 99–125). Berlin: Springer.

Rynearson, T. A., Richardson, K., Lampitt, R. S., Sieracki, M. E., Poulton, A. J., Lyngsgaard, M. M., & Perry, M. J. (2013). Major contribution of diatom resting spores to vertical flux in the sub-polar North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, *82*, 60–71.

Škaloud, P., Řezáčová, M., & Ellegaard, M. (2006). Spatial distribution of phytoplankton in Spring 2004 along a transect in the eastern part of the North Sea. *Journal of oceanography*, *62*, 717–729.

Smetacek, V. S. (1985). Role of sinking in diatom life-history cycles: Ecological, evolutionary and geological significance. *Marine Biology*, *84*, 239–251.

Sukhanova, I. N., Flint, M. V., Sergeeva, V. M., & Kremenetskiy, V. V. (2011). Phytoplankton of the south-western part of the Kara Sea. *Oceanology*, *51*, 978–992.

Takahashi, T., & Sutherland, S. (2009). Climatological mean and decadal change in surface ocean pCO 2, and net sea–air CO 2 flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*, 554–577.

Thuiller, W., Georges, D., & Engler, R. (2014). *biomod2: Ensemble platform for species distribution modeling*, R package version 3.1-64. Retrieved from http://CRAN.R-project.org/package=biomod2

Toro, J. (1989). The growth rate of two species of microalgae used in shellfish hatcheries cultured under two light regimes. *Aquaculture Research*, *20*, 249–254.

Uysal, Z., & Sur, H. (1995). Net phytoplankton discriminating patches along the southern Black Sea coast in winter 1990. *Oceanologica Acta*, *18*, 639–647.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In E. T. Sundquist, & W. S. Broecker (Eds.), *The Carbon Cycle and Atmospheric CO: Natural Variations Archean to Present* (pp. 99–110). Washington, DC: American Geophysical Union.

Wasmund, N., Göbel, J., & Bodungen, B. V. (2008). 100-years-changes in the phytoplankton community of Kiel Bight (Baltic Sea). *Journal of Marine Systems*, *73*, 300–322.

Williams, K. M. (1986). Recent Arctic marine diatom assemblages from bottom sediments in Baffin Bay and Davis Strait. *Marine Micropaleontology*, *10*, 327–341.

Zielinski, U., & Gersonde, R. (1997). Diatom distribution in Southern Ocean surface sediments (Atlantic sector): implications for paleoenvironmental reconstructions. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *129*, 213–250.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.