



A robust empirical model for deriving depth-integrated N₂ fixation rates from surface measurements in the global open ocean

Wei-Li Chen¹, Zhibo Shao², Yangchun Xu¹, Ya-Wei Luo^{1,3,4}

¹State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, 361102, China

²School of Ocean and Earth Science, University of Southampton, Southampton, SO14 3ZH, UK

³Fujian Ocean Innovation Center, Xiamen 361102, China

⁴China-ASEAN College of Marine Sciences, Xiamen University Malaysia, 43900 Sepang, Selangor Darul Ehsan, Malaysia

Correspondence to: Ya-Wei Luo (ywluo@xmu.edu.cn)

Abstract. Marine dinitrogen (N₂) fixation (NF) is a critical nitrogen source supporting ocean productivity, yet its global quantification remains challenging because vertically resolved measurements are logistically demanding. To address this, we developed an empirical model to predict depth-integrated NF (INF, $\mu\text{mol N m}^{-2} \text{d}^{-1}$) within the euphotic zone directly from surface NF (SNF, $\mu\text{mol N m}^{-3} \text{d}^{-1}$) measurements. Using 690 measured NF profiles from the global open ocean, we established a robust relationship: $\log_{10}\text{INF} = 1.85 + 0.84 \times \log_{10}\text{SNF}$ ($R^2 = 0.82$, RMSE = 0.31). The model performed well globally, particularly for typical profiles in which NF decreased monotonically with depth. It tended to underestimate $\log_{10}\text{INF}$ for profiles with substantial subsurface NF maxima, especially in the subtropical South Pacific (bias = -0.20), where such maxima were likely sustained by deeper iron supply, and in the northwestern Pacific (bias = -0.33), where the maxima were likely driven by photoinhibition of surface unicellular diazotrophs. Overestimation of $\log_{10}\text{INF}$ occurred in the eastern Atlantic (bias = +0.24), likely reflecting steep depth declines in NF driven by rapid decreases in light and iron availability. A small subset of profiles had uniformly low NF rates throughout the euphotic zone, associated with low iron and/or phosphorus concentrations. When applied to historical surface measurements lacking corresponding INF observations, model-derived INF estimates were statistically consistent with available INF measurements, supporting its preliminary application. Our model enables more efficient global surveys of N₂ fixation, thereby improving our ability to map this critical process and refine understanding of ocean nutrient cycles.

1 Introduction

Marine dinitrogen (N₂) fixation (NF), primarily mediated by prokaryotic diazotrophs, contributes nearly half of the bioavailable nitrogen input to the ocean (Gruber, 2008) and plays an important role in supporting carbon sequestration (Karl et al., 2012). Reliable assessments of marine NF are essential for understanding marine ecosystems and global biogeochemical cycles. However, substantial uncertainties still exist in current estimates of global marine NF, ranging from less than 100 Tg



30 N yr⁻¹ to over 200 Tg N yr⁻¹, largely due to limited spatiotemporal coverage of in situ measurements (Zhang et al., 2020; Shao et al., 2023).

NF rates are typically measured using ¹⁵N₂ tracer methods or the acetylene reduction assay. The ¹⁵N₂ tracer methods involve incubating seawater samples with ¹⁵N₂ gas and measuring the ¹⁵N/¹⁴N ratio of particulate nitrogen before and after incubation to determine NF rates (Capone and Montoya, 2001; Montoya et al., 1996). The acetylene reduction assay estimates
35 NF rates by converting measurements of acetylene reduction (Capone, 1993), although significant uncertainties are associated with the conversion factor (Mulholland et al., 2006; Wilson et al., 2012). Given the significant diel variations in NF rates among diazotrophic groups, a 24-hour incubation period is recommended for these methods (White et al., 2020). Such time-consuming and labor-intensive methods impose significant logistical constraints on NF assessments.

Additionally, estimating NF in the upper ocean, where it predominantly occurs, requires integrating NF measurements
40 across multiple depths (usually 3 to 7), typically from the surface to reference depths such as the deep chlorophyll maximum (DCM) (e.g., Hallstrøm et al., 2022), the base of the euphotic zone (often 1 % of surface irradiance) (e.g., Bonnet et al., 2009), or a fixed depth below the euphotic zone (e.g., Goebel et al., 2010). Such vertical profiling further increases the logistical challenges of NF assessments.

Previous studies have identified relationships between surface NF (SNF) and depth-integrated NF (INF) rates in limited
45 ocean regions (Großkopf et al., 2012; Yu et al., 2024; Tang et al., 2019). The recent release of a comprehensive global marine NF database (Shao et al., 2023) now enables possible analysis of this relationship from basin to global scales, as well as identification of the factors that govern it.

The vertical distribution pattern of NF, which shapes the SNF-INF relationship, varies in the global ocean. NF typically decreases with depth due to the light dependence of cyanobacterial diazotrophs (Luo et al., 2014; Zehr and Capone, 2020).
50 The atmospheric deposition of iron (Fe), a critical nutrient for NF enzymes, may also enhance surface NF (Moore et al., 2009; Fonseca-Batista et al., 2017). Additionally, upwelled waters with low phosphate-to-nitrate ratios may provide a nutrient environment that further suppresses subsurface NF (Singh et al., 2017; Hallstrøm et al., 2022). Meanwhile, other vertical patterns of NF have been observed. For instance, NF can remain relatively constant throughout the euphotic zone (e.g., Shiozaki et al., 2018; Sato et al., 2022), suggesting well-mixed resource dynamics for diazotrophs. Subsurface NF maxima
55 have also been reported (e.g., Shiozaki et al., 2014; Selden et al., 2019). It has been proposed that these subsurface NF maxima may be driven by factors such as photoinhibition near the surface (Shen et al., 2024a; Shen et al., 2024b), enriched subsurface phosphate concentrations (Voss et al., 2004), and mesoscale eddies that vertically transport diazotrophs (Hoerstmann et al., 2024). Temperature may also contribute to vertical NF patterns by affecting diazotrophic community structure (Moisander et al., 2010; Tang et al., 2020). Despite these insights, the factors controlling NF vertical distribution remain poorly understood.

60 In this study, we develop an empirical model to derive INF from SNF using historical measurements spanning diverse open ocean regimes. We also evaluate environmental controls on this relationship, particularly those governing vertical NF distribution patterns, to characterize its regional variability and operational constraints. Our model provides a tool for more efficient large-scale assessments of marine NF.



65 2 Methods

2.1 Empirical model for surface and depth-integrated N₂ fixation

In this study, we used in situ historical NF measurements compiled in the Global Oceanic Diazotroph Database version 2 (GODDv2) (Shao et al., 2023). Vertical NF profiles were selected using the following quality-control procedures. First, we retained only measurements that employed the ¹⁵N₂ assimilation-based methods with a 24-hour incubation to control for potential biases introduced by methodological differences and the diel cycles of diazotrophs (White et al., 2020). Second, our analysis was confined to the open ocean to preclude the complexities associated with coastal dynamics and shallow sampling depths. We operationally define the open ocean as areas where the bathymetry exceeds 250 m. Third, even though some measurements extended into deeper, dark ocean layers, we focused on the upper 250 m to ensure consistency across vertical profiles. Vertical profiles were included to evaluate INF if they contained measurements at three or more depths within the upper 250 m, assuming that scientists extended their measurements to depths with measurable NF activity. This assumption was later validated in our analysis. Finally, to analyze the relationship between SNF and INF, each selected profile was required to have at least one measurement within the upper 10 m to represent SNF (Fig. 1a). For profiles with multiple measurements in the upper 10 m, the mean NF in this layer was used as SNF.

After applying these criteria, 690 vertical NF profiles, comprising 3483 volumetric measurements, were selected. Two hundred eight volumetric NF measurements below detection limits were assigned a conservative value of 0.001 $\mu\text{mol N m}^{-3} \text{d}^{-1}$, as typical laboratory detection limits range from 0.01 to 1 $\mu\text{mol N m}^{-3} \text{d}^{-1}$ (Tanita et al., 2021; White et al., 2020). We analyzed the model sensitivity by testing to use 0, 0.01, and 0.1 $\mu\text{mol N m}^{-3} \text{d}^{-1}$ for NF rates below detection limits. Notably, there were no profiles in which all NF measurements fell below the detection limit. For each profile, we calculated INF using trapezoidal integration of NF from the surface to the deepest measurement not exceeding 250 m depth (Fig. 1b).

Both SNF and INF rates were approximately log-normally distributed. Thereby, linear regressions were applied to log-transformed SNF and INF to derive their relationship.

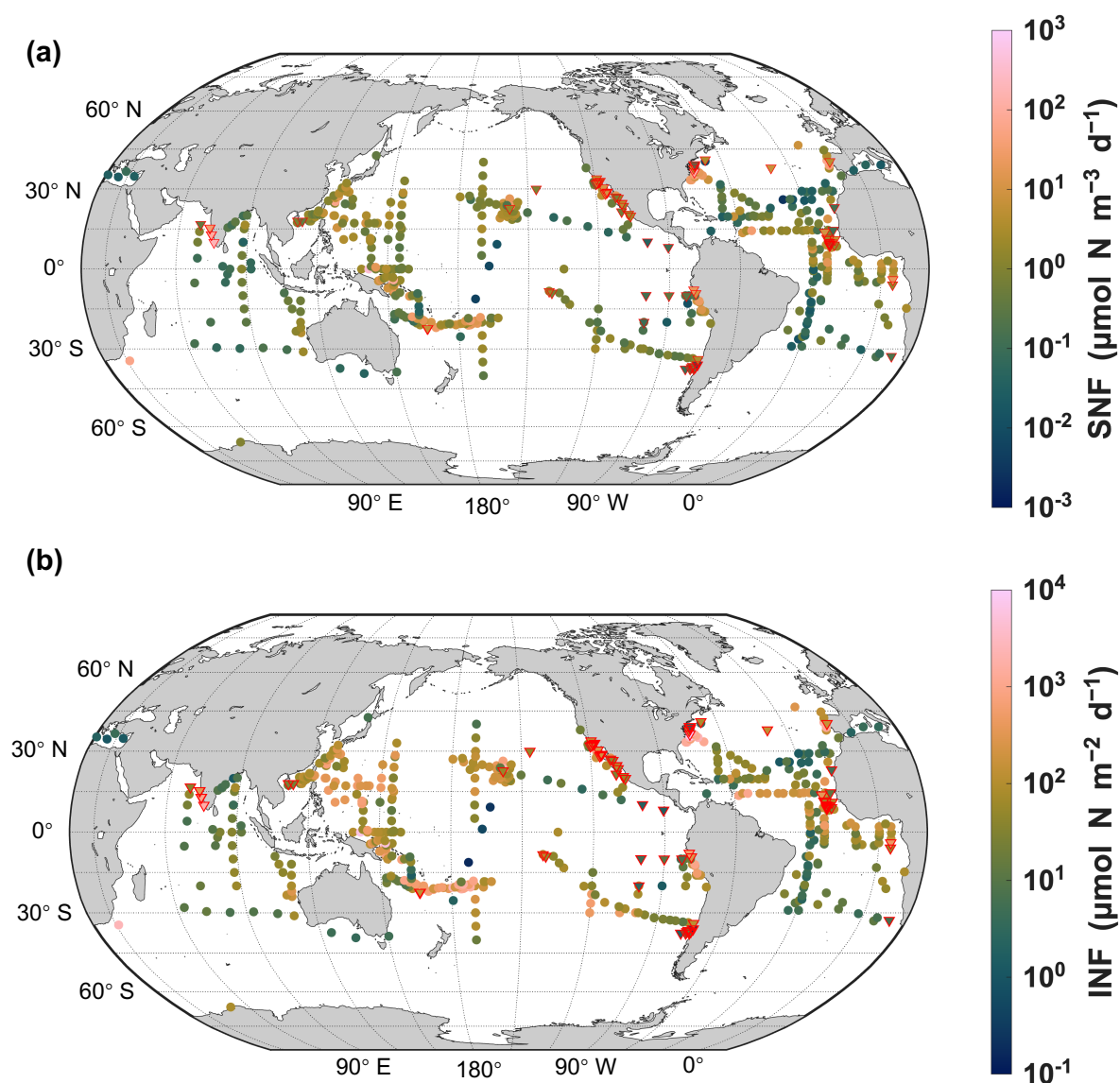


Figure 1. Global distribution of marine N_2 fixation observations. (a) Surface N_2 fixation (SNF) and (b) depth-integrated N_2 fixation (INF) rates were from GODDv2 and were quality-filtered. Red triangles denote excluded profiles where shallow sampling depths may cause potentially underestimated INF observations (see details in Sect. 3.1).



2.2 Environmental controls on the SNF-INF relationship

To explore potential environmental controls on vertical NF distribution patterns, we compared environmental parameter profiles (Table A1) with NF observations. Nitrate and phosphate concentrations were obtained from the World Ocean Atlas (WOA) 2023 (Garcia et al., 2024) (<https://www.ncei.noaa.gov/products/world-ocean-atlas>, last access: 19 September 2026). Where corresponding WOA data were unavailable, nearby values within 2° of the sampling location were averaged. Excess phosphate (P^*) was calculated using the Redfield ratio ($P^* = [\text{phosphate}] - [\text{nitrate}]/16$). Dissolved iron (Fe) concentrations were the outputs of the Community Earth System Model (CESM) (Kay et al., 2021). Mixed layer depth (MLD) was used to characterize upper-ocean mixing and was downloaded from Ifremer (<https://cerweb.ifremer.fr/mld>, last access: 19 September 2026) using a density threshold criterion of 0.03 kg m^{-3} (De Boyer Montégut et al., 2004). Given their log-normal distributions, the concentrations of nitrate, phosphate, dissolved Fe and mixed layer depth were log-transformed. Monthly temperature profiles were obtained from the Met Office Hadley Centre EN4 gridded ocean analyses (EN.4.2.2, bias corrections .g10) (Good et al., 2013; Gouretski and Reseghetti, 2010) (<https://www.metoffice.gov.uk/hadobs/en4/>, last access: 19 September 2026).

Surface photosynthetically active radiation (PAR) was sourced from 8-day MODIS-Aqua composites (<https://oceandata.sci.gsfc.nasa.gov/I3/>, last access: 19 September 2026). Vertical PAR profiles were generated by calculating the exponential attenuation of light with depth, with the attenuation coefficients of 490-nm light and PAR being estimated subsequently from 8-day composite MODIS-Aqua chlorophyll-*a* data using the method of Morel et al. (2007). The euphotic zone depth was determined as the depth where PAR reached 1 % of its surface value (Lee et al., 2007).

To explore the impact of mesoscale eddies on NF vertical distribution, we extracted the satellite-merged daily sea level anomaly (SLA) from Copernicus Marine Environment Monitoring Service (CMEMS) (Copernicus, 2018). Sampling points located inside closed SLA contours were identified as either cyclonic-eddy (negative SLA) or anticyclonic-eddy (positive SLA) samples.

To determine whether environmental factors could improve the predictive performance of the SNF-INF model, we experimentally incorporated surface environmental conditions, MLD and the euphotic zone depth (Table A1) as additional predictors for INF. We employed stepwise multiple linear regression (MLR) (Draper and Smith, 1998) to iteratively select significant predictors in ascending order of their p values, retaining only those with p value ≤ 0.05 . We also utilized a generalized additive model (GAM) in R with the “mgcv” package to capture possible nonlinear relationships between environmental factors and INF (Wood, 2017). Smoothing parameters in the GAM were estimated using the restricted maximum likelihood (REML) method, with the basis function number (k) set to 9 for each predictor (Wood et al., 2016). A double-penalization approach was applied for variable selection in the GAM, shrinking weak smooth terms towards zero (Marra and Wood, 2011).



2.3 Regional performance of the SNF-INF relationship

To identify regions of potential systematic biases using our model, we applied Getis-Ord G_i^* analysis to detect spatial clusters of strong positive or negative model residuals (Getis and Ord, 1992; Ord and Getis, 1995):

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{x} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \quad (1)$$

where n is the number of observations, x_j is the residual at location j , $\bar{x}$ and S are the global mean and standard deviation of the residuals, and $w_{i,j}$ represents the spatial weight between locations i and j . Spatial weights $w_{i,j}$ were first set to 1 for locations within a threshold distance and 0 otherwise, and were then normalized so that the sum of weights for each location equals 1 (Getis and Ord, 1992; Ord and Getis, 1995). The optimal threshold distance was determined to reflect the scale at which similar residuals are strongly clustered (Fig. A1): Global Moran's I z-scores were assessed across incremental threshold distances from 100 km to 2000 km, and all tested distances showed statistically significant spatial clustering ($p < 0.05$). The first significant peak in Global Moran's I z-scores occurred at 1000 km, which was selected to reflect the dominant local clustering scale (Memisoglu Baykal, 2025).

For each local G_i^* , the p value was evaluated using a Monte Carlo simulation with 1000 permutations (Rubinstein and Kroese, 2016), in which the residuals were randomly reassigned to build an empirical null distribution for significance testing. A high positive or negative G_i^* with $p < 0.05$ indicates a statistically significant hot spot or cold spot, respectively, meaning that a residual of high positive or negative value is surrounded by other high positive or negative residuals. The identified significant hot and cold spot points were then used to delineate regions of systematic overestimation and underestimation by constructing α -shape polygons, with the α parameter set to produce the convex hull (Edelsbrunner et al., 1983).

The Getis-Ord G_i^* analysis, including G_i^* statistics, corresponding p values, and Global Moran's I z-scores, was conducted using the R package “spdep” (Bivand and Wong, 2018).

2.4 Preliminary application of the empirical model

To demonstrate a preliminary application of our model, we predicted INF from historical SNF measurements without paired INF observations and evaluated these predictions against direct INF observations. The dataset included 938 quality-controlled INF measurements (note that some were excluded from the SNF-INF model calibration due to a lack of concomitant SNF data), plus 560 INF predictions derived from those SNF measurements without paired INF observations. Among these, 248 were directly reported INF measurements without corresponding profile data, of which 64 fell below the detection limit. We assigned a value of $0.1 \mu\text{mol N m}^{-2} \text{d}^{-1}$ to these data points (calculated as 100 meters multiplied by the assumed below-



detection NF rate of $0.001 \mu\text{mol N m}^{-3} \text{d}^{-1}$). Additionally, 46 SNF measurements were below the detection limit, for which we also predicted INF values of $0.1 \mu\text{mol N m}^{-2} \text{d}^{-1}$. All INF rates were $\log_{10}$ -transformed prior to analysis.

155

3 Results

3.1 Surface to depth-integrated N_2 fixation model

This study analyzed 690 NF profiles distributed across the tropical and subtropical regions in the Pacific, Atlantic, and Indian Oceans and the Mediterranean Sea (Figs. 1a, b). Each profile typically contained measurements at 3–7 depths above
160 250 m, with some profiles containing measurements at up to 12 depths (Fig. A2). SNF rates ranged from below detection limitations (assigned $10^{-3} \mu\text{mol N m}^{-3} \text{d}^{-1}$; see *Methods*) to $10^3 \mu\text{mol N m}^{-3} \text{d}^{-1}$ and INF ranged from 10^{-1} to $10^5 \mu\text{mol N m}^{-2} \text{d}^{-1}$, both exhibiting approximately log-normal distributions (Fig. A3).

A linear regression of $\log_{10}$ -transformed INF ($\mu\text{mol N m}^{-2} \text{d}^{-1}$) to $\log_{10}$ -transformed SNF ($\mu\text{mol N m}^{-3} \text{d}^{-1}$) yielded the following relationship (Fig. 2a):

165

$$\text{Log}_{10}\text{INF} = 1.81 + 0.84 \times \log_{10}\text{SNF} (R^2 = 0.81, \text{RMSE} = 0.32), \quad (2)$$

where RMSE denotes the root-mean-squared error. This relationship can be expressed as a power function:

170 $\text{INF} = 64 \times \text{SNF}^{0.84}. \quad (3)$

The power coefficient of 0.84 (< 1) indicates that the ratio of INF to SNF decreases as SNF increases. This implies that favorable conditions enhance NF rates disproportionately more in surface waters than in deeper layers.

To evaluate potential methodological biases, we conducted separate regression analyses for NF measurements derived
175 from the $^{15}\text{N}_2$ bubble and dissolution protocols (Mohr et al., 2010; Großkopf et al., 2012). The resulting regressions were nearly identical (Fig. A4), indicating that the SNF-INF relationship is not significantly influenced by the measurement protocol. We therefore pooled all $^{15}\text{N}_2$ -based NF measurements for subsequent analyses.

Our empirical regression model (Eq. 2) tended to overestimate INF in profiles where the deepest sample was collected at ≤ 50 m (Fig. 2b). We interpret this not as model failure, but as a limitation of the observational data: INF rates derived from
180 such shallow observational profiles are likely lower than true values by omitting significant NF activity in deeper layers. Consequently, even when the model correctly predicts true INF, its estimates will appear elevated relative to the shallow-profile values. To mitigate this bias, we excluded all profiles with maximum sampling depths ≤ 50 m. We made an exception



for two profiles where the deepest NF is less than 1 % of SNF (Singh et al., 2017), suggesting negligible NF beneath. This further quality control step removed 86 SNF-INF data pairs (remaining $n = 604$) (Fig. 2a). The updated regression equation
185 was:

$$\text{Log}_{10}\text{INF} = 1.85 + 0.84 \times \log_{10}\text{SNF} \quad (R^2 = 0.82, \text{RMSE} = 0.31). \quad (4)$$

This adjustment slightly improved the model performance and, more importantly, avoided systematic biases related to
190 sampling depths. The RMSE of 0.31 on a $\log_{10}$ scale corresponds to an uncertainty factor of 2.0. Sensitivity analysis showed that using different values (0, 0.001, 0.01, and $0.1 \mu\text{mol N m}^{-3} \text{d}^{-1}$) for below-detection NF rates had a negligible impact on the model performance (Table A2; Fig. A5).

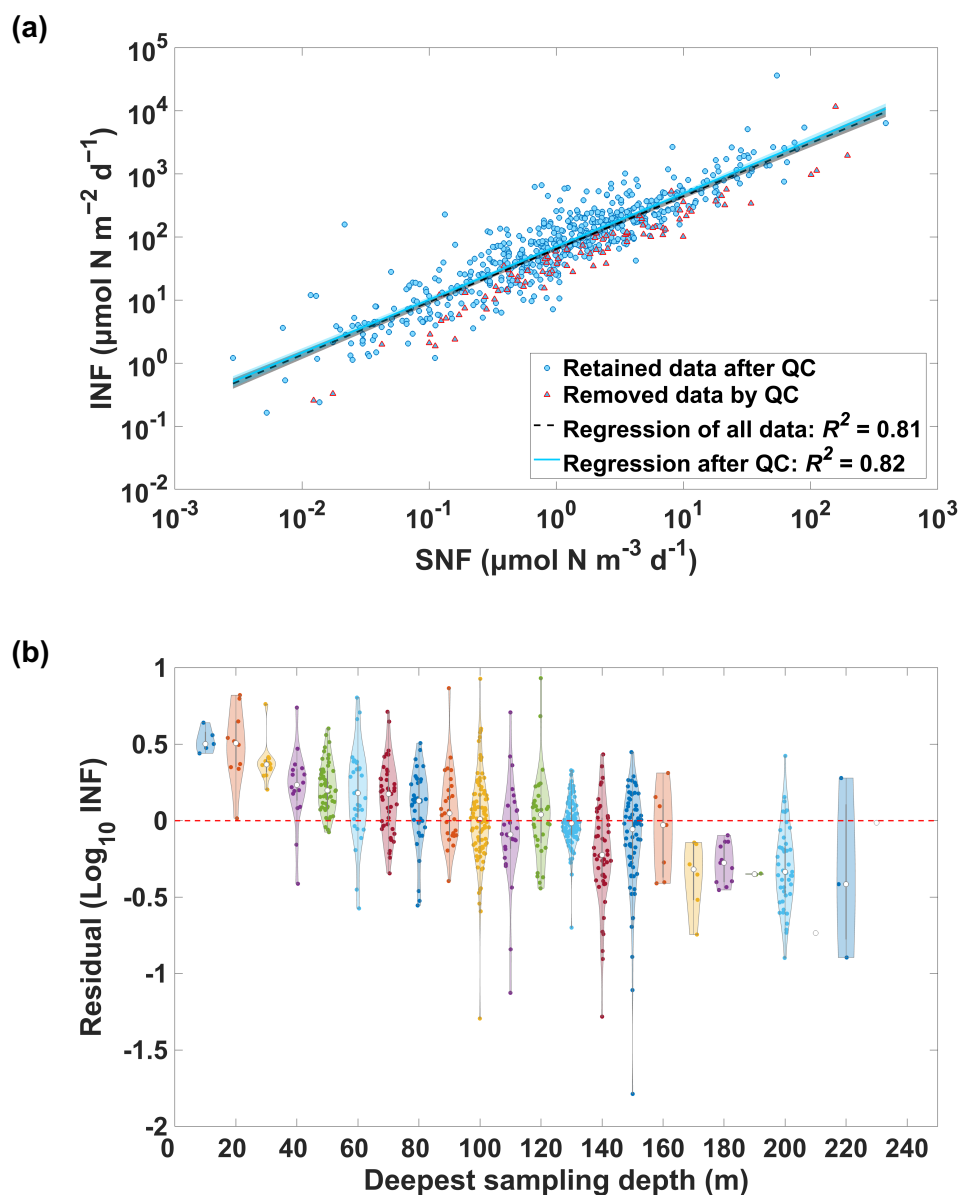
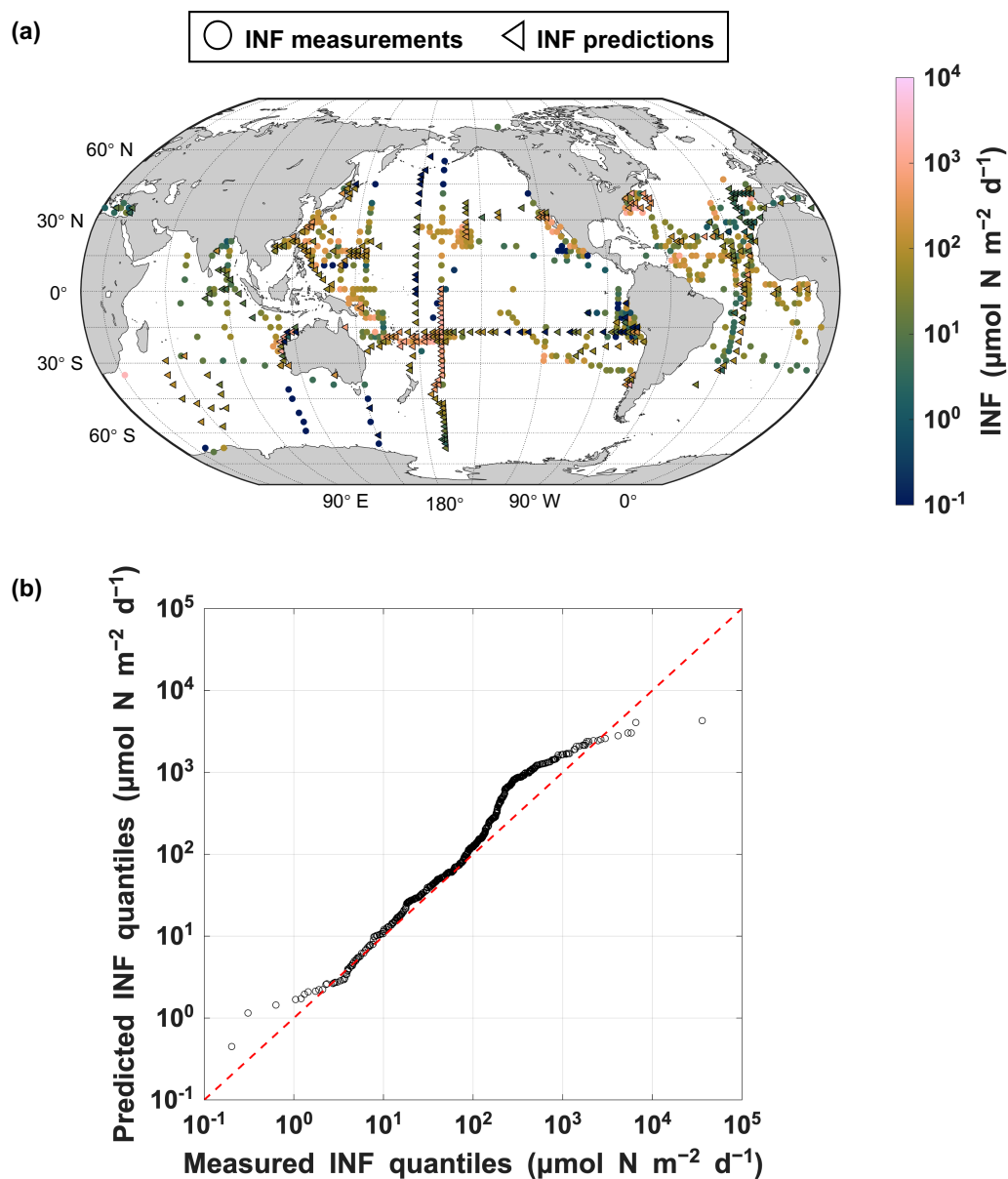


Figure 2. Relationship between observed SNF and INF rates. **(a)** Regression between log_{10} -transformed INF and SNF measurements. The black dashed regression includes all data points, while the blue solid regression line was for the data after a quality control (QC) procedure that excludes potentially underestimated INF observations (red triangles) from profiles with shallow maximum sampling depths (≤ 50 m). Shaded areas denote 95 % confidence intervals. **(b)** Violin plot of regression residuals (predicted minus observed INF) versus deepest sampling depth, demonstrating how shallow sampling may systematically bias INF estimates. The white dots in the violin plots indicate medians.



205 There have been 938 historical INF measurements distributed in the global open oceans (Fig. 3). The inclusion of 560
INF predictions derived from SNF measurements (using Eq. 4) substantially improved spatial coverage, particularly in the
Pacific and Southern Oceans (Fig. 3a). Additionally, the quantile–quantile plot indicated good agreement between INF
measurements and predictions (Fig. 3b). The proportions of values below detection limits were also comparable in the two
datasets (6.9 % and 8.2 % for INF measurements and predictions, respectively). These results suggest that INF predictions are
broadly consistent with direct measurements, demonstrating the value of our empirical model in leveraging existing SNF
measurements to improve the spatial coverage of INF.



210

Figure 3. Comparison of INF measurements and INF predictions. (a) Global distributions of direct INF measurements (circles) and model-predicted INF values (triangles) from SNF. INF rates are binned to $2^\circ \times 2^\circ$ for a clear demonstration. (b) Quantile-quantile (QQ) plot comparing distributions of measured and predicted INF values.



3.2 Vertical patterns of N₂ fixation profiles

215 In the upper 200 m, where the majority of NF measurements were concentrated, the mean NF exhibited an exponential decrease with depth (Fig. 4a). The NF profiles were categorized into three types with distinct vertical patterns: (1) vertically uniform, characterized by minimal variation in NF with depth (ratio of maximal to minimal NF < 2); (2) monotonic decrease, where NF generally declined with depth and there was no distinct subsurface maximum (SNF exceeded 90 % of the maximal NF); and (3) subsurface maximum, featuring a distinct subsurface peak in NF. This classification yielded 44 vertically uniform
220 profiles, 330 monotonically decreasing profiles, and 230 profiles with subsurface maxima (Fig. 4a).

When applied to each profile type, the model (Eq. 4, derived from all quality-controlled profiles) performed robustly. It yielded R^2 values of 0.89, 0.85, and 0.75 for the vertically uniform, monotonically decreasing, and subsurface-maximum profiles, respectively (Fig. 4b). The corresponding RMSE values were 0.25, 0.28, and 0.36, with biases of -0.12, 0.16, and -0.21. For most profiles, predicted values agreed with observations to within an order of magnitude (Fig. 4b). The only
225 exceptions were five profiles with subsurface maxima, where INF values were underestimated by more than one order of magnitude (Fig. 4b) due to pronounced subsurface NF maxima (Fig. A6). Specifically, we found a strong negative correlation ($r = -0.75$, $p < 0.01$) between the ratio of subsurface maximal NF to SNF and the model residuals (predicted minus observed INF; Eq. 4), indicating that the magnitude of INF underestimation increased with the intensity of subsurface NF maxima (Fig. A6). The potential mechanisms underlying these deviations are explored in Sect. 4.1.

230

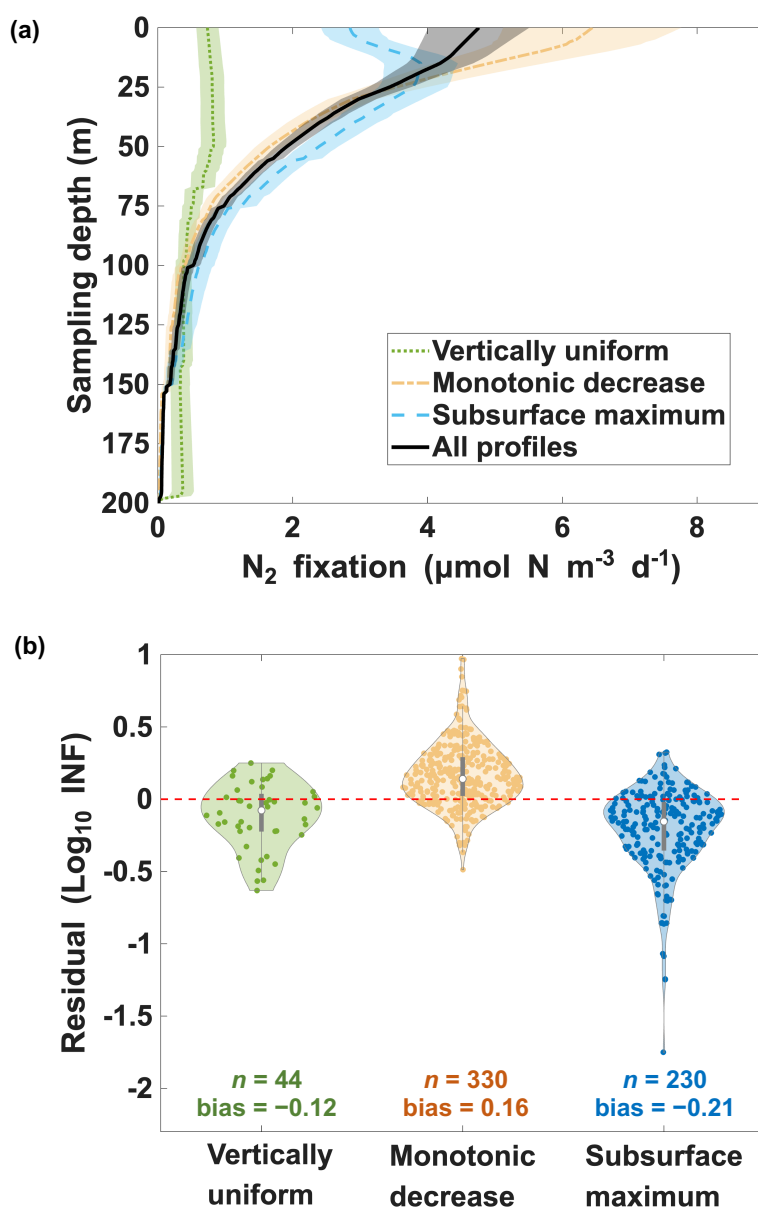


Figure 4. Types of vertical N_2 fixation profiles and model performance. **(a)** Mean vertical profiles for each type (colored lines) and the overall mean profiles (black line). Shaded areas represent one standard error. To better represent the general characteristics of NF profiles, two profiles with exceptionally high rates [23–28 $\mu\text{mol N m}^{-3} \text{d}^{-1}$ from Hoerstmann et al. (2024); 54–568 $\mu\text{mol N m}^{-3} \text{d}^{-1}$ from Selden et al. (2021)] were excluded from averaging. **(b)** Model residuals (predicted minus observed values) by profile type when applying the regression model (Eq. 4). The white dots and gray lines in the violin plots indicate medians and the ranges of 25th to 75th percentiles, respectively.



4 Discussion

240 4.1 Model applicability across N₂ fixation profile types

The robust performance of our empirical model across the global open oceans, despite variations in vertical NF distributions and NF levels, warrants detailed examination.

Overall, the mean profiles of both the monotonically decreasing and subsurface-maximum types closely aligned with the mean profile of all data (Fig. 4a). Specifically, most subsurface NF maxima occurred at shallow depths (< 50 m; Fig. A7) and
245 typically showed only moderate elevation above SNF rates (Fig. 4a). Therefore, most INF underestimations for subsurface-maximum profiles were generally minor, except for five profiles in which the subsurface maxima were pronounced (Fig. 4b; Fig. A6).

While vertically uniform profiles differ significantly from other types (Fig. 4a), our model still performed acceptably. To better understand this, we evaluate the magnitude of residuals when applying our model to vertically uniform profiles. Given
250 a SNF rate, we derive an ideal NF profile whose depth-integrated value matches the predicted INF (using Eq. 4; equivalent form: $INF = 71 \text{ SNF}^{0.84}$). By assuming the ideal NF profile decreases exponentially with depth z :

$$NF(z) = SNF \cdot e^{-\alpha z}, \quad (5)$$

255 where α is the exponential coefficient, the integration of this profile equals the predicted INF:

$$INF = \int_0^{z_{\max}} SNF \cdot e^{-\alpha z} dz = SNF \cdot \frac{1 - e^{-\alpha z_{\max}}}{\alpha} = 71 \text{ SNF}^{0.84}, \quad (6)$$

where $z_{\max}$ is the maximum sampling depth. By further setting $z_{\max} = 200$ m, α can be calculated at the given SNF, and the
260 ideal NF profiles (i.e., Eq. 6) can be derived. The results reveal that lower SNF rates lead to slower exponential decay of NF over depth, making integrated values approach those of vertically uniform profiles (Fig. 5a). Consequently, the generally lower SNF rates in vertically uniform profiles than the other two types (Fig. 5b) partially mitigated the INF underestimations using our empirical model. It is worth noting that, although NF rates in vertically uniform profiles are relatively low (Fig. 4a), they were all above detection limits.

265 These insights demonstrate how the model inherently compensates for variations in vertical NF distributions, explaining its broad applicability.

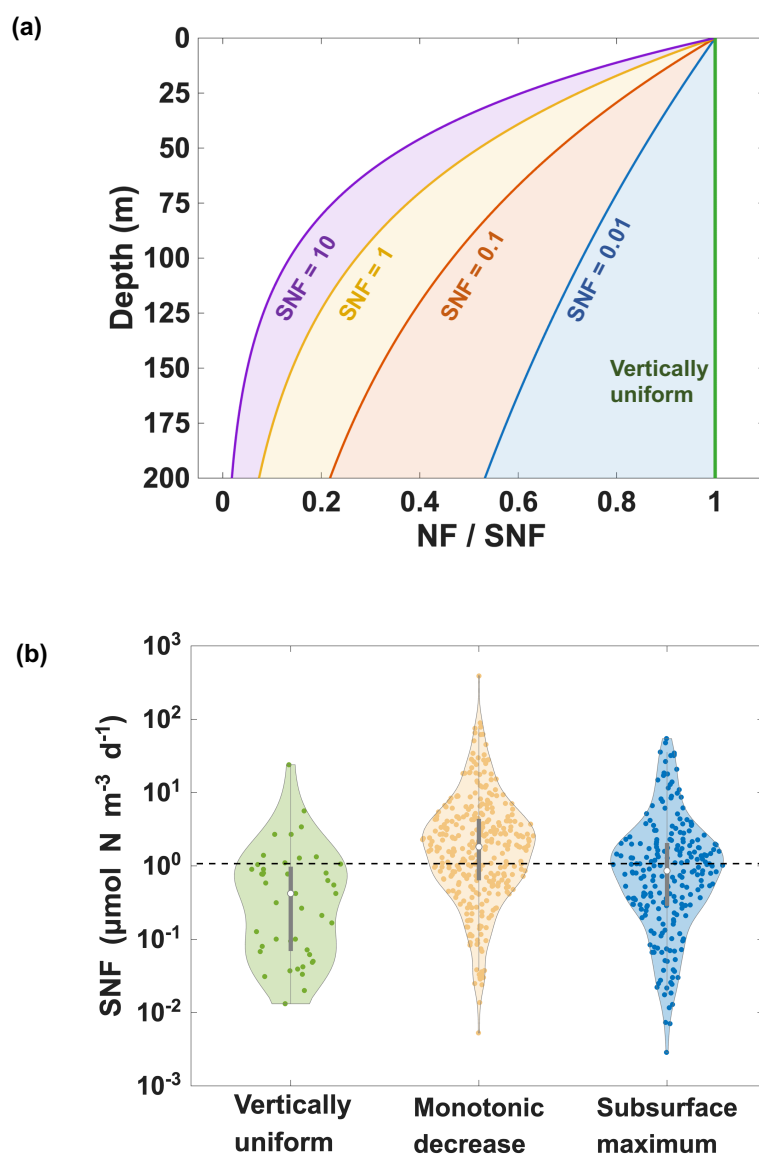


Figure 5. Model performance evaluation for vertically uniform N_2 fixation profiles. (a) The plotted idealized profiles represent cases where INF matches the model predictions (Eq. 4) at given SNF rates ($\mu\text{mol N m}^{-3} \text{d}^{-1}$) marked for each line. Hence, color-shaded areas indicate residuals when the model is applied to vertically uniform N_2 fixation profiles. To facilitate comparison, the x-axis shows SNF-normalized values. (b) Violin plots of SNF measurements for three profile types. Violin plot denotations follow Fig. 4, and the dashed black line represents the mean value for all data.



Our regression model exhibited predictable biases based on profile types, overestimating INF in monotonically decreasing
275 profiles ($\log_{10}$ -scale bias = 0.16) while underestimating it in both vertically uniform and subsurface-maximum profiles ($\log_{10}$ -
scale bias = -0.12 and -0.21, respectively) (Fig. 4b). These systematic errors highlight the need to evaluate here whether
environmental factors can predict the occurrence of these distinct profile types.

Profiles with monotonically decreasing NF had the highest dissolved Fe, DIP, and P^* concentrations; hence, their vertical
patterns are likely governed by the attenuation of irradiance and/or temperature with depth (Fig. 6). Vertically uniform profiles
280 exhibited low NF rates throughout the euphotic zone, likely constrained by the lowest concentrations of dissolved Fe, DIP,
and P^* (Fig. 6). MLD did not differ significantly among the three profile types (one-way ANOVA, $p = 0.84$; Fig. A8),
suggesting that vertical mixing did not drive the uniform NF distribution.

The key feature of subsurface-maximum profiles—suppressed surface NF—can be attributed to two regional mechanisms.
The first, and likely dominant, mechanism is limited Fe supplies from atmospheric deposition. Increases in Fe concentration
285 with depth in a majority of subsurface-maximum profiles ($n = 137$) indicate that deep water is the primary Fe source (Fig. 6a;
Fig. A9). The remaining subsurface-maximum profiles ($n = 93$), in which Fe does not increase with depth, may involve
photoinhibition (Fig. A9). This mechanism is evident in regions where unicellular diazotrophs often dominate surface waters,
such as the northwestern Pacific (Shen et al., 2024a; Jiang et al., 2025; Wen et al., 2022), northeastern Atlantic (Subramaniam
et al., 2013; Benavides and Voss, 2015) and eastern Indian Ocean off Australia (Sato et al., 2022). This supports previous
290 findings that light inhibition, rather than Fe scarcity, limits the surface activity of unicellular diazotrophs (Shen et al., 2024a;
Shen et al., 2024b). These analyses reveal regional differences in the mechanisms that reduce surface NF. Additionally,
temperature did not differ clearly among the three profile types, suggesting a limited role in distinguishing NF vertical patterns
(Fig. 6e). Overall, the factors controlling the magnitude of subsurface NF maxima remain unclear.

Our findings are based on environmental conditions with limited temporal resolution, reflecting the current sparsity of
295 observations. To test the influence of transient dynamics, we analyzed profile occurrence against daily satellite SLA data. This
analysis showed no significant link between profile types and specific mesoscale eddy fields (cyclonic, anticyclonic, or outside
eddies; χ^2 test, $p = 0.37$; Table A3). Unraveling these complex mechanisms will require future studies designed for coordinated
sampling of NF rates, diazotroph abundances, and vertical environmental profiles.

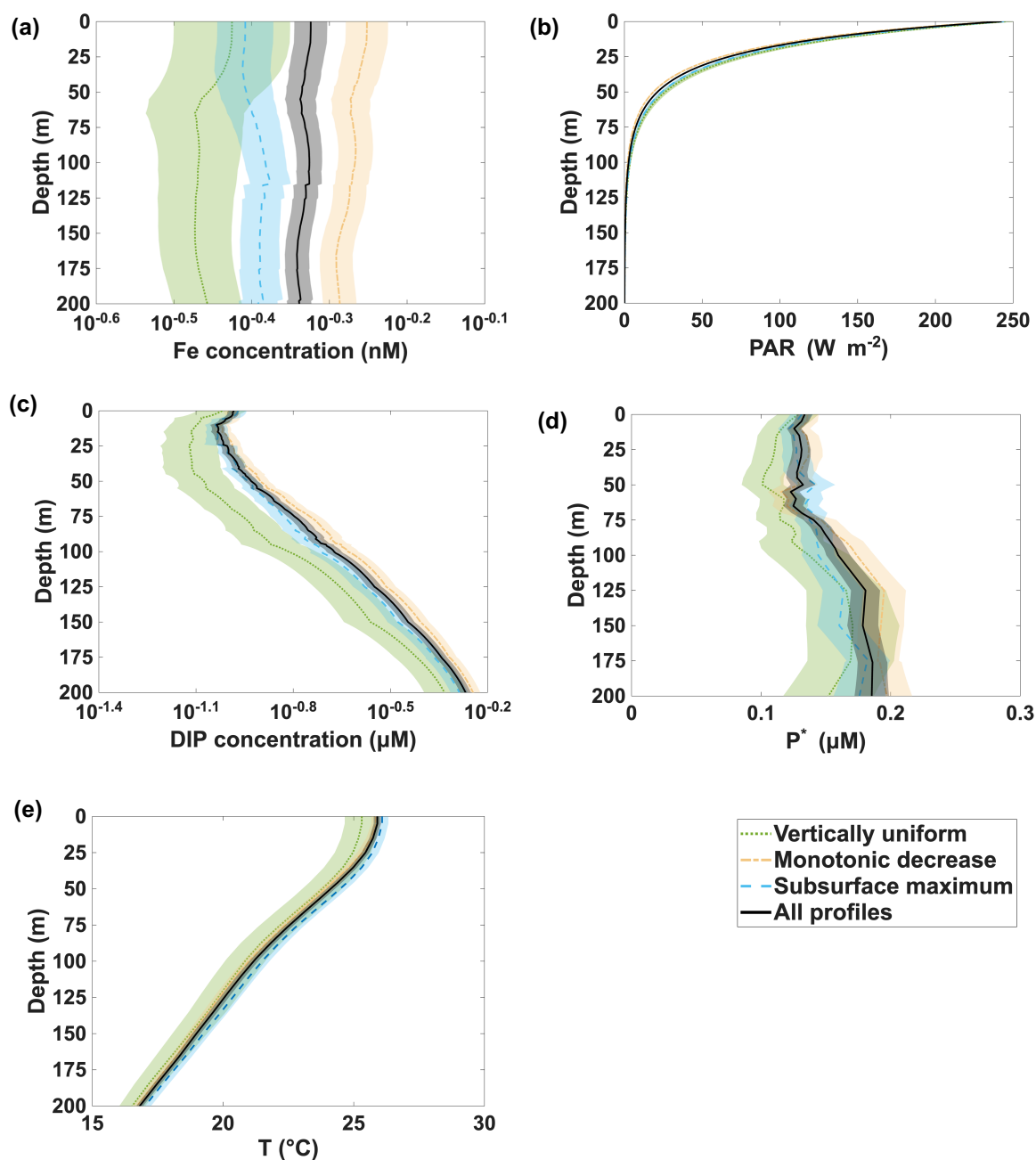
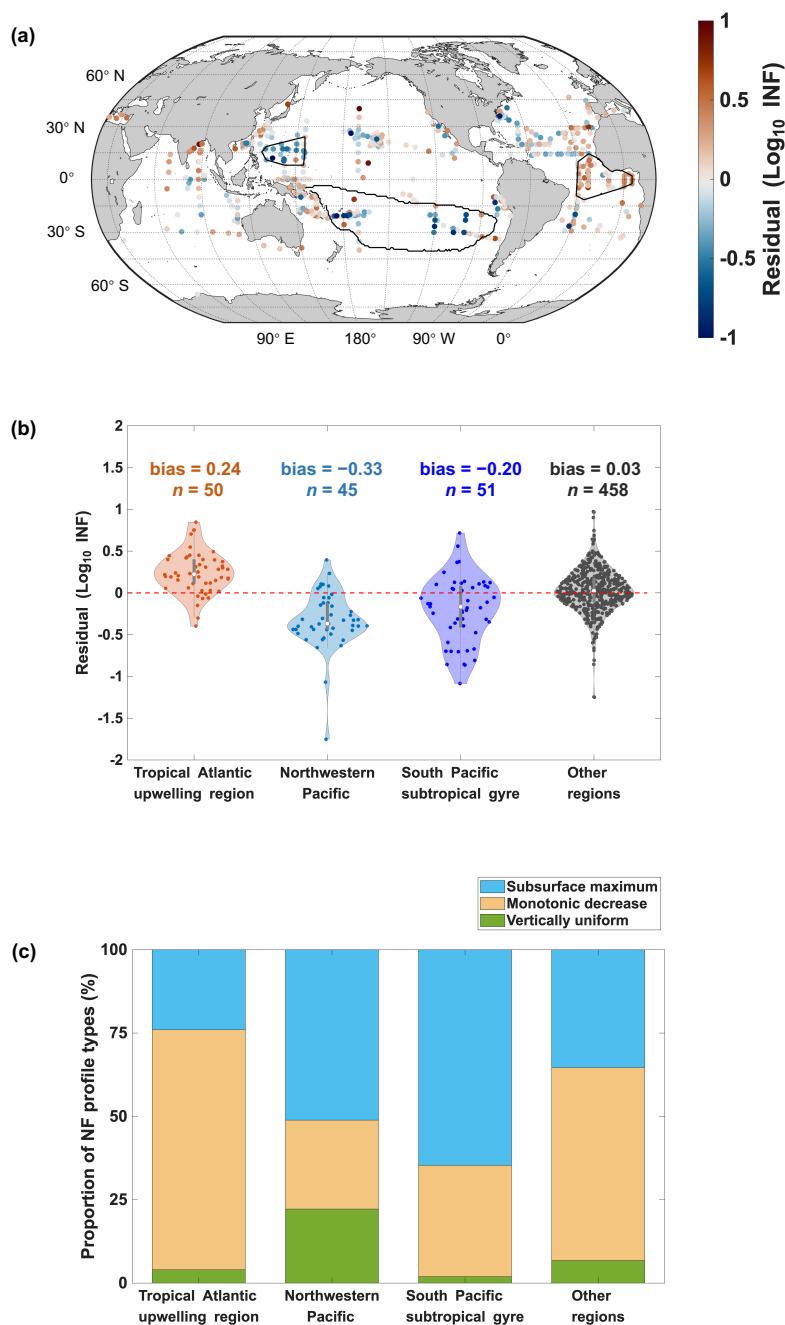


Figure 6. Vertical distribution of environmental conditions for vertically uniform ($n = 44$), monotonic decrease ($n = 330$), subsurface maximum ($n = 230$) profile types, and all profiles combined. Shaded areas represent one standard error.



4.2 Regional model performance

Our empirical model demonstrated robust global performance, with residuals distributed randomly across most of the ocean (Fig. 7). However, the Getis-Ord G_i^* analysis identified three regions with significant spatial clustering of residuals, indicating systematic biases: overestimation in the tropical Atlantic upwelling region and underestimation in the northwestern Pacific and the South Pacific subtropical gyre (SPSG) (Fig. A10). The SPSG was defined using the criterion of annual mean surface chlorophyll-a concentrations $\leq 0.1 \text{ mg m}^{-3}$ (Dai et al., 2023).



310

Figure 7. Regional model biases and associated NF profile types. The INF residuals (predicted minus observed values) of the regression model (Eq. 4) are shown in **(a)** the global ocean and **(b)** analyzed for three regions with systematic regional biases [solid outlines in panel (a)] and for the remaining regions. **(c)** Proportions of the three NF profile types across these regions.



315 In the tropical Atlantic upwelling region, the model consistently overestimated INF rates by a factor of approximately 1.7
(log₁₀-scale bias: +0.24; Figs. 7a, b). This bias is linked to the prevalence of monotonically decreasing NF profiles in this
region (72 % of profiles), with NF declining rapidly with depth, which diverges from the global average (Figs. 7c and 8a; Fig.
A11). We attribute this sharp vertical gradient to a combination of factors that promote surface NF while limiting subsurface
activity. While abundant Saharan dust supplies Fe that sustains elevated surface NF, primarily by *Trichodesmium* (Moore et
320 al., 2009; Schlosser et al., 2014), this Fe is rapidly scavenged in the subsurface, creating a steep vertical gradient that is
consistent with observed subsurface Fe minima in this region (Fig. 8b) (Conway and John, 2014; Fitzsimmons et al., 2015).
Concurrently, the same dust deposition may enhance light attenuation (Fig. 8c) by both direct mineral scattering (Ohde, 2016)
and indirectly by stimulating phytoplankton blooms (Dansie et al., 2022), which increase the biomass of light-scattering
particles in the water column.

325 Conversely, the model systematically underestimated INF rates by factors of approximately 2.1 and 1.6 in the
northwestern Pacific and the SPSG (log₁₀-scale bias: -0.33 and -0.20, respectively; Figs. 7a, b). In these regions, NF declines
more gradually with depth, and subsurface maxima are common (51 % and 65 % of profiles, respectively) (Figs. 7c and 8a;
Fig. A11). In the northwestern Pacific, the model's underestimation is driven by subsurface NF maxima (Fig. 8a), which, as
discussed above, we ascribe to the photoinhibition of unicellular diazotrophs (Shen et al., 2024a; Shen et al., 2024b). In the
330 SPSG, atmospheric Fe deposition is among the lowest reported for the open ocean (Wagener et al., 2008), and surface dissolved
Fe concentrations are generally below 0.2 nM (Tanita et al., 2021), potentially constraining surface NF (Fig. 8a). Meanwhile,
substantial NF rates (often close to 1 $\mu\text{mol N m}^{-3} \text{d}^{-1}$) frequently persist near or even below the base of the euphotic zone in
the SPSG (Fig. A12). This deepwater NF activity appears to be supported by deep Fe sources (Fig. 8b), including shallow
hydrothermal inputs observed in the west (Guieu et al., 2018; Bonnet et al., 2023) and benthic sources and westward plumes
335 from the Peruvian slope in the east (Kittu et al., 2023; Resing et al., 2015).

Collectively, these three regions aligned with the distinct areas where particular NF profile types prevailed, exemplifying
how the global mechanisms discussed above operate at more localized scales.

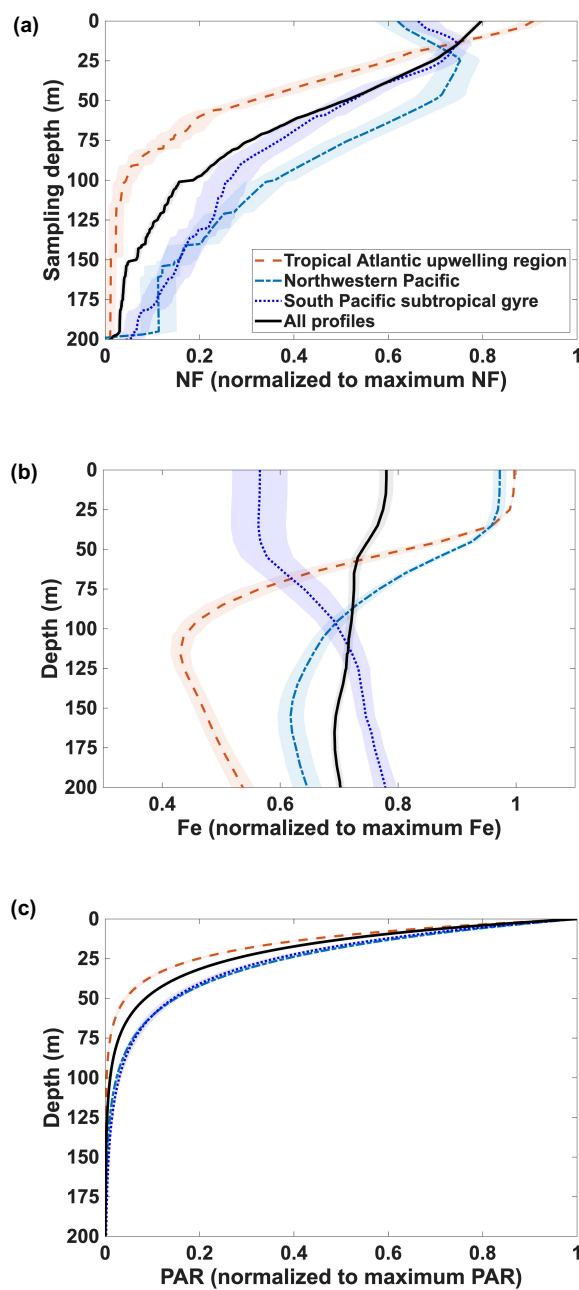


Figure 8. Regional profile comparisons of NF, Fe concentrations, and PAR. Mean profiles of (a) NF rates, (b) dissolved Fe concentrations, and (c) PAR in three regions (marked by solid outlines in Fig. 7a) and for all profiles. NF, Fe, and PAR are normalized to the maximum level in each profile before their means and errors are calculated. Shaded areas represent one standard error.



For these three regions, we revised the INF predictions by simply subtracting the corresponding biases, which improved the fit to the observed INF rates and reduced RMSE from 0.35 to 0.25, 0.47 to 0.34, and 0.43 to 0.39, respectively (Fig. A13). This indicates that the bias-corrected predictions are practically useful (Alexander et al., 2015). Whether to apply these bias corrections, however, is left to the discretion of the users, as the sample sizes are relatively small ($n = 50, 45$, and 51 in the three regions, respectively).

Finally, to assess whether environmental variables could enhance model performance, we conducted stepwise MLR and GAM incorporating more potential predictors, including surface environmental conditions, MLD, and euphotic zone (EZ) depth (Table A1). The MLR model retained only the EZ depth as an additional predictor and had better performance (Table A4):

$$\text{Log}_{10}\text{INF} = 1.43 + 0.87 \times \text{log}_{10}\text{SNF} + 0.0047 \times \text{EZ} \quad (R^2 = 0.85, \text{RMSE} = 0.28). \quad (7)$$

Similarly, the GAM identified EZ depth as a highly significant smooth term ($p < 2 \times 10^{-16}$), with weaker statistical support for log_{10}Fe and temperature (both $p \approx 0.04$) ($R^2 = 0.85$, $\text{RMSE} = 0.28$; Fig. A14). The similar performance between the linear MLR model and the non-linear GAM supports the sufficiency of using MLR in our application.

Given the additional EZ measurements required, the usage of the updated MLR model (Eq. 7) is also at the users' discretion. Nevertheless, including the EZ depth reduced the model's underestimations, particularly in the SPSG and northwestern Pacific (Table A4). This improvement is attributed to the deeper EZ in these regions: The positive coefficient for EZ in the model (Eq. 7) indicates that greater light penetration in these oligotrophic waters sustains substantial NF at deeper layers.

5 Summary and future directions

Our study establishes a robust empirical model to derive depth-integrated N_2 fixation rates within the euphotic zone from surface N_2 fixation, providing a practical tool for global ocean assessment. While previous regional studies have examined this relationship, our global-scale analysis benefits from substantially expanded datasets, stricter quality-control criteria, and comparable or superior model performance despite broader geographical applicability. The model significantly reduces sampling effort while maintaining accuracy across most open ocean regions. However, the model should be applied with caution, or be corrected with regional biases (see above), in regions with (1) unique nutrient supply dynamics and/or light regimes (the tropical Atlantic upwelling region and the South Pacific subtropical gyre), and (2) photoinhibition of surface N_2 fixation when unicellular diazotrophs dominate (the northwestern Pacific).

Incorporating the model-derived depth-integrated N_2 fixation improved the spatial coverage of N_2 fixation rates, suggesting its applicability for large-scale or time-series observations, where traditional profile-based methods remain



logistically constrained. Nevertheless, the spatial coverage is still sparse. Seasonal variability in N_2 fixation profiles also remains poorly constrained due to uneven temporal sampling. Future efforts should prioritize obtaining in situ measurements of N_2 fixation rates, diazotroph abundances, and environmental factors, especially in underrepresented regions, such as the Southern Ocean, subtropical and subpolar North Atlantic, the equatorial Pacific, and the western Indian Ocean (Fig. 3a). These new measurements can contribute to refining model parameterization and better elucidate the complex mechanisms controlling the vertical distributions of N_2 fixation.

The utility of our empirical model can be further enhanced by ongoing developments of high-frequency shipboard underway N_2 fixation measurements. For instance, a method based on acetylene reduction assays has already been developed and applied to continuous in situ surface N_2 fixation measurements during a research cruise across the North Atlantic (Cassar et al., 2018; Tang et al., 2020). Similar approaches using more reliable $^{15}N_2$ uptake techniques may also become available in the future (Zehr and Capone, 2024), further expanding the potential for efficient and widespread surface and depth-integrated N_2 fixation assessments and improving our understanding of the biogeochemical role of N_2 fixation.



Appendix A

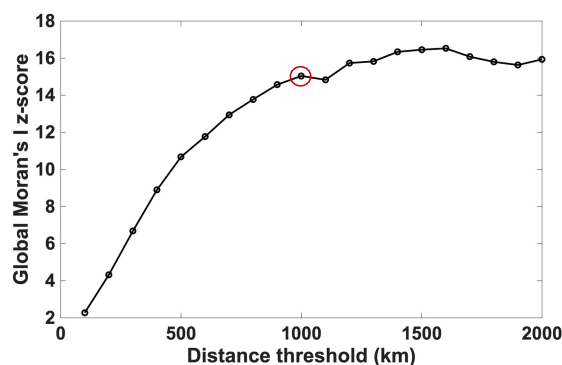


Figure A1. Global Moran's I z-scores for the model residuals (predicted minus observed INF; Eq. 4) across increasing fixed distance bands, with the optimal threshold distance (1000 km) marked with a red circle.

390

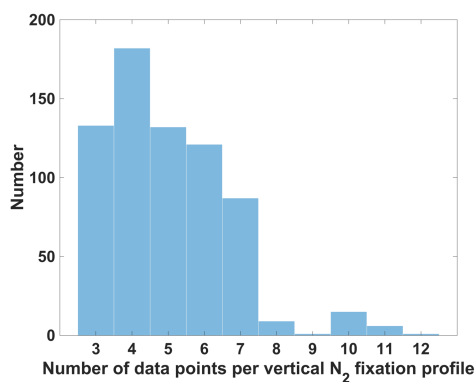


Figure A2. Distribution of data point counts per vertical NF profile.

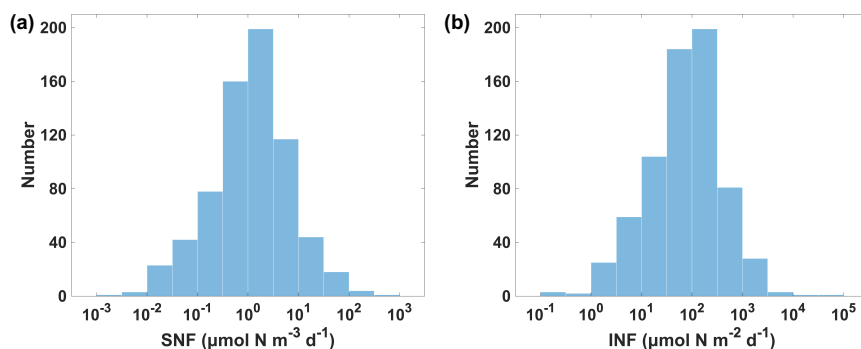
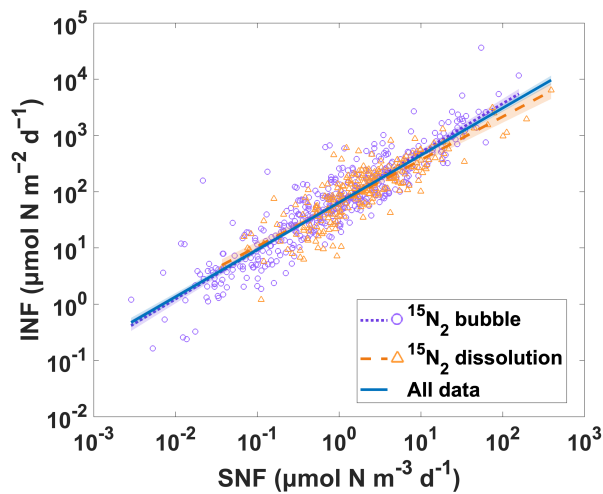


Figure A3. Histograms of (a) surface and (b) depth-integrated NF measurements.



395

Figure A4. Regression between INF and SNF for data measured using the $^{15}\text{N}_2$ bubble protocol, the $^{15}\text{N}_2$ dissolution protocol, and all data combined. The shaded areas represent 95 % confidence intervals of the regressions.

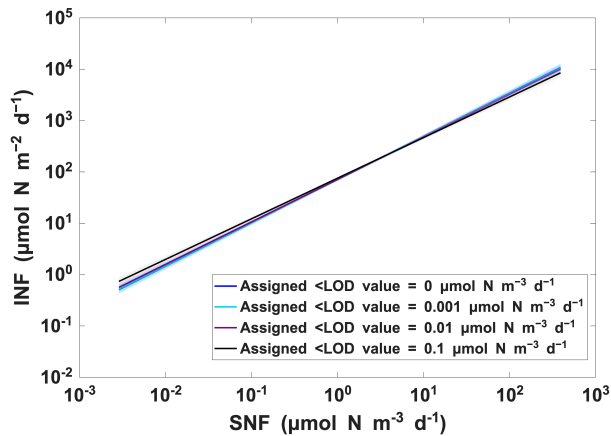


Figure A5. Comparison of the quality-controlled regression between SNF and INF under alternative assignments (0, 0.001, 0.01, and 0.1 $\mu\text{mol N m}^{-3} \text{d}^{-1}$) for NF rates below detection limits. Shaded areas denote 95 % confidence intervals.

400

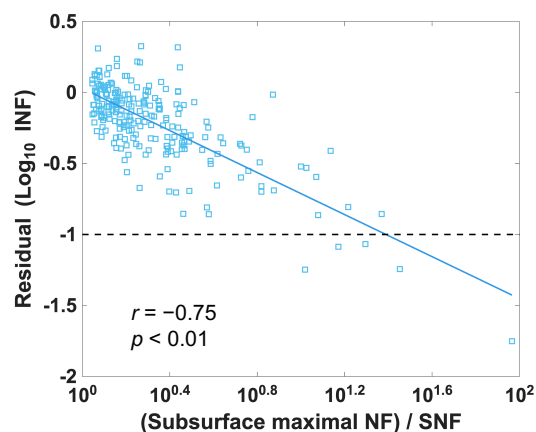
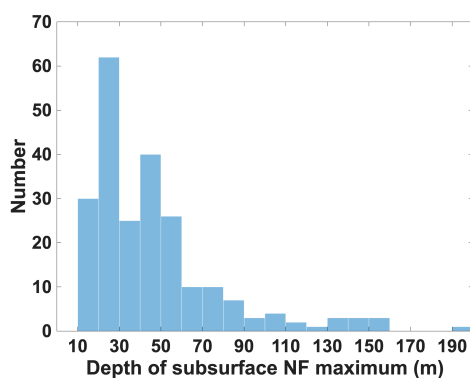


Figure A6. Regression between residuals (predicted minus observed INF values using Eq. 4) and the ratios of subsurface maximal NF to SNF.



405 **Figure A7.** Histogram of the depths of subsurface NF maxima.

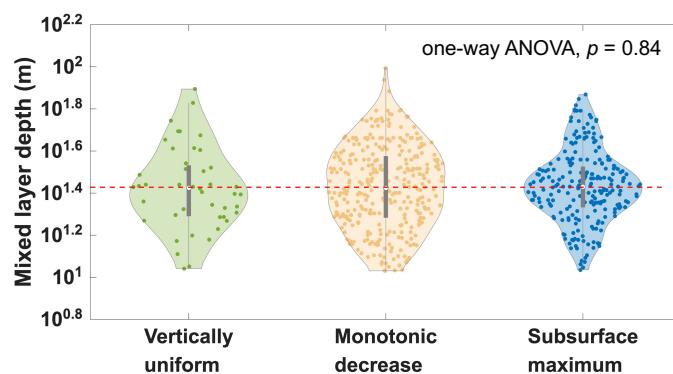
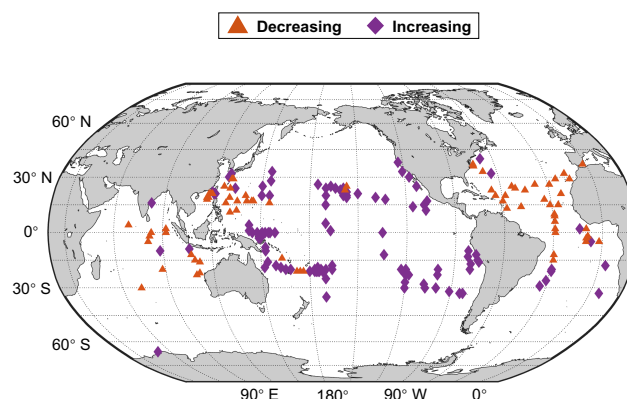


Figure A8. Violin plots of mixed layer depth for three profile types. Violin plot denotations follow Fig. 4, and the dashed line represents the mean value for all data.



410 **Figure A9.** Comparison of vertical dissolved Fe patterns (decreasing or increasing with depth) in subsurface-maximum NF profiles.

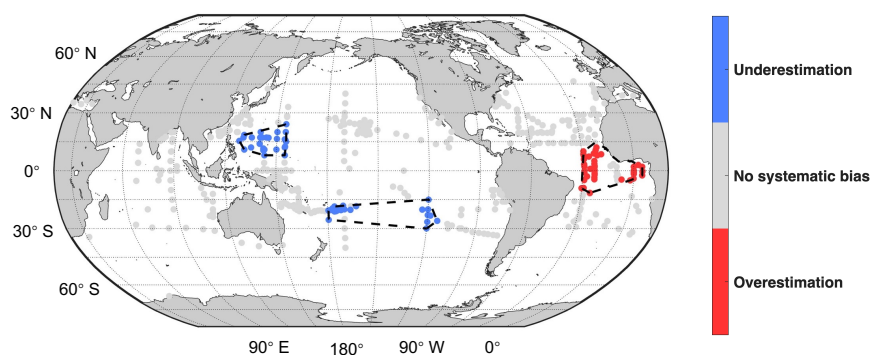
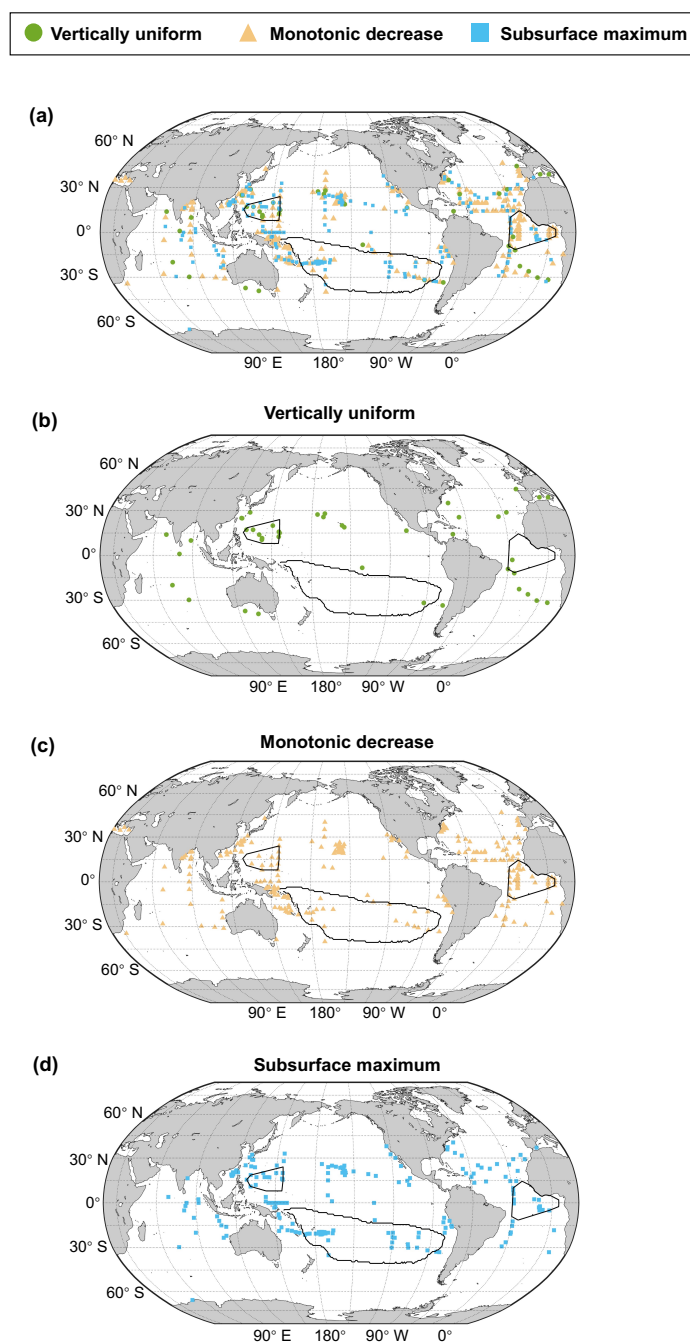


Figure A10. Regions with systematic biases identified by Getis-Ord G_i^* analysis. Dashed outlines indicate regions of significant overestimation or underestimation ($p < 0.05$), and areas without significant clustering are shown in gray.



415 **Figure A11.** Global distribution of NF profile types. (a) All profile types; (b–d) separate plots for each profile type.

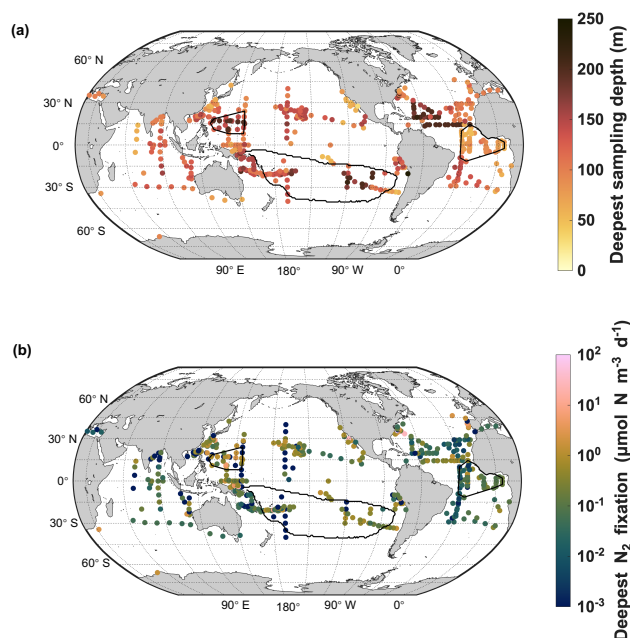


Figure A12. Global distribution of **(a)** deepest sampling depth and **(b)** deepest rates in observational NF profiles.

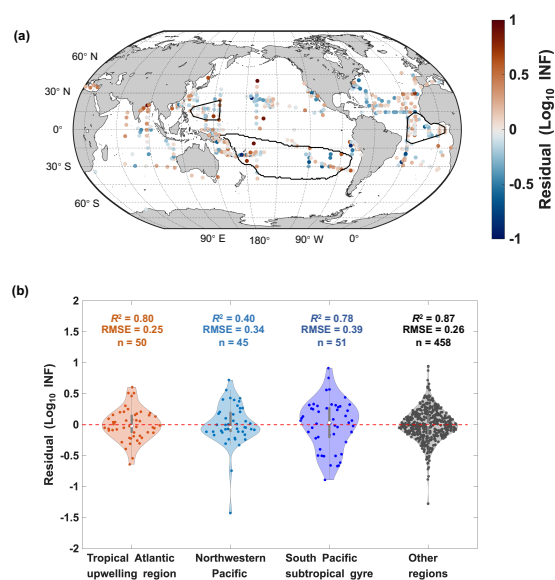


Figure A13. Performance of INF predictions revised by subtracting region-specific biases. Updated residuals (predicted minus observed values) are shown as **(a)** global distribution and **(b)** violin plots. The three specific regions analyzed in panel **(b)** are marked in panel **(a)** using solid outlines.

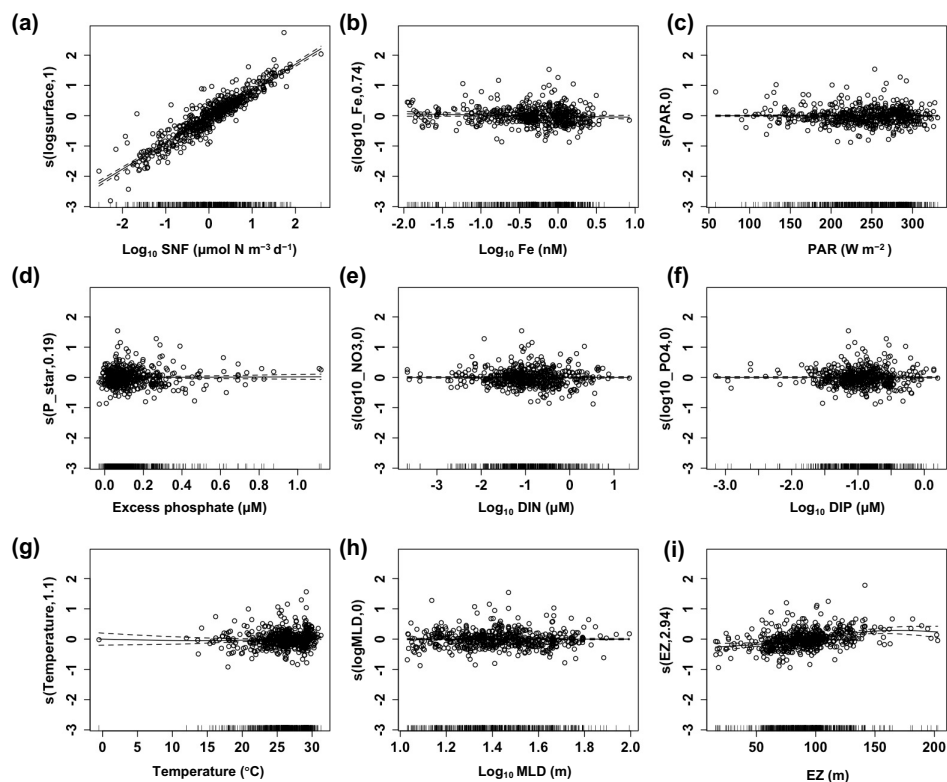


Figure A14. Partial effects of SNF and environmental predictors on INF estimated by a generalized additive model (GAM). Smooth functions (solid lines) and their 95 % confidence intervals (dashed lines) are shown for each predictor. Circles represent partial residuals, and tick marks along the x-axis indicate the distribution of observations. Numbers in parentheses on the y-axis denote the effective degrees of freedom of the smooth terms, which reflect the complexity of each partial effect.



430 **Table A1.** Data sources of environmental parameters.

Data description	Symbol	Source	Spatial resolution	Temporal resolution
Dissolved iron (nM)	Fe	CESM	1°	Monthly climatology
Nitrate (μM)	DIN	World Ocean Atlas 2023	1°	Monthly climatology
Phosphate (μM)	DIP	World Ocean Atlas 2023	1°	Monthly climatology
Excess phosphate (μM)	P*	DIP – DIN / 16	1°	Monthly climatology
Mixed layer depth (m)	MLD	Ifremer	1°	Monthly climatology
Temperature (°C)	T	EN4	1°	Monthly
Surface photosynthetically active radiation (W m ⁻²)	PAR	MODIS-Aqua	0.083°	8 days
Euphotic zone depth (m)	EZ	1 % surface PAR depth	0.083°	8 days
Sea level anomaly (m)	SLA	CMEMS	0.25°	Daily

Table A2. Sensitivity analysis results for alternative assignments of NF rates below detection limits.

Assigned < LOD value (μmol N m ⁻³ d ⁻¹)	Regression after quality control	R ²	RMSE
0	Log ₁₀ INF = 1.85 + 0.84 × log ₁₀ SNF	0.82	0.31
0.001	Log ₁₀ INF = 1.85 + 0.84 × log ₁₀ SNF	0.82	0.31
0.01	Log ₁₀ INF = 1.86 + 0.83 × log ₁₀ SNF	0.82	0.30
0.1	Log ₁₀ INF = 1.88 + 0.79 × log ₁₀ SNF	0.82	0.29



435 **Table A3.** Number of NF observational profiles inside and outside mesoscale eddies.

	Cyclonic eddies	Anticyclonic eddies	Outside eddies
Well-mixed profiles	10	6	28
Monotonic decreasing profiles	57	57	198
Subsurface-maximum profiles	33	50	165

Table A4. Comparison of model performance by using the SNF alone (Eq. 4) and both SNF and EZ (Eq. 7) as predictors. The comparisons are conducted for different NF profile types and specific oceanic regions.

	Predictor: SNF		Predictor: SNF and EZ	
	RMSE	bias	RMSE	bias
Vertically uniform profiles	0.25	−0.12	0.20	−0.08
Monotonic increasing profiles	0.28	0.16	0.25	0.15
Subsurface-maximum profiles	0.36	−0.21	0.33	−0.19
Tropical Atlantic upwelling region	0.35	0.24	0.30	0.15
Northwestern Pacific	0.47	−0.33	0.37	−0.18
South Pacific subtropical gyre	0.43	−0.20	0.38	−0.06

440 **Data availability**

The datasets generated in this study, based on the Global Oceanic Diazotroph Database (version 2) (Shao et al., 2023), include: (1) 690 vertical profiles of N₂ fixation and corresponding environmental factors used for model development; (2) 560 INF predictions derived by applying our empirical model to historical SNF measurements. These datasets are available at the Figshare repository via <https://figshare.com/s/5e9ebba17fcc614194af>, with the Creative Commons Attribution 4.0 International License (CC BY 4.0). Environmental data used in this study are publicly available, including surface PAR (<https://oceandata.sci.gsfc.nasa.gov/l3/>), dissolved Fe concentrations (Kay et al., 2021), World Oceans Atlas 2023 data (Garcia et al., 2024) (<https://www.ncei.noaa.gov/products/world-ocean-atlas>), MLD (De Boyer Montégut et al., 2004) (<https://cerweb.ifremer.fr/mld>), temperature (Good et al., 2013) (<https://www.metoffice.gov.uk/hadobs/en4/>) and SLA data (Copernicus, 2018).

450



Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

Wei-Li Chen: methodology, formal analysis, writing – original draft, writing – review & editing. Zhibo Shao: formal analysis, writing – review & editing. Yangchun Xu: formal analysis, writing – review & editing. Ya-Wei Luo: funding acquisition, conceptualization, supervision, methodology, formal analysis, writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the scientists and crew for sampling the data used in this study. We also thank Hua Wang, Weicheng Luo, Lice Wang, and Yuhong Huang from XMU for their efforts in data collection and computation support. Generative artificial intelligence tools were used only to improve language clarity.

Financial support

This study was partially supported by the National Natural Science Foundation of China (grant nos. 42376140 and 42076153) and the PhD Fellowship of the State Key Laboratory of Marine Environmental Science at Xiamen University.

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Alexander, D. L. J., Tropsha, A., and Winkler, D. A.: Beware of R^2 : simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR Models, *Journal of Chemical Information and Modeling*, 55, 1316-1322, 10.1021/acs.jcim.5b00206, 2015.



- Benavides, M. and Voss, M.: Five decades of N₂ fixation research in the North Atlantic Ocean, *Frontiers in Marine Science*, 2, 40, 10.3389/fmars.2015.00040, 2015.
- 475 Bivand, R. S. and Wong, D. W. S.: Comparing implementations of global and local indicators of spatial association, *TEST*, 27, 716-748, 10.1007/s11749-018-0599-x, 2018.
- Bonnet, S., Biegala, I. C., Dutrieux, P., Slemmons, L. O., and Capone, D. G.: Nitrogen fixation in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class distribution, and biogeochemical significance, *Global Biogeochemical Cycles*, 23, GB3012, 10.1029/2008gb003439, 2009.
- 480 Bonnet, S., Guieu, C., Taillandier, V., Boulart, C., Bouruet-Aubertot, P., Gazeau, F., Scalabrin, C., Bressac, M., Knapp, A. N., and Cuypers, Y.: Natural iron fertilization by shallow hydrothermal sources fuels diazotroph blooms in the ocean, *Science*, 380, 812-817, 10.1126/science.abq4654, 2023.
- Capone, D. G.: Determination of nitrogenase activity in aquatic samples using the acetylene reduction procedure, in: *Handbook of methods in aquatic microbial ecology*, edited by: Kemp, P. F., Sherr, B. F., Sherr, E. B., and Cole, J. J., Lewis Publishers, 485 Boca Raton, FL, 621-631, 1993.
- Capone, D. G. and Montoya, J. P.: Nitrogen fixation and denitrification, in: *Methods in Microbiology*, Academic Press, 501-515, 10.1016/S0580-9517(01)30060-0, 2001.
- Cassar, N., Tang, W., Gabathuler, H., and Huang, K.: Method for high frequency underway N₂ fixation measurements: flow-through incubation acetylene reduction assays by cavity ring down laser absorption spectroscopy (FARACAS), *Analytical chemistry*, 90, 2839-2851, 10.1021/acs.analchem.7b04977, 2018.
- 490 Conway, T. M. and John, S. G.: Quantification of dissolved iron sources to the North Atlantic Ocean, *Nature*, 511, 212-215, 10.1038/nature13482, 2014.
- Copernicus: Sea level gridded data from satellite observations for the global ocean from 1993 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [dataset], 10.24381/cds.4c328c78, 2018.
- 495 Dai, M., Luo, Y. W., Achterberg, E. P., Browning, T. J., Cai, Y., Cao, Z., Chai, F., Chen, B., Church, M. J., Ci, D., Du, C., Gao, K., Guo, X., Hu, Z., Kao, S. J., Laws, E. A., Lee, Z., Lin, H., Liu, Q., Liu, X., Luo, W., Meng, F., Shang, S., Shi, D., Saito, H., Song, L., Wan, X. S., Wang, Y., Wang, W. L., Wen, Z., Xiu, P., Zhang, J., Zhang, R., and Zhou, K.: Upper Ocean Biogeochemistry of the Oligotrophic North Pacific Subtropical Gyre: From Nutrient Sources to Carbon Export, *Reviews of Geophysics*, 61, 10.1029/2022rg000800, 2023.
- 500 Dansie, A. P., Thomas, D. S., Wiggs, G. F., Baddock, M., and Ashpole, I.: Plumes and blooms—Locally-sourced Fe-rich aeolian mineral dust drives phytoplankton growth off southwest Africa, *Science of the Total Environment*, 829, 154562, 10.1016/j.scitotenv.2022.154562, 2022.
- de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *Journal of Geophysical Research: Oceans*, 109, 505 10.1029/2004JC002378, 2004.
- Draper, N. R. and Smith, H.: *Applied regression analysis*, John Wiley & Sons, Hoboken, NJ1998.



- Edelsbrunner, H., Kirkpatrick, D., and Seidel, R.: On the shape of a set of points in the plane, *IEEE Transactions on information theory*, 29, 551-559, 10.1109/TIT.1983.1056714, 1983.
- Fitzsimmons, J. N., Carrasco, G. G., Wu, J., Roshan, S., Hatta, M., Measures, C. I., Conway, T. M., John, S. G., and Boyle, E. A.: Partitioning of dissolved iron and iron isotopes into soluble and colloidal phases along the GA03 GEOTRACES North Atlantic Transect, *Deep Sea Research Part II: Topical Studies in Oceanography*, 116, 130-151, 10.1016/j.dsr2.2014.11.014, 2015.
- Fonseca-Batista, D., Dehairs, F., Riou, V., Fripiat, F., Elskens, M., Deman, F., Brion, N., Qu  rou  , F., Bode, M., and Auel, H.: Nitrogen fixation in the eastern Atlantic reaches similar levels in the Southern and Northern Hemisphere, *Journal of Geophysical Research: Oceans*, 122, 587-601, 10.1002/2016JC012335, 2017.
- Garcia, H. E., Bouchard, C., Cross, S. L., Paver, C. R., Reagan, J. R., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Baranova, O. K., and Seidov, D.: *World Ocean Atlas 2023, volume 4: dissolved inorganic nutrients (phosphate, nitrate, and silicate)*, NOAA National Centers for Environmental Information [dataset], 10.25923/39qw-7j08, 2024.
- Getis, A. and Ord, J. K.: The analysis of spatial association by use of distance statistics, *Geographical analysis*, 24, 189-206, 10.1111/j.1538-4632.1992.tb00261.x, 1992.
- Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., and Zehr, J. P.: Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N₂ fixation in the tropical Atlantic Ocean, *Environmental Microbiology*, 12, 3272-3289, 10.1111/j.1462-2920.2010.02303.x, 2010.
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *Journal of Geophysical Research: Oceans*, 118, 6704-6716, 10.1002/2013JC009067, 2013.
- Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database, *Deep Sea Research Part I: Oceanographic Research Papers*, 57, 812-833, 10.1016/j.dsr.2010.03.011, 2010.
- Gro  kopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M. M. M., Lavik, G., Schmitz, R. A., Wallace, D. W. R., and Laroche, J.: Doubling of marine dinitrogen-fixation rates based on direct measurements, *Nature*, 488, 361-364, 10.1038/nature11338, 2012.
- Gruber, N.: The marine nitrogen cycle: overview and challenges, in: *Nitrogen in the marine environment*, 1-50, 2008.
- Guieu, C., Bonnet, S., Petrenko, A., Menkes, C., Chavagnac, V., Desboeufs, K., Maes, C., and Moutin, T.: Iron from a submarine source impacts the productive layer of the Western Tropical South Pacific (WTSP), *Scientific Reports*, 8, 10.1038/s41598-018-27407-z, 2018.
- Hallstr  m, S., Benavides, M., Salamon, E. R., Ar  stegui, J., and Riemann, L.: Activity and distribution of diazotrophic communities across the Cape Verde Frontal Zone in the Northeast Atlantic Ocean, *Biogeochemistry*, 160, 49-67, 10.1007/s10533-022-00940-w, 2022.



- 540 Hoerstmann, C., Aguiar-González, B., Barrillon, S., Carpaneto Bastos, C., Grosso, O., Pérez-Hernández, M. D., Doglioli, A. M.,
Petrenko, A. A., Carracedo, L. I., and Benavides, M.: Nitrogen fixation in the North Atlantic supported by Gulf Stream eddy-
borne diazotrophs, *Nature Geoscience*, 17, 1141-1147, 10.1038/s41561-024-01567-2, 2024.
- Jiang, R., Hong, H., Wen, Z., Yu, X., Browning, T. J., Chen, Z., Shang, Y., Liu, X., Cao, Z., and Achterberg, E. P.: Significant
contribution of the unicellular cyanobacterium UCYN-B to oceanic nitrogen fixation, *National Science Review*, nwaf337,
545 10.1093/nsr/nwaf337, 2025.
- Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M., and Mahaffey, C.: Predictable and efficient carbon sequestration in the
North Pacific Ocean supported by symbiotic nitrogen fixation, *Proceedings of the National Academy of Sciences*, 109, 1842-
1849, 10.1073/pnas.1120312109, 2012.
- Kay, J. E., Deser, C., Phillips, A. S., and Simpson, I. R.: CESM1 Large Ensemble Community Project, NSF National Center for
550 Atmospheric Research [dataset], 10.5065/D6J101D1, 2021.
- Kittu, L. R., Paul, A. J., Fernández-Méndez, M., Hopwood, M. J., and Riebesell, U.: Coastal N₂ fixation rates coincide spatially
with nitrogen loss in the Humboldt Upwelling System off Peru, *Global Biogeochemical Cycles*, 37, e2022GB007578,
10.1029/2022GB007578, 2023.
- Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K. L., and Davis, C.: Euphotic zone depth: Its derivation and implication
555 to ocean-color remote sensing, *Journal of Geophysical Research: Oceans*, 112, 10.1029/2006JC003802, 2007.
- Luo, Y.-W., Lima, I., Karl, D., Deutsch, C., and Doney, S.: Data-based assessment of environmental controls on global marine
nitrogen fixation, *Biogeosciences*, 11, 691-708, 10.5194/bg-11-691-2014, 2014.
- Marra, G. and Wood, S. N.: Practical variable selection for generalized additive models, *Computational Statistics & Data Analysis*,
55, 2372-2387, 10.1016/j.csda.2011.02.004, 2011.
- 560 Memisoglu Baykal, T.: Performance assessment of GIS-based spatial clustering methods in forest fire data, *Natural Hazards*, 121,
8445-8477, 10.1007/s11069-025-07135-0, 2025.
- Mohr, W., Grosskopf, T., Wallace, D. W., and LaRoche, J.: Methodological underestimation of oceanic nitrogen fixation rates,
PloS one, 5, e12583, 10.1371/journal.pone.0012583, 2010.
- Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., and Zehr, J. P.:
565 Unicellular cyanobacterial distributions broaden the oceanic N₂ fixation domain, *Science*, 327, 1512-1514,
10.1126/science.1185468, 2010.
- Montoya, J. P., Voss, M., Kahler, P., and Capone, D. G.: A simple, high-precision, high-sensitivity tracer assay for N₂ fixation,
Applied and Environmental Microbiology, 62, 986-993, 10.1128/aem.62.3.986-993.1996, 1996.
- Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., McDonagh, E. L., Pan, X., Poulton, A.
570 J., and Rijkenberg, M. J.: Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability, *Nature*
Geoscience, 2, 867-871, 10.1038/ngeo667, 2009.



- Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., and Franz, B. A.: 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, 69-88, 10.1016/j.rse.2007.03.012, 2007.
- 575 Mulholland, M. R., Bernhardt, P. W., Heil, C. A., Bronk, D. A., and O'Neil, J. M.: Nitrogen fixation and release of fixed nitrogen by *Trichodesmium* spp. in the Gulf of Mexico, *Limnology and Oceanography*, 51, 1762-1776, 10.4319/lo.2006.51.4.1762, 2006.
- Ohde, T.: Saharan dust effects in different trophic areas off Northwest Africa, *Arabian Journal of Geosciences*, 9, 216, 10.1007/s12517-015-2031-0, 2016.
- 580 Ord, J. K. and Getis, A.: Local spatial autocorrelation statistics: distributional issues and an application, *Geographical analysis*, 27, 286-306, 10.1111/j.1538-4632.1995.tb00912.x, 1995.
- Resing, J. A., Sedwick, P. N., German, C. R., Jenkins, W. J., Moffett, J. W., Sohst, B. M., and Tagliabue, A.: Basin-scale transport of hydrothermal dissolved metals across the South Pacific Ocean, *Nature*, 523, 200-203, 10.1038/nature14577, 2015.
- Rubinstein, R. Y. and Kroese, D. P.: *Simulation and the Monte Carlo method*, John Wiley & Sons 2016.
- 585 Sato, T., Shiozaki, T., Hashihama, F., Sato, M., Murata, A., Sasaoka, K., Umeda, S. i., and Takahashi, K.: Low nitrogen fixation related to shallow nitracline across the eastern Indian Ocean, *Journal of Geophysical Research: Biogeosciences*, 127, 10.1029/2022JG007104, 2022.
- Schlosser, C., Klar, J. K., Wake, B. D., Snow, J. T., Honey, D. J., Woodward, E. M. S., Lohan, M. C., Achterberg, E. P., and Moore, C. M.: Seasonal ITCZ migration dynamically controls the location of the (sub) tropical Atlantic biogeochemical divide, *Proceedings of the National Academy of Sciences*, 111, 1438-1442, 10.1073/pnas.1318670111, 2014.
- 590 Selden, C., Mulholland, M., Bernhardt, P., Widner, B., Macías-Tapia, A., Ji, Q., and Jayakumar, A.: Dinitrogen fixation across physico-chemical gradients of the Eastern Tropical North Pacific oxygen deficient zone, *Global Biogeochemical Cycles*, 33, 1187-1202, 10.1029/2019GB006242, 2019.
- Selden, C. R., Chappell, P. D., Clayton, S., Macías-Tapia, A., Bernhardt, P. W., and Mulholland, M. R.: A coastal N₂ fixation hotspot at the Cape Hatteras front: Elucidating spatial heterogeneity in diazotroph activity via supervised machine learning, *Limnology and Oceanography*, 66, 1832-1849, 10.1002/lno.11727, 2021.
- Shao, Z., Xu, Y., Wang, H., Luo, W., Wang, L., Huang, Y., Agawin, N. S. R., Ahmed, A., Benavides, M., Bentzon-Tilia, M., Berman-Frank, I., Berthelot, H., Biegala, I. C., Bif, M. B., Bode, A., Bonnet, S., Bronk, D. A., Brown, M. V., Campbell, L., Capone, D. G., Carpenter, E. J., Cassar, N., Chang, B. X., Chappell, D., Chen, Y. L., Church, M. J., Cornejo-Castillo, F. M.,
- 600 Detoni, A. M. S., Doney, S. C., Dupouy, C., Estrada, M., Fernandez, C., Fernández-Castro, B., Fonseca-Batista, D., Foster, R. A., Furuya, K., Garcia, N., Goto, K., Gago, J., Gradoville, M. R., Hamersley, M. R., Henke, B. A., Hörstmann, C., Jayakumar, A., Jiang, Z., Kao, S. J., Karl, D. M., Kittu, L. R., Knapp, A. N., Kumar, S., LaRoche, J., Liu, H., Liu, J., Lory, C., Löscher, C. R., Maraño, E., Messer, L. F., Mills, M. M., Mohr, W., Moisander, P. H., Mahaffey, C., Moore, R., Mouriño-Carballido, B., Mulholland, M. R., Nakaoka, S., Needoba, J. A., Raes, E. J., Rahav, E., Ramírez-Cárdenas, T., Reeder, C. F., Riemann, L.,
- 605 Riou, V., Robidart, J. C., Sarma, V. V. S. S., Sato, T., Saxena, H., Selden, C., Seymour, J. R., Shi, D., Shiozaki, T., Singh, A.,



- Sipler, R. E., Sun, J., Suzuki, K., Takahashi, K., Tan, Y., Tang, W., Tremblay, J. É., Turk-Kubo, K., Wen, Z., White, A. E., Wilson, S. T., Yoshida, T., Zehr, J. P., Zhang, R., Zhang, Y., and Luo, Y. W.: Global oceanic diazotroph database version 2 and elevated estimate of global oceanic N₂ fixation, *Earth System Science Data*, 15, 3673-3709, 10.5194/essd-15-3673-2023, 2023.
- 610 Shen, H., Wan, X. S., Zou, W., Dai, M., Xu, M. N., and Kao, S.-J.: Light-driven integration of diazotroph-derived nitrogen in euphotic nitrogen cycle, *Nature Communications*, 15, 10.1038/s41467-024-53067-x, 2024a.
- Shen, H., Wan, X. S., Zou, W., Chen, M., Hu, Z., Tong, S., Zhou, K., Jiang, Z.-P., Zhang, Y., and Dai, M.: Physical optima for nitrogen fixation in cyclonic eddies in the Subtropical Northwestern Pacific, *Progress in Oceanography*, 226, 103298, 10.1016/j.pocean.2024.103298, 2024b.
- 615 Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., and Furuya, K.: Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean, *Global Biogeochemical Cycles*, 28, 1096-1110, 10.1002/2014GB004886, 2014.
- Shiozaki, T., Bombar, D., Riemann, L., Sato, M., Hashihama, F., Kodama, T., Tanita, I., Takeda, S., Saito, H., and Hamasaki, K.: Linkage between dinitrogen fixation and primary production in the oligotrophic South Pacific Ocean, *Global Biogeochemical Cycles*, 32, 1028-1044, 10.1029/2017GB005869, 2018.
- 620 Singh, A., Bach, L. T., Fischer, T., Hauss, H., Kiko, R., Paul, A. J., Stange, P., Vandromme, P., and Riebesell, U.: Niche construction by non-diazotrophs for N₂ fixers in the eastern tropical North Atlantic Ocean, *Geophysical Research Letters*, 44, 6904-6913, 10.1002/2017GL074218, 2017.
- Subramaniam, A., Mahaffey, C., Johns, W., and Mahowald, N.: Equatorial upwelling enhances nitrogen fixation in the Atlantic Ocean, *Geophysical Research Letters*, 40, 1766-1771, 10.1002/grl.50250, 2013.
- 625 Tang, W., Cerdán-García, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, A., Whitby, H., Planquette, H., Mowlem, M., and Robidart, J.: New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods, *The ISME Journal*, 14, 2514-2526, 10.1038/s41396-020-0703-6, 2020.
- Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A. G., Gallinari, M., Planquette, H., Sarthou, G., and Cassar, N.: Revisiting the distribution of oceanic N₂ fixation and estimating diazotrophic contribution to marine production, *Nature Communications*, 10, 10.1038/s41467-019-08640-0, 2019.
- 630 Tanita, I., Shiozaki, T., Kodama, T., Hashihama, F., Sato, M., Takahashi, K., and Furuya, K.: Regionally variable responses of nitrogen fixation to iron and phosphorus enrichment in the Pacific Ocean, *Journal of Geophysical Research: Biogeosciences*, 126, e2021JG006542, 10.1029/2021JG006542, 2021.
- Voss, M., Croot, P., Lochte, K., Mills, M., and Peeken, I.: Patterns of nitrogen fixation along 10°N in the tropical Atlantic, *Geophysical Research Letters*, 31, L23S09, 10.1029/2004GL020127, 2004.
- 635 Wagener, T., Guieu, C., Losno, R., Bonnet, S., and Mahowald, N.: Revisiting atmospheric dust export to the Southern Hemisphere ocean: Biogeochemical implications, *Global Biogeochemical Cycles*, 22, 10.1029/2007GB002984, 2008.



- Wen, Z., Browning, T. J., Cai, Y., Dai, R., Zhang, R., Du, C., Jiang, R., Lin, W., Liu, X., Cao, Z., Hong, H., Dai, M., and Shi, D.: Nutrient regulation of biological nitrogen fixation across the tropical western North Pacific, *Science Advances*, 8, 10.1126/sciadv.abl7564, 2022.
- White, A. E., Granger, J., Selden, C., Gradoville, M. R., Potts, L., Bourbonnais, A., Fulweiler, R. W., Knapp, A. N., Mohr, W., Moisaner, P. H., Tobias, C. R., Caffin, M., Wilson, S. T., Benavides, M., Bonnet, S., Mulholland, M. R., and Chang, B. X.: A critical review of the $^{15}\text{N}_2$ tracer method to measure diazotrophic production in pelagic ecosystems, *Limnology and Oceanography: Methods*, 18, 129-147, 10.1002/lom3.10353, 2020.
- Wilson, S. T., Böttjer, D., Church, M. J., and Karl, D. M.: Comparative assessment of nitrogen fixation methodologies, conducted in the oligotrophic North Pacific Ocean, *Applied and Environmental Microbiology*, 78, 6516-6523, 10.1128/AEM.01146-12, 2012.
- Wood, S. N.: Generalized additive models: an introduction with R, CRC press, 10.1201/9781315370279, 2017.
- Wood, S. N., Pya, N., and Säfken, B.: Smoothing parameter and model selection for general smooth models, *Journal of the American Statistical Association*, 111, 1548-1563, 10.1080/01621459.2016.1180986, 2016.
- Yu, X., Wen, Z., Jiang, R., Yang, J. Y. T., Cao, Z., Hong, H., Zhou, Y., and Shi, D.: Assessing N_2 fixation flux and its controlling factors in the (sub) tropical western North Pacific through high-resolution observations, *Limnology and Oceanography Letters*, 10.1002/lol2.10404, 2024.
- Zehr, J. P. and Capone, D. G.: Changing perspectives in marine nitrogen fixation, *Science*, 368, eaay9514, 10.1126/science.aay9514, 2020.
- Zehr, J. P. and Capone, D. G.: Unsolved mysteries in marine nitrogen fixation, *Trends in Microbiology*, 32, 532-545, 10.1016/j.tim.2023.08.004, 2024.
- Zhang, X., Ward, B. B., and Sigman, D. M.: Global nitrogen cycle: critical enzymes, organisms, and processes for nitrogen budgets and dynamics, *Chemical reviews*, 120, 5308-5351, 10.1021/acs.chemrev.9b00613, 2020.