12-2016

Net primary productivity estimates and environmental variables in the Arctic Ocean; an assessment of coupled physical-biogeochemical models

Younjoo J. Lee

Patrica A. Matrai

Marjorie A.M. Friedrichs
Virginia Institute of Marine Science, marjy@vims.edu

et al

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2016JC011993

Key Points:
- Arctic models underestimated net primary productivity (NPP) but overestimated nitrate, mixed layer depth, and euphotic depth
- Arctic NPP model skill was greatest in low production regions
- Arctic NPP model skill was constrained by different environmental factors in different Arctic Ocean regions

Correspondence to:
Y. J. Lee,
ylee@bigelow.org and P. A. Matrai,
pmatrai@bigelow.org

Citation:
Lee, Y. J., et al. (2016), Net primary productivity estimates and environmental variables in the Arctic Ocean: An assessment of coupled physical-biogeochemical models, *J. Geophys. Res. Oceans*, *121*, 8635–8669, doi:10.1002/2016JC011993.

Received 25 MAY 2016
Accepted 8 NOV 2016
Accepted article online 14 NOV 2016
Published online 9 DEC 2016

Net primary productivity estimates and environmental variables in the Arctic Ocean: An assessment of coupled physical-biogeochemical models

Younjoo J. Lee[1,2], Patricia A. Matrai[1], Marjorie A. M. Friedrichs[3], Vincent S. Saba[4], Olivier Aumont[5], Marcel Babin[6], Erik T. Buitenhuis[7], Matthieu Chevallier[8], Lee de Mora[9], Morgane Dessert[10], John P. Dunne[11], Ingrid H. Ellingsen[12], Doron Feldman[13], Robert Frouin[14], Marion Gehlen[15], Thomas Gorgues[10], Tatiana Ilyina[16], Meibing Jin[17,18], Jasmin G. John[11], Jon Lawrence[19], Manfredi Manizza[20], Christophe E. Menkes[5], Coralie Perruche[21], Vincent Le Fouest[22], Ekaterina E. Popova[19], Anastasia Romanou[23], Annette Samuelsen[24], Jörg Schwinger[25], Roland Séférian[8], Charles A. Stock[11], Jerry Tjiputra[25], L. Bruno Tremblay[26], Kyozo Ueyoshi[14], Marcello Vichi[27,28], Andrew Yool[19], and Jinlun Zhang[29]

[1]Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA, [2]Now at Department of Oceanography, Naval Postgraduate School, Monterey, California, USA, [3]Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia, USA, [4]National Ocean and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, USA, [5]Laboratoire Océan, Climat, Exploitation et Application Numérique/Institut Pierre-Simon Laplace, CNRS/IRD/UPMC, Université Pierre et Marie Curie, Paris, France, [6]Takuvik Joint International Laboratory, CNRS-Université Laval, Québec, Canada, [7]School of Environmental Sciences, University of East Anglia, Norwich, UK, [8]Centre National de Recherches Météorologiques, Unite mixte de recherche 3589 Météo-France/CNRS, Toulouse, France, [9]Plymouth Marine Laboratory, Plymouth, UK, [10]Laboratoire d'Océanographie Physique et Spatiale CNRS/IFREMER/IRD/UBO, Institut Universitaire et Européen de la Mer, Plouzané, France, [11]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA, [12]SINTEF Fisheries and Aquaculture, Trondheim, Norway, [13]NASA Goddard Institute for Space Studies, New York, USA, [14]Climate, Atmospheric Science, and Physical Oceanography Division, Scripps Institution of Oceanography, University of California, La Jolla, California, USA, [15]Laboratoire des Sciences du Climat et de l'Environnement/Institut Pierre-Simon Laplace, Gif-sur-Yvette, France, [16]Max Planck Institute for Meteorology, Hamburg, Germany, [17]International Arctic Research Center, University of Alaska, Fairbanks, Alaska, USA, [18]Laboratoty for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [19]National Oceanography Centre, University of Southampton, Southampton, UK, [20]Geosciences Research Division, Scripps Institution of Oceanography, University of California, La Jolla, California, USA, [21]Mercator-Ocean, Toulouse, France, [22]LIttoral ENvironnement et Sociétés, Université de La Rochelle, La Rochelle, France, [23]Department of Applied Physics and Applied Mathematics, Columbia University and NASA Goddard Institute for Space Studies, New York, USA, [24]Nansen Environmental and Remote Sensing Centre and Hjort Centre for Marine Ecosystem Dynamics, Bergen, Norway, [25]Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway, [26]Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Canada, [27]Department of Oceanography, University of Cape Town, Cape Town, South Africa, [28]Marine Research Institute, University of Cape Town, Cape Town, South Africa, [29]Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

Abstract The relative skill of 21 regional and global biogeochemical models was assessed in terms of how well the models reproduced observed net primary productivity (NPP) and environmental variables such as nitrate concentration (NO_3), mixed layer depth (MLD), euphotic layer depth (Z_{eu}), and sea ice concentration, by comparing results against a newly updated, quality-controlled in situ NPP database for the Arctic Ocean (1959–2011). The models broadly captured the spatial features of integrated NPP (iNPP) on a pan-Arctic scale. Most models underestimated iNPP by varying degrees in spite of overestimating surface NO_3, MLD, and Z_{eu} throughout the regions. Among the models, iNPP exhibited little difference over sea ice condition (ice-free versus ice-influenced) and bottom depth (shelf versus deep ocean). The models performed relatively well for the most recent decade and toward the end of Arctic summer. In the Barents and Greenland Seas, regional model skill of surface NO_3 was best associated with how well MLD was reproduced. Regionally, iNPP was relatively well simulated in the Beaufort Sea and the central Arctic Basin, where in situ NPP is low and nutrients are mostly depleted. Models performed less well at simulating iNPP in the Greenland and Chukchi Seas, despite the higher model skill in MLD and sea ice concentration,

respectively. iNPP model skill was constrained by different factors in different Arctic Ocean regions. Our study suggests that better parameterization of biological and ecological microbial rates (phytoplankton growth and zooplankton grazing) are needed for improved Arctic Ocean biogeochemical modeling.

1. Introduction

Primary production provides the energy that fuels the Arctic Ocean ecosystem. However, given the regional heterogeneity, the dramatic seasonal changes, and the limited access due to harsh environments, it is difficult to obtain a comprehensive picture of the Arctic Ocean's in situ primary production regime. Models of biogeochemical carbon cycling at various spatial/temporal scales also rely on quantification of the first level of the marine food chain—phytoplankton, their biomass, and production. Current projected changes in future climate and responses of the Arctic Ocean may indicate either (i) an increase in net primary production (NPP) due to enhanced light availability as a direct result of sea ice and snow cover loss as well as increased nutrient fluxes from sub-Arctic waters, bottom water by vertical mixing and greater upwelling as sea ice extent decreases, as well as nutrient input by Arctic rivers and terrestrial ecosystems as temperature rises [*Slagstad et al*., 2015], or (ii) a decrease of NPP due to nutrient limited conditions with enhanced stratification (due to ice melt, freshwater inflow, and thermal warming), increased denitrification, and decreased light availability (due to more clouds and higher water turbidity from permafrost melt, beach erosion, riverine input, and wind-driven resuspension) [*Wassmann and Reigstad*, 2011]. But, not all drivers listed above are commonly incorporated in numerical models that are widely used for evaluating future changes in NPP over the entire Arctic region today and under a changing climate.

There are two main types of numerical models that have been applied to estimate NPP in the Arctic Ocean: satellite-based diagnostic models (reviewed in *Babin et al*. [2015] and *International Ocean Colour Coordinating Group* (*IOCCG*) [2015], and assessed in *Lee et al*. [2015]) and process-based coupled dynamic physical-biological models run in either retrospective or predictive modes. This latter class of models has been applied to improve our understanding of the Arctic Ocean carbon sinks and sources using regional [e.g., *Deal et al*., 2011; *Manizza et al*., 2013; *Slagstad et al*., 2015; *Zhang et al*., 2010c,2015] to global [e.g., *Popova et al*., 2010; *Vancoppenolle et al*., 2013] ice-ocean-atmosphere carbon cycle simulations. In general, the magnitude, vertical extent, and seasonal evolution of NPP are simulated largely as a function of temperature, light, nutrient availability, and physical forcing (advection, mixing, and stratification) as well as associated planktonic and sea ice food web processes. The complexity of the models differs not only in the number of processes/drivers included in model simulations but also in their degree of interactive coupling among oceanic, sea ice, atmospheric, and/or hydrologic processes as well as in their biogeochemical component and spatial resolution; on one end, this allows for year-round simulations in ice-free and ice-covered waters and for detailed comparisons with satellite-derived observations at certain times and places. At the other end, certain models will lack ecosystem, biogeochemical and physical (i.e., freshwater inputs, sea ice concentration/extent/thickness, and stratification) parameterizations and feedbacks important to the Arctic Ocean, which in turn impact controls on NPP. For instance, regional models may have more flexibility to include processes that are specific to the Arctic Ocean, but are limited by the boundary conditions. Global models, on the other hand, are independent of the boundary conditions, but are often lacking in regional complexity in representation of Arctic processes.

Previous primary production intercomparison exercises matched in situ carbon uptake rates with NPP estimated mostly from ocean color models (OCMs) [e.g., *Balch et al*., 1992; *Campbell et al*., 2002; *Saba et al*., 2011] and, more recently, from some coupled biogeochemical ocean general circulation models (BOGCMs) [*Carr et al*., 2006; *Friedrichs et al*., 2009; *Saba et al*., 2010] in various regions of the ocean except the Arctic Ocean. *Popova et al*. [2012] and *Vancoppenolle et al*. [2013] reported the results from various BOGCMs and Earth System Models (ESMs) that were compared for the Arctic domain against satellite-derived NPP albeit with known issues. These latter two studies found reasonable agreement on the main features of pan-Arctic annual NPP and highlighted the disagreement among the models with respect to nutrient availability and source, as well as which factor, light or nutrients, controls the present-day and future Arctic Ocean NPP. Furthermore, recent individual model studies emphasized the need to accurately parameterize specific

processes to adequately represent NPP within the Arctic Ocean system, such as sea ice physics and biogeochemistry [e.g., *Jin et al*., 2012, 2016; *Tedesco and Vichi*, 2014], circulation patterns [e.g., *Popova et al*., 2013], and secondary production [*Slagstad et al*., 2015].

Models that estimate Arctic Ocean NPP have previously been assessed using annual mean and seasonal variations in surface nutrients from the World Ocean Atlas and, as indicated above, ocean color-derived NPP. *Babin et al*. [2015] and *IOCCG* [2015] extensively review the strengths (pan-Arctic, weekly, or monthly regular coverage) and limitations (time and space composite imagery; sea ice, cloud/fog; seasonal darkness; bio-optical and physiological considerations; mostly surface observations) of remotely sensed data in this seasonally dark and ice-covered region. More recently, *Lee et al*. [2015] assessed the skill of 32 OCMs to reproduce the observed NPP across the Arctic Ocean. Overall, OCMs were most sensitive to uncertainties in surface chlorophyll-*a* concentration (hereafter, chlorophyll), generally performing better with in situ chlorophyll than with ocean color-derived values. Regardless of type or complexity, most OCMs were not able to fully reproduce the variability of in situ NPP, whereas some of them exhibited almost no bias. As a group, OCMs overestimated mean NPP, however, this was partly offset by some models underestimating NPP when a subsurface chlorophyll maximum was present. OCMs performed better reproducing NPP values when using Arctic-relevant parameter values.

Sampling difficulties and the resulting biogeochemical data scarcity in the Arctic Ocean affect vertical, horizontal, and temporal interpolations frequently used for BOGCM and ESM validation and skill assessment, such that only basin-wide, temporally averaged (i.e., annual) data are often used to evaluate their NPP estimates. While sampling irregularity in the Arctic Ocean, even in summer, has a strong effect on the spatio-temporal distribution of in situ biogeochemical data, it is the increasing availability of such data in recent years that currently makes the assessment of BOGCM and ESM simulated NPP at smaller and shorter scales in the Arctic Ocean possible. We have assembled a pan-Arctic, multidecadal data set of in situ NPP and nitrate (NO_3) vertical profiles as well as associated biophysical parameters that allow model skill to be assessed, especially at regional scales.

NPP observations are herein compared against an available ensemble of physical-biological simulations presently being applied for a range of global and Arctic-specific carbon and ecosystem questions (Table 1). While model performance of multiple models should ideally be assessed using simulations generated with identical forcing, initialization, and boundary conditions, such restrictions would severely limit the number of models considered and fail to sample the variety of present modeling systems. Thus in this study, in an opportunistic approach, 21 model simulations were analyzed within the Arctic domain, without forcing the participants to use the same setup, with the aim of elucidating large-scale similarities and contrasts in NPP and associated drivers. Model skill of existing simulations was quantified by assessing modeled integrated NPP (hereafter, iNPP), vertical profiles of NPP and NO_3, euphotic layer depth (Z_{eu}; defined here as the depth at which photosynthetically available radiation (PAR) is 1% of the surface value), mixed layer depth (MLD), and sea ice concentration against in situ or climatological (only MLD) data. The participating BOGCMs and ESMs include biogeochemical and ecosystem modules of varying complexity that may or may not have been optimized for the Arctic Ocean. Only Models 9 and 14 include an ice algae component, but ice algal production was not included in NPP for this study. Spatial resolution also varies among the participating models. Despite these caveats, BOGCMs and ESMs are the only tools that account for complex processes of land-ocean-atmosphere-sea ice responses to future climate change in the Arctic Ocean. Their steady improvement requires a regular assessment of their skill to reproduce observations. Here, we report the first intercomparison study of NPP in the Arctic Ocean, based on 21 simulations from various BOGCMs and ESMs, against in situ measurements. It is a companion study to our earlier assessment of NPP estimates derived from OCMs [*Lee et al*., 2015].

2. Participating Models

A total of 21 models (Table 1) participated in this exercise: 18 of these were coupled ocean-ice-ecosystem models (4 Arctic regional and 14 global) and 3 were ESMs. The participating models have a horizontal resolution between 0.1 and 200 km with 25–75 vertical levels; they used various vertical mixing schemes, atmospheric forcing, and sea ice physics (Table 1). Only Model 9 assimilated sea ice concentration and sea surface temperature. In terms of biogeochemistry, each model has between one and six phytoplankton and

Journal of Geophysical Research: Oceans

10.1002/2016JC011993

Table 1. A Brief Description of Participating Models and Output Variables

Model	Name	Domain/ Type	Horizontal Resolution (km)	Vertical Levels	Vertical Mixing	Atmospheric Forcing	Sea Ice	Nutrients	# of Phytoplankton Classes	# of Zooplankton Classes	# of Detritus	Sea Ice Biology/ River Nutrients/ Pelagic-Benthic Coupling	Simulation Period	Model Output Variables					
														Integrated NPP	NPP (0–100 m)	NO_3 (0–100 m)	Z_{eu} (1%)	Sea Ice	MLD
1	HYCOM-NORWECOM	Regional	10–16.5	28	GISS	ERA-Interim reanalysis	*Hunke and Dukowicz* [1997]	N, P, Si	2	2	3	N/Y/N	1997–2011	•	•	•	•	•	•
2	NEMO-PISCES	Global	10–16	75	TKE	ERA-Interim reanalysis	LIM2	N, P, Si, Fe	2	2	3	N/Y/Y	1998–2014	•	•	•	•	•	
3	PlankTOM5.3	Global	0.13–192	30	TKE	NCEP reanalysis	LIM2	N, Si, Fe	3	2	7	N/Y/N	1956–2013	•	•	•		•	
4	PlankTOM10	Global	0.13–192	30	TKE	NCEP reanalysis	LIM2	N, P, Si, Fe	6	3	7	N/Y/N	1948–2013	•	•	•		•	
5	MOM-SIS-COBALT	Global	25–111	50	KPP	CORE2	GFDL-SIS	N, P, Si, Fe	3	3	1	N/Y/Y	1948–2007	•	•	•			•
6	MOM-SIS-TOPAZ	Global	25–111	50	KPP	CORE2	GFDL-SIS	N, P, Si, Fe, NH_4	3	1	7	N/Y/Y	1948–2007	•	•	•	•	•	•
7	SINMOD	Regional	20	25	Based on the Richardson Number	ERA-Interim reanalysis	*Hibler* [1979]; *Hunke and Dukowicz* [1997]	N, Si	2	2	2	N/Y/Y	1979–2014	•	•	•	•	•	•
8	NorESM-OC	Global	18–65	53	TKE	CORE2	CICE4	N, P, Si, Fe	1	1	3	N/N//N	1850–2012	•	•	•		•	•
9	BIOMAS	Regional	3–70	30	KPP	NCEP reanalysis	Thickness and enthalpy distribution	N, Si	2	3	2	Y/N/N	1971–2015	•	•	•		•	•
10	NEMO-MEDUSA	Global	6.8–15.4	64	TKE	DFS4.1 reanalysis	LIM2	N, Si, Fe	2	2	1	N/N/N	1988–2006	•	•	•	-	•	•
11	NEMO-ERSEM (xhonp)	Global	25–68	75	TKE	CORE2	CICE	N, P, Si, Fe	4	3	6	N/N/Y	1890-2007	•	•	•	•	•	•
12	NEMO-ERSEM (xhonc)	Global	25-68	75	TKE	CORE2	CICE	N, P, Si, Fe	4	3	6	N/N/Y	1890–2007	•	•	•	•	•	•
13	PELAGOS	Global	120–160	30	TKE	ERA-Interim reanalysis	LIM2	N, P, Si, Fe	3	3	1	N/Y/N	1988–2010	•	•	•	•	•	•
14	IARC POP-CICE	Global	30–45	40	KPP	CORE2	CICE 5.0	N, P, Si, Fe	3	1	1	Y/N/N	1958–2009	•	•	•		•	•
15	NEMO/PISCES	Global	63–131	31	TKE	DFS5.2 reanalysis	LIM2	N, P, Si, Fe	2	2	2	N/Y/Y	1979–2010	•	•		•	•	
16	NEMO/PISCES	Global	32	46	TKE	CORE2 + NCEP/ DOE AMIP-II	LIM3	N, P, Si, Fe	2	2	2	N/Y/N	2002–2011	•	•			•	
17	NEMO-GELATO-PISCES	Global	40–70	42	TKE	NCEP reanalysis	GELATO5	N, P, Si, Fe	2	2	2	N/Y/Y	1948–2013	•	•	•	•	•	•
18	MITgcm	Regional	18	50	KPP	JRA25 reanalysis	*Hibler* [1979]	N	2	2	2	N/Y/N	1979–2012	•	•	•		•	•
19	NorESM	Global/ESM	18–65	53	TKE	CMIP5 + RCP8.5	CICE4	N, P, Si, Fe	1	1	3	N/N/N	1850–2012	•	•	•		•	•
20	GISS-E2-R-CC	Global /ESM	100–125	32	KPP	20th century climate forcing	Sea ice modelE2	N, Si, Fe	4	1	3	N/N/N	1850–2010	•		•		•	•
21	MPI-ESM (MPIOM-HAMOCC)	Global /ESM	15–65	40	*Pacanowski and Philander* [1981] +wind mixing	CMIP5 + RCP4.5	*Notz et al.* [2013]	N, P, Si, Fe	1	1	3	N/N/Y	1850–2012	•					

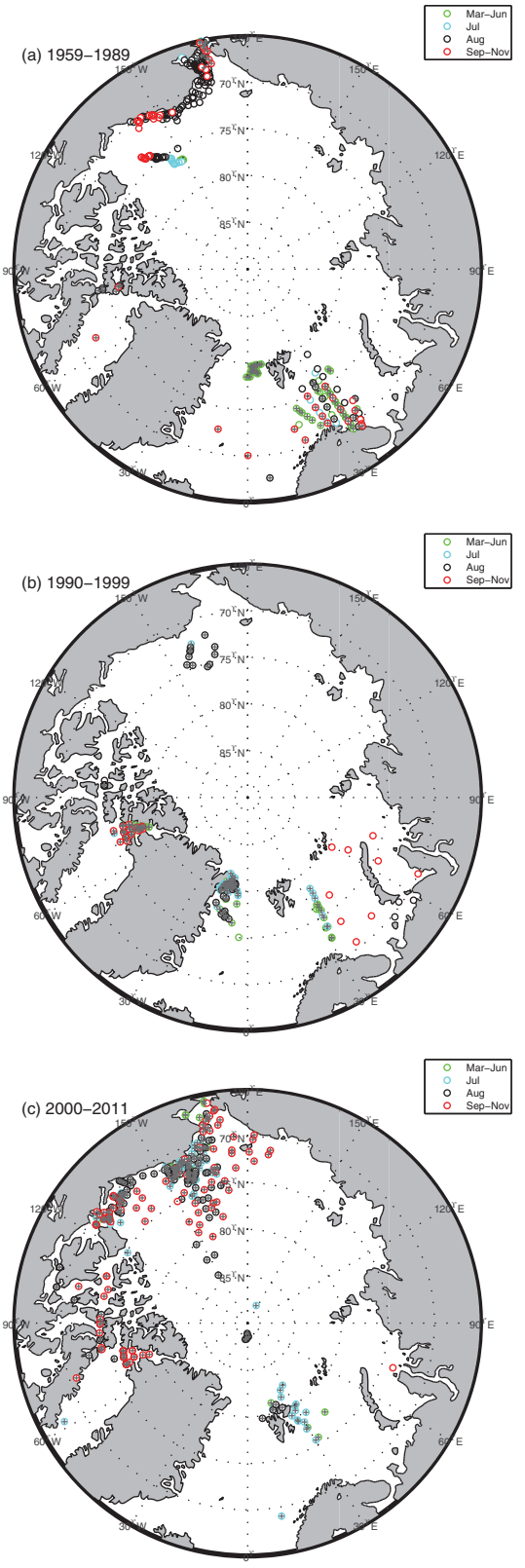

Figure 1. The sampling stations where in situ NPP (open circle, N = 928, 1959–2011) and nitrate (+, N = 663, 1979–2011) were measured during the periods of (a) 1959–1989, (b) 1990–1999, and (c) 2000–2011. In each plot, the stations were also grouped by seasons: March–June (green), July (blue), August (black), and September–November (red).

zooplankton groups. Models use nitrogen (N) or carbon (C) as currency in their simulations, while phosphorus (P), silica (Si), and/or iron (Fe) are included as state variables in some cases as well. Some models additionally include sea ice biology, riverine nutrients, and/or benthic-pelagic coupling. A brief description of the models can be found in Appendix A.

3. Data and Methods

3.1. In Situ Net Primary Productivity (NPP; 1959–2011)

We built a quality-controlled in situ NPP (mgC m^{-3} d^{-1}) database over five decades (1959–2011) that were sampled mostly in the Chukchi, Beaufort, and Barents Seas as well as the Canadian Archipelago (Figure 1). This database consisted of the ARCSS-PP data set [*Matrai et al.*, 2013], available at NOAA/National Oceanographic Data Center (NODC) (http://www.nodc.noaa.gov/cgi-bin/OAS/prd/accession/details/63065), and more recent international polar research activities such as CASES (2004) [*Brugel et al.*, 2009; *Tremblay et al.*, 2011], ArcticNet (2005–2008 and 2010–2011) (M. Gosselin, personal communication, 2014), K-PORT (2007, 2008, and 2010) (S. H. Lee, personal communication, 2014), CFL (2008) [*Mundy et al.*, 2009; *Sallon et al.*, 2011], ICE-CHASER (2008) [*Charalampopoulou et al.*, 2011], JOIS (2009) [*Yun et al.*, 2012], RUSALCA (2009) (S. H. Lee, personal communication, 2014), JAMSTEC (2009–2010; http://www.godac.jamstec.go.jp/darwin/e), Malina (2009; http://malina.obs-vlfr.fr/), and ICESCAPE (2010–2011; http://seabass.gsfc.nasa.gov/) [*Arrigo et al.*, 2014]. These recent data were primarily collected in the Pacific-sector of the Arctic Ocean (Figure 1c). The in situ sampling stations were selected such that discrete NPP measurements, using ^{13}C or ^{14}C-labeled compounds, were available for at least four discrete depths at each station where the minimum depth of in situ NPP was between 0 and 5 m in the surface layer. Most of the in situ NPP using a ^{14}C-labeled

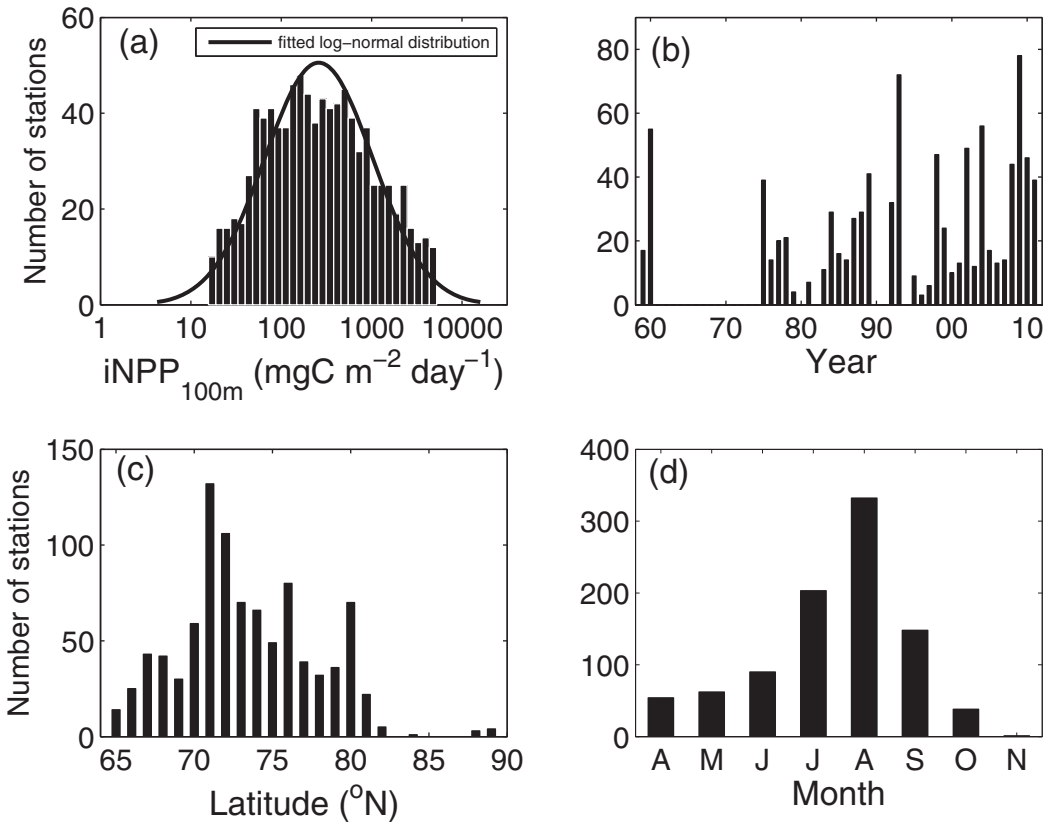

Figure 2. (a) The log-transformed distribution of in situ iNPP (mgC m^{-2} d^{-1}) down to 100 m (N = 927). (b), (c), and (d) show when and where those in situ NPP were measured in terms of years, latitude, and month, respectively.

compound were measured after 24 ± 2 h in situ or simulated in situ incubations (N = 480) and the rest of the ^{14}C samples were measured after less than a 12 h incubation or an unknown time period (N = 397). In situ carbon uptake rates were also measured using ^{13}C-labeled compounds (N = 90) after 3–6 h on-deck incubations [*Lee and Whitledge*, 2005; *Lee et al*., 2010]. A good agreement between estimates of ^{13}C and ^{14}C productivity was previously described [*Cota et al*., 1996]. Most of the NPP data were reported as daily values, but, if hourly in situ production rates were reported, they were converted to daily production rates by multiplying with a calculated day length [*Forsythe et al*., 1995], based on the date and location of the sampling stations. Note that we did not assess ice algal production in this study due to the scarcity of in situ measurements, although sea ice algae may contribute significantly to NPP early in the growth season and in certain localities [*Fernández-Méndez et al*., 2015; *Leu et al*., 2015].

Prior to any integration, minimum NPP was set to 0.25 mgC m^{-3} d^{-1} due to reported methodological precision [*Richardson*, 1991]. If the NPP value at the surface (0 m, Z_0) was not measured, it was assumed to be the same as the closest measurement taken from the uppermost 5 m (Z_{min}) of the water column. In addition, if the maximum sampling depth (Z_{max}) was shallower than 100 m (or bottom depth, Z_{bot}), NPP was extrapolated from Z_{max} to 100 m (or Z_{bot}) by assuming NPP decreased exponentially. Using the multiple segment trapezoidal rule, NPP was integrated from Z_0 to 100 m depth as well as, alternatively, over the euphotic layer that was determined during the incubation or calculated from available light penetration data; but, in situ Z_{eu} was reported at approximately half of the NPP sampling stations (N = 519): ice-influenced or ice-free. In order to maximize the number of stations available for skill assessment, iNPP down to 100 m (or Z_{bot}) was used for analysis since it exhibited only a small difference when compared to iNPP down to Z_{eu} ($r = 0.99$, $p < 0.01$).

Although all the available in situ iNPP ranged between 5.45 and 18,200 mgC m^{-2} d^{-1} at 967 stations, outliers (N = 39; 4%) at a confidence level of 95% were removed prior to any analysis due to accuracy issues with in situ NPP measurements. The remaining in situ iNPP data (N = 928; Figure 2) were between 15.7 and

Table 2. Mean (μ) and Standard Deviation (σ) of In Situ Log-Transformed iNPP (1959–2011; N = 928) and Depth-Averaged NO_3 (1979–2011; N = 663) in Different Regions (Except for Sea Ice Condition During the Period of 1979–2011), Periods (Except for NO_3 During the Period of 1979–2011), and Months (Except for March and November)

		log(iNPP)		iNPP (mgC m^{-2} d^{-1})		Depth-Averaged NO_3 (mmol m^{-3})	
		N	$\mu \pm \sigma$	10^{μ}	Range	N	$\mu \pm \sigma$
Regions	Sea ice-free	349	2.42 ± 0.58	263	18.1–5260	273	3.90 ± 3.05
	Sea ice-influenced	413	2.47 ± 0.63	295	15.7–5210	390	3.76 ± 2.77
	Shelf	432	2.51 ± 0.60	321	18.1–5080	244	3.47 ± 2.91
	Deep	496	2.34 ± 0.58	217	15.7–5260	419	4.03 ± 2.86
Periods	Year 1959–1989	344	2.50 ± 0.58	316	19.4–5260	284	5.09 ± 3.10
	Year 1990–1999	193	2.61 ± 0.55	404	17.4–5080		
	Year 2000–2011	391	2.25 ± 0.59	177	15.7–4460	379	2.87 ± 2.30
Months	April–June	201	2.64 ± 0.59	436	24.0–5260	176	5.89 ± 3.32
	July	203	2.61 ± 0.52	408	24.9–4420	160	3.27 ± 2.23
	August	332	2.39 ± 0.60	244	16.0–5080	182	2.86 ± 2.28
	September–October	186	2.03 ± 0.45	108	15.7–1570	139	2.86 ± 1.96
	All	928	2.42 ± 0.60	260	15.7–5260	663	3.82 ± 2.89

5,260 mgC m^{-2} d^{-1} (Table 2) with a median value of 246 mgC m^{-2} d^{-1}. These iNPP data were log-transformed (base 10) throughout the analysis, with a mean of 2.42 and a standard deviation of 0.60 (Table 2); they exhibited a log-normal distribution (Figure 2a) that was confirmed with the Kolmogorov-Smirnov (K-S) test. The majority of available NPP data in the Arctic Ocean was measured after 1975 (Figure 2b) and collected in the region between 70 and 80°N (Figure 2c), mainly from June to September (Figure 2d).

3.2. In Situ Nitrate (1979–2011)

Vertical profiles (0–100 m) of in situ NO_3 were extracted from the NOAA/NODC ArcNut database [*Codispoti et al*., 2013] (http://www.nodc.noaa.gov/archive/arc0034/0072133/) as well as additional polar research cruises at the stations where NPP was measured. However, in situ NO_3 profiles were available only at 663 stations (Figure 1). Note that, if only combined values of in situ NO_3 and nitrite concentration (hereafter, NO_2) were available, those were assumed to be the same as NO_3 because NO_2 in the Arctic Ocean is extremely low [*Codispoti et al*., 2013]. Vertical profiles were sorted by layers (0–10, 10–20, 20–30, 30–50, and 50–100 m); the deepest layer varied depending on station bottom depth, and in most cases (approximately 70%) was measured between 50 and 100 m.

3.3. Sea Ice Concentration (1979–2011)

Monthly sea ice concentrations (%) at the corresponding location and date of each NPP station were determined at the nearest grid cell (25 km × 25 km polar stereographic projection), using NOAA/National Snow and Ice Data Center (NSIDC) Climate Data Record of Passive Microwave Sea Ice Concentration, Version 2 (ftp://sidads.colorado.edu/pub/DATASETS/NOAA/G02202_v2/north/monthly/) [*Meier et al*., 2013, updated 2015; *Peng et al*., 2013]. These data files include not only monthly sea ice concentration from July 1987 to the present, but also the merged Goddard Space Flight Center NASA Team [*Cavalieri et al*., 1984] and Bootstrap [*Comiso*, 1986] sea ice concentrations from late 1978 to mid-1987, as an ancillary data set. Sea ice concentration was used to distinguish ice-free (sea ice concentration $<$10% which is the NSIDC threshold cut-off for sea ice edge detection in *Meier et al*. [2013, updated 2015]) and ice-influenced (sea ice concentration $\geq$10%) regions. Roughly 46% of the NPP stations were located in ice-free waters (N = 349, Table 2). Among the ice-influenced stations (N = 413), 75% were in marginal ice zones (N = 175 where sea ice concentration was 10–50%; N = 136 where sea ice concentration was 50–80%) and the rest (N = 102) were in ice-covered regions (sea ice concentration $\geq$80%).

3.4. Mixed Layer Depth (MLD)

Because temperature and salinity profile data were often missing or unavailable at the in situ NPP stations (i.e., separate files archived at different locations), monthly climatological salinity, and temperature from the Polar Science Center Hydrographic Climatology (http://psc.apl.washington.edu/nonwp_projects/PHC/Climatology.html) [*Steele et al*., 2001] were used to estimate MLD. After linearly interpolating the climatological temperature and salinity data in the vertical, the monthly climatological MLD was determined at the nearest grid cell (1° × 1° resolution) for each NPP station by using a threshold criterion of $\Delta\sigma_\theta = 0.1$ kg m^{-3}

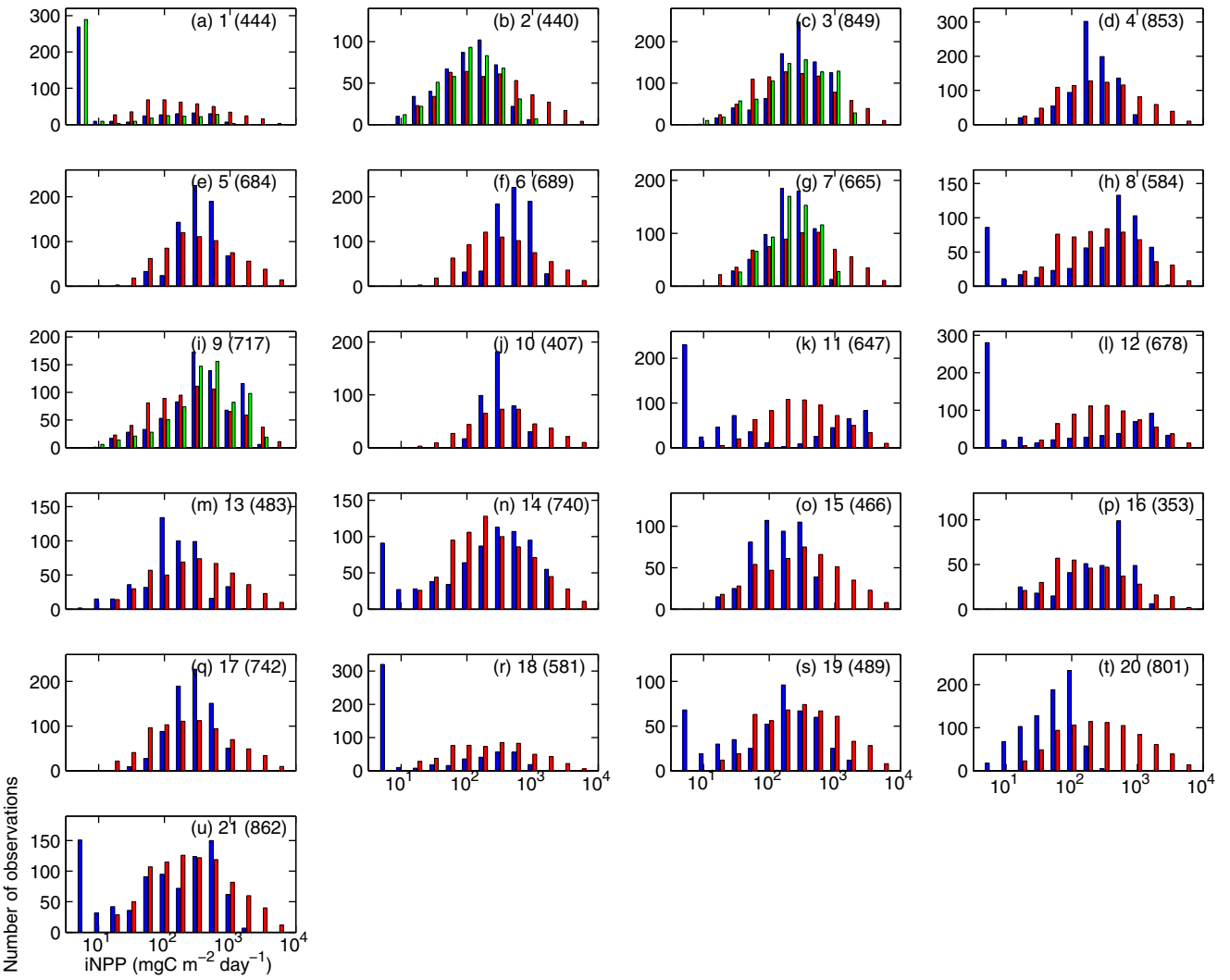

Figure 3. (a–u) Histogram of simulated (blue, monthly mean; green, daily mean) and in situ iNPP (red; mgC m^{-2} d^{-1}). Indicated are model numbers in the upper right with the number of sampling stations in parenthesis, which varies due to different simulation periods.

[*Peralta-Ferriz and Woodgate*, 2015], where $\Delta\sigma_\theta = \sigma_\theta(z) - \sigma_\theta$ (5 m); $\sigma_\theta(z)$ and σ_θ(5 m) are the potential density anomaly ($\sigma_\theta = \rho_\theta$ – 1000 kg m^{-3}, with ρ_θ being potential density) at a given depth (z) and 5 m, respectively.

3.5. Model Output

Location (latitude and longitude) and date of the in situ NPP station data were provided to all modeling teams. Each modeling team was then asked to generate monthly mean estimates for vertical profiles (0–100 m) of NPP and NO_3, iNPP, MLD, Z_{eu}, and sea ice concentration. Note that a minimum value for model NPP was set to 0.25 mgC m^{-3} d^{-1}, similar to the minimum in situ NPP. Five models (two global and three regional) provided daily mean iNPP as well (Models 1, 2, 3, 7, and 9). Unlike iNPP, other output variables were provided by fewer models due to their model setup and configuration (Table 1). For instance, Z_{eu} was only available in nine participating models. Moreover, because the simulation period differed among the participating models, the number of stations used for model-data comparisons varied between N = 853 and N = 353 (Figure 3). Regardless of the number of in situ stations represented by each model, the corresponding in situ iNPP were log-normally distributed whereas model estimates were not, based on a single sample K-S test, except for Models 2, 6, 10, and 17 (not shown). Note that the minimum value for model iNPP was set similar to the minimum in situ value of 5.45 mgC m^{-2} d^{-1}, instead of removing extremely low or negative values. Then, analogous to the in situ iNPP approach, statistical outliers at a confidence level of 95% were discarded in each set of model iNPP output (i.e., 0–10% of model results).

3.6. Model Skill Assessment

A model-data intercomparison of NPP (log-transformed) and NO_3 was carried out by assessing the root-mean-square difference (*RMSD*) from 21 participating models using Target [*Jolliff et al*., 2009] and Taylor [*Taylor*, 2001] diagrams, where N is the number of observations in each variable:

$$RMSD^2 = \frac{\sum_{i=1}^{N} \left(X^i_{model} - X^i_{\text{in situ}}\right)^2}{N} = bias^2 + uRMSD^2$$

$$bias = \overline{X_{model}} - \overline{X_{\text{in situ}}}$$

$$uRMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(X^i_{model} - \overline{X_{model}}\right) - \left(X^i_{\text{in situ}} - \overline{X_{\text{in situ}}}\right)\right]^2}$$

Bias represents the difference between the means of in situ measurement and model output, and *uRMSD* represents the difference in the variability between in situ data and model results. Hence, bias and *uRMSD* provide measures of how well mean and variability are reproduced, respectively.

To visualize bias and *uRMSD* on a single plot, a Target diagram [*Jolliff et al*., 2009] is used after normalizing by the standard deviation (σ) of in situ data, i.e., normalized bias (bias*) is defined as:

$$bias^* = bias/\sigma_{\text{in situ}}$$

Although *uRMSD* is a positive quantity mathematically, normalized *uRMSD* (*uRMSD**) is also defined in Target diagram as:

$$uRMSD^* = uRMSD/\sigma_{\text{in situ}} \quad (when\ \sigma_{model} > \sigma_{\text{in situ}})$$

$$= -uRMSD/\sigma_{\text{in situ}} \quad (when\ \sigma_{model} < \sigma_{\text{in situ}})$$

where σ_{model} is the standard deviation of X_{model}. If *uRMSD** is positive (negative), the model overestimates (underestimates) the variability of the in situ data. In this diagram, the closer to the observational reference (the origin), the higher the skill of the model.

The Taylor diagram [*Taylor*, 2001] illustrates a different set of statistics in terms of *uRMSD** that is comprised of standard deviation (σ) of the model output and in situ data as well as the Pearson's correlation coefficient (r) between model estimates and *in situ* measurements:

$$uRMSD^* = \sqrt{1 + \frac{\sigma_{model}{}^2}{\sigma_{\text{in situ}}{}^2} - 2 \cdot \frac{\sigma_{model}}{\sigma_{\text{in situ}}} \cdot r}$$

Note that, unlike the Target diagram, bias is not illustrated in the Taylor diagram.

In addition, the Willmott skill (*WS*) scores [*Willmott*, 1981] were used to quantify an overall regional agreement between multimodel mean (i.e., iNPP, NO_3, Z_{eu}, sea ice concentration, and MLD) and observations, and computed as:

$$WS = 1 - \frac{N \cdot RMSD^2}{\sum_{i=1}^{N}\left[\left(X^i_{model} - \overline{X_{\text{in situ}}}\right) - \left(X^i_{\text{in situ}} - \overline{X_{\text{in situ}}}\right)\right]^2}$$

The highest value, $WS = 1$, means perfect agreement between model and observation, while the lowest value, $WS = 0$, indicates disagreement. In this study, *WS* is used to quantify model performance in simulating different parameters from various model runs at regional scales. However, *WS* may not be useful when in situ data have extremely low variability and zero mean, i.e., when nutrients are depleted, because $\sum_{i=1}^{N}\left[\left(X^i_{model} - \overline{X_{\text{in situ}}}\right) - \left(X^i_{\text{in situ}} - \overline{X_{\text{in situ}}}\right)\right]^2$ becomes close to $N \cdot RMSD^2$. Moreover, since *WS* is often criticized for producing high skill values for entirely uncorrelated signals [*Ralston et al*., 2010], we provide the modeling efficiency (*ME*) as an alternative validation metric which determines the relative magnitude of the residual variance compared to the measured data variance [*Nash and Sutcliffe*, 1970]. *ME* indicates how well the plot of observed versus simulated regionally averaged data fits the 1:1 line [*Moriasi et al*., 2007], but it can be sensitive to a number of factors such as sample size, outliers, and magnitude bias [*McCuen et al*., 2006].

$$ME = 1 - \frac{RMSD^2}{\sigma^2_{\text{in situ}}}$$

The range of *ME* lies between 1.0 (perfect fit) and $-\infty$ [*Stow et al.*, 2009]. If *ME* is zero, the model estimates are as accurate as the mean of the observed data, whereas an efficiency less than zero ($-\infty < ME < 0$) indicates that the observed mean is a better predictor than the model. *ME* can also be illustrated by drawing a circle with a radius equal to one, i.e., $ME = 0$, from the origin on the Target diagram. For example, *ME* is negative (positive) when a model symbol appears outside (inside) the circle with a radius of one.

4. Results

The models broadly captured the spatial features of iNPP on a pan-Arctic scale. A majority of the models underestimated iNPP by varying degrees in spite of overestimating surface NO_3, MLD, and Z_{eu} throughout the regions. Model skill of iNPP exhibited little difference over sea ice condition and bottom depth. The models performed equally well on daily versus monthly scales and relatively well for the most recent decade and toward the end of Arctic summer. Much complexity can be seen from the high and sometimes opposing interaction between iNPP and selected physical variables, as shown by the variable degree of skill provided by the different participating models.

4.1. Comparison Between Simulated Daily and Monthly Mean iNPP

Most of the models in this intercomparison were only able to provide monthly mean rather than daily mean iNPP. In order to determine how the results of this analysis would be affected by using monthly mean model output instead of simulated daily iNPP, those five models that provided both daily and monthly estimates (Models 1, 2, 3, 7, and 9) were compared to each other. The distributions of simulated daily and monthly mean iNPP were very similar in each model (Figures 3a–3c, 3f, and 3h) and they were strongly correlated ($r = 0.83$ to 0.91, $p < 0.01$). On average, the maximum simulated iNPP was up to 6% higher when computed from daily mean values compared to the monthly mean estimates (log-transformed). The standard deviation was higher (4–23%) when computed from the daily mean iNPP estimates, except Model 1. However, the simulated mean iNPP from those five models computed from the monthly averaged output was almost equal or slightly higher (up to 7%) than that computed from the daily mean estimates. This demonstrates that, as expected, the monthly averaging smoothed the daily variability while the mean value exhibited little change. Hence, results based on the simulated monthly mean iNPP are representative of higher temporal resolution data. Although the simulated daily mean iNPP exhibited little difference over monthly averaging, the simulated monthly mean iNPP could miss details of the dynamics of NPP on daily time scales. For example, no highest estimate from the bloom peak should be shown in model monthly mean output, relative to in situ data.

4.2. Model Skill Assessment of iNPP and Depth-Averaged NO_3

iNPP was underestimated by most models on a station-by-station basis though not consistently. Log-mean values of simulated iNPP were mostly negatively biased (Table 3) because the log-distribution of iNPP from many models was negatively skewed with a longer tail on the left, toward low values. While the bias was small for some models, it was very large, i.e., a factor of 10 for others. Thirteen out of 21 models reproduced iNPP up to 1500 mgC m^{-2} d^{-1} and five others estimated iNPP up to 3000 mgC m^{-2} d^{-1}; the remaining three models simulated iNPP up to 4300 mgC m^{-2} $d-^{1}$ (see Figure 3) while the maximum in situ value was 5255 mgC m^{-2} d^{-1} (Table 2). Generally, models with large *uRMSD*, i.e., overestimated variability of iNPP or depth-averaged NO_3, exhibited high standard deviation due to underestimating iNPP (Figure 3) or overestimating depth-averaged NO_3 (not shown), respectively. Unlike iNPP, depth-averaged NO_3 was positively biased in most of the models, indicating those models overestimated the mean NO_3 in the top 100 m, and their correlation coefficients were relatively high compared to those of iNPP (Table 3). However, no evident relationship was found that better model skill in estimating NO_3 effected better reproduction of iNPP, or vice versa, in terms of *RMSD* (i.e., combining mean and variability).

4.3. Model Skill Assessment of iNPP as a Function of Ice Cover, Depth, Decade, and Season

Based on Table 3, model skill of estimating iNPP was not a function of model domain, type, or complexity in terms of *RMSD* on a pan-Arctic scale. Hence, the model skill was examined spatially as a function of sea ice

Table 3. Mean (μ) and Standard Deviation (σ) of Estimated iNPP (Log-Transformed) and Depth-Averaged NO_3 (mmol m^{-3})[a]

	log(iNPP)									Depth-Averaged NO_3								
		μ		σ							μ		σ					
Model	N	In Situ	Modeled	In Situ	Modeled	RMSD	Bias	uRMSD	r	N	In Situ	Modeled	In Situ	Modeled	RMSD	Bias	uRMSD	r
1	444	2.27	1.30	0.59	0.76	1.32	−0.98	0.89	0.16	439	3.21	4.05	2.63	1.71	2.64	0.86	2.50	0.39
2	440	2.31	2.02	0.60	0.44	0.67	−0.29	0.60	0.37	445	3.22	5.23	2.65	2.28	3.26	2.02	2.56	0.46
3	849	2.40	2.43	0.58	0.42	0.59	0.03	0.59	0.34	663	3.82	5.43	2.89	2.29	4.03	1.61	3.70	−0.01
4	853	2.40	2.32	0.58	0.36	0.63	−0.08	0.62	0.20	663	3.82	6.04	2.89	3.96	4.75	2.23	4.19	0.28
5	684	2.53	2.51	0.54	0.29	0.59	−0.01	0.59	0.10	461	4.22	3.67	3.09	3.47	2.70	−0.54	2.65	0.68
6	689	2.52	2.71	0.54	0.27	0.59	0.19	0.56	0.16	461	4.22	0.41	3.09	0.41	4.73	−3.80	2.82	0.66
7	665	2.47	2.31	0.60	0.34	0.59	−0.16	0.57	0.37	621	3.89	5.60	2.90	2.76	3.27	1.72	2.78	0.52
8	584	2.42	2.31	0.59	0.83	0.92	−0.11	0.92	0.20	436	3.92	9.79	3.03	3.20	7.37	5.87	4.46	−0.02
9	717	2.44	2.57	0.60	0.51	0.68	0.13	0.67	0.27	663	3.82	5.01	2.89	4.88	5.67	1.20	5.54	0.05
10	407	2.58	2.50	0.53	0.22	0.55	−0.08	0.55	0.12	392	4.11	3.78	3.09	2.97	3.06	−0.33	3.04	0.49
11	647	2.50	1.81	0.54	1.08	1.43	−0.69	1.25	−0.09	448	4.26	6.53	3.07	2.09	3.93	2.27	3.21	0.27
12	678	2.51	1.82	0.55	1.07	1.42	−0.69	1.24	−0.06	461	4.22	9.33	3.09	3.18	6.36	5.12	3.77	0.28
13	483	2.45	2.13	0.60	0.47	0.71	−0.32	0.63	0.32	493	3.69	1.93	2.93	1.74	2.96	−1.74	2.39	0.57
14	740	2.36	2.18	0.58	0.78	0.88	−0.18	0.86	0.22	578	3.92	6.86	2.96	3.05	4.61	2.95	3.54	0.30
15	466	2.44	2.13	0.61	0.36	0.68	−0.31	0.61	0.28	-	-	-	-	-	-	-	-	-
16	353	2.26	2.39	0.59	0.50	0.60	0.14	0.59	0.44	-	-	-	-	-	-	-	-	-
17	742	2.39	2.43	0.58	0.31	0.60	0.03	0.60	0.23	550	3.82	7.29	3.00	2.86	4.62	3.48	3.04	0.46
18	581	2.37	1.42	0.60	0.83	1.45	−0.94	1.10	−0.17	548	3.21	5.82	2.45	2.77	4.06	2.62	3.10	0.30
19	489	2.47	1.98	0.58	0.72	0.99	−0.49	0.86	0.13	436	3.92	12.51	3.03	2.45	9.38	8.59	3.78	0.06
20	730	2.39	1.23	0.60	0.39	1.37	−1.16	0.72	−0.03	585	3.81	0.66	2.78	0.83	3.97	−3.15	2.42	0.56
21	862	2.40	1.96	0.59	0.76	0.98	−0.44	0.87	0.18	-	-	-	-	-	-	-	-	-

[a]*RMSD*, bias, *uRMSD*, and Pearson's correlation coefficient (*r*) are computed between each model estimate and in situ measurement (see Appendix A for details). The number of stations (N) varies mainly due to different model simulation periods and in situ data availability. Note that Models 15, 16, and 21 did not provide model estimates of NO_3, which is indicated by a dash (-).

presence, defined from satellite measurements (see section 3.3), and bottom depth as well as temporally by seasons and simulation periods (Figure 4). While more individual models overestimated mean iNPP (Figure 4a) and exhibited a higher correlation coefficient (Figure 4b) in ice-free regions than in ice-influenced regions, most models performed nearly equally well in terms of standard deviation between ice-influenced and ice-free regions. Indeed, as a group, model performance was similar regardless of sea ice presence or absence in terms of both bias and variability, i.e., *RMSD*. Similarly, in the case of bottom depth (Z_{bot}) (shelf, $Z_{bot} < 200$ m versus deep ocean, $Z_{bot} > 200$ m), little difference in model skill was exhibited in terms of bias (Figure 4c) and variability (Figure 4d). Therefore, based on the K-S test at a 5% level of significance, model performance was not significantly different between spatially defined regions (i.e., ice-free versus ice-influenced and shelf versus deep ocean) in terms of *RMSD* (not shown).

With respect to temporal scales, the performance of models was significantly different between two simulation periods: 1959–1989 and 2000–2011. Overall, the models performed better in the most recent decade (2000–2011) in terms of *RMSD* whereas most of them significantly underperformed in terms of correlation coefficient during the earliest decade ($r = -0.24$ to 0.24) compared to the latter period ($r = -0.19$ to 0.53) (Figure 4f). However, no significant difference was found between the two periods in terms of bias only (Figure 4e) based on the K-S test for 5% significance. This may be partly due to the fact that more in situ measurements have become available in recent years when ice-free conditions have become more pronounced. On the other hand, seasonally estimated iNPP was more negatively biased in the growth period (mean bias of −1.12 in April–June) compared to other seasons (mean bias of −0.72 in July, −0.24 in August, and −0.23 in September–October), but the models performance was not significantly different between seasons in terms of variability (Figures 4g and 4h). This is probably because the models have a seasonal cycle that, although slightly shifted early in phase for the Arctic Ocean, is still within a reasonable range. In addition, the models include a range of approaches to estimating the vertical distribution of irradiance (i.e., light transmission through sea ice). Some models simply scale down shortwave radiation by sea ice concentration, whereas others include radiation transmission as a fully coupled and bidirectional formulation (e.g., CICE-based models). Still other models compute light transmission through sea ice as a function of sea ice physics and biogeochemistry. For example, the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (PISCES)-based models have a complex formulation of the vertical penetration of PAR through the water column, which takes into account incoming radiation, wavelength, chlorophyll-dependent attenuation

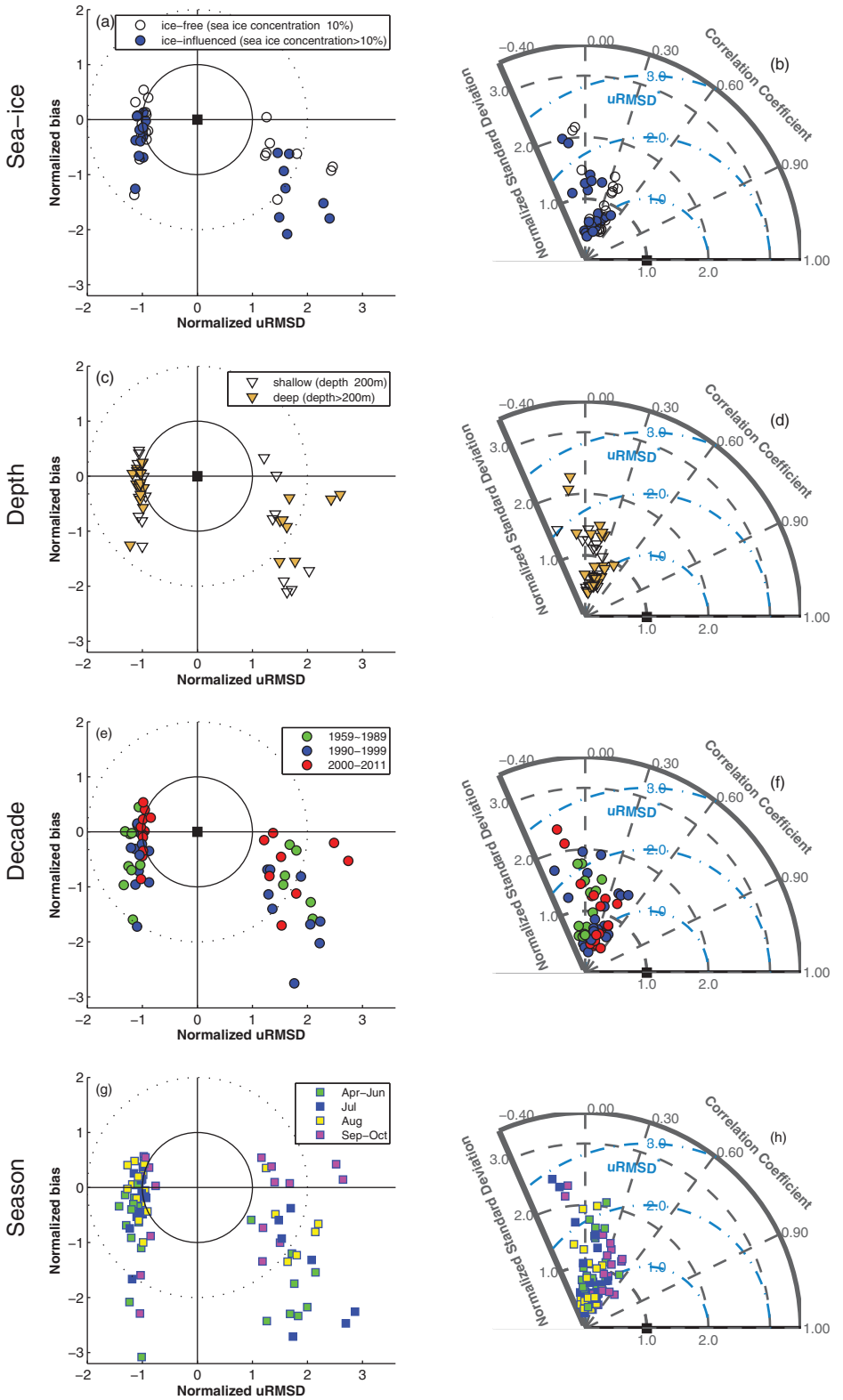

Figure 4. (a, c, e, and g) Target and (b, d, f, and h) Taylor diagrams illustrating relative model performance in reproducing iNPP as a function of (Figures 4a and 4b) sea ice condition: ice-free region (sea ice concentration ≤15%) versus ice-influenced region (sea ice concentration >15%), (Figures 4c and 4d) depth: shelf (≤200 m) versus deep (>200 m), (e and f) simulation period (1959–1989, 1990–1999, and 2000–2011), and (g and h) month (April–June, July, August, and September–October).

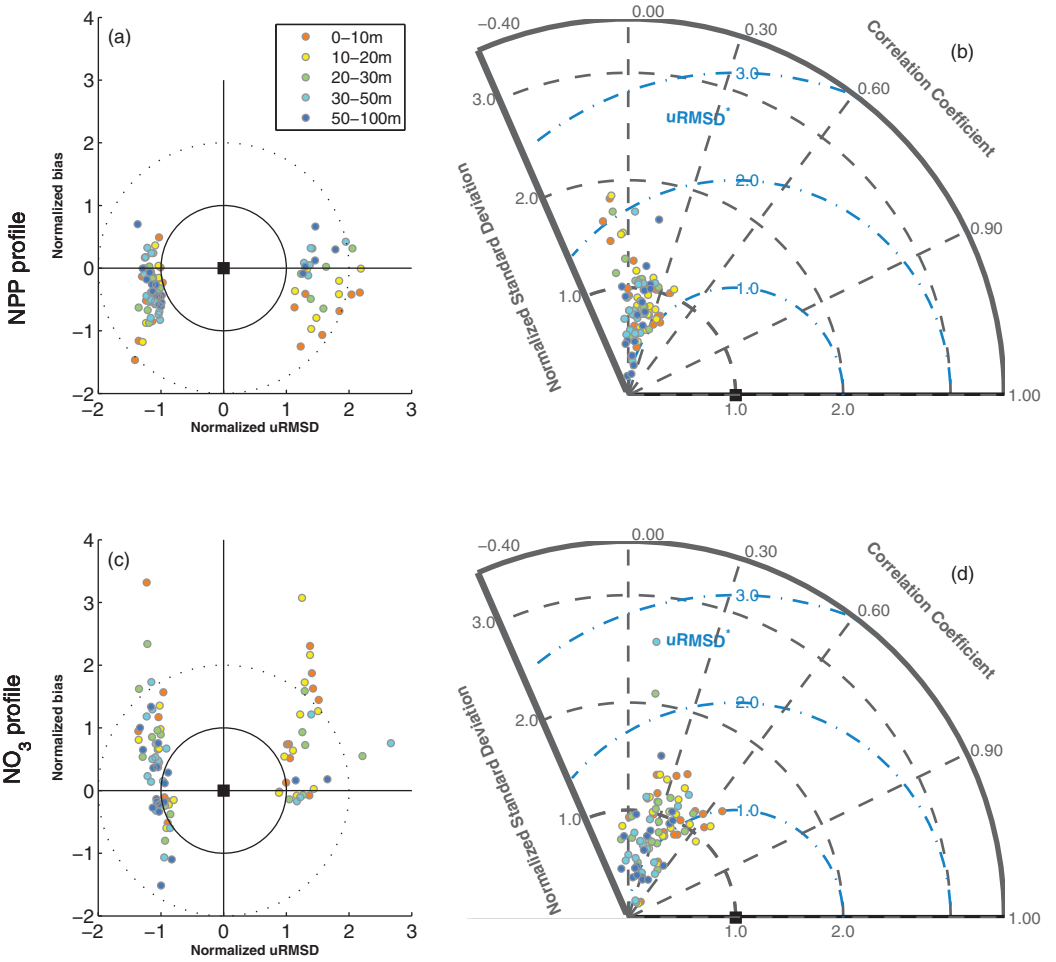

Figure 5. (a and c) Target and (b and d) Taylor diagrams of vertical profiles in (Figures 5a and 5b) NPP (mgC m^{-3} d^{-1}) and (Figures 5c and 5d) (NO_3) (mmol m^{-3}), which were grouped at given depth layers: 0–10 m, 10–20 m, 20–30 m, 30–50 m, and 50–100 m.

coefficients assigned to distinct phytoplankton groups represented in the model [e.g., *Aumont et al.*, 2015] and thus result in differing Z_{eu} for the same incoming PAR. These differences may account for the early negative bias in the participating model runs, but do not (yet) point to either phytoplankton bloom physiology or phenology processes being preferentially affected.

4.4. Model Skill Assessment of Vertical Profiles of NPP and NO_3

On a pan-Arctic scale, the models had a strong tendency to underestimate the observed mean NPP at various depths; eight out of 19 models had negative bias at all depth layers (Figure 5a). Note that not all the models provided vertical profiles of NPP and NO_3 (see Table 1). When the model standard deviation was smaller than the observed one (−*uRMSD**), mean NPP was mostly underestimated (Figure 5a). But, when the model standard deviation was larger than the observed one (+*uRMSD**), mean NPP was mostly underestimated only between 0 and 20 m and overestimated between 30 and 100 m (Figure 5a). Deep NPP (50–100 m; mean *RMSD* of 1.27) was better estimated than surface NPP (0–10 m; mean RMSD of 1.48) in terms of *RMSD* since both bias and *uRMSD* became less at depth.

Unlike NPP, nine of 19 models overestimated mean NO_3 at all depth layers while four models (Models 5, 10, 13, and 20) underestimated mean NO_3, except at the deepest layer (50–100 m) (Figure 5c). Model bias was lowest at depth such that mean NO_3 concentration in the deep layer was better reproduced than in the surface layer (mean bias of 0.29 and 1.31, respectively), while the correlation coefficient was slightly higher at the surface (Figure 5d). Furthermore, the range of model bias for NO_3 became wider (i.e., more scattered both in negative and positive *y* axis) when the model standard deviations were smaller than the observed one (−*uRMSD**). However,

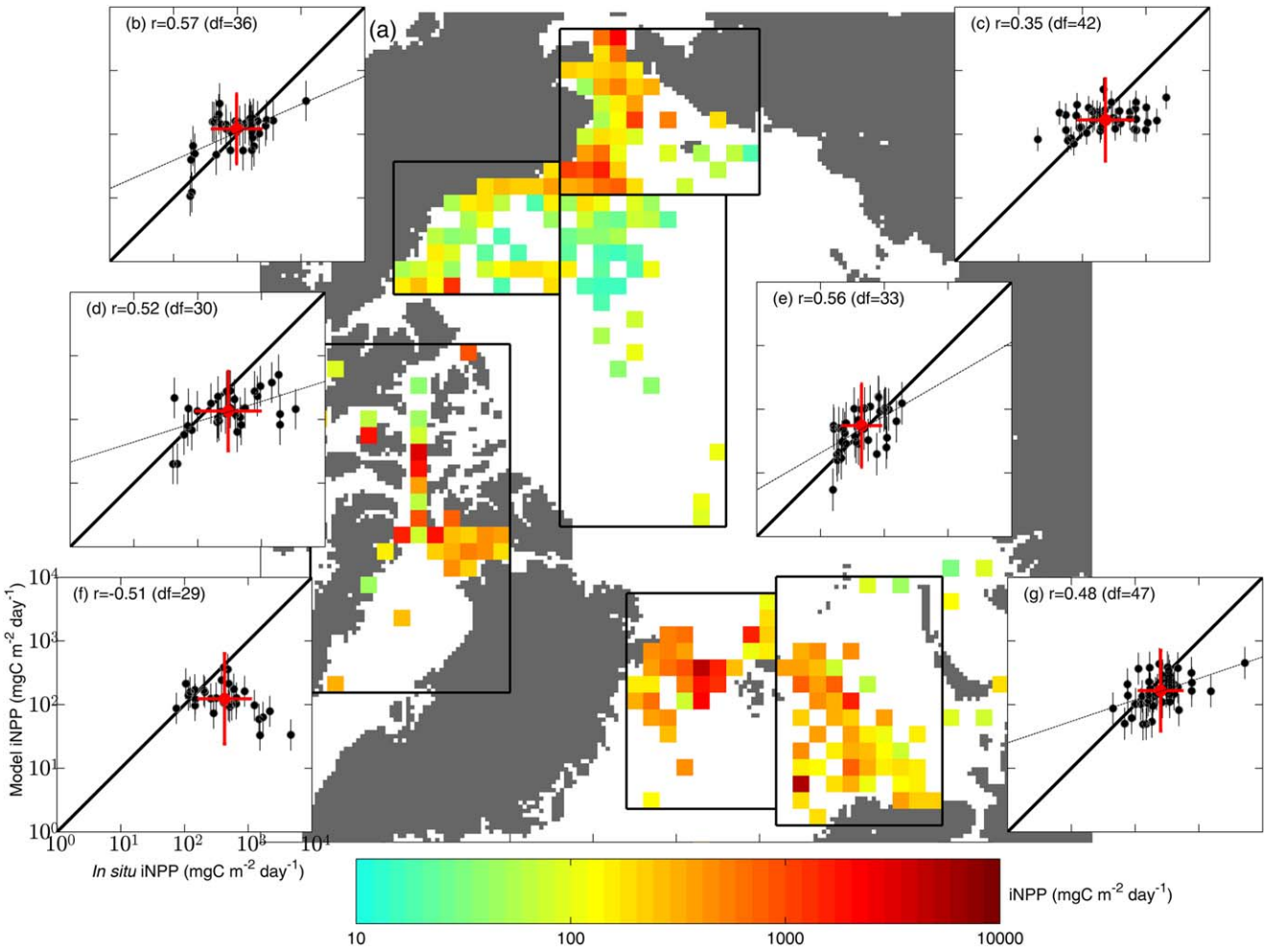

Figure 6. (a) In situ iNPP (mgC m^{-2} d^{-1}) projected on the EASE-Grid map is shown. (b–g) Average model estimates with an error bar (±1 standard deviation from the multimodel mean) in each grid cell were regionally compared to in situ values in Figure 6b the Beaufort Sea, 6c the Chukchi Sea, 6d Canadian Archipelago, 6e the central Arctic Basin, 6f the Greenland Sea, and 6g the Barents Sea; all inserts have the same axis units as in Figure 6f. In each plot, the solid-line shows a slope of 1.0 and the dashed line is a linear regression fit if significant ($p < 0.05$). The correlation coefficient (r) with degrees of freedom (df = number of grid cells−2) are shown in the upper left. A red circle with error bars indicates the regional average of in situ (x axis) and modeled (y axis) iNPP.

no significant difference in terms of *uRMSD* was exhibited between the depth layers (K-S test at a 5% level of significance). Overall performance of the models for NPP and NO_3 was not significantly different within each depth layer, although correlation coefficients were higher for NO_3 (Figures 5b and 5d). In general, the models performed better in reproducing NO_3 and NPP in the deepest layer (50–100 m) than at the surface (0–10 m).

4.5. Regional iNPP and NO_3 Climatologies

To illustrate climatological spatial patterns and compare with regional iNPP estimated from each model, available in situ iNPP was projected onto the Equal-Area Scalable Earth Grid (EASE-Grid) map (100 km spatial resolution) by averaging data within each grid cell (Figure 6a). EASE-Gridded model iNPP was then averaged across all models in each grid cell and compared to in situ climatological iNPP in six different regions of the Arctic Ocean where NPP was frequently measured between 1959 and 2011 (Figures 6b–6g). In situ and model iNPP exhibited no significant linear relationship on 10,000 km^2 basis (Figures 6c and 6f) and, regionally, the lowest *WS* (Table 4) in the Chukchi and Greenland Seas; both are influenced by inflow from subpolar latitudes. In addition, mean iNPP was most underestimated in the Canadian Archipelago and the Greenland Sea (Figures 6d and 6f) where *ME* was lowest (Table 4). Hence, iNPP was least reproduced regionally in the Greenland Sea in terms of intermodel mean (i.e., bias) as well as variability (i.e., *WS* and *ME*). As a group, the models performed best in the interior regions of the Arctic Ocean, such as the Beaufort Sea and the central Arctic Basin (Figures 6b and 6e) where NPP is relatively low (Figure 6a) and surface nutrients are

Table 4. Willmott Skill (*WS*) Score Ranging Between 0 and 1 for a Perfect Model, and Modeling Efficiency (*ME*) Between -∞ and 1 for a Perfect Model of Multimodel Mean iNPP, Surface, and Deep NO_3, Z_{eu}, Sea Ice Concentration, and MLD in Six Regions of the Arctic Ocean

		Variables					
Regions	Model Skill	iNPP	Surface NO_3	Deep NO_3	Z_{eu}	Sea ice	MLD
Chukchi	*WS*	0.49	0.46	0.33	0.67	0.90	0.08
	ME	−0.06	−3.65	−0.27	−0.55	0.72	−264
Canadian Archipelago	*WS*	0.61	0.49	0.59	0.31	0.54	0.46
	ME	−0.32	−2.31	0.30	−12.8	−0.04	−4.49
Beaufort	*WS*	0.77	0.09	0.51	0.65	0.85	0.25
	ME	0.36	−135	0.15	0.01	0.64	−7.64
Central	*WS*	0.75	0.13	0.46	0.45	0.67	0.34
Arctic Basin	*ME*	0.09	−38.3	0.15	0.01	0.39	−2.25
Barents	*WS*	0.63	0.80	0.25	0.63	0.72	0.93
	ME	−0.10	0.42	−0.14	−0.38	0.15	0.79
Greenland	*WS*	0.33	0.74	0.74	0.27	0.40	0.90
	ME	−2.84	0.09	0.41	−24.7	−0.19	0.62
Pan−Arctic	*WS*	0.67	0.63	0.47	0.67	0.80	0.92
	ME	0.12	−0.84	0.07	−0.25	0.50	0.74

depleted toward the end of the growing season [*Codispoti et al*., 2013]. At the larger pan-Arctic scale, model skill for iNPP was higher than for most individual regions, except where iNPP was best estimated, i.e., the Beaufort Sea (*WS* and *ME*) and the Central Arctic Basin (*WS*, Table 4).

Analogously, in situ and modeled NO_3 were also processed onto the EASE-Grid map in the deep layer (50–100 m; Figure 7a) as well as at the surface (0–10 m; Figure 8a). Again, the models better reproduced the regional mean NO_3 at the deepest layer overall (Figures 7b–7g). However, models notably underestimated deep NO_3 variability in the inflow Chukchi (Figure 7c) and Barents (Figure 7g) Seas ($r = -0.08$ and $r = 0.23$, respectively) (see Table 4). On the contrary, at the surface, the models exhibited a tendency to overestimate mean NO_3 (Figures 8b–8g). This bias was smaller in the Greenland and Barents Seas (Figures 8f and 8g, respectively) and highest in the interior regions of the Arctic Ocean (Figures 8b and 8e) where in situ surface NO_3 is depleted at the end of summer and NPP is relatively low compared to that in other regions (Figures 8a and 6a, respectively) [*Codispoti et al*., 2013]. Model skill scores (Table 4) indicate that the regions where the models have the best skill in reproducing iNPP are also the ones where they exhibited the least skill in reproducing mean surface NO_3, i.e., the Beaufort Sea and central Arctic Basin.

4.6. Regional MLD, Z_{eu}, and Sea Ice Concentration Climatologies

Using the same approach as in section 4.5, the regional, mean MLD was relatively better estimated in the Barents and Greenland Seas (i.e., highest *WS* and *ME* in Table 4) and least reproduced in the Chukchi Sea where MLD was consistently overestimated (Figure 9a). Moreover, the highest regional model skills in surface NO_3 were also calculated in the Barents and Greenland Seas. But they do not necessarily appear to influence iNPP estimates, i.e., poorly simulated iNPP, especially in those regions. Compared to the variables previously analyzed, Z_{eu} had the least amount of field data, especially in the Barents and Greenland Seas, and its mean was largely overestimated (Figure 9b); indeed, incoming PAR onto the surface ocean will be very different in open waters versus snow and ice-covered waters in the same latitude and season. In the absence of additional biological light-relevant parameters, the assessment of model estimates of Z_{eu} provides a limited understanding of phytoplankton light utilization, as it does not take into account that model phytoplankton will differ in their response to the same light field. Sea ice concentration is a logical next factor to consider; its model skill response is different from region to region, and is mostly underestimated in the Greenland Sea and overestimated in the Barents Sea (Figure 9c). In the Chukchi Sea, mean sea ice concentration was best estimated, but iNPP was still rather poorly estimated when compared to other regions (Table 4), as indicated earlier. While more high WS and ME model skill scores were shown for MLD and sea ice concentration among the different regions, no single environmental factor showed consistently high model skill for all regions; indeed, correlation analysis (not shown) between iNPP and any one environmental factor ME skill suggests a possible robust relationship with Z_{eu} instead, despite individual lower skill scores. Hence, iNPP model skill was constrained by different factors in different Arctic Ocean regions. Pan-Arctic averaging did not necessarily improve model skill for iNPP and associated variables over regional (Table 4) and subregional (Figure 6) scales.

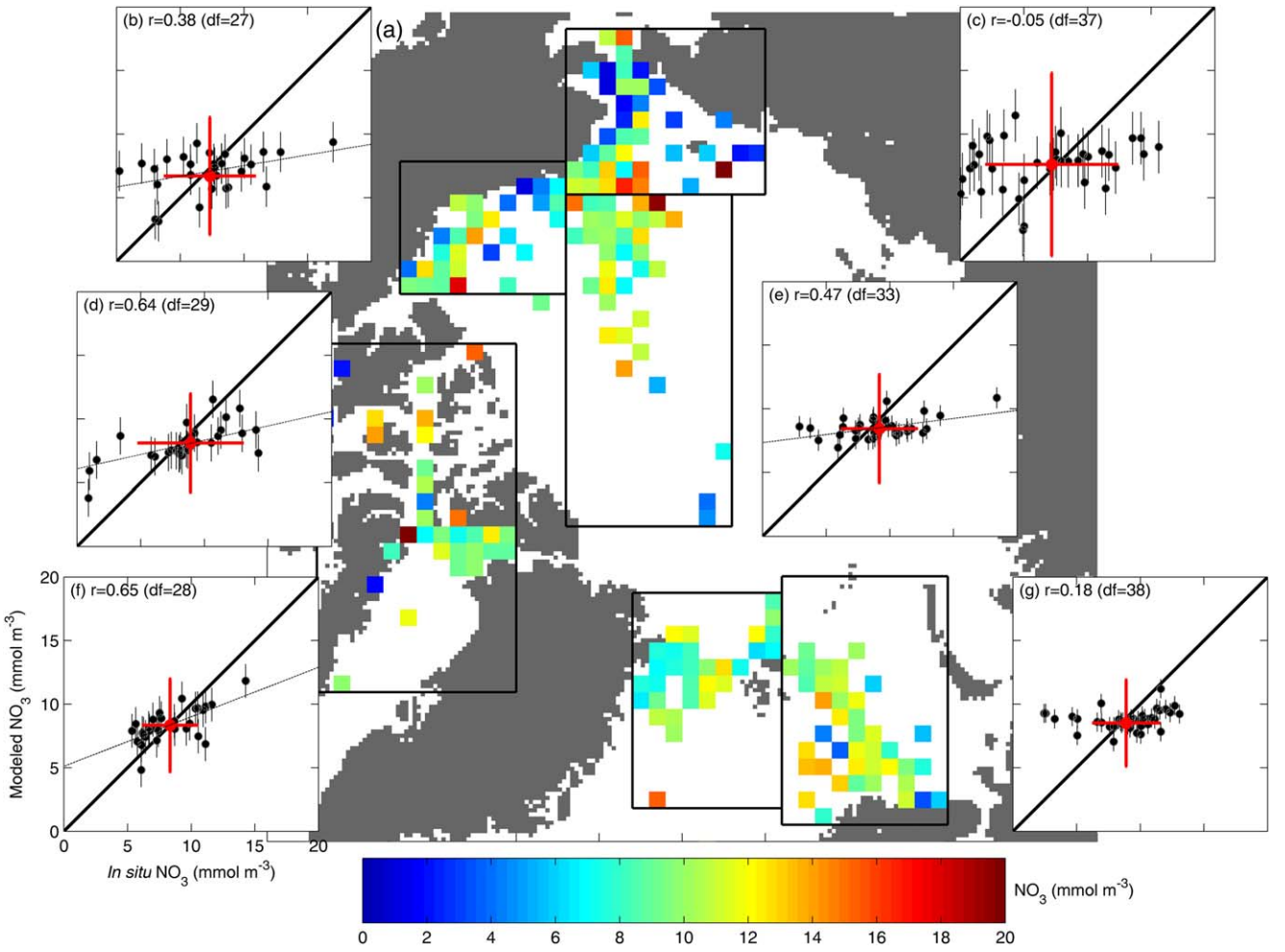

Figure 7. Same as Fig. 6, but for NO_3 (mmol m^{-3}) in the deep layer (50–100 m).

5. Discussion

It has long been known that marine primary production is controlled by the interplay of the bottom-up factors of light and nutrients, whose availability is a function of physical and/or biological processes. Earlier intercomparison studies for the Arctic Ocean emphasized the importance of a realistic representation of ocean physics [*Popova et al*. 2012], in particular vertical mixing and nutrient concentration, for accurate Arctic Ocean ecosystem modeling. Five BOGCMs manifested similar levels of light limitation owing to a general agreement on the simulated sea ice distribution [*Popova et al*. 2012]. On the other hand, the ESMs participating in the *Vancoppenolle et al*. [2013] study showed that Arctic phytoplankton growth was more limited by light related to sea ice than by nutrient availability for the period of 1980–1999. Furthermore, with the large uncertainties on present-day nitrate in the sea ice zone, ESMs used in the framework of the CMIP5 exercise were inconsistent in their future projections of primary production and oligotrophic condition in the Arctic Ocean [*Vancoppenolle et al*., 2013]. The need to quantify the relative role of environmental and ecological controls on NPP in the heterogeneous, seasonally ice-covered Arctic Ocean remains.

In our study, iNPP model skill was linked to different environmental variables in the various Arctic Ocean regions. The models exhibited relatively lower skill in reproducing iNPP in high to very high production Arctic Ocean regions [in the sense of *Codispoti et al*., 2013] (i.e., the Chukchi and Greenland Seas in Table 4 and Figure 6), despite higher model skill in physical parameters such as sea ice concentration or MLD, respectively. Model performance in estimating iNPP was also more related to the model skill of sea ice conditions (and Z_{eu}) than NO_3 (and MLD) in the extremely low production regions [in the sense of *Codispoti et al*., 2013] (i.e., the Beaufort Sea and central Arctic Basin in Table 4 and Figure 8), where in situ NO_3 is virtually

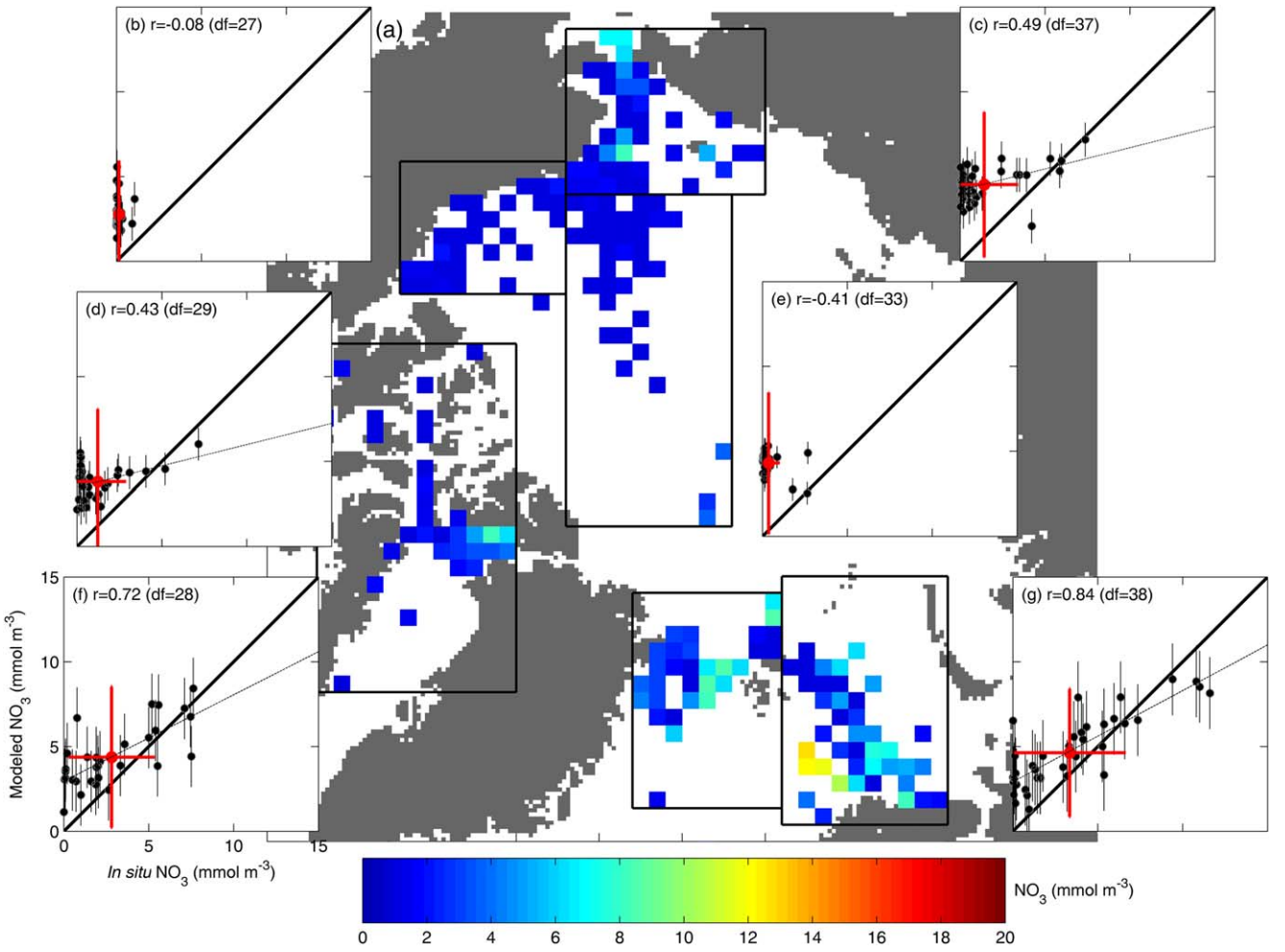

Figure 8. Same as Fig. 6, but for NO_3 (mmol m^{-3}) in the surface layer (0–10 m).

depleted (i.e., almost no variability in the in situ data). Field studies in the Arctic Ocean have failed to demonstrate any statistical relationship between phytoplankton biomass or productivity and ambient nutrient concentrations or nutrient utilization rates in eastern Arctic waters [*Harrison and Platt*, 1986; *Harrison and Cota*, 1991], while phytoplankton biomass and productivity were significantly related to incident radiation [*Harrison and Cota*, 1991; but see *Tremblay et al*., 2015]. Hence, simulated iNPP was close to in situ values in such low production regions because light fields were likely better reproduced, even though surface NO_3 was systematically biased (Figures 8b and 8e).

Our study also showed that simulations of surface NO_3 (nutrient) and Z_{eu} (light) associated with MLD (controlling the depth horizon where phytoplankton cells are) exhibited low model skill in most of the regions (Table 4), being lowest in the interior of the Arctic Ocean. This observation is not necessarily new [*Popova et al*., 2012; *Vancoppenolle et al*., 2013], but one should be cautious when interpreting this result because these three simulated parameters were positively biased (i.e., means usually overestimated) (Figures 8 and 9). For example, the simulated, mean surface NO_3 was vastly overestimated, especially in the low-productivity, nutrient-depleted interior regions of the Arctic Ocean, i.e., the Beaufort Sea and the central Arctic Basin where model and in situ iNPP agreed best (Figures 6b and 6e; Table 4). At least on a regional scale, the overestimation of surface NO_3 possibly stems from spurious vertical mixing between the upper mixed layer and deeper water in the model simulations, resulting in excessive nutrient supply into the surface layer. For example, except in the Greenland and Barents Seas, MLD mean and variability were poorly estimated (Table 4) and mostly overestimated (Figure 9a) by as much as twofold, possibly resulting in higher nutrients at the surface. Alternatively, a physiological perspective may be examined, based on the

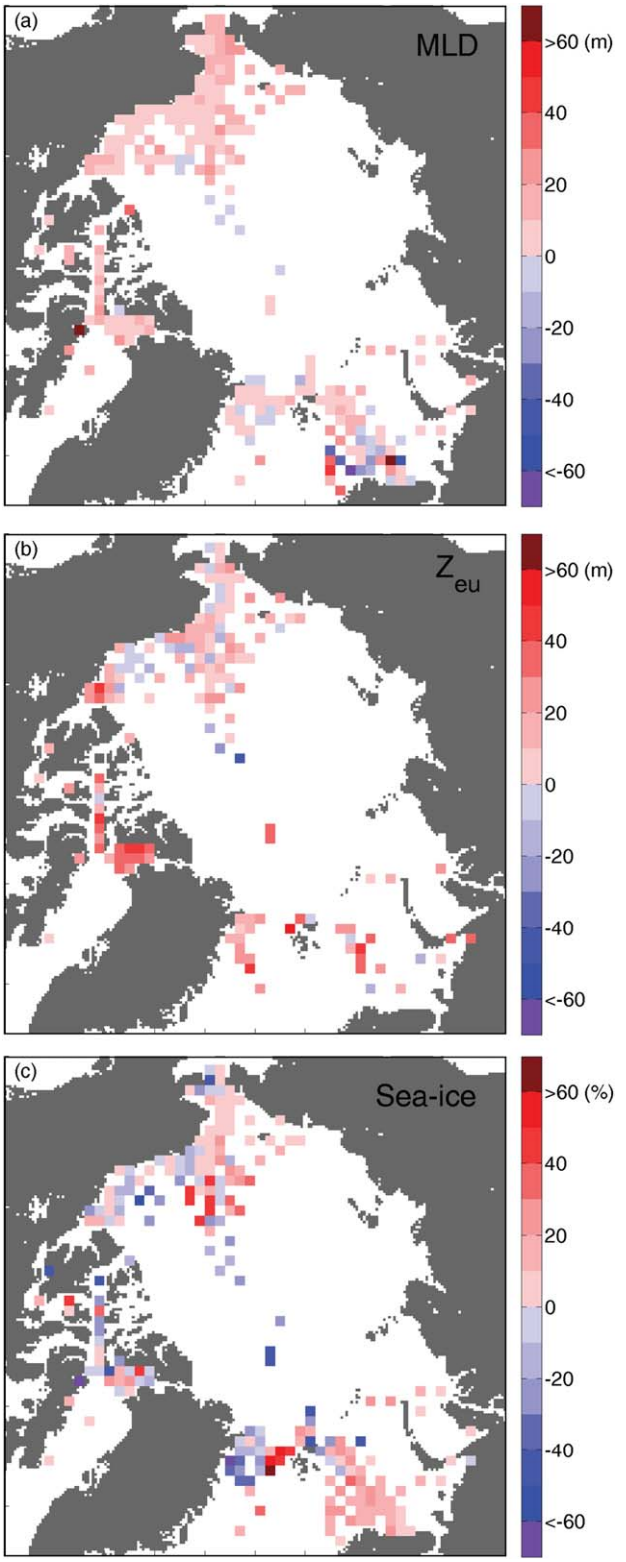

Figure 9. Modeled MLD, Z_{eu}, and sea ice concentration bias: (a) MLD bias between 15 model mean and monthly climatological MLD (m), (b) Z_{eu} bias between 9 model mean and in situ Z_{eu} (m), and (c) sea ice concentration bias between 19 model mean and satellite-derived values (%).

most common model parameterization of phytoplankton growth (i.e., a combination of a nutrient term, a light term, a temperature-dependent growth term, and a maximum growth rate over a daily period). Then, the overestimated NO_3 may be caused by insufficient phytoplankton nutrient uptake (i.e., too low a rate for Arctic waters) since NPP was overall underestimated by most of the models. Phytoplankton nitrate uptake is strongly dependent on both temperature (i.e., less at colder temperatures) [*Lomas and Glibert*, 1999] and irradiance (i.e., photoinhibited at higher levels) [*Hu and Smith*, 1998] and becomes more dependent on their combined effect in the simultaneously cold and dark Arctic waters during the transition to spring conditions [*Tremblay et al.*, 2006]. It is likely that several of the models have globally relevant phytoplankton physiological parameter values that are too high and may thus not correspond to the lower and narrower ranges required by Arctic (or polar) phytoplankton. Some Arctic phytoplankton species are shade-acclimated, especially in spring, though most available data are for summer, light-adjusted phytoplankton [*Sakshaug*, 2004; *Harrison et al.*, 1982]. Arctic phytoplankton community structure appears to change toward smaller cells in response to warming both in the field and in experiments [*Li et al.*, 2009; *Tremblay et al.*, 2009; *Coello-Camba et al.*, 2014, 2015] that may have different nutrient uptake kinetics [*Lovejoy et al.*, 2007], while individual phytoplankton species may have narrower temperature ranges for optimal growth, usually in combination with light requirements [*Harrison and Platt*, 1986; *Vernet et al.*, 2012]. These phenological changes will require the inclusion of multiple phytoplankton size classes in models and/or different physiological rates.

The distribution of NO_3—hence, its potential availability for NPP—differed by depth (Figure 5c) as well as region (Table 4). A less ecological and more pragmatic explanation for the overestimation of surface NO_3 may be that the models are incapable of reproducing the vertical mixing dynamics required (as shown by the ubiquitous MLD overestimation; Figure 9a); no evident

relationship was found between vertical resolution and skill in surface NO_3 among the participating models (not shown). Indeed, the future Arctic Ocean may produce less NPP per unit area due to less surface NO_3 [e.g., *Vancoppenolle et al*., 2013] resulting from enhanced freshening [*Coupel et al*., 2015], as also hypothesized by *McLaughlin and Carmack* [2010]. Unlike surface NO_3, deep NO_3 was more realistically reproduced pan-Arctic wide in terms of the multimodel ensemble mean values at 100 km grid scales (average symbols in Figure 7). Indeed, deep NO_3 and sea ice concentration (i.e., measured as *ME*) were relatively well explained by the multimodel ensemble in at least four regions (Table 4). Still, deep NO_3 (50–100 m) was relatively poorly estimated in the inflow Barents and Chukchi Seas in terms of correlation coefficients (Figures 7c and 7g) and regional means (*WS* in Table 4). Those two regions are strongly influenced by horizontal advection of sub-Arctic inflow waters which contribute to the nutrient supply into the Arctic Ocean via the Bering Strait and Fram Strait/Barents Sea openings [e.g., *Carmack and Wassmann*, 2006; *Wassmann et al*., 2015]. *Torres-Valdé et al*. [2013] estimated the volume and NO_3 transports into the Arctic Ocean based on a combination of modeled and in situ velocity fields and nutrient sections: total volume and NO_3 transport into the Arctic Ocean at depth corresponded to 20% and 80% through the Bering Strait and the Barents Sea opening, respectively (i.e., a significant contribution). *Smith et al*. [2011] also showed the greater relative importance of this advective nutrient input compared to vertical mixing in the Arctic Ocean.

Involvement of a light-driven component of the simulation of NPP is provided by the fact that Z_{eu} was generally overestimated in all Arctic regions by most of the nine models that provided this field (Figure 9b). The simulated Z_{eu} (9 model mean of 56 m), compared to the mean in situ Z_{eu} (37 m), was deeper than MLD (18 model mean of 23 m and climatological mean of 16 m). Possible explanations for the Z_{eu} overestimation may include the prescription of lower turbidity levels than the typical values in the Arctic Ocean. In addition (or instead), the parametrization of light extinction in the water column may be associated with sea ice conditions, i.e., the Z_{eu} overestimation, coupled with underestimation of sea ice concentration, may result in light being able to penetrate deeper into the water column during a simulation. From a biophysical perspective, the self-shading effect caused by the presence of phytoplankton is considered in many of the participating models already. Although we focused on model skill in estimating NPP, not phytoplankton biomass, the self-shading effect has been reported as negligible for high latitudes in sensitivity analyses [*Manizza et al*., 2005]. Yet, other biogeochemical species like colored dissolved organic matter that could reduce the light penetration into the water column are not represented in the models used in this study [*Granskog et al*., 2012; *Dutkiewicz et al*., 2015; *Kim et al*., 2015]. Nonetheless, NPP was still underestimated despite apparent light availability, emphasizing a physiological coupling between light availability and light requirements by phytoplankton; alternatively, it suggests that low NPP would generate low phytoplankton biomass which would result in a deeper Z_{eu} simply due to lack of light absorption.

Given that the models overestimated surface NO_3 and Z_{eu}, the simulated phytoplankton populations would experience more nutrient and light availability, which should have led to higher iNPP. However, the simulated NPP was mostly underestimated in all regions of the Arctic Ocean, clearly indicating that surface nutrients were not the primary limiting factor for the underestimated iNPP in these model simulations. Because the regions with higher model skill at estimating iNPP coincided with better model skill in Z_{eu} and sea ice concentration, especially in low production regions, the amount of shortwave radiation or PAR transmitted through sea ice could account partly for underestimating iNPP [*Babin et al*., 2015]. Our findings further suggest that biological processes (i.e., phytoplankton growth and photosynthesis) were possibly misrepresented by the model configurations run for this Arctic Ocean exercise. A previous model intercomparison study for lower latitudes pointed out that an improved understanding of the temperature effect on photosynthesis and a better parameterization of the maximum photosynthetic rate are required [*Carr et al*., 2006]. Hence, one might argue that phytoplankton growth functions were not adequately parameterized especially for colder high latitudes in many models, resulting in lesser utilization of surface nutrients. This is debatable in the Beaufort Sea where the effect of temperature on photosynthetic parameters over the in situ range observed (−2 to 8°C) was found not significant [*Huot et al*., 2013]. Furthermore, a large variability in photosynthetic parameters and half-saturation constants exists in Arctic ecosystem models, as recently reviewed by *Babin et al*. [2015]; during a model sensitivity exercise, they showed that temperature was not as strong a limiting factor for phytoplankton growth as others (i.e., light and nitrate). Although it could be that nutrients, light and temperature relationships need to be adjusted to improve model skill in the Arctic Ocean, such a regional tuning is, however, not possible for global BOGCMs and ESMs.

Simulated NPP may be low due to not only limiting factors, but also because of too little biomass, e.g., through excessive loss of carbon from or in the upper water column. Vertical export of organic matter is parameterized in all models. The organic carbon to organic nitrogen stoichiometric ratio in sinking particulate matter has been shown not to be constant in Arctic waters [*Tamelander et al*., 2013; *Michel et al*., 2006], as assumed in most models. Arctic phytoplankton are also known to have large exudation levels during certain phases of their growth cycle [*Vernet et al*., 1998; *Laroche et al*., 1999] which promotes the formation of sinking aggregates, resulting in significant carbon export from the surface layer when large phytoplankton cells dominate [*Reigstad et al*., 2011; *Wassmann et al*., 2006]. Changes in phytoplankton species composition and bloom phenology in Arctic waters are also being reported. These changes include shifts toward smaller cells, possibly with different physiology, especially in the freshening Beaufort Sea and Canadian Arctic Archipelago [*Li et al*., 2009; *Tremblay et al*., 2009], which will again affect vertical export fluxes. Changes in bloom phenology have been described as lengthening bloom periods or even the presence of fall blooms (reviewed by *Babin et al*. [2015]; *Michel et al*. [2006, 2015]). However, these changes are fairly recent and likely not yet simulated by models nor captured by the in situ data used for our exercise, despite covering five decades back (i.e., diminishing the likelihood of an effect in our field data set).

If bottom-up factors were not the main or sole controls on phytoplankton productivity, grazing could be: not only may grazers reduce NPP but they also regulate vertical flux. If simulated grazing rates were too high, then the resulting nutrient concentrations would be much higher than the values of the high saturation constants used. This scenario was ruled out for high latitudes by a PISCES-based model simulation in which bottom-up controls did dominate [*Aumont et al*., 2015]. On the other hand, the northward advective inflow of sub-Arctic water has been reported to also carry significant zooplankton biomass [*Wassmann et al*., 2015]. Hence, it is possible that the Arctic Ocean primary production in model simulations might be reduced under increased grazing pressure from such sub-Arctic (expatriate) zooplankton. Such a hypothetical extension of sub-Arctic grazing influence to the interior regions of the Arctic Ocean is not currently supported by observations [*Kosobokova and Hopcroft*, 2010]. Much is changing with respect to trophic structure in the Arctic Ocean, including grazer community composition and biodiversity [*Bluhm et al*., 2015], all of which may or may not be simulated by the models, with consequences for estimating iNPP. However, we currently lack a comprehensive review and climatology for in situ micro, meso, and macrozooplankton growth and grazing rates in the Arctic Ocean to broadly assess model relative skill.

Our use of an in situ NPP database allowed the assessment of model skill in ice-influenced as well as in ice-free waters, overcoming a limitation encountered by previous model intercomparison exercises for the Arctic Ocean [e.g., *Popova et al*., 2012; *Vancopenolle et al*., 2013] that depended on satellite-derived NPP, which is not available for ice-covered regions. Regardless of whether sea ice was present or not, the mean and variance of in situ iNPP were almost identical, confirming that the ice-influenced regions could be as productive as the ice-free regions (Table 2). It should be noted that the similarity could be due to biased seasonal coverage by field studies in ice-free waters, introducing more measurements collected in late summer and fall (i.e., September), thus lowering the ice-free iNPP average (not shown). Alternatively, the similarity in simulated iNPP values in ice-free and ice-influenced areas could result from the averaging of regional differences previously observed in field and satellite-derived NPP data [e.g., *Arrigo et al*., 2012; *Hill et al*., 2013; *Ardyna et al*., 2013] as well as in model simulations [e.g., *Zhang et al*., 2015]. Recently, *Jin et al*. [2016] showed that simulated under-ice production is also regionally different (i.e., higher in the Arctic shelves, possibly due to enhanced nutrient supply) [*Zhang et al*., 2015], but not necessarily related to the recent decrease in sea ice concentration, especially in marginal ice zones. Regardless of whether NPP was measured in deep versus shallow areas or ice-influenced versus open water regions, model skill for iNPP differed little on such spatial scales.

Time did make a difference, however. Model skill varied from decade to decade, and higher model skill was found for the most recent decade (Figures 4e and 4f) when in situ iNPP and NO_3 were lower (Table 2), since the models performed better with low iNPP. For example, on a pan-Arctic scale, in situ carbon uptake rates decreased to 177 mg C m^{-2} d^{-1} (N = 344; 2000–2011) compared to 316 mg C m^{-2} d^{-1} in the earlier decades (N = 391; 1959–1989). Similarly, in situ depth-averaged NO_3 decreased from 5.1 mmol m^{-3} (N = 284; 1979–1999) to 2.9 mmol m^{-3} (N = 379; 2000–2011). These decreases likely resulted from differences in geographical sampling distribution between decades (Figure 1). For example, during 2000–2011, more data were available from the large Canadian and International Polar Year programs in low production regions,

such as the Beaufort Sea and the central Arctic Basin (Figure 1c). Not all of the participating models captured these decreasing changes in simulated iNPP and NO_3 (not shown). However, regardless of whether model iNPP estimates increased or decreased over the most recent decade, simulated iNPP values became more comparable to in situ values on a pan-Arctic scale, such that model skill was seen to be improved in more recent years (Figures 4e and 4f). On a seasonal time-scale, the models generally performed poorly in the beginning of the growing season (April–June) compared to later months (Figures 4g and 4h). For some models, this could be due to insufficient penetration of winter mixing, as the main mechanism controlling basin-scale patterns of nutrient supply [e.g., *Popova et al*., 2012]; though, as stated earlier, MLD was overestimated overall (Figure 9a). Also occurring at this time is the critical balance between the onset and phenology of sea ice melting, ice algal bloom, and/or phytoplankton bloom; small mismatches in timing among them resulted in a large variability of simulated phytoplankton availability, especially at the Arctic marginal seas [*Ji et al*., 2013].

Model quantification of biogeochemical processes is highly dependent on how accurately physical fields are simulated [*Friedrichs et al*., 2006; *Popova et al*., 2012]. Submesoscale physical features are especially important to biogeochemical processes affecting the horizontal and vertical distribution of nutrients and phytoplankton biomass [*Watanabe et al*., 2014]. Most of the participating models are developed on global scales and only four of them are regional models of the order of 10 km resolution (Table 1). The Rossby radius of deformation is dramatically reduced with increasing latitude and it becomes less than 10 km in the Arctic Ocean. Hence, those models with resolution of the order of 10–100 km may not adequately resolve submesoscale physical features at the surface layer that are likely critical for Arctic biogeochemical processes—in particular—regionally [*Timmermans et al*., 2012], rather than larger-scale physical processes determining the response of the pan-Arctic marine ecosystem. High spatial resolution models or more Arctic-appropriate parameterizations and closures are necessary to capture small-scale physical processes, such as eddies, tides, fronts, internal waves as well as interactions between shelf and open ocean or sea and ice processes, that can influence phytoplankton growth, especially in high-productivity, coastal shelf regions adjacent to sub-Arctic waters. Indeed, regionally, model skill patterns are contradictory, such that the Greenland Sea has a very closely simulated surface and deep NO_3 and MLD (Table 4) but the least skill in simulated sea ice concentration (Figure 9c); yet it has the lowest skill in iNPP, underestimating it (Figure 6f). Conversely, the Beaufort Sea shows medium skill in deep NO_3, least skill in surface NO_3 (overestimating concentrations) and yet the best skill in iNPP (Table 4)—still underestimating iNPP and its variability (Figure 6b), though less than other Arctic Ocean subregions. Assuming improved parameterization of nutrient, light, and temperature relationships of phytoplankton growth, a higher spatial resolution model could potentially produce significant subregional features [e.g., *Yool et al*., 2015], as mesoscale and submesoscale processes have been shown to stimulate biological productivity in lower latitudes [*McGillicuddy et al*., 1998; *Oschlies and Garcon*, 1998; *Lévy et al*., 2001]. Finally, coastal regions, characterized by elevated primary productivity are also not properly resolved by a coarse model grid, in general [*Aumont et al*., 2015]. Higher resolution in ecosystem complexity has already been shown to lead to improved realism of BOGCM high latitude simulations, with the addition of slow growing macrozooplankton leading to the successful representation of the different bloom dynamics in the Southern Ocean [*Le Quéré et al*., 2016]. Much effort is now being applied in the modeling community to introduce high-resolution strategies into BOGCMs, especially those regionally focused on the Arctic Ocean [*Interagency Arctic Research Policy Committee*, 2013], with the hope that reproducing meso and submesoscale processes and features will improve their predictive skill for iNPP and other biogeochemical properties in the Arctic Ocean.

6. Summary and Recommendation

Using in situ, satellite-derived, and climatological data, we have analyzed model estimates of iNPP and NO_3 associated with physical variables such as MLD, Z_{eu}, and sea ice concentration from 21 BOGCMs and ESMs during the period of 1959–2011, although the time span of simulation as well as the initialization and forcing were different among the participating models. No models fully reproduced the variability of in situ measurements, whereas some of them exhibited almost no bias. Generally, the models captured the spatial features of the iNPP distribution: lower iNPP in the interior regions and higher iNPP in the inflow regions. The models exhibited different regional characteristics in estimating iNPP since each model's skill was a function of environmental factors with distinctive sensitivity in one region versus another. For example,

MLD and NO_3 were reproduced well in the Atlantic-side inflow and outflow regions, Z_{eu} had greater skill in the interior regions, and sea ice concentration was reasonably well estimated everywhere, except in the Greenland Sea. Most of the models generally underestimated mean iNPP, in spite of overestimating Z_{eu} and surface NO_3, especially in the Chukchi and Greenland Seas where mean in situ iNPP is high. Regionally, model iNPP was best estimated in the regions where simulated surface NO_3 was most overestimated relative to in situ values and where both sea ice concentration and Z_{eu} were reproduced well, such as the Beaufort Sea and the central Arctic Basin. By contrast, iNPP exhibited the largest model-data disagreement in the Greenland Sea where the simulated surface NO_3 and MLD had greater model skill, but sea ice concentration had lower model skill.

The Arctic Ocean appears to be experiencing a fundamental shift from a polar to a temperate mode, which is likely to alter its marine ecosystem [*Yool et al*., 2015; *Wassmann et al*., 2015]. Under a changing climate, sea ice loss, shifts in composition and distribution of phytoplankton functional types, and bloom phenology changes already play a role in the Arctic Ocean marine ecosystem [e.g., *Li et al*., 2009; *Ardyna et al*., 2014]. Globally, changes in nutrient and light availability have emerged as the most important determinants in models for alterations in simulated growth rates, with light generally playing a lesser role, except for the very high latitudes, particularly the Arctic Ocean [*Laufkötter et al*., 2015]. At these high latitudes, temperature does not appear to play a physiological role but could well play an ecological role of controlling phytoplankton bloom composition and phenology [*Babin et al*., 2015]. Thus, ecosystem processes and algal physiology in models need to be carefully parameterized for the Arctic Ocean in order to better quantify uncertainties in estimating primary production, including under-ice primary production and ice algae production [*Leu et al*., 2015; *Jin et al*., 2016]; enhanced functional types for all levels of the marine food web [*Babin et al*., 2015]; and, the role of horizontal advection of planktonic organisms versus their vertical export [*Wassmann et al*., 2015]. Furthermore, it is necessary to resolve submesoscale physical processes to decide the apparently contradictory bottom-up controls of NPP and the resulting biogeochemical cycles in the various subregions of the Arctic Ocean, with the hope that BOGCMs and ESMs perform within the uncertainty due to natural variability. Last but not least, models will continue to improve their skill when additional validating data sets and climatologies are gathered and openly shared for relevant biogeochemical, physiological, and ecological variables in the Arctic Ocean.

Appendix A: Brief Description of Participating Models

A1. Model 1

HYCOM-NORWECOM [*Samuelsen et al*., 2015] is a regional coupled physical-biogeochemical model with 11 compartments in the biogeochemical model. The physical model is the Hybrid Coordinate Ocean Model (HYCOM) [*Bleck*, 2002] adapted to the North Atlantic and Arctic Ocean [*Bertino and Lisaeter*, 2008]. The model encompasses the entire North Atlantic and has open lateral boundaries at the Bering Strait in the Pacific Ocean and near the equator between 3°S and 16°S in the Atlantic Ocean; the model is relaxed toward climatological values at the lateral boundaries. The model is forced by ERA-Interim [*Dee et al*., 2011], hydrological forcing from TRIP [*Oki and Sud*, 1998], but does not include tidal forcing. The sea ice model uses an elastic-viscous-plastic (EVP) rheology [*Hunke and Dukowicz*, 1997] while thermodynamics are described in *Drange and Simonsen* [1996]. NORWEgian ECOlogical Model (NORWECOM) [*Aksnes et al*., 1995; *Skogen and Søiland*, 1998] is coupled online to HYCOM and uses HYCOM facility for advection of tracers and has the same time-step as the physical model. The current model version includes two classes of phytoplankton (diatom and flagellates), two classes of zooplankton (meso and microzooplankton) derived with the same formulation from the model ECOHAM4 [*Pätsch et al*., 2009], three types of nutrients (N, P, and Si) and detritus (N and P), biogenic Si, and oxygen (O_2). Additionally, riverine nutrient loads are based on the results from the Global-NEWS model [*Seitzinger et al*., 2005]. The model assumes a constant element ratio (i.e., Redfield) as well as constant chlorophyll:N ratio in phytoplankton. In the model, light does not propagate through sea ice; therefore, the model is not expected to yield realistic primary production and nutrients in areas with thin ice cover where light actually is available for primary production.

A2. Model 2

This global biogeochemical model (1998–2014) uses the PISCES formulation on the horizontal curvilinear grid ORCA025 and 75 vertical levels. It is forced offline by daily fields of an ocean circulation simulation:

Nucleus for European Modelling of the Ocean (NEMO)-Océan PArallélisé (OPA) (version 3.1) [*Madec*, 2008]. Atmospheric forcings are 3 hourly ERA-Interim reanalysis product with the CORE bulk formulation. There is no assimilation of physical data. The sea ice model is the Louvain la Neuve sea Ice Model version 2 (LIM2) with EVP rheology [*Fichefet and Morales Maqueda*, 1997]. PISCES (volume 2) [*Aumont et al*., 2015] is a model of intermediate complexity designed for global ocean applications and is a component of the NEMO modeling system. It has 24 prognostic variables and simulates biogeochemical cycles of O_2, C, and the main nutrients controlling phytoplankton growth (NO_3, ammonium (NH_4), phosphate (PO_4), silicate (SiO_4), and Fe). The model distinguishes four plankton functional types based on size: two phytoplankton groups (nanophytoplankton and diatoms) and two zooplankton groups (micro and mesozooplankton). PISCES traces three nonliving pools for small and big particulate organic carbon (POC), and semilabile dissolved organic carbon (DOC). Version 3.5 of NEMO-PISCES with default parameter set is used here. The configuration is described and assessed in the Product User Manual and Quality Information Document (http://marine.copernicus.eu/documents/PUM/CMEMS-GLO-PUM-001-018.pdf and http://marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-018.pdf). Daily and monthly mean outputs are available on Copernicus Marine Service website (http://marine.copernicus.eu).

A3. Model 3

This model is the Plankton Types Ocean Model 5.3 (PlankTOM5.3) biogeochemical model coupled to the NEMO (version 2.3) global ocean circulation model and sea ice model (LIM2). The biological model is a 26 component marine pelagic ecosystem model, including three phytoplankton components (diatoms, coccolithophores, and mixed phytoplankton), two zooplankton components (micro and mesozooplankton), DOC, small and large detrital POC and particulate organic Fe, detrital particulate Si and calcite, and dissolved NO_3, SiO_4, and Fe [*Buitenhuis et al*., 2013]. Each phytoplankton functional type has C, chlorophyll, and Fe components, linked with a dynamic photosynthesis model [*Buitenhuis and Geider*, 2010]. The standard simulation was spun up from 1920 to 1955, and for the purpose of this paper the optimized simulation was run from 1956 to 2013 [*Buitenhuis et al*., 2013].

A4. Model 4

This model is the Plankton Types Ocean Model 10 (PlankTOM10) biogeochemical model coupled to the NEMO (version 3.1)-LIM2 ocean circulation and sea ice model. The biological model is a 37 component marine pelagic ecosystem model, including six phytoplankton components (diatoms, coccolithophores, picophytoplankton, N_2-fixers, phaeocystis, and mixed phytoplankton), three zooplankton components (micro, meso, and macrozooplankton), bacteria, DOC, small and large detrital POC and particulate organic Fe, detrital particulate Si and calcite, and dissolved NO_3, SiO_4, and Fe [*Le Quéré et al*., 2016]. PlankTOM10 has the same dynamic photosynthesis model as PlankTOM5.3 (Model 3). In addition, it has an improved representation of the difference between the high latitude Northern and Southern Hemisphere phytoplankton blooms through the representation of slow growing macrozooplankton. The different physical settings result in different overwintering macrozooplankton populations, in turn leading to different trophic cascades and more realistic phytoplankton blooms [*Le Quéré et al*., 2016].

A5. Model 5

The Carbon, Ocean Biogeochemistry, and Lower Trophics (COBALT) version 1.0 marine ecosystem model uses 33 state variables to resolve global-scale cycles of C, N, P, Si, Fe, calcium carbonate ($CaCO_3$), O_2, and lithogenic material [*Stock et al*., 2014a]. COBALT derives many elements of its biogeochemical formulation from the TOPAZ model (see Model 6). However, COBALT enhances the resolution of planktonic food web dynamics to better understand the climate impacts on the flow of energy from phytoplankton to fish [*Stock et al*., 2014b]. The model includes three phytoplankton groups (small, large, and diazotrophs) and uses the variable chlorophyll:C formulation of *Geider et al*., [1997] to estimate NPP. Results contributed to this study were conducted with the Modular Ocean Model (MOM) version 4p1 [*Griffies*, 2012] with sea ice dynamics as described in *Winton* [2000]. The model was run on a global domain with a nominal spatial resolution of 1° and a tripolar Arctic grid. The model uses 50 vertical layers with a nominal resolution of 10 m over the top 200 m. Simulations are forced with the Coordinated Ocean-ice Reference Experiment version 2 (CORE2) data set [*Large and Yeager*, 2009]. The 60 year forcing period was looped twice and results from the second loop provided for this study. Surface salinity is relaxed to observations with 2 month relaxation time-scale in order to avoid salinity drifts characteristic of long integrations of ocean-ice simulations without fully

coupled atmosphere dynamics [*Griffies et al*., 2009]. Complete details of the simulation and model validation can be found in *Stock et al*. [2014a].

A6. Model 6

Tracers of Ocean Phytoplankton with Allometric Zooplankton version 2.0 (TOPAZ2) is an ocean biogeochemical and ecological model developed at NOAA's Geophysical Fluid Dynamics Laboratory for global Earth System Model simulations contributed to the 5th assessment report of the Intergovernmental Panel on Climate Change [*Dunne et al*., 2012b,2013]. TOPAZ2 includes 30 tracers to describe the cycles of C, N, P, Si, Fe, O_2, alkalinity, and lithogenic material as well as surface sediment calcite dynamics [*Dunne et al*., 2012a]. TOPAZ2 considers three explicit phytoplankton groups (small, large, and diazotrophs) that utilize a modified growth physiology after *Geider et al*. [1997] with *Eppley* [1972] temperature functionality and iron colimitation with luxury uptake after *Sunda and Huntsman* [1997]. CO_2:NO_3:O_2 stoichiometry is 106:16:150 after *Anderson* [1995]. N:P is based on optimal allocation after *Klausmeier et al*. [2004]. Complete technical details are available in the supplementary information of *Dunne et al*. [2013]. Results contributed to this study were conducted with the MOM (version 4p1) [*Griffies*, 2012] with sea ice dynamics as described in *Winton* [2000]. The model was run on a global domain with a nominal spatial resolution of 1° and a tripolar Arctic grid. The model uses 50 vertical layers with a nominal resolution of 10 m over the top 200 m. Simulations are forced with the CORE2 data set [*Large and Yeager*, 2009]. The 60 year forcing period was looped twice and results from the second loop provided for this study. Surface salinity is relaxed to observations with 2 month relaxation time-scale in order to avoid salinity drifts characteristic of long integrations of ocean-ice simulations without fully coupled atmosphere dynamics [*Griffies et al*., 2009].

A7. Model 7

SINMOD is a model system that has been continuously developed over 30 years and includes a hydrodynamic and a sea ice module [*Slagstad and McClimans*, 2005], an ecosystem module [*Wassmann et al*., 2006], and stage resolved zooplankton models [*Slagstad and Tande*, 2007]. The hydrodynamic model is based on the primitive Navier-Stokes equations and is established on a Z-grid [*Slagstad and McClimans*, 2005]. The ice model is similar to that of *Hibler* [1979] and has two state variables: ice thickness and ice concentration. The ice momentum equation is solved together with an equation for the ice internal stress, using the EVP dynamical model of *Hunke and Dukowicz* [1997]. The ecosystem model includes nutrients (NO_3, NH_4, and SiO_4), phytoplankton (diatoms and flagellates), bacteria, heterotrophic nanoflagellates, microzooplankton, and two mesozooplankters: the Atlantic *Calanus finmarchicus* and the arctic *C. glacialis, two* compartments for fast and slow sinking detritus, DOC, and sediments. The model uses constant stoichiometry with a C:N ratio equal 7.6, based upon average data from the Barents Sea [*Reigstad et al*., 2002]. In the present study, SINMOD data are from a pan-Arctic configuration with 20 km horizontal grid resolution and with 25 vertical layers. Forcing fields includes atmospheric forcing from the ERA-Interim reanalysis (http://www.ecmwf.int) [*Dee et al*., 2011], tides from TPXO 6.2 model prescribed at open boundaries, and freshwater run-off from R-ArcticNet (http://www.r-arcticnet.sr.unh.edu/). More details of the recent updates of SINMOD and the model setup may be found in *Slagstad et al*. [2015].

A8. Model 8

The NorESM-OC is the ocean component of the Norwegian Earth System Model (NorESM). It is a coupled BOGCM, which comprises the Miami Isopycnic Coordinate Ocean Model (MICOM) coupled with the Hamburg Oceanic Carbon Cycle (HAMOCC)−5 model [*Schwinger et al*., 2016]. HAMOCC consist of a Nutrient-Phytoplankton-Detritus-Zooplankton (NPDZ)-type ecosystem model, where the primary production is formulated as a function of phytoplankton growth and nutrient concentration within the euphotic layer (fixed at 100 m). In addition to multinutrients (i.e., NO_3, PO_4, and dissolved Fe) colimitation, the phytoplankton growth is regulated by light availability and temperature. The model adopts a generic single class phytoplankton and zooplankton compartment. The particulate materials (i.e., POC, biogenic opal, and $CaCO_3$) produced at the surface sink with constant velocities and are remineralized at depth. The production of surface opal and $CaCO_3$ is determined by the SiO_4 availability at the surface. Nonremineralized particles reaching the seafloor undergo chemical reactions with sediment pore waters, bioturbation, and vertical advection within the sediment module. For the simulation produced in this study, the NorESM-OC was forced by the CORE2 atmospheric reanalysis data.

A9. Model 9

This model is the Biology/Ice/Ocean Modeling and Assimilation System (BIOMAS), a coupled biophysical model [*Zhang et al*., 2010a] with four model elements: an ocean circulation model, a sea ice model, a pelagic biological model, and a sea ice algae model. The ocean circulation model is based on the Parallel Ocean Program (POP), developed at Los Alamos National Laboratory [*Smith et al*., 1992] and modified by *Zhang and Steele* [2007] so that open boundary conditions can be specified. The POP ocean model is further modified by *Zhang et al*. [2010b] to incorporate tidal forcing from eight primary constituents [*Gill*, 1982]. The sea ice model is a thickness and enthalpy distribution (TED) sea ice model [*Zhang and Rothrock*, 2003; *Hibler*, 1980], with eight categories each for ice thickness, ice enthalpy, and snow depth. It is adopted from the Pan-arctic Ice/Ocean Modeling and Assimilation System (PIOMAS) [*Zhang and Rothrock*, 2003] and capable of assimilating satellite observations of sea ice concentration [*Lindsay and Zhang*, 2006] and sea surface temperature [*Manda et al*., 2005; *Schweiger et al*., 2011]. The pelagic biological model is an 11 component marine pelagic ecosystem model, including two phytoplankton components (diatoms and flagellates), three zooplankton components (microzooplankton, copepods, and predatory zooplankton), dissolved organic nitrogen (DON), detrital particulate organic nitrogen (PON), particulate organic Si, NO_3, NH_4, and SiO_4 (see Figure 3 in *Zhang et al*. [2014]; also see *Kishi et al*. [2007]). Values for key biological parameters used in the model are listed in *Zhang et al*. [2010a]. The sea ice algal component of BIOMAS represents colonies in a 2 cm layer at the bottom of each of eight sea ice thickness categories, coupled to the pelagic model through nutrient and biotic fluxes. The sea ice algae model has two ice algae components (diatoms and flagellates), with NO_3, NH_4, and SiO_4 as limiting nutrients for ice algal growth. Nutrient and biotic exchange at the sea ice and water interface is based on *Jin et al*. [2006].

A10. Model 10

This model is NEMO (version 3.2)-MEDUSA, a coupled biophysics model comprised of an ocean general circulation model (OGCM), a sea ice model, and a representation of marine biogeochemistry [*Popova et al*., 2010, 2012]. Ocean circulation is modeled using OPA version 9 (OPA9) [*Madec*, 2008], a primitive equation model configured at global-scale, with a horizontal resolution of approximately 0.25° and 64 grid levels in the vertical [*Popova et al*., 2010]. The model grid is tripolar with cell sizes ranging approximately 6.8–15.4 km horizontally, and increasing in thickness from 6 m at the surface to approximately 250 m at abyssal depths. OPA9 includes partial level thicknesses at the seafloor to improve bottom circulation, and a parameterized turbulent kinetic energy scheme for vertical mixing [*Gaspar et al*., 1990; *Madec*, 2008]. The model is forced at the surface using bulk formulae [*Large and Yeager*, 2004] fed using DFS4.1 climatology fields of atmospheric properties and downward heat and freshwater fluxes [*DRAKKAR Group*, 2007]. Sea ice is represented with LIM2 [*Timmermann et al*., 2005], based on a viscous-plastic ice rheology [*Hibler*, 1979] with three layer (two ice and one snow) thermodynamics [*Semtner*, 1976]. LIM2 is coupled to OPA9 every five ocean time steps through a nonlinear quadratic drag law of ocean-ice shear [*Timmermann et al*., 2005], and represents detailed heat and freshwater fluxes [*Fichefet and Morales Maqueda*, 1997]. OPA9 and LIM2 are combined as the NEMO physical model [*Madec*, 2008]. Marine biogeochemistry is represented using the Model of Ecosystem Dynamics, nutrient Utilization, Sequestration, and Acidification (MEDUSA) [*Yool et al*., 2011]. This includes the N, Si, and Fe cycles in a model of the pelagic ecosystem that resolves nanophytoplankton, diatoms, and micro and mesozooplankton, as well as size-resolved (small and large) detrital material, including biomineral ballasting [*Armstrong et al*., 2002]. MEDUSA has simplified seafloor remineralization of organic material and does not explicitly include sea ice algae.

A11. Model 11 and 12

NEMO (version 3.2)-European Regional Seas Ecosystem Model (ERSEM) is a global ocean model that includes the NEMO ocean circulation model, the ERSEM marine biogeochemistry model, and the Los Alamos sea ice model (CICE). NEMO (version 3.2) is a framework of ocean related engines, ocean dynamics, thermodynamics, sinks and sources, and transport [*Madec*, 2008]. The CICE model [*Hunke et al*., 2013] has several interacting components: a thermodynamic submodel, an ice dynamics submodel, a vertical and a horizontal transport submodel. ERSEM is a lower trophic level biogeochemical cycling, carbon-based process model that uses the functional-group approach [*Blackford et al*., 2004; *Butenschön et al*., 2016]. The C, N, P, Si, and Fe cycles are explicitly resolved with variable stoichiometry. The pelagic ERSEM model simulates four phytoplankton functional types, three zooplankton functional types, one bacterial functional type, and six detritus size classes. The benthic ERSEM is a simple parameterization of remineralization where

sedimented organic matter is recycled to the water column in inorganic form. The simulations were run as part of the iMarNet project: an intercomparison project of six UK biogeochemical models [*Kwiatkowski et al*., 2014]. The simulation of Model 12 (xhonc) uses an updated configuration of the ERSEM model, with a new iron model, updated food web, and reduced detrital sinking rate relative to the simulation of Model 11 (xhonp).

A12. Model 13

PELAgic biogeochemistry for Global Ocean Simulations (PELAGOS) [*Vichi et al*., 2007a,2007b] is a global ocean biogeochemical model that consists of the coupling between the NEMO (version 3.4) [*Madec*, 2008] and the Biogeochemical Flux Model (BFM, version 5.1, http://bfm-community.eu) [*Vichi et al*., 2015a]. The ocean model is computed on the ORCA2 grid that has a nominal resolution of 2° with a refinement of the latitudinal grid size toward the equator down to 0.5°, with 30 vertical levels and partial step parameterization to increase the resolution of bottom gradients. The physical parameterizations are as in *Vichi et al*. [2007b]. The sea ice model is LIM2 and the BFM model can directly be coupled with NEMO [*Vichi et al*., 2015b]. The BFM is based on a biomass continuum description of the lower trophic levels of the marine system [*Vichi et al*., 2015b]. The model implements a set of biomass-based differential equations that solve the fluxes of C, N, P, Si, and Fe among selected biological functional groups representing the major components of the lower trophic levels. The functional groups in the pelagic environment are represented by unicellular planktonic autotrophs (pico and nanophytoplankton and diatoms), zooplankton (nano, micro, and mesozooplankton) and heterotrophic oxic and anoxic bacterioplankton. The model also simulates the dynamics of NO_3, NH_4, PO_4, biogenic Si, Fe, O_2, and it has an explicit parameterization of the biochemical cycling of dissolved/particulate nonliving organic matter. Another peculiarity of the model is that the nutrient stoichiometry is not fixed and therefore the ratios of C to N, P, and Si can vary within the living functional groups as well as the chlorophyll:C ratio in phytoplankton. This simulation includes nutrient river inputs from *Cotrim da Cunha et al*. [2007] but no benthic remineralization.

A13. Model 14

A marine biogeochemical (mBGC) module [*Moore et al*., 2004; *Jin et al*., 2012, 2016] is incorporated in a global version of the POP-CICE model developed at Los Alamos National Laboratory. A subgrid-scale ocean mixing scheme [*Jin et al*., 2012] was added to reduce model errors in simulated mixed layer depth, salinity, and nutrients in the Arctic Ocean. The pelagic component of mBGC includes 26 state variables of multiple nutrients, phytoplankton, zooplankton, and detritus. The ice algal component of mBGC represents colonies in a 3 cm layer at the bottom of each sea ice thickness category, coupled to the pelagic model through nutrient and biotic fluxes. This submodel was based on biophysical ice core data collected in land-fast ice offshore from Barrow, Alaska [*Jin et al*., 2006] and coupled with a pelagic ecosystem model in vertically 1-D models [*Jin et al*., 2007] and global POP-CICE model settings [*Deal et al*., 2011; *Jin et al*., 2012, 2016].

A14. Model 15

This model is based on the ocean circulation model OPA, the dynamical component of NEMO (version 3.5). It solves the primitive equations on the Arakawa C-grid, with a second-order centered finite difference scheme [*Madec*, 2008]. It assumes the Boussinesq and hydrostatic approximations, the incompressibility hypothesis, and uses a free-surface formulation [*Roullet and Madec*, 2000]. The density is computed from potential temperature, salinity, and pressure using the *Jacket and Mcdougall* [1995] equation of state. Vertical mixing is computed from a turbulence closure scheme based on a prognostic vertical turbulent kinetic equation, which performs well in the tropics [*Blanke and Delecluse*, 1993]. Partial step formulation of bottom topography is used instead of a full step one [*Bernard et al*., 2006]. LIM2 is a two-level thermodynamic-dynamic sea ice model in which sensible heat storage and vertical heat conduction within snow and ice are determined by a three-layer model (one layer for snow and two layers for ice) [*Fichefet and Morales Maqueda*, 1997; *Bouillon et al*., 2009]. PISCES is the biogeochemical component of NEMO and it simulates the lower trophic levels of marine ecosystem and the biogeochemical cycles of carbon and of the main nutrients (P, N, Fe, and Si) [*Aumont and Bopp*, 2006]. There are 24 prognostic variables including two phytoplankton compartments (diatoms and nanophytoplankton), two zooplankton size-classes (micro and mesozooplankton) and a description of the carbonate chemistry. Formulations in PISCES are based on a mixed Monod-Quota formalism: on one hand, stoichiometry of C, N, and P is fixed and growth rate of phytoplankton is limited by the external availability in N, P, and Si. On the other hand, the Fe and Si quotas are variable

and growth rate of phytoplankton is limited by the internal availability in Fe. There are three nonliving compartments: semilabile-dissolved organic matter, small and big sinking particles.

A15. Model 16

This simulation is based on the global ocean biogeochemical model NEMO-PISCES. NEMO (version 3.4) was run with horizontal resolution of up to 1° by 1° (globally 362 longitude × 292 latitude grid points) and with 46 vertical levels. The model results are available for 2002–2011. For those years, external ocean-surface forcing was specified from the CORE2 (2002–2007) and NCEP/DOE AMIP-II Reanalysis (Reanalysis-2) data (2008–2011). The ocean circulation model is based on the hydrostatic, primitive equation model OPA [*Madec et al*., 1998], and is a finite difference OGCM with a free sea surface. OPA uses a global tripolar orthogonal curvilinear grid to avoid singularity points inside the computational domain and attain the ratio of anisotropy of nearly one [*Madec and Imbard*, 1996]. The variables are distributed on a 3-D Arakawa-C type grid. The sea ice model (LIM3) is a C-grid dynamic-thermodynamic model, including the representation of the subgrid-scale distributions of five-category sea ice thickness, enthalpy, salinity, and age distributions. The representation of sea ice is done through a velocity vector and a series of ice state variables described in *Vancoppenolle et al*. [2009a]. Brine entrapment and drainage as well as brine impact on ice thermodynamics and a snow ice formation scheme are explicitly included [*Vancoppenolle et al*., 2009a, 2009b]. The biogeochemical model used is PISCES (volume 2) [*Aumont et al*., 2015]. It has 24 prognostic variables and simulates the lower trophic levels of marine ecosystems and the biogeochemical cycles of O_2, C, and the main nutrients (P, N, Fe, and Si). The model distinguishes four plankton functional types based on size: two phytoplankton groups (nanophytoplankton and diatoms) and two zooplankton groups (micro and mesozooplankton). Prognostic phytoplankton variables are total biomass in C, Fe, Si (for diatoms) and chlorophyll and hence the Fe:C, Si:C, chlorophyll:C ratios are fully predicted by the model. For zooplankton, all O, C, N, and P ratios are kept constant and total biomass in C is the only prognostic variable. The bacterial pool is not modeled explicitly.

A16. Model 17

This simulation is based on the NEMO-PISCES model. The ocean component uses NEMO (version 3.2) [*Madec*, 2008] in the ORCA1 configuration with horizontal resolution from 1° in polar regions to 1/3° near the equator and 42 vertical levels. This configuration includes an improved Turbulent Kinetic Energy (TKE) closure scheme based on *Blanke and Delecluse* [1993]. The ocean radiative scheme is represented by three-wave-band scheme of *Lengaigne et al*. [2007], which accounts for influence of phytoplankton chlorophyll. The sea ice model is GELATO5 [*Salas Mélia*, 2002], which considers four ice thickness categories (0–0.3 m, 0.3–0.8 m, 0.8–3 m, and over 3 m). Each category has 10 vertical layers with enhanced resolution near the top of the slab. The sea ice dynamics is represented by the EVP rheology [*Hunke and Dukowicz*, 1997]. The biogeochemical model PISCES [*Aumont and Bopp*, 2006] simulates the biogeochemical cycles of O_2, C, and the main nutrients: NO_3, NH_4, PO_4, SiO_4, Fe, two sizes of phytoplankton, two sizes of zooplankton, semilabile-dissolved organic matter, and small and big sinking particles. Three different sources supply nutrients: atmospheric dust deposition [*Tegen and Fung*, 1995; *Jickells and Spokes*, 2001; *Moore et al*., 2004], rivers [*Ludwig et al*., 1996], and sediment mobilization [*Johnson et al*., 1999; *de Baar and de Jong*, 2001]. The model input of riverine-dissolved inorganic carbon (DIC) and DOC is taken from the annual estimates of the Global Erosion Model [*Ludwig et al*., 1996]. The DOC from rivers is assumed to be labile and is directly converted to DIC upon its delivery to the ocean. Inputs of dissolved Fe, NO_3, PO_4, and SiO_4 are computed from the sum of DIC and DOC river input using a constant set of ratios for C:N:P:Si:Fe as computed from *Meybeck* [1982] for C:N, from *Takahashi et al*. [1985] for N:P, from *de Baar and de Jong* [2001] for Fe:C, and from *Tréguer et al*. [1995] for Si:C. More complete descriptions of the whole ocean-sea ice and marine biogeochemistry component are provided by *Voldoire et al*. [2013] and *Séférian et al*. [2013], respectively.

A17. Model 18

This model is the MIT general circulation model (MITgcm) [*Marshall et al*., 1997] coupled with a sea ice model and with a pelagic biogeochemical model. The coupled model is configured on a "cubed-sphere" grid encompassing the Arctic domain with open boundaries at around 55°N in the Atlantic and Pacific sectors. Prescribed boundary conditions for potential temperature, salinity, flow, and sea surface elevation are provided from previous integrations of a global configuration of the same model [*Menemenlis et al*., 2005]. The sea ice component follows the viscous-plastic rheology formulation of *Hibler* [1979] with momentum

equations solved implicitly on a C-grid [*Arakawa and Lamb*, 1977] using a procedure based on *Zhang and Hibler* [1997]. The biogeochemical component, improved from previous applications in Arctic waters [*Le Fouest et al*., 2011, 2013], has 10 state variables [see *Le Fouest et al*., 2015, Figure 1]. Equations and biological parameters are listed in *Le Fouest et al*. [2005]. It includes NO_3, NH_4, 2 size-fractionated phytoplankton and zooplankton, DON, PON, bacteria, and riverine DON (RDON) that couples the marine and terrestrial cycling of N. The RDON flux is based on the approach developed by *Manizza et al*. [2009] using seasonally explicit regression relationships quantifying the covariation between freshwater discharge and measured RDON.

A18. Model 19

The NorESM is a fully coupled climate-biogeochemical model [*Tjiputra et al*., 2013]. It is the fully coupled version of Model 8 described above. In this configuration, the NorESM-OC is coupled with atmospheric general circulation (modified from Community Atmospheric Model, CAM4), land (Community Land Model, CLM4), and sea ice (CICE version 4). For the simulation performed in this study, we apply the standard Coupled Model Intercomparison Project phase 5 (CMIP5) protocol for the historical and future high CO_2 (RCP8.5) scenario.

A19. Model 20

The GISS-E2-R-CC is a coupled atmosphere-land-ocean-ice model with ocean biogeochemistry explicitly simulated [*Romanou et al*., 2013, 2014]. GISS-E2 is the GISS atmospheric model within the coupled climate model, which was submitted in the CMIP5 [*Schmidt et al*., 2014], and which employs a very accurate and nondiffusive quadratic upstream differencing scheme for the advection of tracers such as CO_2. The GISS-E2-R ocean model [*Russell et al*., 1995; *Hansen et al*., 2007] is a non-Boussinesq mass-conserving ocean model with a free surface and natural surface boundary fluxes of freshwater and heat. It uses the *Gent and Mcwilliams* [1990] scheme for isopycnal eddy fluxes and isopycnal thickness diffusion and the K-Profile Parameterization (KPP) [*Large et al*., 1994] for vertical mixing, with nonlocal mixing effects included. The GISS-E2 sea ice model includes improvements from earlier model versions mostly in the formulation of temperature profiles in the sea ice, which greatly improved Arctic sea ice distributions. Salinity is treated as thermodynamically passive in the sea ice, though it does impact basal ice formation/melt and heat diffusion through a new glacial-melt and melt-pond fraction [*Schmidt et al*., 2004]. The interactive ocean carbon cycle model consists of the NASA Ocean Biogeochemistry Model (NOBM) [*Gregg and Casey*, 2007], a gas exchange parameterization for the computation of the CO_2 flux between the ocean and the atmosphere and radiative coupling between the atmosphere and the ocean biology. NOBM utilizes ocean temperature and salinity, MLD, ocean circulation fields and the horizontal advection and vertical mixing schemes from the host ocean model as well as shortwave radiation (direct and diffuse) and surface wind speed obtained from the atmospheric model to produce horizontal and vertical distributions of several biogeochemical constituents. There are four phytoplankton groups (diatoms, chlorophytes, cyanobacteria, and coccolithophores), with distinct maximum growth rates, sinking rates, nutrient requirements, and optical properties, and four nutrients (NO_3, NH, SiO_4, and Fe), three detrital pools (for NO_3:C, SiO_4, and Fe) and a single herbivore component.

A20: Model 21

The HAMOCC [*Ilyina et al*., 2013] is the global ocean biogeochemical model, which is part of the Max Planck Institute's Earth System Model (MPI-ESM) [*Giorgetta et al*., 2013]. The model HAMOCC is coupled to the MPI-ESM's ocean general circulation model MPI-OM [*Junglcaus et al*., 2013] with an embedded sea ice submodel [*Notz et al*., 2013]. The ocean physics, biogeochemistry, and sea ice are thus resolved on the same spatial grid. In this study, the low-resolution configuration of MPI-ESM from the CMIP5 was used with a nominal resolution of 1.5°. The grid has a bipolar structure with one pole located over Greenland and another one over Antarctica. In the Arctic Ocean, the model resolution varies between 15 and 65 km depending on the proximity to the pole. The model has 40 levels with varying thicknesses increasing toward the bottom. HAMOCC represents the biogeochemical cycles of carbon and nutrients in the water column and in the upper sediments (represented by an interactive sediment module) [*Heinze et al*., 1999]. The biological part of HAMOCC is based on the NPZD approach (extended by dissolved organic matter) [*Six and Maier-Reimer*, 1996] so that the concentrations of phytoplankton and zooplankton (both represented by 1 tracer) are functions of, solar radiation, temperature, and colimiting nutrients (NO_3, PO_4, SiO_4, and Fe). The model separates detrital material into POC, $CaCO_3$, and opal, each exported from the euphotic layer with individually distinct

sinking speeds. Phytoplankton growth and air-sea exchange of CO_2 are modulated by the presence of sea ice; no air-sea exchange takes place through sea ice. In this model configuration biogeochemistry in sea ice has not been considered. Zero concentrations of nutrients, DIC, and total alkalinity in riverine fluxes are assumed. The model considers globally uniform weathering fluxes of $CaCO_3$ and SiO_4.

Acknowledgments
This project was funded by the National Aeronautics and Space Agency (NASA) Ocean Biology and Biogeochemistry (OBB) program (NNX13AE81G). We thank the anonymous reviewers for careful reading and constructive comments, and also thank Charlotte Laufkötter for thoughtful suggestions. The project "Green Mercator" provided funding to M. Gehlen and C. Perruche through the national program CNRS/LEFE/INSU. M. Jin's contribution was supported by the NSF Office of Polar Programs (PLR-1417925, and PLR-1416920). A. Samuelsen acknowledges the projects FP7 MyOcean2 (project number 283367) and PAVE (Polish-Norwegian Research Program) and a grant of CPU time from the Norwegian Supercomputing Project (NOTUR2). J. Tjiputra acknowledges the Research Council of Norway funded project ORGANIC (239965/RU). J. Zhang's contribution was supported by the NASA Cryosphere program (NNX15AG68G). R. Séférian and M. Chevallier thank Météo-France/DSI supercomputing center and the support of the team in charge of the CNRM-CM climate model. Upon publication, the in situ data will be available for academic purposes through the NASA SeaWiFS Bio-optical Archive and Storage System (http://seabass.gsfc.nasa.gov/), including NPP, NO_3, and Z_{eu}.

References

Aksnes, D. L., K. B. Ulvestad, B. M. Baliño, J. Berntsen, J. K. Egge, and E. Svendsen (1995), Ecological modelling in coastal waters: Towards predictive physical-chemical-biological simulation models, *Ophelia*, *41*(1), 5–36, doi:10.1080/00785236.1995.10422035.

Anderson, L. A. (1995), On the hydrogen and oxygen content of marine phytoplankton, *Deep Sea Res., Part I*, *42*(1), 1675–1680, doi:10.1016/0967-0637(95)00072-E.

Arakawa, A., and V. R. Lamb (1977), Computational design of the basic dynamical processes of the UCLS general circulation model, *Methods Comput. Phys.*, *17*, 173–265.

Ardyna, M., M. Babin, M. Gosselin, E. Devred, S. Bélanger, A. Matsuoka, and J. É. Tremblay (2013), Parameterization of vertical chlorophyll a in the Arctic Ocean: Impact of the subsurface chlorophyll maximum on regional, seasonal, and annual primary production estimates, *Biogeosciences*, *10*(6), 4383–4404, doi:10.5194/bg-10-4383-2013.

Ardyna, M., M. Babin, M. Gosselin, E. Devred, L. Rainville, and J. É. Tremblay (2014), Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms, *Geophys. Res. Lett.*, *41*, 6207–6212, doi:10.1002/2014GL061047.

Armstrong, R. A., C. Lee, J. I. Hedges, S. Honjo, and S. G. Wakeham (2002), A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals, *Deep. Sea Res., Part II*, *49*(1–3), 219–236, doi:10.1016/S0967-0645(01)00101-1.

Arrigo, K. R. et al. (2012), Massive phytoplankton blooms under Arctic Sea Ice, *Science*, *336*(6087), 1408, doi:10.1126/science.1215065.

Arrigo, K. R. et al. (2014), Phytoplankton blooms beneath the sea ice in the Chukchi Sea, *Deep Sea Res., Part II*, *105*, 1–16.

Aumont, O., and L. Bopp (2006), Globalizing results from ocean in situ iron fertilization studies, *Global Biogeochem. Cycles*, *20*, GB2017, doi:10.1029/2005GB002591.

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen (2015), PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies, *Geosci. Model Dev.*, *8*(8), 2465–2513, doi:10.5194/gmd-8-2465-2015.

Babin, M., S. Bélanger, I. Ellingsen, A. Forest, V. Le Fouest, T. Lacour, M. Ardyna, and D. Slagstad (2015), Estimation of primary production in the Arctic Ocean using ocean colour remote sensing and coupled physical-biological models: Strengths, limitations and how they compare, *Progr. Oceanogr.*, *139*, 197–220.

Balch, W., R. Evans, J. Brown, C. McLain, and W. Esaias (1992), The remote sensing of ocean primary productivity: Use of a new data compilation to test satellite algorithms, *J. Geophys. Res.*, *97*(C2), 2279–2293.

Bernard, B. et al. (2006), Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution, *Ocean Dyn.*, *56*(5-6), 543–567, doi:10.1007/s10236-006-0082-1.

Bertino, L., and K. A. Lisaeter (2008), The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans, *J. Oper. Oceanogr.*, *1*(2), 15–18, doi:10.1080/1755876X.2008.11020098.

Blackford, J. C., J. I. Allen, and F. J. Gilbert (2004), Ecosystem dynamics at six contrasting sites: A generic modelling study, *J. Mar. Syst.*, *52*(1–4), 191–215, doi:10.1016/j.jmarsys.2004.02.004.

Blanke, B., and P. Delecluse (1993), Variability of the tropical Atlantic ocean simulated by a general circulation model with two different mixed-layer physics, *J. Phys. Oceanogr.*, *23*(7), 1363–1388, doi:10.1175/1520-0485(1993)023 < 1363:VOTTAO>2.0.CO;2.

Bleck, R. (2002), An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, *Ocean Modell.*, *4*(1), 55–88, doi:10.1016/S1463-5003(01)00012-9.

Bluhm, B. A., K. N. Kosobokova, and E. C. Carmack (2015), A tale of two basins: An integrated physical and biological perspective of the deep Arctic Ocean, *Prog. Oceanogr.*, *139*, 89–121, doi:10.1016/j.pocean.2015.07.011.

Bouillon, S., M. A. Morales Maqueda, V. Legat, and T. Fichefet (2009), An elastic-viscous-plastic sea ice model formulated on Arakawa B and C grids, *Ocean Modell.*, *27*(3–4), 174–184, doi:10.1016/j.ocemod.2009.01.004.

Brugel, S., C. Nozais, M. Poulin, J.-E. Tremblay, L. A. Miller, K. G. Simpson, Y. Gratton, and S. Demers (2009), Phytoplankton biomass and production in the southeastern Beaufort Sea in autumn 2002 and 2003, *Mar. Ecol. Prog. Ser.*, *377*, 63–77, doi:10.3354/meps07808.

Buitenhuis, E. T., and R. J. Geider (2010), A model of phytoplankton acclimation to iron-light colimitation, *Limnol. Oceanogr.*, *55*(2), 714–724, doi:10.4319/lo.2009.55.2.0714.

Buitenhuis, E. T., T. Hashioka, and C. Le Quéré (2013), Combined constraints on global ocean primary production using observations and models, *Global Biogeochem. Cycles*, *27*, 847–858, doi:10.1002/gbc.20074.

Butenschön, M. et al. (2016), ERSEM 15.06: A generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, *Geosci. Model Dev.*, *9*, 1293–1339, doi:10.5194/gmd-9-1293-2016.

Campbell, J. et al. (2002), Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance, *Global Biogeochem. Cycles*, *16*(3), 1035, doi:10.1029/2001GB001444.

Carmack, E., and P. Wassmann (2006), Food webs and physical-biological coupling on pan-Arctic shelves: Unifying concepts and comprehensive perspectives, *Prog. Oceanogr.*, *71*(2–4), 446–477, doi:10.1016/j.pocean.2006.10.004.

Carr, M. E., et al. (2006), A comparison of global estimates of marine primary production from ocean color, *Deep. Sea Res., Part II*, *53*(5–7), 741–770, doi:10.1016/j.dsr2.2006.01.028.

Cavalieri, D. J., P. Gloersen, and W. J. Campbell (1984), Determination of sea ice parameters with the NIMBUS 7 SMMR, *J. Geophys. Res.*, *89*(D4), 5355–5369, doi:10.1029/JD089iD04p05355.

Charalampopoulou, A., A. J. Poulton, T. Tyrrell, and M. I. Lucas (2011), Irradiance and pH affect coccolithophore community composition on a transect between the North Sea and the Arctic Ocean, *Mar. Ecol. Prog. Ser.*, *431*, 25–43.

Codispoti, L. A., V. Kelly, A. Thessen, P. Matrai, S. Suttles, V. Hill, M. Steele, and B. Light (2013), Synthesis of primary production in the Arctic Ocean: III. Nitrate and phosphate based estimates of net community production, *Prog. Oceanogr.*, *110*, 126–150, doi:10.1016/j.pocean.2012.11.006.

Coello-Camba, A., S. Agustí, J. Holding, J. M. Arrieta, and C. M. Duarte (2014), Interactive effect of temperature and CO_2 increase in Arctic phytoplankton, *Front. Mar. Sci.*, *1*, 1–10.

Coello-Camba, A., S. Agustí, D. Vaqué, D., J. Holding, J. M. Arrieta, P. Wassmann, and C. M. Duarte (2015), Experimental assessment of temperature thresholds for arctic phytoplankton communities, *Estuaries Coasts*, *38*(3), 873–885.
Comiso, J. C. (1986), Characteristics of Arctic winter sea ice from satellite multispectral microwave observations, *J. Geophys. Res.*, *91*(C1), 975–994, doi:10.1029/JC091iC01p00975.
Cota, G. F., L. R. Pomeroy, W. G. Harrison, E. P. Jones, F. Peters, W. M. Sheldon, and T. R. Weingartner (1996), Nutrients, primary production and microbial heterotrophy in the southeastern Chukchi Sea: Arctic summer nutrient depletion and heterotrophy, *Mar. Ecol. Prog. Ser.*, *135*, 247–258, doi:10.3354/meps135247.
Cotrim da Cunha, L., E. T. Buitenhuis, C. Le Quéré, X. Giraud, and W. Ludwig (2007), Potential impact of changes in river nutrient supply on global ocean biogeochemistry, *Global Biogeochem. Cycles*, *21*, GB4007, doi:10.1029/2006GB002718.
Coupel, P., D. Ruiz-Pino, M. A. Sicre, J. F. Chen, S. H. Lee, N. Schiffrine, H. L. Li, and J. C. Gascard (2015), The impact of freshening on phytoplankton production in the Pacific Arctic Ocean, *Prog. Oceanogr.*, *131*, 113–125, doi:10.1016/j.pocean.2014.12.003.
Deal, C. J., M. Jin, S. Elliott, E. Hunke, M. Maltrud, and N. Jeffery (2011), Large-scale modeling of primary production and ice algal biomass within arctic sea ice in 1992, *J. Geophys. Res.*, *116*, C07004, doi:10.1029/2010JC006409.
de Baar, H. J. W., and J. T. M. de Jong (2001), Distribution, Sources and Sinks of Iron in Seawater, in *The Biogeochemistry of Iron in Seawater*, edited by D. R. Turner and K. A. Hunter, pp. 123–253, John Wiley, Hoboken, N. J.
Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.
DRAKKAR Group (2007), Eddy-permitting ocean circulation hindcasts of the past decades, *CLIVAR Exch.*, *12*, 8–10.
Drange, H., and K. Simonsen (1996), Formulation of air-sea fluxes in the ESOP2 version of MICOM, *Tech. Rep. 125*, Nansen Environ. and Remote Sens. Cent., Bergen, Norway.
Dunne, J. P., B. Hales, and J. R. Toggweiler (2012a), Global calcite cycling constrained by sediment preservation controls, *Global Biogeochem. Cycles*, *26*, GB3023, doi:10.1029/2010GB003935.
Dunne, J. P., et al. (2012b), GFDL's ESM2 global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics, *J. Clim.*, *25*(19), 6646–6665, doi:10.1175/JCLI-D-11-00560.1.
Dunne, J. P., et al. (2013), GFDL's ESM2 global coupled climate-carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics, *J. Clim.*, *26*(7), 2247–2267, doi:10.1175/JCLI-D-12-00150.1.
Dutkiewicz, S., A. E. Hickman, O. Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows (2015), Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, *Biogeosciences*, *12*(14), 4447–4481, doi:10.5194/bg-12-4447-2015.
Eppley, R. W. (1972), Temperature and phytoplankton growth in the sea, *Fish. Bull.*, *70*, 1063–1085.
Fernández-Méndez, M., C. Katlein, B. Rabe, M. Nicolaus, I. Peeken, K. Bakker, H. Flores, and A. Boetius (2015), Photosynthetic production in the central Arctic Ocean during the record sea-ice minimum in 2012, *Biogeosciences*, *12*(11), 3525–3549, doi:10.5194/bg-12-3525-2015.
Fichefet, T., and M. A. Morales Maqueda (1997), Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, *J. Geophys. Res.*, *102*(C6), 12,609–12,646, doi:10.1029/97JC00480.
Forsythe, W. C., E. J. Rykiel, R. S. Stahl, H. I. Wu, and R. M. Schoolfield (1995), A model comparison for day length as a function of latitude and day of year, *Ecol. Modell.*, *80*(1), 87–95.
Friedrichs, M. A. M., R. R. Hood, and J. D. Wiggert (2006), Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data, *Deep. Sea Res., Part II*, *53*(5–7), 576–600, doi:10.1016/j.dsr2.2006.01.026.
Friedrichs, M. A. M., et al. (2009), Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean, *J. Mar. Syst.*, *76*(1–2), 113–133, doi:10.1016/j.jmarsys.2008.05.010.
Gaspar, P., Y. Grégoris, and J.-M. Lefevre (1990), A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site, *J. Geophys. Res.*, *95*(C9), 16,179–16,193, doi:10.1029/JC095iC09p16179.
Geider, R., H. MacIntyre, and T. Kana (1997), Dynamic model of phytoplankton growth and acclimation: Responses of the balanced growth rate and the chlorophyll a: Carbon ratio to light, nutrient-limitation and temperature, *Mar. Ecol. Prog. Ser.*, *148*, 187–200, doi:10.3354/meps148187.
Gent, P. R., and J. C. Mcwilliams (1990), Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, *20*(1), 150–155, doi:10.1175/1520-0485(1990)020 < 0150:IMIOCM>2.0.CO;2.
Gill, A. E. (1982), *Atmosphere-Ocean Dynamics*, Academic, San Diego, Calif.
Giorgetta, M. A., et al. (2013), Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5, *J. Adv. Model. Earth Syst.*, *5*, 572–597, doi:10.1002/jame.20038.
Granskog, M. A., C. A. Stedmon, P. A. Dodd, R. M. W. Amon, A. K. Pavlov, L. De Steur, and E. Hansen (2012), Characteristics of colored dissolved organic matter (CDOM) in the Arctic outflow in the Fram Strait: Assessing the changes and fate of terrigenous CDOM in the Arctic Ocean, *J. Geophys. Res.*, *117*, C12021, doi:10.1029/2012JC008075.
Gregg, W. W., and N. W. Casey (2007), Modeling coccolithophores in the global oceans, *Deep Sea Res., Part II*, *54*(5), 447–477, doi:10.1016/j.dsr2.2006.12.007.
Griffies, S. M. (2012), Elements of the Modular Ocean Model (MOM): 2012 Release. GFDL Ocean Group, *Tech. Rep. 7*, NOAA/Geophys. Fluid Dyn. Lab., Princeton, N. J.
Griffies, S. M., et al. (2009), Coordinated Ocean-ice Reference Experiments (COREs), Ocean *Modell.*, *26*(1), 1–46, doi:10.1016/j.ocemod.2008.08.007.
Hansen, J., et al. (2007), Climate simulations for 1880–2003 with GISS modelE, *Clim. Dyn.*, *29*(7-8), 661–696, doi:10.1007/s00382-007-0255-8.
Harrison, W. G., and G. F. Cota (1991), Primary production in polar waters: Relation to nutrient availability, *Polar Res.*, *10*(1), 87–104, doi:10.1111/j.1751-8369.1991.tb00637.x.
Harrison, W. G., and T. Platt (1986), Photosynthesis-irradiance relationships in polar and temperate phytoplankton populations, *Polar Biol.*, *5*(3), 153–164, doi:10.1007/BF00441695.
Harrison, W. G., T. Platt, and B. Irwin (1982), Primary production and nutrient assimilation by natural phytoplankton populations of the eastern Canadian Arctic, *Can. J. Fish. Aquat. Sci.*, *39*(2), 335–345.
Heinze, C., E. Maier-Reimer, A. M. E. Winguth, and D. Archer (1999), A global oceanic sediment model for long-term climate studies, *Global Biogeochem. Cycles*, *13*(1), 221–250, doi:10.1029/98GB02812.
Hibler, W. D. (1979), A dynamic thermodynamic sea ice model, *J. Phys. Oceanogr.*, *9*(4), 815–846, doi:10.1175/1520-0485(1979)009 < 0815:ADTSIM>2.0.CO;2.
Hibler, W. D. (1980), Modeling a variable thickness sea ice cover, *Mon. Weather Rev.*, *108*(12), 1943–1973, doi:10.1175/1520-0493(1980)108 < 1943:MAVTSI>2.0.CO;2.

Hill, V. J., P. A. Matrai, E. Olson, S. Suttles, M. Steele, L. A. Codispoti, and R. C. Zimmerman (2013), Synthesis of integrated primary production in the Arctic Ocean: II. In situ and remotely sensed estimates, *Prog. Oceanogr.*, *110*, 107–125, doi:10.1016/j.pocean.2012.11.005.

Hu, S., and W. O. Smith (1998), The effects of irradiance on nitrate uptake and dissolved organic nitrogen release by phytoplankton in the Ross Sea, *Cont. Shelf Res.*, *18*(9), 971–990, doi:10.1016/S0278-4343(98)00021-1.

Hunke, E. C., and J. K. Dukowicz (1997), An elastic–viscous–plastic model for sea ice dynamics, *J. Phys. Oceanogr.*, *27*, 1849–1867, doi:10.1175/1520-0485(1997)027 < 1849:AEVPMF>2.0.CO;2.

Hunke, E. C., W. H. Lipscomb, A. K. Turner, N. Jeffery, and S. Elliott (2013), CICE: The Los Alamos Sea Ice Model, Documentation and Software, version 5.0, *Tech. Rep. LA-CC-06-012*, Los Alamos Natl. Lab., N. M.

Huot, Y., M. Babin, and F. Bruyant (2013), Photosynthetic parameters in the Beaufort Sea in relation to the phytoplankton community structure, *Biogeosciences*, *10*(5), 3445–3454, doi:10.5194/bg-10-3445-2013.

Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li, and I. Núñez-Riboni (2013), Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, *J. Adv. Model. Earth Syst.*, *5*, 287–315, doi:10.1029/2012MS000178.

Interagency Arctic Research Policy Committee (2013), Arctic Research Plan: FY2013–2017, Washington, D. C. [Available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/2013_arctic_research_plan.pdf.]

International Ocean Colour Coordinating Group (IOCCG) (2015), Ocean Colour Remote Sensing in Polar Seas, edited by M. Babin et al., IOCCG Rep. Ser. 16, Dartmouth, Canada.

Jackett, D. R., and T. J. Mcdougall (1995), Minimal Adjustment of Hydrographic Profiles to Achieve Static Stability, *J. Atmos. Ocean. Technol.*, *12*(2), 381–389, doi:10.1175/1520-0426(1995)012 < 0381:MAOHPT>2.0.CO;2.

Ji, R., M. Jin, and Ø. Varpe (2013), Sea ice phenology and timing of primary production pulses in the Arctic Ocean, *Global Chang. Biol.*, *19*(3), 734–741, doi:10.1111/gcb.12074.

Jickells, T., and L. Spokes (2001), Atmospheric iron inputs to the oceans, in *The Biogeochemistry of Iron in Seawater*, edited by D. R. Turner and K. A. Hunter, pp. 85–121, John Wiley, Hoboken, N. J.

Jin, M., C. J. Deal, J. Wang, K.-H. Shin, N. Tanaka, T. E. Whitledge, S. H. Lee, and R. R. Gradinger (2006), Controls of the landfast ice–ocean ecosystem offshore Barrow, Alaska, *Ann. Glaciol.*, *44*(1), 63–72, doi:10.3189/172756406781811709.

Jin, M., C. Deal, J. Wang, V. Alexander, R. Gradinger, S. Saitoh, T. Iida, Z. Wan, and P. Stabeno (2007), Ice-associated phytoplankton blooms in the southeastern Bering Sea, *Geophys. Res. Lett.*, *34*, L06612, doi:10.1029/2006GL028849.

Jin, M., C. Deal, S. H. Lee, S. Elliott, E. Hunke, M. Maltrud, and N. Jeffery (2012), Investigation of Arctic sea ice and ocean primary production for the period 1992-2007 using a 3-D global ice-ocean ecosystem model, *Deep. Sea Res., Part II*, *81–84*, 28–35, doi:10.1016/j.dsr2.2011.06.003.

Jin, M., E. E. Popova, J. Zhang, R. Ji, D. Pendleton, Ø. Varpe, A. Yool, and Y. J. Lee (2016), Ecosystem model intercomparison of under-ice and total primary production in the Arctic Ocean, *J. Geophys. Res. Oceans*, *121*, 934–948, doi:10.1002/2015JC011183.

Johnson, K. S., F. P. Chavez, and G. E. Friederich (1999), *Continental-shelf sediment as a primary source of iron for coastal phytoplankton, Nature*, *398*(6729), 697–700, doi:10.1038/19511.

Jolliff, J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber, and R. A. Arnone (2009), Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, *J. Mar. Syst.*, *76*, 64–82.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewiczi, D. Notz, and J. S. vonStorch (2013), Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-Earth system model, *J. Adv. Model. Earth Syst.*, *5*, 422–446, doi:10.1002/jame.20023.

Kim, G. E., M.-A. Pradal, and A. Gnanadesikan (2015), Quantifying the biological impact of surface ocean light attenuation by colored detrital matter in an ESM using a new optical parameterization, *Biogeosciences*, *12*(16), 5119–5132, doi:10.5194/bg-12-5119-2015.

Kishi, M. J., et al. (2007), NEMURO-a lower trophic level model for the North Pacific marine ecosystem, *Ecol. Modell.*, *202*(1–2), 12–25, doi:10.1016/j.ecolmodel.2006.08.021.

Klausmeier, C. A., E. Litchman, T. Daufresne, and S. A. Levin (2004), Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton, *Nature*, *429*(6988), 171–174, doi:10.1038/nature02454.

Kosobokova, K. N., and R. R. Hopcroft (2010), Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin, *Deep Sea Res., Part II*, *57*(1), 96–110, doi:10.1016/j.dsr2.2009.08.009.

Kwiatkowski, L., et al. (2014), iMarNet: An ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, *Biogeosciences*, *11*(24), 7291–7304, doi:10.5194/bg-11-7291-2014.

Large, W. G., and S. G. Yeager (2004), Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies, NCAR Tech. Note NCAR/TN-460+STR, Natl. Cent. for Atmos. Res., Boulder, Colo.

Large, W. G., and S. G. Yeager, (2009), The global climatology of an interannually varying air-sea flux data set, *Clim. Dyn.*, *33*, 341–364, doi:10.1007/s00382-008-0441-3.

Large, W. G., J. C. McWilliams, and S. C. Doney (1994), Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, *Rev. Geophys.*, *32*(4), 363–403, doi:10.1029/94RG01872.

Laroche, D., A. F. Vézina, M. Levasseur, M. Gosselin, J. Stefels, M. D. Keller, P. A. Matrai, and R. L. J. Kwint (1999), DMSP synthesis and exudation in phytoplankton: A modeling approach, *Mar. Ecol. Prog. Ser.*, *180*, 37–49, doi:10.3354/meps180037.

Laufkötter, C., et al. (2015), Drivers and uncertainties of future global marine primary production in marine ecosystem models, *Biogeosciences*, *12*, 6955–6984, doi:10.5194/bg-12-6955-2015.

Le Fouest, V., B. Zakardjian, F. J. Saucier, and M. Starr (2005), Seasonal versus synoptic variability in planktonic production in a high-latitude marginal sea: The Gulf of St. Lawrence (Canada), *J. Geophys. Res.*, *110*, C09012, doi:10.1029/2004JC002423.

Le Fouest, V., C. Postlethwaite, M. A. Morales Maqueda, S. Bélanger, and M. Babin (2011), On the role of tides and strong wind events in promoting summer primary production in the Barents Sea, *Cont. Shelf Res.*, *31*(17), 1869–1879, doi:10.1016/j.csr.2011.08.013.

Le Fouest, V., B. Zakardjian, H. Xie, P. Raimbault, F. Joux, and M. Babin (2013), Modeling plankton ecosystem functioning and nitrogen fluxes in the oligotrophic waters of the Beaufort Sea, Arctic Ocean: A focus on light-driven processes, *Biogeosciences*, *10*(7), 4785–4800, doi:10.5194/bg-10-4785-2013.

Le Fouest, V., M. Manizza, B. Tremblay, and M. Babin (2015), Modelling the impact of riverine DON removal by marine bacterioplankton on primary production in the Arctic Ocean, *Biogeosciences*, *12*(11), 3385–3402, doi:10.5194/bg-12-3385-2015.

Lee, S. H., and T. E. Whitledge (2005), Primary production in the deep Canada Basin during summer 2002, *Polar Biol.*, *28*, 190–197, doi:10.1007/s00300-004-0676-3.

Lee, S. H., D. Stockwell, and T. E. Whitledge (2010), Uptake rates of dissolved inorganic carbon and nitrogen by under-ice phytoplankton in the Canada Basin in summer 2005, *Polar Biol.*, *33*, 1027–1036, doi:10.1007/s00300-010-0781-4.

Lee, Y. J., et al. (2015), An assessment of phytoplankton primary productivity in the Arctic Ocean from satellite ocean color/in situ chlorophyll-a basedmodels, *J. Geophys. Res. Oceans*, *120*, 6508–6541, doi:10.1002/2015JC011018.
Lengaigne, M., C. Menkes, O. Aumont, T. Gorgues, L. Bopp, J.-M. André, and G. Madec (2007), Influence of the oceanic biology on the tropical Pacific climate in a coupled general circulation model, *Clim. Dyn.*, *28*(5), 503–516, doi:10.1007/s00382-006-0200-2.
Le Quéré, C., et al. (2016), Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles, *Biogeosci.*, *13*, 4111–4133, doi:10.5194/bg-13-4111-2016.
Leu, E., C. J. Mundy, P. Assmy, K. Campbell, T. M. Gabrielsen, M. Gosselin, T. Juul-Pedersen, and R. Gradinger (2015), Arctic spring awakening: Steering principles behind the phenology of vernal ice algal blooms, *Prog. Oceanogr.*, *139*, 151–170, doi:10.1016/j.pocean.2015.07.012.
Lévy, M., P. Klein, and A.-M. Treguier (2001), Impact of sub-mesoscale physics on production and subduction of phytoplankton in an oligotrophic regime, *J. Mar. Res.*, *59*(4), 535–565, doi:10.1357/002224001762842181.
Li, W. K. W., F. McLaughlin, C. Lovejoy, and E. C. Carmack (2009), Smallest algae thrive as the Arctic Ocean freshens., *Science*, *326*(5952), 539, doi:10.1126/science.1179798.
Lindsay, R. W., and J. Zhang (2006), Assimilation of ice concentration in an ice–ocean model, *J. Atmos. Ocean. Technol.*, *23*(5), 742–749, doi:10.1175/JTECH1871.1.
Lomas, M. W., and P. M. Glibert (1999), Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms, *Limnol. Oceanogr.*, *44*(3), 556–572, doi:10.4319/lo.1999.44.3.0556.
Lovejoy, C., W. F. Vincent, S. Bonilla, S. Roy, M. J. Martineau, R. Terrado, M. Potvin, R. Massana, and C. Pedrós-Alió (2007), Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas, *J. Phycol.*, *43*(1), 78–89, doi:10.1111/j.1529-8817.2006.00310.x.
Ludwig, W., J.-L. Probst, and S. Kempe (1996), Predicting the oceanic input of organic carbon by continental erosion, *Global Biogeochem. Cycles*, *10*(1), 23–41, doi:10.1029/95GB02925.
Madec, G. (2008), NEMO ocean engine, *Note du Pôle de modélisation*, 27, Inst. Pierre-Simon Laplace, France.
Madec, G., and M. Imbard (1996), A global ocean mesh to overcome the North Pole singularity, *Clim. Dyn.*, *12*(6), 381–388, doi:10.1007/BF00211684.
Madec, G., P. Delecluse, M. Imbard, and C. Lévy (1998), OPA 8.1 Ocean General Circulation Model reference manual, *Note du Pôle de modélisation*, 11, 91 pp., Inst. Pierre-Simon Laplace, France.
Manda, A., N. Hirose, and T. Yanagi (2005), Feasible Method for the Assimilation of Satellite-Derived SST with an Ocean Circulation Model, *J. Atmos. Ocean. Technol.*, *22*(6), 746–756, doi:10.1175/JTECH1744.1.
Manizza, M., C. Le Quéré, A. J. Watson, and E. T. Buitenhuis (2005), Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model, *Geophys. Res. Lett.*, *32*, L05603, doi:10.1029/2004GL020778.
Manizza, M., M. J. Follows, S. Dutkiewicz, J. W. Mcclelland, D. Menemenlis, C. N. Hill, and B. J. Peterson (2009), Modeling transport and fate of riverine dissolved organic carbon in the Arctic Ocean, *Global Biogeochem. Cycles*, *23*, GB4006, doi:10.1029/2008GB003396.
Manizza, M., M. J. Follows, S. Dutkiewicz, D. Menemenlis, C. N. Hill, and R. M. Key (2013), Changes in the Arctic Ocean CO_2 sink (1996–2007): A regional model analysis, *Global Biogeochem. Cycles*, *27*, 1108–1118, doi:10.1002/2012GB004491.
Marshall, J., C. Hill, L. Perelman, and A. Adcroft (1997), Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, *J. Geophys. Res.*, *102*(C3), 5733–5752, doi:10.1029/96JC02776.
Matrai, P. A., E. Olson, S. Suttles, V. Hill, L. A. Codispoti, B. Light, and M. Steele (2013), Synthesis of primary production in the Arctic Ocean: I. Surface waters, 1954-2007, *Prog. Oceanogr.*, *110*, 93–106, doi:10.1016/j.pocean.2012.11.004.
McCuen, R. H., Z. Knight, and A. G. Cutter (2006), Evaluation of the Nash–Sutcliffe efficiency index, *J. Hydrol. Eng.*, *11*(6), 597–602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597).
McGillicuddy, D. J., A. R. Robinson, D. A. Siegel, H. W. Jannasch, R. Johnson, T. D. Dickey, J. McNeil, A. F. Michaels, and A. H. Knap (1998), Influence of mesoscale eddies on new production in the Sargasso Sea, *Nature*, *394*(6690), 263–266.
McLaughlin, F. A., and E. C. Carmack (2010), Deepening of the nutricline and chlorophyll maximum in the Canada Basin interior, 2003-2009, *Geophys. Res. Lett.*, *37*, L24602, doi:10.1029/2010GL045459.
Meier, W., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve (2013, updated 2015), *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration*, version 2, January 1979 to December 2011, Natl. Snow and Ice Data Cent., Boulder, Colo., doi:10.7265/N55M63M1.
Menemenlis, D., et al. (2005), NASA supercomputer improves prospects for ocean climate research, *Eos Trans. AGU*, *86*(9), 89–96, doi:10.1029/2005EO090002.
Meybeck, M. (1982), Carbon, nitrogen, and phosphorus transport by world rivers, *Am. J. Sci.*, *282*(4), 401–450, doi:10.2475/ajs.282.4.401.
Michel, C., R. G. Ingram, and L. R. Harris (2006), Variability in oceanographic and ecological processes in the Canadian Arctic Archipelago, *Prog. Oceanogr.*, *71*(2–4), 379–401, doi:10.1016/j.pocean.2006.09.006.
Michel, C., J. Hamilton, E. Hansen, D. Barber, M. Reigstad, J. Iacozza, L. Seuthe, and A. Niemi (2015), Arctic Ocean outflow shelves in the changing Arctic: A review and perspectives, *Prog. Oceanogr.*, *139*, 66–88, doi:10.1016/j.pocean.2015.08.007.
Moore, J. K., S. C. Doney, and K. Lindsay (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, *Global Biogeochem. Cycles*, *18*, GB4028, doi:10.1029/2004GB002220.
Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Binger, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE*, *50*(3), 885–900, doi:10.13031/2013.23153.
Mundy, C. J., et al. (2009), Contribution of under-ice primary production to an ice-edge upwelling phytoplankton bloom in the Canadian Beaufort Sea, *Geophys. Res. Lett.*, *36*, L17601, doi:10.1029/2009GL038837.
Nash, J., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I: A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.
Notz, D., F. A. Haumann, H. Haak, J. H. Jungclaus, and J. Marotzke (2013), Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model, *J. Adv. Model. Earth Syst.*, *5*, 173–194, doi:10.1002/jame.20016.
Oki, T., and Y. C. Sud (1998), Design of total runoff integrating pathways (TRIP): A global river channel network, *Earth Interact.*, *2*(1), 1–37, doi:10.1175/1087-3562(1998)002 < 0001:DOTRIP>2.3.CO;2.
Oschlies, A., and V. Garcon (1998), Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean, *Nature*, *394*(6690), 266–269.
Pacanowski, R. C., and S. G. H. Philander (1981), Parameterization of vertical mixing in numerical models of tropical oceans, *J. Phys. Oceanogr.*, *11*(11), 1443–1451, doi:10.1175/1520-0485(1981)011<1443:POVMIN>2.0.CO;2.
Pätsch, J., W. Kühn, A. Moll, and H. Lenhart (2009), ECOHAM4 User Guide: Ecosystem Model, Hamburg, version 4, *Tech. Rep. 01−2009*, Inst. für Meereskunde, Hamburg, Germany.
Peng, G., W. Meier, D. Scott, and M. Savoie (2013), A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring, *Earth Syst. Sci. Data*, *5*, 311–318.

Peralta-Ferriz, C., and R. A. Woodgate (2015), Seasonal and interannual variability of pan-Arctic surface mixed layer properties from 1979 to 2012 from hydrographic data, and the dominance of stratification for multiyear mixed layer depth shoaling, *Prog. Oceanogr.*, *134*, 19–53, doi:10.1016/j.pocean.2014.12.005.

Popova, E. E., A. Yool, A. C. Coward, Y. K. Aksenov, S. G. Alderson, B. A. De Cuevas, and T. R. Anderson (2010), Control of primary production in the Arctic by nutrients and light: Insights from a high resolution ocean general circulation model, *Biogeosciences*, *7*(11), 3569–3591, doi:10.5194/bg-7-3569-2010.

Popova, E. E., A. Yool, A. C. Coward, F. Dupont, C. Deal, S. Elliott, E. Hunke, M. Jin, M. Steele, and J. Zhang (2012), What controls primary production in the Arctic Ocean? Results from an intercomparison of five general circulation models with biogeochemistry, *J. Geophys. Res.*, *117*, C00D12, doi:10.1029/2011JC007112.

Popova, E. E., A. Yool, Y. Aksenov, and A. C. Coward (2013), Role of advection in Arctic Ocean lower trophic dynamics: A modeling perspective, *J. Geophys. Res. Ocean.*, *118*, 1571–1586, doi:10.1002/jgrc.20126.

Ralston, D. K., W. R. Geyer, and J. A. Lerczak (2010), Structure, variability, and salt flux in a strongly forced salt wedge estuary, *J. Geophys. Res.*, *115*, C06005, doi:10.1029/2009JC005806.

Reigstad, M., P. Wassmann, C. Wexels Riser, S. Øygarden, and F. Rey (2002), Variations in hydrography, nutrients and chlorophyll a in the marginal ice-zone and the central Barents Sea, *J. Mar. Syst.*, *38*(1–2), 9–29, doi:10.1016/S0924-7963(02)00167-7.

Reigstad, M., J. Carroll, D. Slagstad, I. Ellingsen, and P. Wassmann (2011), Intra-regional comparison of productivity, carbon flux and ecosystem composition within the northern Barents Sea, *Prog. Oceanogr.*, *90*(1–4), 33–46, doi:10.1016/j.pocean.2011.02.005.

Richardson, K. (1991), Comparison of 14C primary production determinations made by different laboratories, *Mar. Ecol. Prog. Ser.*, *72*(1-2), 189–201, doi:10.3354/meps072189.[10.3354/meps072189]

Romanou, A., et al. (2013), Natural air-sea flux of CO_2 in simulations of the NASA-GISS climate model: Sensitivity to the physical ocean model formulation, *Ocean Modell.*, *66*, 26–44, doi:10.1016/j.ocemod.2013.01.008.

Romanou, A., J. Romanski, and W. W. Gregg (2014), Natural ocean carbon cycle sensitivity to parameterizations of the recycling in a climate model, *Biogeosciences*, *11*(4), 1137–1154, doi:10.5194/bg-11-1137-2014.

Roullet, G., and G. Madec (2000), Salt conservation, free surface, and varying levels: A new formulation for ocean general circulation models, *J. Geophys. Res.*, *105*(C10), 23,927–23,942, doi:10.1029/2000JC900089.

Russell, G. L., J. R. Miller, and D. Rind (1995), A coupled atmosphere-ocean model for transient climate change studies, *Atmos. Ocean*, *33*(4), 683–730, doi:10.1080/07055900.1995.9649550.

Saba, V. S., et al. (2010), Challenges of modeling depth-integrated marine primary productivity over multiple decades: A case study at BATS and HOT, *Global Biogeochem. Cycles*, *24*, GB3020, doi:10.1029/2009GB003655.

Saba, V. S., et al. (2011), An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe, *Biogeosciences*, *8*, 489–503, doi:10.5194/bg-8-489-2011.

Sakshaug, E. (2004), Primary and secondary production in the Arctic seas, in *The Organic Carbon Cycle in the Arctic Ocean*, edited by R. Stein and R. MacDonald, pp. 57–81, Springer, Berlin.

Salas Mélia, D. (2002), A global coupled sea ice–ocean model, *Ocean Modell.*, *4*(2), 137–172, doi:10.1016/S1463-5003(01)00015-4.

Sallon, A., C. Michel, and M. Gosselin (2011), Summertime primary production and carbon export in the southeastern Beaufort Sea during the low ice year of 2008, *Polar Biol.*, *34*, 1989–2005, doi:10.1007/s00300-011-1055-5.

Samuelsen, A., C. Hansen, and H. Wehde (2015), Tuning and assessment of the HYCOM-NORWECOM V2.1 biogeochemical modeling system for the North Atlantic and Arctic oceans, *Geosci. Model Dev.*, *8*(7), 2187–2202, doi:10.5194/gmd-8-2187-2015.

Schmidt, G. A., C. M. Bitz, U. Mikolajewicz, and L. B. Tremblay (2004), Ice-ocean boundary conditions for coupled models, *Ocean Modell.*, *7*(1–2), 59–74, doi:10.1016/S1463-5003(03)00030-1.

Schmidt, G. A., et al. (2014), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, *6*, 141–184, doi:10.1002/2013MS000265.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok (2011), Uncertainty in modeled Arctic sea ice volume, *J. Geophys. Res.*, *116*, C00D06, doi:10.1029/2011JC007084.

Schwinger, J., N. Goris, J. Tjiputra, I. Kriest, M. Bentsen, I. Bethke, M. Ilicak, K. M. Assmann, and C. Heinze (2016), Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1), *Geosci. Model Dev.*, *9*, 2589–2622, doi:10.5194/gmd-9-2589-2016.

Séférian, R., L. Bopp, M. Gehlen, J. C. Orr, C. Ethé, P. Cadule, O. Aumont, D. Salas y Mélia, A. Voldoire, and G. Madec (2013), Skill assessment of three earth system models with common marine biogeochemistry, *Clim. Dyn.*, *40*(9–10), 2549–2573, doi:10.1007/s00382-012-1362-8.

Seitzinger, S. P., J. A. Harrison, E. Dumont, A. H. W. Beusen, and A. F. Bouwman (2005), Sources and delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application, *Global Biogeochem. Cycles*, *19*, GB4S01, doi:10.1029/2005GB002606.

Semtner, A. J., Jr. (1976), A model for the thermodynamic growth of sea ice in numerical investigations of climate, *J. Phys. Oceanogr.*, *6*(3), 379–389.

Six, K. D., and E. Maier-Reimer (1996), Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model, *Global Biogeochem. Cycles*, *10*(4), 559–583, doi:10.1029/96GB02561.

Skogen, M. D., and H. Søiland (1998), *A User's Guide to NORWECOM v2.0*, *Tech. Rep. Fisken og Havet 18/98*, Inst. of Mar. Res., Bergen, Norway.

Slagstad, D., and T. A. McClimans (2005), Modeling the ecosystem dynamics of the Barents sea including the marginal ice zone: I. Physical and chemical oceanography, *J. Mar. Syst.*, *58*(1–2), 1–18, doi:10.1016/j.jmarsys.2005.05.005.

Slagstad, D., and K. S. Tande (2007), Structure and resilience of overwintering habitats of Calanus finmarchicus in the Eastern Norwegian Sea, *Deep. Sea Res., Part II*, *54*(23–26), 2702–2715, doi:10.1016/j.dsr2.2007.08.024.

Slagstad, D., P. F. J. Wassmann, and I. Ellingsen (2015), Physical constrains and productivity in the future Arctic Ocean, *Front. Mar. Sci.*, *2*, 1–23, doi:10.3389/fmars.2015.00085.

Smith, J. N., F. A. McLaughlin, W. M. Smethie, S. B. Moran, and K. Lepore (2011), Iodine-129, ^{137} Cs, and CFC-11 tracer transit time distributions in the Arctic Ocean, *J. Geophys. Res.*, *116*, C04024, doi:10.1029/2010JC006471.

Smith, R. D., J. K. Dukowicz, and R. C. Malone (1992), Parallel ocean general circulation modeling, *Physica D*, *60*(1–4), 38–61, doi:10.1016/0167-2789(92)90225-C.

Steele, M., R. Morley, and W. Ermold (2001), PHC: A global ocean hydrography with a high quality Arctic Ocean, *J. Clim.*, *14*, 2079–2087.

Stock, C. A., J. P. Dunne, and J. G. John (2014a), Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical–biological model, *Prog. Oceanogr.*, *120*, 1–28, doi:10.1016/j.pocean.2013.07.001.

Stock, C. A., J. P. Dunne, and J. G. John (2014b), Drivers of trophic amplification of ocean productivity trends in a changing climate, *Biogeosciences*, *11*(24), 7125–7135, doi:10.5194/bg-11-7125-2014.

Stow, C. A., J. Jolliff, D. J. McGillicuddy Jr., S. C. Doney, J. Icarus Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead (2009), Skill assessment for coupled biological/physical models of marine systems, *J. Mar. Syst.*, *76*, 4–15.
Sunda, W. G., and S. A. Huntsman (1997), Interrelated influence of iron, light and cell size on marine phytoplankton growth, *Nature*, *390*(6658), 389–392, doi:10.1038/37093.
Takahashi, T., W. S. Broecker, and S. Langer (1985), Redfield ratio based on chemical data from isopycnal surfaces, *J. Geophys. Res.*, *90*(C4), 6907–6924, doi:10.1029/JC090iC04p06907.
Tamelander, T., M. Reigstad, K. Olli, D. Slagstad, and P. Wassmann (2013), New Production Regulates Export Stoichiometry in the Ocean, edited by H. Browman, *PLoS One*, *8*(1), e54027, doi:10.1371/journal.pone.0054027.
Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*,*106*(D7), 7183–7192.
Tedesco, L., and M. Vichi (2014), Sea ice biogeochemistry: A guide for modellers, edited by J. M. *Dias, PLoS One*, *9*(2), e89217, doi:10.1371/journal.pone.0089217.
Tegen, I., and I. Fung (1995), Contribution to the atmospheric mineral aerosol load from land surface modification, *J. Geophys. Res.*, *100*(D9), 18,707–18,726, doi:10.1029/95JD02051.
Timmermann, R., H. Goosse, G. Madec, T. Fichefet, C. Ethe, and V. Dulière (2005), On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model, *Ocean Modell.*, *8*(1–2), 175–201, doi:10.1016/j.ocemod.2003.12.009.
Timmermans, M.-L., S. Cole, and J. Toole (2012), Horizontal Density Structure and Restratification of the Arctic Ocean Surface Layer, *J. Phys. Oceanogr.*, *42*(4), 659–668, doi:10.1175/JPO-D-11-0125.1.
Tjiputra, J. F., C. Roelandt, M. Bentsen, D. M. Lawrence, T. Lorentzen, J. Schwinger, Ø. Seland, and C. Heinze (2013), Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM), *Geosci. Model Dev.*, *6*(2), 301–325, doi:10.5194/gmd-6-301-2013.
Torres-Valdés, S., T. Tsubouchi, S. Bacon, A. C. Naveira-Garabato, R. Sanders, F. A. McLaughlin, B. Petrie, G. Kattner, K. Azetsu-Scott, and T. E. Whitledge (2013), Export of nutrients from the Arctic Ocean, *J. Geophys. Res. Oceans*, *118*, 1625–1644, doi:10.1002/jgrc.20063.
Tréguer, P., D. M. Nelson, A. J. Van Bennekom, D. J. Demaster, A. Leynaert, and B. Quéguiner (1995), The silica balance in the world ocean: A reestimate., *Science*, *268*(5209), 375–379, doi:10.1126/science.268.5209.375.
Tremblay, G., C. Belzile, M. Gosselin, M. Poulin, S. Roy, and J. É. Tremblay (2009), Late summer phytoplankton distribution along a 3500 km transect in Canadian Arctic waters: Strong numerical dominance by picoeukaryotes, *Aquat. Microb. Ecol.*, *54*(1), 55–70, doi:10.3354/ame01257.
Tremblay, J.-É., C. Michel, K. A. Hobson, M. Gosselin, and N. M. Price (2006), Bloom dynamics in early opening waters of the Arctic Ocean, *Limnol. Oceanogr.*, *51*(2), 900–912, doi:10.4319/lo.2006.51.2.0900.
Tremblay, J.-É., et al. (2011), Climate forcing multiplies biological productivity in the coastal Arctic Ocean, *Geophys. Res. Lett.*, *38*, L18604, doi:10.1029/2011GL048825.
Tremblay, J.-É., L. G. Anderson, P. Matrai, P. Coupel, S. Bélanger, C. Michel, and M. Reigstad (2015), Global and regional drivers of nutrient supply, primary production and CO_2 drawdown in the changing Arctic Ocean, *Prog. Oceanogr.*, *139*, 171–196, doi:10.1016/j.pocean.2015.08.009.
Yool, A., E. E. Popova, and T. R. Anderson (2011), MEDUSA-1.0: A new intermediate complexity plankton ecosystem model for the global domain, *Geosci. Model Dev.*, *4*(2), 381–417, doi:10.5194/gmd-4-381-2011.
Yool, A., E. E. Popova, and A. C. Coward (2015), Future change in ocean productivity: Is the Arctic the new Atlantic?, *J. Geophys. Res. Oceans*, *120*, 7771–7790, doi:10.1002/2015JC011167.
Yun, M. S., K. H. Chung, S. Zimmermann, J. Zhao, H. M. Joo, and S. H. Lee (2012), Phytoplankton productivity and its response to higher light levels in the Canada Basin, *Polar Biol.*, *35*(2), 257–268.
Vancoppenolle, M., T. Fichefet, H. Goosse, S. Bouillon, G. Madec, and M. A. M. Maqueda (2009a), Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation, *Ocean Modell.*, *27*(1–2), 33–53, doi:10.1016/j.ocemod.2008.10.005.
Vancoppenolle, M., T. Fichefet, and H. Goosse (2009b), Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 2. Importance of sea ice salinity variations, *Ocean Modell.*, *27*(1–2), 54–69, doi:10.1016/j.ocemod.2008.11.003.
Vancoppenolle, M., L. Bopp, G. Madec, J. Dunne, T. Ilyina, P. R. Halloran, and N. Steiner (2013), Future Arctic Ocean primary productivity from CMIP5 simulations: Uncertain outcome, but consistent mechanisms, *Global Biogeochem. Cycles*, *27*, 605–619, doi:10.1002/gbc.20055.
Vernet, M., P. A. Matrai, and I. Andreassen (1998), Synthesis of particulate and extracellular carbon by phytoplankton at the marginal ice zone in the Barents Sea, *J. Geophys. Res.*, *103*(C1), 1023–1037, doi:10.1029/97JC02288.
Vernet, M., W. A. Kozlowski, L. R. Yarmey, A. T. Lowe, R. M. Ross, L. B. Quetin, and C. H. Fritsen (2012), Primary production throughout austral fall, during a time of decreasing daylength in the western Antarctic Peninsula. *Mar. Ecol. Prog. Ser.*, *452*, 45–61.
Vichi, M., N. Pinardi, and S. Masina (2007a), A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, *J. Mar. Syst.*, *64*(1–4), 89–109, doi:10.1016/j.jmarsys.
Vichi, M., N. Pinardi, and S. Masina (2007b), A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: Numerical simulations, *J. Mar. Syst.*, *64*(1–4), 110–134.
Vichi, M.,et al. (2015a), The Biogeochemical Flux Model (BFM): Equation Description and User Manual, BFM version 5.1. Release 1.1, *BFM Rep. Ser. 2*, The BFM Syst. Team, Bologna, Italy.
Vichi, M., T. Lovato, E. Gutierrez Mlot, and W. McKiver (2015b), Coupling BFM with ocean models: The NEMO model (Nucleus for the European Modelling of the Ocean), Release 1.0, *BFM Rep. Ser. 2*, The BFM Syst. Team, Bologna, Italy.
Voldoire, A., et al. (2013), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, *40*(9–10), 2091–2121, doi:10.1007/s00382-011-1259-y.
Wassmann, P., and M. Reigstad (2011), Future Arctic Ocean seasonal ice zones and implications for pelagic-benthic coupling, *Oceanography*, *24*(3), 220–231.
Wassmann, P., et al. (2006), Food webs and carbon flux in the Barents Sea, *Prog. Oceanogr.*, *71*(2-4), 232–287, doi:10.1016/j.pocean.2006.10.003.
Wassmann, P., et al. (2015), The contiguous domains of Arctic Ocean advection: Trails of life and death, *Prog. Oceanogr.*, *139*, 42–65, doi:10.1016/j.pocean.2015.06.011.
Watanabe, E., et al. (2014), Enhanced role of eddies in the Arctic marine biological pump., *Nat. Commun.*, *5*, 1–9, doi:10.1038/ncomms4950.
Willmott, C. J. (1981), On the validation of models, *Phys. Geogr.*, *2*(2), 184–194, doi:10.1080/02723646.1981.10642213.
Winton, M. (2000), A reformulated three-layer sea ice model, *J. Atmos. Ocean. Technol.*, *17*(4), 525–531, doi:10.1175/1520-0426(2000)017 < 0525:ARTLSI>2.0.CO;2.
Zhang, J., and W. D. Hibler (1997), On an efficient numerical method for modeling sea ice dynamics, *J. Geophys. Res.*, *102*(C4), 8691–8702, doi:10.1029/96JC03744.

Zhang, J., and D. A. Rothrock (2003), Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, *Mon. Weather Rev.*, *131*(Hibler 1980), 845–861, doi:10.1175/1520-0493(2003)131 < 0845:MGSIWA>2.0.CO;2.

Zhang, J., and M. Steele (2007), Effect of vertical mixing on the Atlantic Water layer circulation in the Arctic Ocean, *J. Geophys. Res.*, *112*, C04S04, doi:10.1029/2006JC003732.

Zhang, J., Y. H. Spitz, M. Steele, C. Ashjian, R. Campbell, L. Berline, and P. Matrai (2010a), Modeling the impact of declining sea ice on the Arctic marine planktonic ecosystem, *J. Geophys. Res. Ocean.*, *115*, C10015, doi:10.1029/2009JC005387.

Zhang, J., R. Woodgate, and R. Moritz (2010b), Sea ice response to atmospheric and oceanic forcing in the Bering sea, *J. Phys. Oceanogr.*, *40*(8), 1729–1747, doi:10.1175/2010jpo4323.1.

Zhang, J., Y. H. Spitz, M. Steele, C. Ashjian, R. Campbell, L. Berline, P. Matrai (2010c), Modeling the impact of declining sea ice on the Arctic marine planktonic ecosystem, *J. Geophys. Res.*, *115*, C10015, doi:10.1029/2009JC005387.

Zhang, J., C. Ashjian, R. Campbell, V. Hill, Y. H. Spitz, and M. Steele (2014), The great 2012 Arctic Ocean summer cyclone enhanced biological productivity on the shelves, *J. Geophys. Res. Oceans*, *119*, 297–312, doi:10.1002/2013JC009301.

Zhang, J., C. Ashjian, R. Campbell, Y. H. Spitz, M. Steele, and V. Hill (2015), The influence of sea ice and snow cover and nutrient availability on the formation of massive under-ice phytoplankton blooms in the Chukchi Sea, *Deep Sea Res., Part*, *118*, 122–135, doi:10.1016/j.dsr2.2015.02.008.