JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2018JG004828

Machine Learning Estimates of Global Marine Nitrogen Fixation

Weiyi Tang[1], Zuchuan Li[1], and Nicolas Cassar[1,2]

[1]Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC, USA, [2]Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR 6539 UBO/CNRS/IRD/IFREMER, Institut Universitaire Européen de la Mer, Brest, France

Key Points:
- Global marine N_2 fixation rates of 68 to 90 Tg N/year are estimated based on two machine learning algorithms
- No single environmental factor examined in this study is a strong predictor of marine N_2 fixation rates at global scales
- Comparison of various models shows large discrepancies in the estimated magnitude and distribution of N_2 fixation in the world's oceans

Supporting Information:
- Supporting Information S1
- Movie S1
- Movie S2

Correspondence to:
N. Cassar,
nicolas.cassar@duke.edu

Citation:
Tang, W., Li, Z., & Cassar, N. (2019). Machine learning estimates of global marine nitrogen fixation. *Journal of Geophysical Research: Biogeosciences*, *124*, 717–730. https://doi.org/10.1029/2018JG004828

Received 21 SEP 2018
Accepted 24 FEB 2019
Accepted article online 1 MAR 2019
Published online 28 MAR 2019

Abstract Marine nitrogen (N_2) fixation supplies "new" nitrogen to the global ocean, supporting uptake and sequestration of carbon. Despite its central role, marine N_2 fixation and its controlling factors remain elusive. In this study, we compile over 1,100 published observations to identify the dominant predictors of marine N_2 fixation and derive global estimates based on the machine learning algorithms of random forest and support vector regression. We find that no single environmental property predicts N_2 fixation at global scales. Our random forest and support vector regression algorithms, trained with sampling coordinates and month, solar radiation, wind speed, sea surface temperature, sea surface salinity, surface nitrate, surface phosphate, surface excess phosphorus, minimum oxygen in upper 500 m, photosynthetically available radiation, mixed layer depth, averaged photosynthetically available radiation in the mixed layer, and chlorophyll-*a* concentration, estimate global marine N_2 fixation ranging from 68 to 90 Tg N/year. Comparison of our machine learning estimates and 11 other model outputs currently available in literature shows substantial discrepancies in the global magnitude and spatial distribution of marine N_2 fixation, especially in the tropics and in high latitudes. The large uncertainties in marine N_2 fixation highlighted in our study argue for increased and more coordinated efforts using geochemical tracers, modeling, and observations over broad ocean regions.

1. Introduction

Vast regions of the world's oceans are nitrogen depleted. In these regions, the supply of bioavailable nitrogen by N_2 fixation, conducted by a diverse group of prokaryotes called diazotrophs, supports a significant fraction of primary production (Karl & Letelier, 2008; Sohm et al., 2011). Over long time scales, the balance between marine N_2 fixation flux and denitrification has been hypothesized to control the marine nitrogen inventory, the strength of the biological pump, and ultimately atmospheric CO_2 and the Earth's climate (Altabet, 2007; Falkowski, 1997; Michaels et al., 2001; Somes et al., 2013). However, estimating marine N_2 fixation and its environmental controls remains challenging due to the limited number of observations, with sampling efforts heretofore focusing on tropical and subtropical oceans (Karl et al., 2002; Luo et al., 2012).

Current estimates of global marine N_2 fixation vary widely, with most ranging from less than 100 to over 200 Tg N/year (Gruber & Galloway, 2008; Luo et al., 2012). These estimates are derived using a variety of approaches including direct extrapolation from scarce field observations (Luo et al., 2012) and models (Riche & Christian, 2018). In some prognostic models, N_2 fixation is parameterized relying on N_2 fixers' ecophysiology (Hood et al., 2001; Monteiro et al., 2010; Paulsen et al., 2017). While most of the current models simulate diazotrophy based on *Trichodesmium* (Hood et al., 2004), there is increasing evidence that other diazotrophs with distinct niches, such as unicellular cyanobacteria group A (UCYN-A) and noncyanobacterial diazotrophs, contribute significantly to N_2 fixation (Bombar et al., 2016; Delmont et al., 2018; Martínez-Pérez et al., 2016). Because the biogeographies and ecophysiologies of diazotrophs are poorly defined, the parameterization of N_2 fixation is challenging. Alternatively, N_2 fixation can be derived based on diagnostic models by scaling the relations of N_2 fixation to environmental properties obtained from field observations. Luo et al. (2014) derived a statistical algorithm of marine N_2 fixation using a data set compiled by Luo et al. (2012). Surface solar radiation and subsurface minimum oxygen were identified as the strongest predictors of the geographical distribution of marine N_2 fixation. Since the meta-analysis of Luo et al. (2012), sampling efforts have intensified and the geographical distribution of observations has expanded to broader ocean domains, including higher latitudes and the Indian Ocean (Blais et al., 2012; Shiozaki et al., 2014; Sipler et al., 2017).

Table 1
N_2 Fixation Rates and Environmental Predictors Used in the Regression and Machine Learning Analyses

Data	Symbol	Source	Spatial resolution	Temporal resolution	Range	Log10-transformed	Correlation coefficient	*p* value
Observed N_2 fixation (μmol N·m^{-2}·day^{-1})	NF	Tang et al. (2019)	Point	Daily	0.0011–25750	Yes	/	/
Surface downward solar radiation (W/m^2)	I_s	NCEP/NCAR reanalysis	2°	Daily	60.9–403.9	No	0.0287	0.49
Surface wind speed (m/s)	WS		2.5°		0.7–16.3	No	0.1613	<0.01
Mixed layer depth (m)	MLD	Ifremer	2°	Monthly climatology	10.2–101.7	No	−0.0546	0.19
Sea surface salinity (psu)	SSS	World Ocean	1°	Monthly climatology	21.7–39.2	No	−0.1682	<0.01
Minimum dissolved oxygen in 0–500 m (ml/L)	DO_{min}	Atlas 2013	1°		0–6.6	No	−0.2989	<0.01
Surface nitrate (μM)	DIN		1°		0.0004–13.5	Yes	−0.0419	0.32
Surface phosphate (μM)	DIP		1°		0.0015–1.6	Yes	0.1114	<0.01
Excess phosphate (μM)	P^*		1°		0.0003–1.5	Yes	0.1223	<0.01
Sea surface temperature (°C)	SST	SeaWiFS	0.083°	8 days	2.7–31.8	No	0.2021	<0.01
Photosynthetically available radiation (Einstein·m^{-2}·day^{-1})	PAR	and MODIS	0.083°		11.3–66.4	No	0.0350	0.40
Average PAR in mixed layer (Einstein·m^{-2}·day^{-1})	PAR_{mld}		0.083°		1.8–51.5	No	−0.0445	0.28
Chlorophyll-*a* (mg/m^3)	[Chl]		0.083°		0.016–10.2	Yes	0.1222	<0.01

Note. NCEP = National Centers for Environmental Prediction; NCAR = National Center for Atmospheric Research.

In this study, we build on the work of Luo et al. (2012, 2014) and derive new estimates of N_2 fixation in the world's oceans. To that end, we train two machine learning algorithms with the updated database of Tang et al. (2019), which significantly expands the geographic range of observations presented in Luo et al. (2012), and apply the derived empirical algorithms to map the global biogeography of marine N_2 fixation. We then compare our estimates with the Coupled Model Intercomparison Project Phase 5 (CMIP5) model outputs (Taylor et al., 2012) and other simulations in the literature to identify regions with greater uncertainties. Finally, future projections are presented to further underscore our poor understanding of N_2 fixation and its regulating factors.

2. Materials and Methods

2.1. Data Sources and Transformations

The data sources for the N_2 fixation rates and predictors are listed in Table 1. Our updated database consists of 1,141 observations of depth-integrated N_2 fixation rates, representing ~80% increase on the database used in Luo et al. (2014). Because optical properties are different in coastal (Case II) and open ocean (Case I) waters (Morel & Prieur, 1977), different relations of N_2 fixation to remotely sensed optical properties are expected in Case II and Case I waters. For this reason, we excluded data points at locations with water depth shallower than 200 m. After applying this filter and removing six data points with zero values, 921 observations remained for further analyses (Figure 1). Out of these, 326, 484, and 111 observations were derived based on acetylene reduction assays (Capone & Montoya, 2001), the "N_2 gas addition" method (Montoya et al., 1996), and the "dissolved N_2 gas addition" method (Mohr et al., 2010), respectively. Although both the N_2 gas addition and dissolved N_2 gas addition methods use $^{15}N_2$ as the tracer, $^{15}N_2$ is added in different phase, likely leading to substantial differences in the estimated N_2 fixation rates. For example, incomplete dissolution of $^{15}N_2$ gas in the "$^{15}N_2$ gas addition" method leads to underestimation of N_2 fixation rates (Mohr et al., 2010). We included in our analyses all the observations collected with these various methods despite the known uncertainties and discrepancies. For the analyses, all the N_2 fixation measurements were log10 transformed (Luo et al., 2014).

In contrast to Luo et al. (2014) who modeled N_2 fixation observations based on annual climatologies of predictors, we matched the N_2 fixation observations to contemporaneous satellite predictors when available (e.g., 8-day). When unavailable, we used monthly climatologies instead of annual climatologies in light of the large seasonal variations in N_2 fixation observed in many regions (Böttjer et al., 2016). In addition to

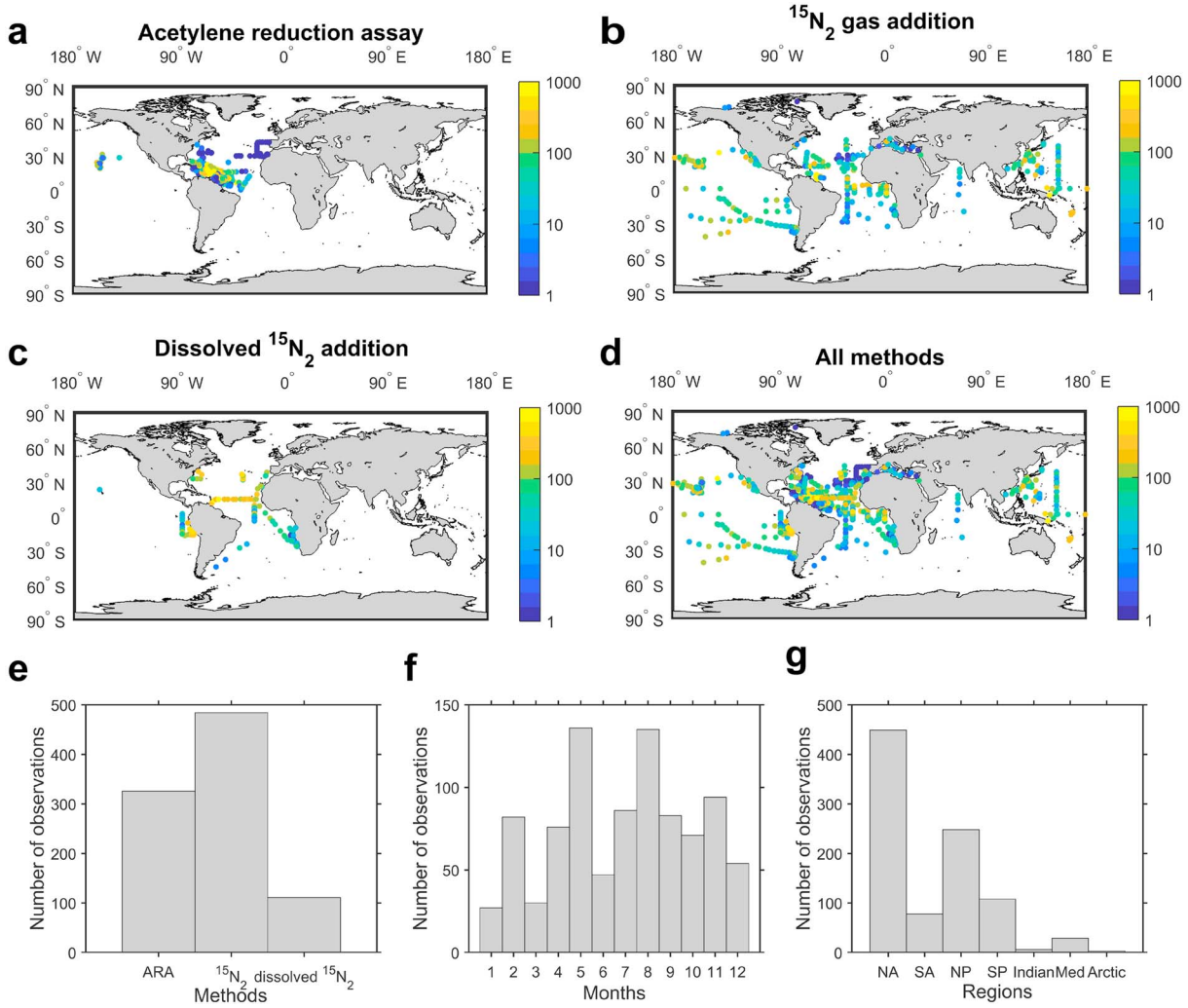

Figure 1. Maps (a–d) and frequency distributions (e–g) of field observations of N_2 fixation rates (unit: μmol N·m^{-2}·day^{-1}) as a function of measurement methods, months, and regions. Only observations used in our analyses are shown. ARA = acetylene reduction assay; NA = North Atlantic; SA = South Atlantic; NP = North Pacific; SP = South Pacific; Indian = Indian Ocean; Med = Mediterranean Sea; Arctic = Arctic Ocean.

depth-integrated N_2 fixation, we also matched the predictors to surface volumetric N_2 fixation. Since it did not significantly improve the prediction of N_2 fixation, the results are not shown here. Predictors used in our analyses include sampling coordinates and month, surface downward solar radiation (I_s), surface wind speed (WS), mixed layer depth (MLD), sea surface temperature (SST), sea surface salinity (SSS), minimum dissolved oxygen in 0–500 m (DO_{min}), surface nitrate (DIN), surface phosphate (DIP), excess phosphate ($P^* = [PO_4^{3-}] - [NO_3^-]/16$), photosynthetically available radiation (PAR), average PAR in the mixed layer (PAR_{mld}), and chlorophyll-*a* concentration ([Chl]; Table 1). These parameters have been hypothesized to directly or indirectly affect N_2 fixation (Karl et al., 2002; Luo et al., 2014; Mahaffey et al., 2005; Severin et al., 2012; Subramaniam et al., 2008). For example, sufficient radiation or PAR is required for autotrophic N_2 fixation as it is an energy-expensive process (Breitbarth et al., 2008; Karl et al., 2002). Excess phosphorus has been used to determine the distribution of N_2 fixation based on the assumption that N_2 fixers would be favored in nitrogen-poor environments (Deutsch et al., 2007). Iron (Fe) was not included in our analyses despite its important role in N_2 fixation (Moore et al., 2009). This is because measurements of iron are limited and modeled Fe fluxes have substantial uncertainties (Tagliabue et al., 2015). The work of Luo et al. (2014) also suggests a weak correlation between N_2 fixation and dust deposition at global scales. In addition, we tested the importance of other properties believed to influence

marine N_2 fixation rates, including Fe:N and P:N nutrient supply ratios as modeled in Ward et al. (2013). A detailed description of the different parameters is presented below.

Sampling coordinates and months were transformed to preserve continuity in the data. For example, the months of January and December are numerically far apart (i.e., months 1 and 12) while they are temporally close (i.e., January is after December). This is also true for geographical distances in coordinate space (e.g., coordinates −179° and 180°). To address this issue, we transformed the latitude, longitude, and month properties to periodic functions simulated with sine and cosine functions as follows (Gade, 2010; Gregor et al., 2017):

$$\text{coordinates} = \begin{pmatrix} \sin\left(\text{latitude}\cdot\frac{\pi}{180}\right) \\ \sin\left(\text{longitude}\cdot\frac{\pi}{180}\right)\cdot\cos\left(\text{latitude}\cdot\frac{\pi}{180}\right) \\ -\cos\left(\text{longitude}\cdot\frac{\pi}{180}\right)\cdot\cos\left(\text{latitude}\cdot\frac{\pi}{180}\right) \end{pmatrix} \quad (1)$$

$$\text{time} = \begin{pmatrix} \cos\left(\text{month}\cdot\frac{2\pi}{12}\right) \\ \sin\left(\text{month}\cdot\frac{2\pi}{12}\right) \end{pmatrix} \quad (2)$$

The I_s and WS data products were downloaded from National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis products (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html; Kalnay et al., 1996). The MLD climatology, using a threshold criterion of 0.03 kg/m^3 compared to density at 10 m, was obtained from Ifremer (http://www.ifremer.fr/cerweb/deboyer/mld/Surface_Mixed_Layer_Depth.php; de Boyer Montégut et al., 2004). Monthly climatologies of SSS (0 m), dissolved oxygen concentration, and nutrient concentrations (0 m) were downloaded from the World Ocean Atlas 2013 (https://www.nodc.noaa.gov/OC5/woa13/woa13data.html; Boyer et al., 2013). DO_{min} was defined as the minimum oxygen concentration in the upper 500 m as in Luo et al. (2014). Eight-day averaged PAR, light attenuation coefficient [K_d(490)], and [Chl] measured by SeaWiFS and MODIS satellite were downloaded from the NASA OceanColor website (https://oceancolor.gsfc.nasa.gov/). PAR denotes the spectral range of the surface downward solar radiation accessible to photosynthetic organisms. The average PAR in the mixed layer was calculated based on the following equations (Morel et al., 2007):

$$K_d(PAR) = 0.0665 + 0.874{*}K_d(490) - 0.00121{*}[K_d(490)]^{-1} \quad (3)$$

$$PAR_{mld} = \frac{1}{mld}\int_0^{mld} PAR{*}e^{-K_d(PAR){*}z}\mathrm{d}z = \frac{PAR}{K_d(PAR){*}mld}{*}\left(1-e^{-K_d(PAR){*}mld}\right) \quad (4)$$

[Chl] and nutrients were log10 transformed. Observations and matched predictors were binned into 2° × 2° grids on monthly scale. After these transformations, we were left with 536 matches of N_2 fixation rates with predictors.

2.2. Linear Regression and Machine Learning Approaches

Regression and statistical analyses were performed using the MATLAB machine learning toolbox (https://www.mathworks.com/solutions/machine-learning.html). Simple regressions were first applied to examine the relationships between N_2 fixation rates and individual predictors. Stepwise multiple linear regression (MLR) was then utilized to build a model between N_2 fixation rates and individual predictors as described in Luo et al. (2014). After these preliminary analyses, two machine learning methods —random forest (RF; Breiman, 2001) and support vector regression (SVR; Vapnik, 2000)—were applied to derive empirical models of N_2 fixation rates as a function of the suite of environmental predictors. These machine learning methods are increasingly used in oceanic studies, for example, recently applied

for building maps of pCO_2 (Gregor et al., 2017) and net community production (Li & Cassar, 2016). See Breiman (2001) and Smola and Schölkopf (2004) for a more thorough description of these machine learning techniques.

For both machine learning methods, we followed the steps described in Gregor et al. (2017) to train and test the models. Briefly, we randomly divided the 536 data matches into a training-validation data set (70%, $n = 375$) and a testing data set (30%, $n = 161$). In the RF and SVR model training, the optimal parameters were determined by minimizing the mean absolute error between the observed and predicted N_2 fixation rates. For the RF algorithm, we built 30 decision trees and set the minimum leaf size to 3. For the SVR algorithm, data were standardized to their z scores ($z = \frac{x-\overline{x}}{\sigma}$) and a Gaussian kernel was used. The trained models were evaluated against the test data set. After the model construction and evaluation, the derived models were used to predict the global distribution of monthly N_2 fixation rates with a 2° × 2° resolution. Sparse and uneven distribution of observations, mismatch of observations with predictors, and methodological biases (i.e., acetylene reduction assays, $^{15}N_2$ gas addition, and dissolved $^{15}N_2$ addition) introduce noise in our estimates. A more detailed discussion of caveats, uncertainties, and future improvements can be found in the supporting information (Bombar et al., 2018; Böttjer et al., 2016; Cassar et al., 2018; Fernandez et al., 2011; Hamersley et al., 2011; Jickells et al., 2017; Landolfi et al., 2015; Luo et al., 2014; Mohr et al., 2010; Monteiro et al., 2011; Mulholland, 2007; Paulsen et al., 2017; Ward et al., 2013; Weber & Deutsch, 2014; Wilson et al., 2012). We tested the sensitivity of our estimates to the limited number of observations used to train the RF and SVR models. To this end, we constructed in silico 100 additional N_2 fixation data sets with the same size as the original data set using a bootstrap approach, that is, randomly sampling N_2 fixation data from the original data set with replacement. For each of these 100 reconstructed data sets, we conducted the same analyses as described above, deriving machine learning models. The errors associated with the sparsity in observations are represented as the standard deviation of the 100 bootstrap global N_2 fixation maps (Figure S1 in the supporting information). Uncertainties can be high, especially in regions where high N_2 fixation rates were observed. This suggests that, in some regions, the derived models are particularly sensitive to the training data. We also tested methodological biases by training the machine learning algorithms using data from each incubation method individually. These analyses show that estimates are also sensitive to the incubation method (Figure S2 in the supporting information). For future projections, we used the environmental factors provided by MPI-ESM-LR from CMIP5 under the Representative Concentration Pathway (RCP) 8.5 scenario to evaluate changes in N_2 fixation rates estimated by RF and SVR from 2006 to 2100.

2.3. Outputs of N_2 Fixation from Other Models

Multiple models estimate the global distribution of marine N_2 fixation. Riche and Christian (2018) underscored the large discrepancies among models, comparing five CMIP5 models and the UVic Earth System Climate Model (CanESM2, MPI-ESM-LR, GFDL-ESM2M, IPSL-CM5A-LR, CESM1-BGC, and UvicESCM). The first two models apply diagnostic relationships between N_2 fixation and environmental properties like excess phosphorus, temperature, and light, while the other models include prognostic diazotrophic groups with nutrient limitations. N_2 fixation in IPSL-CM5A-LR was scaled to balance denitrification, which may have been overestimated (Séférian et al., 2013). We therefore rescaled N_2 fixation in IPSL-CM5A-LR by 0.17 following Riche & Christian, 2018. In our comparisons, we included models obtained from four additional studies: (1) Luo et al. (2014) derived estimates of global N_2 fixation based on a stepwise MLR between observed N_2 fixation rates and predictors; (2) Landolfi et al. (2015) combined resource competition theory and a coupled ecosystem-circulation model to explain the high N_2 fixation rates in nitrogen-rich waters of the North Atlantic but also provided a model for the global ocean; (3) Jickells et al. (2017) evaluated the effects of atmospheric nitrogen deposition on marine N_2 fixation using the output from the PlankTOM model (Buitenhuis et al., 2013; Le Quéré et al., 2005); Finally, (4) Paulsen et al. (2017) incorporated a prognostic representation of diazotrophs in MPI-ESM to map marine N_2 fixation. We also compare CMIP5 RCP 8.5 future projections of N_2 fixation to our two machine learning estimates under the assumption that the semiempirical relationships derived as part of our study hold for the future. All the model outputs were reformatted to 2° × 2° latitude-longitude resolution.

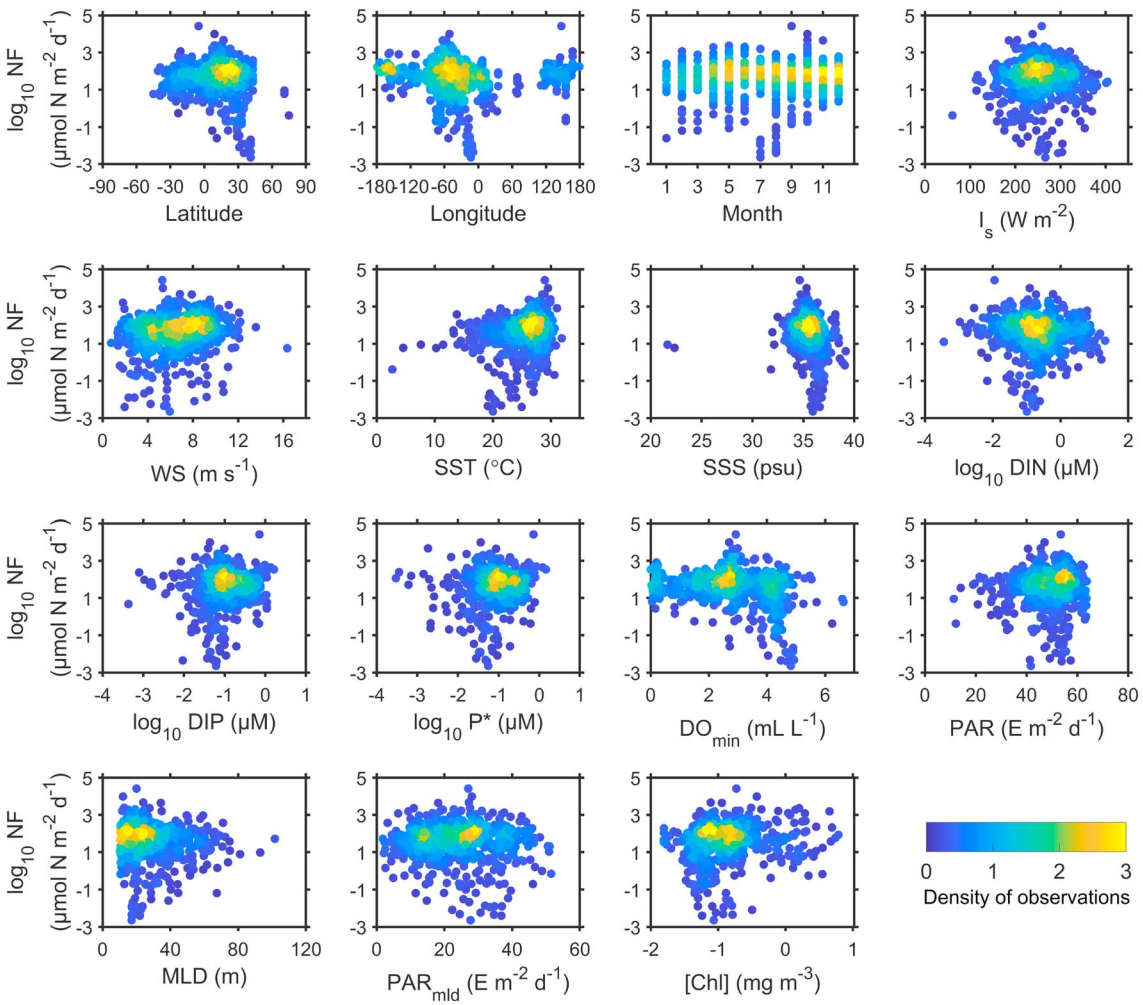

Figure 2. N_2 fixation rates versus environmental predictors. I_s, WS, SST, PAR, and [Chl] are contemporaneous with N_2 fixation observations, while SSS, nutrients, DO_{min}, and MLD are monthly climatologies. The figure is color coded according to the density of observations (Eilers and Goeman (2004)). I_s = surface downward solar radiation; WS = surface wind speed; SST = sea surface temperature; PAR = photosynthetically available radiation; [Chl] = chlorophyll-*a* concentration; SSS = sea surface salinity; DO_{min} = minimum dissolved oxygen in 0–500 m; MLD = mixed layer depth; DIN = surface nitrate; DIP = surface phosphate.

3. Results

3.1. Observations of Marine N_2 Fixation and Relation to Environmental Properties

Observations of marine N_2 fixation range from less than 0.1 to over 10,000 μmol N·m^{-2}·day^{-1} with high rates estimated in the western tropical Atlantic and tropical Pacific (Figure 1). Of the three methods commonly used to measure N_2 fixation rates, the $^{15}N_2$ gas addition method has been applied throughout the global ocean, while the updated dissolved $^{15}N_2$ addition method is increasingly applied but mostly in the Atlantic. The North Atlantic remains the most studied ocean basin while the Indian Ocean, the Arctic, and the Southern Ocean are poorly explored, totaling less than 10 measurements.

Simple linear regressions reveal weak correlations between N_2 fixation rates and environmental properties although some are statistically significant (Figure 2 and Table 1). For instance, DO_{min} explains the highest percentage of variance in the observed N_2 fixation rates showing a negative correlation ($R = -0.3$, $p < 0.01$). SST is positively correlated with N_2 fixation rates ($R = 0.2$, $p < 0.01$). Other statistically significant predictors include SSS, WS, *P**, [Chl], and DIP.

3.2. Machine Learning Estimates

Compared to simple regressions or MLR, machine learning methods are expected to improve predictions by better portraying nonlinearities between the predictors and the predictand. We evaluate the

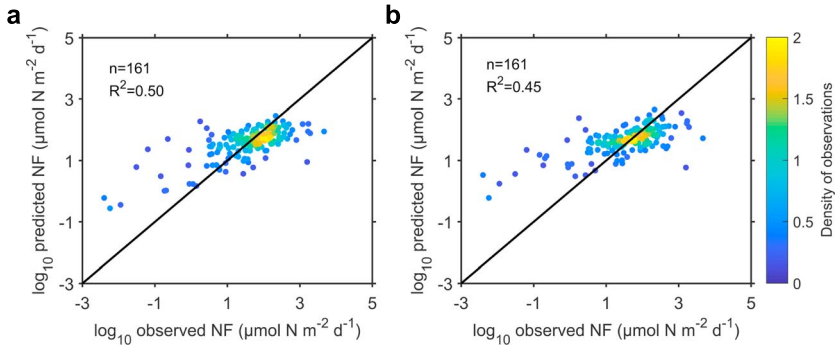

Figure 3. Comparison of observed and predicted N_2 fixation rates for the test data set using (a) random forest and (b) support vector regression with data points color-coded for density of observations. NF = N_2 fixation rate.

performance of our two machine learning algorithms with a test data set (Figure 3). Predicted and observed N_2 fixation rates generally converge onto the identity (1:1) line for both of our models. In contrast, the stepwise MLR (R^2 = 0.22) does not capture as much variance of N_2 fixation as the machine learning methods (Figure S3 in the supporting information). The machine learning methods display biases at the lower (<1 μmol N·m^{-2}·day^{-1}) and upper (>1,000 μmol N·m^{-2}·day^{-1}) end of the range. While this tends to flatten the geographical variability of N_2 fixation, 89% of the N_2 fixation rates in the test data set lies between 1 and 1,000 μmol N·m^{-2}·day^{-1}. Over this range, the mean absolute error is 2.61 μmol N·m^{-2}·day^{-1}, which is 22% (RF) and 25% (SVR) less than over the entire test data set. Comparison of the seasonal climatology of modeled and observed N_2 fixation rates at the Hawaii Ocean Time series shows good agreement, suggesting that our models are able to capture the seasonal cycle of N_2 fixation in some regions (Figure S4 and videos in the supporting information). Longitude, SSS, and DO_{min} were the most important input variables in RF training (Figure S5 in the supporting information).

Global marine N_2 fixation rates of 68 and 90 Tg N/year are estimated based on the RF and SVR models, respectively. The Pacific Ocean contributes most to this global flux (see Table 2 and section 4). The two machine learning models both predict high N_2 fixation rates in the tropical Atlantic and in the western tropical Pacific (Figure 4). In comparison, low N_2 fixation rates are estimated in the subtropical North and South Atlantic, with nonnegligible rates in temperate and polar regions. In these regions, the SVR model predicts slightly higher N_2 fixation rates than the RF model.

Table 2
Estimated N_2 Fixation Fluxes in the World's Oceans (Tg N/year)

Models	North Atlantic	South Atlantic	North Pacific	South Pacific	Indian Ocean	Global
RF (this study)	6.9	3.7	19.6	19.0	11.5	68
SVR (this study)	10.2	4.8	22.6	24.6	12.4	90
CanESM2	18.2	11.6	44.1	33.1	28.5	143
CNRM-CM5	0.64	0.47	1.14	0.87	1.11	5
GFDL-ESM2M	20.9	16.1	38.2	33.0	26	146
IPSL-CM5A-LR	9.4	11.6	36.5	29.6	12.4	106.4
MPI-ESM-LR	12.9	14.9	54.8	56.2	21.5	193
CESM1-BGC	24.7	20.1	39.2	32.5	43.4	170
UvicESCM	30.5	16.3	84.9	90.0	55.0	303
Jickells et al. (2017)	27.4	13.9	51.7	30.5	21.4	164
Paulsen et al. (2017)	9	12.8	63.5	30.5	17.2	135.6
Landolfi et al. (2015)	35	4	34	21	40	134
Luo et al. (2014)	9.1	4.5	23	14	23	74

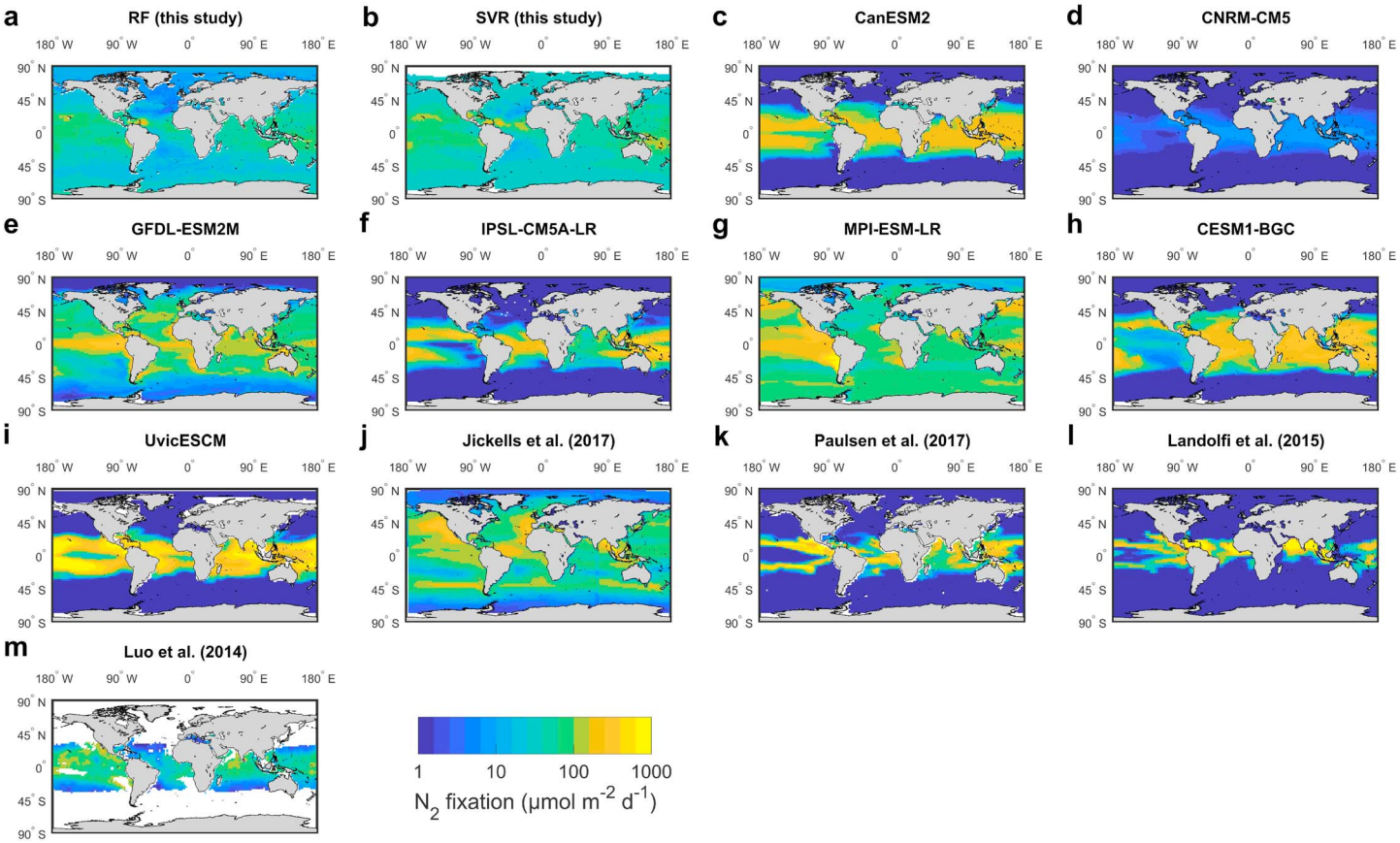

Figure 4. Global distribution of N_2 fixation rates estimated by various models. Daily N_2 fixation rates are calculated by summing monthly N_2 fixation rates and dividing by the number of days in a year. (a) RF (this study), (b) SVR (this study), (c) CanESM2, (d) CNRM-CM5, (e) GFDL-ESM2M, (f) IPSL-CM5A-LR, (g) MPI-ESM-LR, (h) CESM1-BGC, (i) UvicESCM, (j) Jickells et al. (2017), (k) Paulsen et al. (2017), (l) Landolfi et al. (2015), (m) Luo et al. (2014).

4. Discussion

4.1. Environmental Controls on Marine N_2 Fixation

In the study of Luo et al. (2014), DO_{min}, I_s, and SST are found to be the best predictors of N_2 fixation at the global scale. DO_{min} might affect the extent of nitrogen loss in the water column leading to conditions favorable for diazotrophy with an excess of phosphorus ($P^* > 0$). Meanwhile, solar radiation has been hypothesized to constrain the biogeography of diazotrophs in light of their large energy requirement (Mahaffey et al., 2005). In contrast to Luo et al. (2014), we find that none of the properties studied shows a strong correlation with N_2 fixation rates ($R^2 < 0.1$ for all the predictors in Table 1 and Figure 2). This is the case even when applying the MLR approach of Luo et al. (2014). We attribute the lower predictive power to the additional observations in our meta-analysis further confounding the relation of N_2 fixation to predictors. For example, low N_2 fixation rates were recently observed in the eastern South Pacific despite the low DO_{min} and high P^* (Knapp et al., 2016). In addition, organic matter as a source of energy in heterotrophic N_2 fixers may partially offset the light energy requirement (Rahav et al., 2016), thereby broadening the potential niche of N_2 fixation and weakening the correlation to surface solar radiation or to SST. This is in line with the recent discovery of diazotrophs in colder waters (Blais et al., 2012; Fernández-Méndez et al., 2016). We also did not find significant correlations of N_2 fixation rates to modeled Fe:N and P:N nutrient supply ratios or subsurface nutrient concentrations (Figure S6 in the supporting information). Overall, these results challenge some of the traditional assumptions of what regulates N_2 fixation in the ocean. It also implies that simple regression analyses may not capture the complex interplay of environmental factors regulating the distribution and magnitude of N_2 fixation.

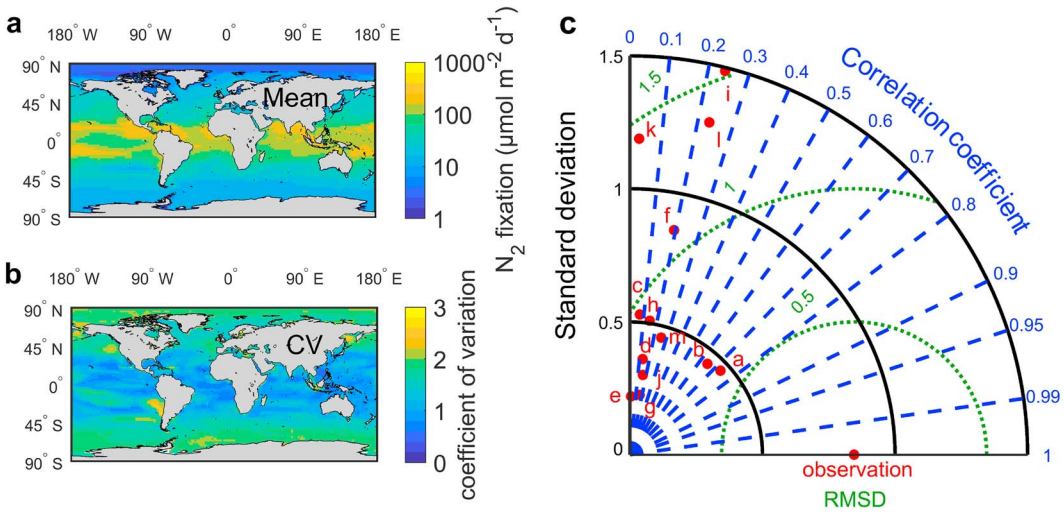

Figure 5. (a) Mean distribution of N_2 fixation calculated based on averaging the 13 model estimates shown in Figure 4. (b) Coefficient of variation (CV) in N_2 fixation estimated by the different models. (c) Taylor diagram of N_2 fixation rates (logarithmic scale) estimated by different models with observations as the reference value. Dashed blue and dotted green lines represent the correlation and centered root-mean-square difference (RMSD) between observations and other estimates, respectively. Solid black lines, the radial distance from the origin, represent the standard deviation of the spatial distribution estimated by each model, with lower values indicating less spatial variability. Note that the observed and simulated N_2 fixation are not well temporally matched when only annually averaged N_2 fixation rates from models are available. a-m represent models with the alphabetical order shown in Figure 4.

4.2. Comparison Between Models

Our machine learning approaches and other models in the literature show substantial discrepancies in the spatial distribution of N_2 fixation rates (Figure 4). Below, we compare the model estimates for different ocean basins.

Global oceans. Modeled global marine N_2 fixation varies widely from 5 to more than 300 Tg N/year (Table 2). Our RF and SVR model estimates of 68 and 90 Tg N/year are in line with others derived from upscaling of field observations (Großkopf et al., 2012; Luo et al., 2012). We however note that our machine learning results represent rough estimates of the magnitude of the global marine N_2 fixation because we extend our estimates to regions outside the range of values of the predictors. The North and South Pacific generally contribute most to the global marine N_2 fixation budget owing to their large surface areas. The majority of models display relatively high N_2 fixation in the tropics and subtropics, such as the western tropical Atlantic and Pacific oceans (Figures 4 and 5a), in line with field observations (Bonnet et al., 2017; Capone et al., 2005). The N_2 fixation rates in temperate midlatitude regions simulated by many of the models is supported by some studies (Needoba et al., 2007; Shiozaki et al., 2015) but will need to be further investigated. Compared to the prognostic models, our two algorithms' estimates are more similar and more in line with MPI-ESM-LR and Luo et al. (2014; Figures 4 and S7 in the supporting information). The models of Paulsen et al. (2017), UvicESCM and Landolfi et al. (2015) show the largest spatial heterogeneity, while our two models display more spatially uniform N_2 fixation, which may result from the flattening of predictions as described above (i.e., underestimation [overestimation] of high [low] N_2 fixation rates). A comparison to field observations further illustrates the current difficulty in modeling N_2 fixation rates (Figure 5c).

Atlantic Ocean. The Atlantic Ocean, which is the most studied ocean basin (Figure 1), displays a small coefficient of variation (CV) among models (Figure 5b). In contrast to the subtropical oceans, the equatorial Atlantic is estimated to harbor intense N_2 fixation rates, consistent with some of the field observations and geochemical estimates (Marconi et al., 2017; Montoya et al., 2007; Snow et al., 2015). However, these hot spots are not consistently identified in all models (e.g., Gulf of Guinea and northeast Atlantic with relatively large CV). Some field observations report different extents of N_2 fixation in the subtropical northeast Atlantic (Benavides et al., 2011; Fonseca-Batista et al., 2017), which may be attributed to differences in methodologies, sampling time, and heterogeneous distribution of N_2 fixation. Many models estimate comparable

levels of N_2 fixation in the South and North Atlantic (e.g., MPI-ESM-LR in Figure 4 and Table 2), in line with the recent study of Fonseca-Batista et al. (2017).

Indian Ocean. In contrast to the Atlantic Ocean, the Indian Ocean has received much less attention which makes it difficult to evaluate the different models. A majority of the models consistently (i.e., low CV) display significant N_2 fixation across the entire ocean basin. Conversely, the limited number of measurements currently available depict a heterogeneous distribution of N_2 fixation (Shiozaki et al., 2014), with some hot spots near the coastal Arabian Sea (Ahmed et al., 2017; Gandhi et al., 2011; Kumar et al., 2017). Some models (e.g., MPI-ESM-LR, CESM1-BGC and Landolfi et al. (2015)) display significant N_2 fixation rates in the northern Indian Ocean where high denitrification rates are associated with the oxygen minimum zone (Ward et al., 2009). In contrast, our two models predict a more homogeneous N_2 fixation distribution annually but with strong seasonal variations (videos in the supporting information). The RF model predicts high N_2 fixation in the western Indian Ocean near Madagascar Island in austral summer, which is supported by some observations (Poulton et al., 2009).

Pacific Ocean. Noticeable discrepancies in models are observed in the Pacific Ocean. On one end of the spectrum are models simulating elevated N_2 fixation rates close to the eastern boundary, as expected should there be a coupling between nitrogen inputs and losses (GFDL-ESM2M, MPI-ESM-LR, and Deutsch et al., 2007). On the other end of the spectrum are models exhibiting low rates around the equator flanked by higher rates in subtropical waters (CanESM2, UvicESCM, Paulsen et al., 2017, and Landolfi et al., 2015). N_2 fixation in some of the models are parameterized to be inhibited by high nitrate concentration, leading to estimates of low rates in the upwelling regions of the eastern Pacific. The remaining models show some combination of the above and/or more complicated distributions. For example, CESM1-BGC displays low N_2 fixation rates in a large portion of the eastern Pacific but increasing N_2 fixation rates westward. Knapp et al. (2016) suggested Fe availability limits diazotrophy in the eastern tropical South Pacific (ETSP). However, Löscher et al. (2016) measured high N_2 fixation signals in anticyclonic eddies in the ETSP especially when accounting for N_2 fixation in the aphotic zone. Our semiempirical approach, which includes the data sets of Knapp et al. (2016) and Löscher et al. (2016), shows high N_2 fixation during some seasons (e.g., austral summer) but midrange annual N_2 fixation rates in the ETSP. More observations of Fe supply and concentration in the eastern Pacific and a better representation of Fe in Earth system models will potentially improve the simulation of N_2 fixation in this region.

Polar oceans. Not too surprisingly, there is a large level of disagreement between models (i.e., high CV) in N_2 fixation rates in polar oceans (Figures 4 and 5). While many models confine N_2 fixation to tropical and subtropical areas (e.g., UviESCM, Paulsen et al., 2017; Landolfi et al., 2015), others show substantial N_2 fixation in colder waters such as in the Southern Ocean, subpolar Pacific, and Arctic Ocean (MPI-ESM-LR; Jickells et al., 2017; and Deutsch et al., 2007). We extrapolate our machine-learning estimates to polar regions with the caveat that they are particularly tenuous because they are outside the range of observations used to train the models. The estimated N_2 fixation rates based on the RF and SVR models are low but not negligible. This is in line with recent observations that suggest that N_2 fixation may not be confined to warm and nutrient-poor waters (Blais et al., 2012; Shiozaki et al., 2017). UCYN-A has recently been found in cold waters (Harding et al., 2018; Shiozaki et al., 2018), and N_2 fixation could provide as much as 3.5 Tg N/year to the Arctic Ocean (Sipler et al., 2017). Differences between model simulations may be attributed to the parameterization of the types of diazotrophs in the models. For example, using *Trichodesmium* as a model organism limits diazotrophy to warm waters (Paulsen et al., 2017). In contrast, the biogeography of N_2 fixation is extended to extratropical regions in models based on geochemical proxies such as the supply of excess phosphorus (Deutsch et al., 2007). Additional measurements in temperate and polar regions and a better model representation of the large diversity in diazotrophic ecophysiologies (in addition to *Trichodesmium*) should help resolve some of these models' discrepancies.

4.3. Future Projections of N_2 Fixation

Since nitrogen, carbon, and climate are tightly coupled (Falkowski, 1997; Falkowski et al., 1998; Gruber & Galloway, 2008), improved simulation of N_2 fixation is critical in assessing future climate. Over the next century, a multitude of factors known to affect N_2 fixation are expected to change, including SST, CO_2, pH, O_2 levels, and many other known and yet-to-be determined factors (Stocker et al., 2013). While our

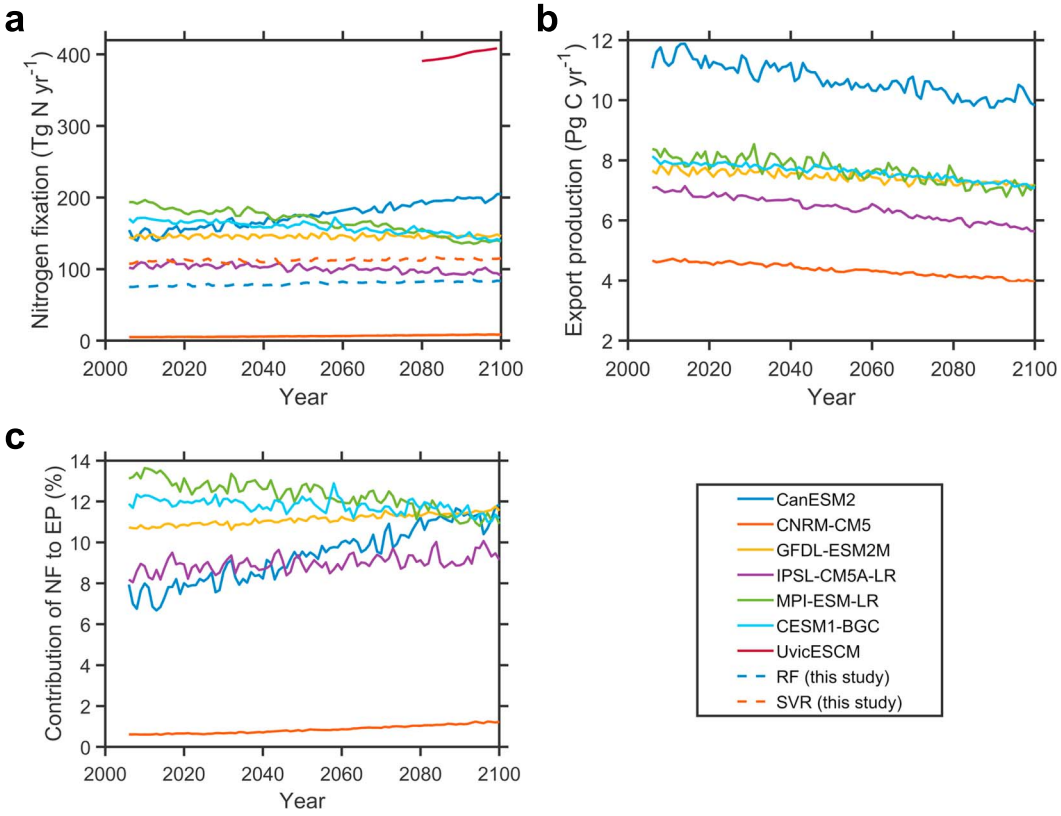

Figure 6. Projections of (a) N_2 fixation rates, (b) export production, and (c) contribution of N_2 fixation to export production under the Representative Concentration Pathway 8.5 scenario. NF = N_2 fixation rate; EP = Export production.

understanding of how these multiple stressors impact the model organism *Trichodesmium* is improving (Hong et al., 2017; Jiang et al., 2018), we now also realize that differences between ecotypes, strains, and species can be significant (Hutchins et al., 2013). The challenge is to estimate how these potentially interactive regulating factors will evolve and how they may influence the broad spectrum of N_2 fixers with their distinct niches. Within this context, it is not surprising to observe substantial differences in future projections of marine N_2 fixation and its contribution to export production. Differences in projections under RCP8.5 based on our machine learning methods, CMIP5, and UvicESCM models are presented in Figure 6. While all CMIP5 models under RCP 8.5 experiments project an average 13% decrease in export production under a warming world (Figure 6b), the magnitude and sign of the change in N_2 fixation varies between models, further highlighting the difficulty in simulating N_2 fixation, especially under a changing climate. Several models project moderate (4–11%; RF, SVR, CNRM-CM5, and UvicESCM) to substantial (33%; CanESM2) rise in marine N_2 fixation by year 2100. In contrast, other models simulate *quasi status quo* (GFDL-ESM2M) or a decline in N_2 fixation (MPI-ESM-LR, IPSL-CM5A-LR, and CESM1-BGC). The large uncertainties in N_2 fixation estimates also lead to substantial variations in the estimated contribution of N_2 fixation to export production, albeit with four models converging on ~12% by year 2100 (Figure 6c).

5. Conclusion

Based on an updated meta-analysis, we found that no single environmental factor can effectively predict marine N_2 fixation at the global scale, suggesting that N_2 fixation may be regulated by an interplay of environmental factors. In light of this, we modeled the marine distribution of N_2 fixation by training two machine-learning algorithms with multiple environmental predictors. While the RF and SVR estimates of global N_2 fixation of 68–90 Tg N/year are in line with previous estimates, the predicted biogeographies are

distinctly different from other simulations. The disagreement between the semiempirical and the process-based models and the sensitivity to the training data sets indicates that current observations may not fully capture the variability in N_2 fixation and/or that we have a poor grasp of the regulating factors. This is in line with the parallel realization that diazotrophs are more diverse and widespread than originally thought (Delmont et al., 2018; Shiozaki et al., 2018). Estimating current and future marine N_2 fixation remains a challenge which may be best addressed with a multipronged approach, combining improved measurement methods, physiological studies, satellite estimates, expanded observations in undersampled regions, and refined models.

Acknowledgments
We thank Olivier G. J. Riche, James R. Christian, Angela Landolfi, Hanna Paulsen, and Ben Ward for providing their model results and helpful discussions. We acknowledge the Coupled Model Intercomparison Project for providing access to the model output. We thank Michael Eby for providing outputs from the UVic Earth System Climate model and T. D. Jickells and coauthors for making some PlankTOM model outputs publically available. We also would like to thank an anonymous reviewer and Y. W. Luo for thorough and insightful comments on our manuscript. Data used in the study can be downloaded from NASA Ocean Color, NCEP/NCAR reanalysis, World Ocean Atlas, and Ifremer. W. T. and N. C. were supported by an NSF-CAREER grant (1350710). N. C. was also supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and co-funded by a grant from the French Government under the program "Investissements d'Avenir."

References

Ahmed, A., Gauns, M., Kurian, S., Bardhan, P., Pratihary, A., Naik, H., et al. (2017). Nitrogen fixation rates in the eastern Arabian Sea. *Estuarine, Coastal and Shelf Science*, *191*, 74–83. https://doi.org/10.1016/j.ecss.2017.04.005

Altabet, M. (2007). Constraints on oceanic N balance/imbalance from sedimentary ^{15}N records. *Biogeosciences*, *4*(1), 75–86. https://doi.org/10.5194/bg-4-75-2007

Benavides, M., Agawin, N. S., Arístegui, J., Ferriol, P., & Stal, L. J. (2011). Nitrogen fixation by *Trichodesmium* and small diazotrophs in the subtropical northeast Atlantic. *Aquatic Microbial Ecology*, *65*(1), 43–53. https://doi.org/10.3354/ame01534

Blais, M., Tremblay, J.-É., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., & Lovejoy, C. (2012). Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. *Global Biogeochemical Cycles*, *26*, GB3022. https://doi.org/10.1029/2011GB004096

Bombar, D., Paerl, R. W., Anderson, R., & Riemann, L. (2018). Filtration via conventional glass fiber filters in $^{15}N_2$ tracer assays fails to capture all nitrogen-fixing prokaryotes. *Frontiers in Marine Science*, *5*(6). https://doi.org/10.3389/fmars.2018.00006

Bombar, D., Paerl, R. W., & Riemann, L. (2016). Marine non-cyanobacterial diazotrophs: Moving beyond molecular detection. *Trends in Microbiology*, *24*(11), 916–927. https://doi.org/10.1016/j.tim.2016.07.002

Bonnet, S., Caffin, M., Berthelot, H., & Moutin, T. (2017). Hot spot of N_2 fixation in the western tropical South Pacific pleads for a spatial decoupling between N_2 fixation and denitrification. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(14), 2800–2801. https://doi.org/10.1073/pnas.1619514114

Böttjer, D., Dore, J. E., Karl, D. M., Letelier, R. M., Mahaffey, C., Wilson, S. T., et al. (2016). Temporal variability of nitrogen fixation and particulate nitrogen export at Station ALOHA. *Limnology and Oceanography*, *62*(1), 200–216. https://doi.org/10.1002/lno.10386

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*, C12003. https://doi.org/10.1029/2004JC002378

Boyer, T. P., Antonov, J. I., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., et al. (2013). World ocean database 2013. NOAA Atlas NESDIS 72. https://doi.org/10.7289/V5NZ85MT

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breitbarth, E., Wohlers, J., Kläs, J., LaRoche, J., & Peeken, I. (2008). Nitrogen fixation and growth rates of *Trichodesmium* IMS-101 as a function of light intensity. *Marine Ecology Progress Series*, *359*, 25–36. https://doi.org/10.3354/meps07241

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

Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., et al. (2005). Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, *19*, GB2024. https://doi.org/10.1029/2004GB002331

Capone, D. G., & Montoya, J. P. (2001). Nitrogen fixation and denitrification. *Methods in Microbiology*, *30*, 501–515. https://doi.org/10.1016/S0580-9517(01)30060-0

Cassar, N., Tang, W., Gabathuler, H., & Huang, K. (2018). Method for high frequency underway N_2 fixation measurements: Flow-through incubation acetylene reduction assays by cavity ring down laser absorption spectroscopy (FARACAS). *Analytical Chemistry*, *90*(4), 2839–2851. https://doi.org/10.1021/acs.analchem.7b04977

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., et al. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(7), 804–813. https://doi.org/10.1038/s41564-018-0176-9

Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N., & Dunne, J. P. (2007). Spatial coupling of nitrogen inputs and losses in the ocean. *Nature*, *445*(7124), 163–167. https://doi.org/10.1038/nature05392

Eilers, P. H., & Goeman, J. J. (2004). Enhancing scatterplots with smoothed densities. *Bioinformatics*, *20*(5), 623–628. https://doi.org/10.1093/bioinformatics/btg454

Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature*, *387*(6630), 272–275. https://doi.org/10.1038/387272a0

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, *281*(5374), 200–206. https://doi.org/10.1126/science.281.5374.200

Fernandez, C., Farías, L., & Ulloa, O. (2011). Nitrogen fixation in denitrified marine waters. *PLoS ONE*, *6*(6), e20539. https://doi.org/10.1371/journal.pone.0020539

Fernández-Méndez, M., Turk-Kubo, K. A., Buttigieg, P. L., Rapp, J. Z., Krumpen, T., Zehr, J. P., & Boetius, A. (2016). Diazotroph diversity in the sea ice, melt ponds, and surface waters of the Eurasian Basin of the Central Arctic Ocean. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01884

Fonseca-Batista, D., Dehairs, F., Riou, V., Fripiat, F., Elskens, M., Deman, F., et al. (2017). Nitrogen fixation in the eastern Atlantic reaches similar levels in the Southern and Northern Hemispheres. *Journal of Geophysical Research: Oceans*, *122*, 587–601. https://doi.org/10.1002/2016JC012335

Gade, K. (2010). A non-singular horizontal position representation. *Journal of Navigation*, *63*(3), 395–417. https://doi.org/10.1017/S0373463309990415

Gandhi, N., Singh, A., Prakash, S., Ramesh, R., Raman, M., Sheshshayee, M., & Shetye, S. (2011). First direct measurements of N_2 fixation during a *Trichodesmium* bloom in the eastern Arabian Sea. *Global Biogeochemical Cycles*, *25*, GB4014. https://doi.org/10.1029/2010GB003970

Gregor, L., Kok, S., & Monteiro, P. M. S. (2017). Empirical methods for the estimation of Southern Ocean CO_2: Support vector and random forest regression. *Biogeosciences*, *14*(23), 5551–5569. https://doi.org/10.5194/bg-14-5551-2017

Großkopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M. M. M., et al. (2012). Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature*, *488*(7411), 361–364. https://doi.org/10.1038/nature11338

Gruber, N., & Galloway, J. N. (2008). An Earth-system perspective of the global nitrogen cycle. *Nature*, *451*(7176), 293–296. https://doi.org/10.1038/nature06592

Hamersley, M. R., Turk, K., Leinweber, A., Gruber, N., Zehr, J., Gunderson, T., & Capone, D. (2011). Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. *Aquatic Microbial Ecology*, *63*(2), 193–205. https://doi.org/10.3354/ame01494

Harding, K., Turk-Kubo, K. A., Sipler, R. E., Mills, M. M., Bronk, D. A., & Zehr, J. P. (2018). Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13,371–13,375. https://doi.org/10.1073/pnas.1813658115

Hong, H., Shen, R., Zhang, F., Wen, Z., Chang, S., Lin, W., et al. (2017). The complex effects of ocean acidification on the prominent N_2-fixing cyanobacterium *Trichodesmium*. *Science*, *356*(6337), 527–531. https://doi.org/10.1126/science.aal2981

Hood, R. R., Bates, N. R., Capone, D. G., & Olson, D. B. (2001). Modeling the effect of nitrogen fixation on carbon and nitrogen fluxes at BATS. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8-9), 1609–1648. https://doi.org/10.1016/S0967-0645(00)00160-0

Hood, R. R., Coles, V. J., & Capone, D. G. (2004). Modeling the distribution of *Trichodesmium* and nitrogen fixation in the Atlantic Ocean. *Journal of Geophysical Research*, *109*, C06006. https://doi.org/10.1029/2002JC001753

Hutchins, D. A., Fu, F.-X., Webb, E. A., Walworth, N., & Tagliabue, A. (2013). Taxon-specific response of marine nitrogen fixers to elevated carbon dioxide concentrations. *Nature Geoscience*, *6*(9), 790–795. https://doi.org/10.1038/ngeo1858

Jiang, H.-B., Fu, F.-X., Rivero-Calle, S., Levine, N. M., Sañudo-Wilhelmy, S. A., Qu, P.-P., et al. (2018). Ocean warming alleviates iron limitation of marine nitrogen fixation. *Nature Climate Change*, *8*(8), 709–712. https://doi.org/10.1038/s41558-018-0216-8

Jickells, T. D., Buitenhuis, E., Altieri, K., Baker, A. R., Capone, D., Duce, R. A., et al. (2017). A re-evaluation of the magnitude and impacts of anthropogenic atmospheric nitrogen inputs on the ocean. *Global Biogeochemical Cycles*, *31*, 289–305. https://doi.org/10.1002/2016GB005586

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Karl, D. M., & Letelier, R. (2008). Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes. *Marine Ecology Progress Series*, *364*, 257–268. https://doi.org/10.3354/meps07547

Karl, D. M., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., et al. (2002). Dinitrogen fixation in the world's oceans. *Biogeochemistry*, *57*(1), 47–98. https://doi.org/10.1023/A:1015798105851

Knapp, A. N., Casciotti, K. L., Berelson, W. M., Prokopenko, M. G., & Capone, D. G. (2016). Low rates of nitrogen fixation in eastern tropical South Pacific surface waters. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(16), 4398–4403. https://doi.org/10.1073/pnas.1515641113

Kumar, P., Singh, A., Ramesh, R., & Nallathambi, T. (2017). N_2 fixation in the eastern Arabian Sea: Probable role of heterotrophic diazotrophs. *Frontiers in Marine Science*, *4*, 80. https://doi.org/10.3389/fmars.2017.00080

Landolfi, A., Koeve, W., Dietze, H., Kähler, P., & Oschlies, A. (2015). A new perspective on environmental controls of marine nitrogen fixation. *Geophysical Research Letters*, *42*, 4482–4489. https://doi.org/10.1002/2015GL063756

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, *11*(11), 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x

Li, Z., & Cassar, N. (2016). Satellite estimates of net community production based on O_2/Ar observations and comparison to other estimates. *Global Biogeochemical Cycles*, *30*, 735–752. https://doi.org/10.1002/2015GB005314

Löscher, C. R., Bourbonnais, A., Dekaezemacker, J., Charoenpong, C. N., Altabet, M. A., Bange, H. W., et al. (2016). N_2 fixation in eddies of the eastern tropical South Pacific Ocean. *Biogeosciences*, *13*(10), 2889–2899. https://doi.org/10.5194/bg-13-2889-2016

Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., et al. (2012). Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, *4*(1), 47–73. https://doi.org/10.5194/essd-4-47-2012

Luo, Y. W., Lima, I. D., Karl, D. M., Deutsch, C. A., & Doney, S. C. (2014). Data-based assessment of environmental controls on global marine nitrogen fixation. *Biogeosciences*, *11*(3), 691–708. https://doi.org/10.5194/bg-11-691-2014

Mahaffey, C., Michaels, A. F., & Capone, D. G. (2005). The conundrum of marine N_2 fixation. *American Journal of Science*, *305*(6–8), 546–595. https://doi.org/10.2475/ajs.305.6-8.546

Marconi, D., Sigman, D. M., Casciotti, K. L., Campbell, E. C., Alexandra Weigand, M., Fawcett, S. E., et al. (2017). Tropical dominance of N_2 fixation in the North Atlantic Ocean. *Global Biogeochemical Cycles*, *31*, 1608–1623. https://doi.org/10.1002/2016GB005613

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, *1*(11), 16163. https://doi.org/10.1038/nmicrobiol.2016.163

Michaels, A. F., Karl, D. M., & Capone, D. G. (2001). Element stoichiometry, new production and nitrogen fixation. *Oceanography*, *14*(4), 68–77. https://doi.org/10.5670/oceanog.2001.08

Mohr, W., Großkopf, T., Wallace, D. W., & LaRoche, J. (2010). Methodological underestimation of oceanic nitrogen fixation rates. *PLoS ONE*, *5*(9), e12583. https://doi.org/10.1371/journal.pone.0012583

Monteiro, F., Dutkiewicz, S., & Follows, M. (2011). Biogeographical controls on the marine nitrogen fixers. *Global Biogeochemical Cycles*, *25*, GB2003. https://doi.org/10.1029/2010GB003902

Monteiro, F. M., Follows, M. J., & Dutkiewicz, S. (2010). Distribution of diverse nitrogen fixers in the global ocean. *Global Biogeochemical Cycles*, *24*, GB3017. https://doi.org/10.1029/2009GB003731

Montoya, J., Voss, M., & Capone, D. (2007). Spatial variation in N_2-fixation rate and diazotroph activity in the tropical Atlantic. *Biogeosciences*, *4*(3), 369–376. https://doi.org/10.5194/bg-4-369-2007

Montoya, J. P., Voss, M., Kahler, P., & Capone, D. G. (1996). A simple, high-precision, high-sensitivity tracer assay for N_2 fixation. *Applied and Environmental Microbiology*, *62*(3), 986–993.

Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., et al. (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geoscience*, *2*(12), 867–871. https://doi.org/10.1038/ngeo667

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (case 1) waters in the perspective of a multi-sensor approach. *Remote Sensing of Environment*, *111*(1), 69–88. https://doi.org/10.1016/j.rse.2007.03.012

Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color 1. *Limnology and Oceanography*, *22*(4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

Mulholland, M. (2007). The fate of nitrogen fixed by diazotrophs in the ocean. *Biogeosciences*, *4*(1), 37–51. https://doi.org/10.5194/bg-4-37-2007

Needoba, J. A., Foster, R. A., Sakamoto, C., Zehr, J. P., & Johnson, K. S. (2007). Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. *Limnology and Oceanography*, *52*(4), 1317–1327. https://doi.org/10.4319/lo.2007.52.4.1317

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*, 438–464. https://doi.org/10.1002/2016MS000737

Poulton, A. J., Stinchcombe, M. C., & Quartly, G. D. (2009). High numbers of *Trichodesmium* and diazotrophic diatoms in the southwest Indian Ocean. *Geophysical Research Letters*, *36*, L15610. https://doi.org/10.1029/2009GL039719

Rahav, E., Giannetto, M., & Bar-Zeev, E. (2016). Contribution of mono and polysaccharides to heterotrophic N_2 fixation at the eastern Mediterranean coastline. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep27858

Riche, O., & Christian, J. (2018). Ocean dinitrogen fixation and its potential effects on ocean primary production in Earth system model simulations of anthropogenic warming. *Elementa: Science of the Anthropocene*, *6*(1). https://doi.org/10.1525/elementa.277

Séférian, R., Bopp, L., Gehlen, M., Orr, J. C., Ethé, C., Cadule, P., et al. (2013). Skill assessment of three Earth system models with common marine biogeochemistry. *Climate Dynamics*, *40*(9–10), 2549–2573. https://doi.org/10.1007/s00382-012-1362-8

Severin, I., Confurius-Guns, V., & Stal, L. J. (2012). Effect of salinity on nitrogenase activity and composition of the active diazotrophic community in intertidal microbial mats. *Archives of Microbiology*, *194*(6), 483–491. https://doi.org/10.1007/s00203-011-0787-5

Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., et al. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *Global Biogeochemical Cycles*, *31*, 996–1009. https://doi.org/10.1002/2017GB005681

Shiozaki, T., Fujiwara, A., Ijichi, M., Harada, N., Nishino, S., Nishi, S., et al. (2018). Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnology and Oceanography*, *63*(5), 2191–2205. https://doi.org/10.1002/lno.10933

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., & Furuya, K. (2014). Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. *Global Biogeochemical Cycles*, *28*, 1096–1110. https://doi.org/10.1002/2014GB004886

Shiozaki, T., Nagata, T., Ijichi, M., & Furuya, K. (2015). Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences*, *12*(15), 4751–4764. https://doi.org/10.5194/bg-12-4751-2015

Sipler, R. E., Gong, D., Baer, S. E., Sanderson, M. P., Roberts, Q. N., Mulholland, M. R., & Bronk, D. A. (2017). Preliminary estimates of the contribution of Arctic nitrogen fixation to the global nitrogen budget. *Limnology and Oceanography Letters*, *2*(5), 159–166. https://doi.org/10.1002/lol2.10046

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Snow, J. T., Schlosser, C., Woodward, E. M. S., Mills, M. M., Achterberg, E. P., Mahaffey, C., et al. (2015). Environmental controls on the biogeography of diazotrophy and *Trichodesmium* in the Atlantic Ocean. *Global Biogeochemical Cycles*, *29*, 865–884. https://doi.org/10.1002/2015GB005090

Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews. Microbiology*, *9*(7), 499–508. https://doi.org/10.1038/nrmicro2594

Somes, C. J., Oschlies, A., & Schmittner, A. (2013). Isotopic constraints on the pre-industrial oceanic nitrogen budget. *Biogeosciences*, *10*(9), 5889–5910. https://doi.org/10.5194/bg-10-5889-2013

Stocker, T. F., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., et al. (2013). Climate change 2013: The physical science basis. Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5). New York.

Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Bjorkman, K., Cooley, S., et al. (2008). Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30), 10,460–10,465. https://doi.org/10.1073/pnas.0710279105

Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., et al. (2015). How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochemical Cycles*, *30*, 149–174. https://doi.org/10.1002/2015GB005289

Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A., et al. (2019). Revisiting the distribution of oceanic N_2 fixation and estimating diazotrophic contribution to marine production. *Nature Communications*, *10*(1), 831. https://doi.org/10.1038/s41467-019-08640-0

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Vapnik, V. (2000). *The nature of statistical learning theory* (2nd ed.). New York: Springer.

Ward, B., Devol, A. H., Rich, J. J., Chang, B. X., Bulow, S. E., Naik, H., et al. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature*, *461*(7260), 78–81. https://doi.org/10.1038/nature08276

Ward, B. A., Dutkiewicz, S., Moore, C. M., & Follows, M. J. (2013). Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnology and Oceanography*, *58*(6), 2059–2075. https://doi.org/10.4319/lo.2013.58.6.2059

Weber, T., & Deutsch, C. (2014). Local versus basin-scale limitation of marine nitrogen fixation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8741–8746. https://doi.org/10.1073/pnas.1317193111

Wilson, S. T., Böttjer, D., Church, M. J., & Karl, D. M. (2012). Comparative assessment of nitrogen fixation methodologies, conducted in the oligotrophic North Pacific Ocean. *Applied and Environmental Microbiology*, *78*(18), 6516–6523. https://doi.org/10.1128/AEM.01146-12