Manuscript egusphere-2025-3763

" Spatiotemporal heterogeneity in diazotrophic communities reveals novel niche zonation on the continental shelf of the East China Sea" by Guangming Mai et al.

We thank the editor and the reviewers for the constructive advice in the previous round of review, which we believe has led to a substantially improved manuscript. We have carefully examined the thoughtful, detailed comments provided by the reviewer #1 in this review, and accommodated them, in so far as possible. Our point-by-point responses are provided below in blue fonts with changes shown in red. Please note that all the line numbers mentioned in the response refer to those in the marked-up manuscript.

Response to comments from Reviewer #1

The manuscript has improved in several respects, particularly in the expansion of the methodological description and clarification of some technical points. However, these additions have not fully resolved my original concerns. Since novelty of this study largely hinges on the modeling framework, the validity, purpose, and advantages of the chosen model must be clearly demonstrated. At present, I am still not convinced by the explanation of MaxEnt and why it is preferable to more direct and standard statistical approaches.

Response:

We thank the Reviewer #1 for the comment. We have further revised the manuscript to justify our choice of the MaxEnt. We have integrated this justification into the Materials and Methods section to demonstrate the model's validity and appropriateness for our framework.

Major comment

It remains unclear why a univariable MaxEnt–GAM framework is more suitable than generalized linear models (GLMs) or generalized additive models (GAMs) with multiple environmental predictors, in which each diazotrophic phylotype is treated as a response variable. Multivariable GLMs and GAMs are widely used to examine species–environment relationships, can accommodate nonlinear responses, and allow explicit treatment of covarying environmental effects (e.g., through variable selection, VIF, and information-theoretic criteria such as AIC).

Response:

We thank the Reviewer #1 for the comment. We have further clarified the use of univariable MaxEnt–GAM framework over multivariable GLMs/GAMs in the Materials and Methods (section 2.5.3).

Line 221-241: "Realized niches of diazotrophs in relation to environmental variables on the ECS shelf were determined in a combined framework implementing the univariate Maximum Entropy (MaxEnt) (Phillips et al., 2006) and GAM, as described by Irwin et al. (2012) and Xiao et al. (2018). This method was chosen over the standard multivariate generalized linear model (GLM) or GAM for several reasons to overcome the challenges in analyzing marine ecological data. While the multivariate models can capture species-environment relationships and nonlinear effects, they require reliable presence/absence records and treat non-detections as true absences, an assumption often invalid in undersampled, patchy marine systems where absences frequently indicate missed detection rather than actual exclusion. The MaxEnt component of our framework circumvents this issue by modeling presence data against background environmental conditions, making it more robust for sparse and patchy biological observations. Additionally, environmental variables in oceanic systems are often highly collinear (e.g., NO_x and SRP), which in a multivariate model can obscure individual predictor effects, destabilize parameter estimates, and produce ecologically ambiguous response curves. Although collinearity may be reduced by variable selection techniques (e.g., variance inflation factors), it comes at the cost of excluding some key drivers that affect species niche traits. In contrast, the univariate approach examines the response along each environmental gradient, providing clear and interpretable niche estimate for robust, cross-phylotype comparisons (Irwin et al., 2012). Finally, the combined framework used here can independently model two ecologically distinct niche components: the probability of occurrence (via MaxEnt) and the expected abundance when present (via GAM), thus avoiding the conflation and bias caused by the zero values in abundance data inherent to multivariate models. This framework has proven robust for sparse and heterogeneous biological data within environmental matrices, as demonstrated in prior studies of phytoplankton niches in the western Pacific marginal seas (Xiao et al., 2018; Zhong et al., 2020).".

At present, the manuscript does not sufficiently explain why a univariable MaxEnt–GAM approach is preferable to a multivariable GAM framework for examining environmental responses, nor what specific ecological insights are gained by using MaxEnt that could not be obtained more directly through these alternative and more standard approaches. While the MaxEnt–GAM framework appears nice approach, I would encourage the authors to more clearly articulate the rationale for its use to strengthens the conclusion.

Response:

We thank the Reviewer #1 for the comment. We have clarified in the Materials and Methods (Section 2.5.3) why a univariable MaxEnt–GAM framework is adopted instead of multivariable GAMs, and have also strengthened the specific ecological insights that can be demonstrated by using this approach but not the other alternative approaches in the Conclusion (Section 5).

Line 221-241: "Realized niches of diazotrophs in relation to environmental variables on the ECS shelf were determined in a combined framework implementing the univariate Maximum Entropy (MaxEnt) (Phillips et al., 2006) and GAM, as described by Irwin et al. (2012) and Xiao et al. (2018). This method was chosen over the standard multivariate generalized linear model (GLM) or GAM for several reasons to overcome the challenges in analyzing marine ecological data. While the multivariate models can capture species-environment relationships and nonlinear effects, they require reliable presence/absence records and treat non-detections as true absences, an assumption often invalid in undersampled, patchy marine systems where absences frequently indicate missed detection rather than actual exclusion. The MaxEnt component of our framework circumvents this issue by modeling presence data against background environmental conditions, making it more robust for sparse and patchy biological observations. Additionally, environmental variables in oceanic systems are often highly collinear (e.g., NO_x and SRP), which in a multivariate model can obscure individual predictor effects, destabilize parameter estimates, and produce ecologically ambiguous response curves. Although collinearity may be reduced by variable selection techniques (e.g., variance inflation factors), it comes at the cost of excluding some key drivers that affect species niche traits. In contrast, the univariate approach examines the response along each environmental gradient, providing clear and interpretable niche estimate for robust, cross-phylotype comparisons (Irwin et al., 2012). Finally, the combined framework used here can independently model two ecologically distinct niche components: the probability of occurrence (via

MaxEnt) and the expected abundance when present (via GAM), thus avoiding the conflation and bias caused by the zero values in abundance data inherent to multivariate models. This framework has proven robust for sparse and heterogeneous biological data within environmental matrices, as demonstrated in prior studies of phytoplankton niches in the western Pacific marginal seas (Xiao et al., 2018; Zhong et al., 2020).".

Line 648-652: "Moreover, our adoption of the univariate MaxEnt-GAM framework provides ecological insights that are less accessible through conventional multivariate approaches, elucidating trade-offs among co-occurring diazotrophic phylotypes (e.g., the separation of UCYN-A from diatom-diazotroph symbioses) and broadening global phytoplankton niche domains.".

A similar concern applies to the use of simple correlation analyses. The authors state that correlations were used to avoid collinearity and to enable direct comparison; however, pairwise correlations do not remove the influence of covariables and therefore cannot disentangle direct from indirect effects. Given the available sample size, a multivariable GLM or GAM framework that quantifies the relative importance of each environmental variable would likely provide a more straightforward and interpretable approach.

Response:

We thank the Reviewer #1 for the constructive suggestion on the statistical approach. Following the recommendation, we have re-analyzed our data using a generalized additive model (GAM) framework to quantify the relative contribution of each environmental variable. The Materials and Methods and Results sections have been updated accordingly. We believe this revision rigorously improves the statistical analysis.

Line 212-218: "Furthermore, a multivariate generalized additive model (GAM) with Tweedie distribution was used to estimate the relative contribution of environmental variables to variations in diazotroph abundances and NFRs, with explained deviance hierarchically partitioned (Lai et al., 2024). Concurvity among variables was assessed, and all pairwise indices below 0.8 were kept to enhance model robustness. To identify key environmental drivers, we applied an automatic smoothness selection method, which effectively shrinks non-significant terms to zero degrees of freedom and thus enables automatic variable selection (Marra and Wood, 2011).".

Line 390-393: "Hierarchical variance partitioning in GAMs further indicated that the measured factors explained 0.6–97.5% of the spatial variation in diazotroph abundances and NFRs (Fig. 7B). Temperature and salinity accounted for the majority (59.2% on average) of the explained variation, followed by nutrients and N:P ratio (37.5% on average).".

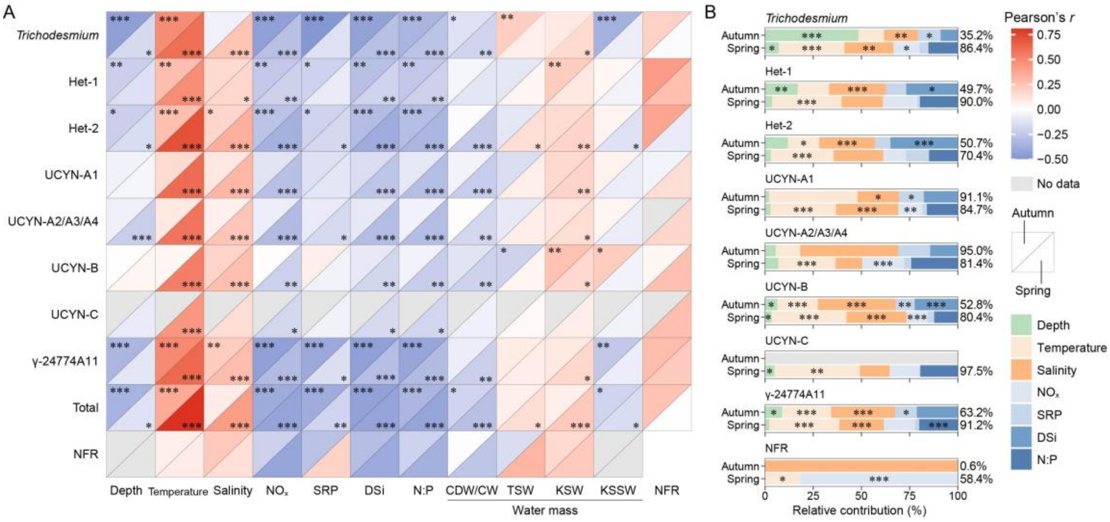

Figure 7. Relative contribution of environmental variables to diazotroph abundances and N_2 fixation rates (NFRs) across distinct water masses on the East China Sea shelf during autumn and spring. (A) Correlation of diazotrophic phylotypes with environmental variables, water masses and NFRs, as indicated with Pearson coefficients. (B) Relative contribution of environmental variables to diazotroph abundances and NFRs, as indicated with the total deviance explained. Total, combined *nifH* gene abundances of the eight diazotrophs detected; CDW, Changjiang diluted water; CW, Coastal water; TSW, Taiwan Strait water; KSW, Kuroshio surface water; KSSW, Kuroshio subsurface water; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Other comments

The interpretation of the niche mean (μ) remains unclear and internally inconsistent. As defined in the manuscript, μ is a weighted mean of the response curve and is described as representing the "central environmental condition." This definition implies a central tendency or preferred condition. However, in the Discussion, μ is reinterpreted as reflecting environmental tolerance when model-derived niche means conflict with observed abundance patterns and simple correlation analyses. This shift in interpretation should be clarified and made conceptually consistent throughout the manuscript.

Response:

We thank the Reviewer #1 for the comment. We have revised the Materials and Methods (Section 2.5.3) to explicitly define μ as the "abundance-weighted mean environmental condition" for the diazotrophic phylotypes within the study area, and reaffirmed that σ represents the tolerance range. We have also added the clarifications regarding μ in the Discussion section in the revised manuscript.

Line 259-260: "where μ represents the abundance-weighted mean environmental condition (i.e., the realized niche occupied by the phylotype in the study area), while σ indicates the phylotype's tolerance range for a given variable.".

Line 565-572: "For instance, while *Trichodesmium* typically dominated in high-temperature, NO_x-deficient waters (e.g., Jiang et al., 2025; Nguyen et al., 2025; Tang and Cassar, 2019), our modeling results indicate its realized niches characterized by cool temperatures ($\mu_{temperature}$ = 23.5°C) and elevated NO_x levels (μ_{NOx} = 6.93 µM). This change in niche means reflects a displacement of *Trichodesmium* population on the shelf during the physical transport of water masses along the temperature and nutrient gradients. Meanwhile, the broad niche breadths ($\sigma_{temperature}$ = 3.5°C; σ_{NOx} = 5.11 µM) also confirm the capacity of *Trichodesmium* to persist across these sub-optimal temperatures and variable nutrient conditions.".

UCYNC and GammaA are treated as nitroplast. However, these relationships remain suggested rather than conclusively demonstrated and should therefore be described more cautiously.

Response:

We thank the Reviewer #1 for the comment. We have described UCYN-C and γ-24774A11 as "putative diatom nitroplasts" in the revised manuscript.

Line 13-18: "An overall spatial heterogeneity among some of the major diazotrophic phylotypes was unveiled, with the filamentous cyanobacteria *Trichodesmium*, diatom-diazotroph symbioses (Het-1 and Het-2), the unicellular cyanobacterial diazotrophs (UCYN-B) and Haptophyta-associated nitroplasts (UCYN-A) dominating the upper 30 m of the warm, nitrogen-limited offshore region intruded by the Kuroshio and Taiwan Strait water, whereas diatom-associated putative nitroplasts (UCYN-C and γ-24774A11) were abundant both at the surface and 50-m depth.".

Line 146-152: "Recent studies have demonstrated UCYN-A2 as a nitroplast in the haptophyte *Braarudosphaera bigelowii* (Coale et al., 2024; Cornejo-Castillo et al., 2024). Similarly, UCYN-C and γ-24774A11 have been proposed as putative nitroplasts in diatoms (Schvarcz et al., 2022, 2024; Tschitschko et al., 2024). In contrast, other UCYN-A sublineages such as UCYN-A1 have not yet

been confirmed to possess the same defining characteristics in their association with haptophytes (Coale et al., 2024; Kantor et al., 2024). Therefore, for clarity and consistency with prior literature, we classified UCYN-A2 as a haptophyte nitroplast, UCYN-C and γ-24774A11 as putative diatom nitroplasts, while designating UCYN-A1 as other UCYN-A sublineages.".

Line 344-345: "The putative diatom nitroplasts (i.e., UCYN-C and γ-24774A11) were most abundant (>10^3 copies L^{-1}) at depths of 0–50 m (Fig. 4, S6U–Z), demonstrating moderately deeper distributions than Hets.".

As noted in my initial review, and as emphasized in White et al. (2020), calculation and reporting of the limit of detection (LOD) for N_2 fixation rates is now standard and strongly recommended practice. The LOD should be calculated and reported following Eq. 5 in White et al. (2020).

Response:

We thank the Reviewer #1 for the comment. After reexamining the calculation of N_2 fixation detection limit, we have adopted a minimum difference of 0.00146 atom% in the ^{15}N enrichment of particulate nitrogen (PN) between the initial and the final measurements. Although White et al. (2020) caution against using this value for PN mass <10 μg, our samples exceeded 16 μg. Thus, this threshold sufficiently accounts for the analytical uncertainty in PN isotopic measurements (White et al., 2020). We have revised the calculation of LOD accordingly.

Line 185-188: "The NFRs were determined following Montoya et al. (1996). The LOD for the NFRs was calculated according to White et al. (2020), defined as a minimum difference of 0.00146 atom% in ^{15}N enrichment of particulate nitrogen between the initial and the final measurements. This LOD corresponds to a range of 0.11–0.76 nmol N L^{-1} d^{-1} across the stations (Table S2).".

References

Brun, P., Vogt, M., Payne, M. R., Gruber, N., O'Brien, C. J., Buitenhuis, E. T., Le Quéré, C., Leblanc, K., and Luo, Y.: Ecological niches of open ocean phytoplankton taxa, Limnol. Oceanogr., 60, 1020–1038, https://doi.org/10.1002/lno.10074, 2015.

Coale, T. H., Loconte, V., Turk-Kubo, K. A., Vanslembrouck, B., Mak, W. K. E., Cheung, S., Ekman, A., Chen, J.-H., Hagino, K., Takano, Y., Nishimura, T., Adachi, M., Le Gros, M., Larabell, C., and Zehr, J. P.: Nitrogen-fixing organelle in a marine alga, Science, 384, 217–222, https://doi.org/10.1126/science.adk1075, 2024.

Cornejo-Castillo, F. M., Inomura, K., Zehr, J. P., and Follows, M. J.: Metabolic trade-offs constrain the cell size ratio in a nitrogen-fixing symbiosis, Cell, S009286742400182X, https://doi.org/10.1016/j.cell.2024.02.016, 2024.

Irwin, A. J., Nelles, A. M., and Finkel, Z. V.: Phytoplankton niches estimated from field data, Limnol. Oceanogr., 57, 787–797, https://doi.org/10.4319/lo.2012.57.3.0787, 2012.

Jiang, R., Hong, H., Wen, Z., Yu, X., Browning, T. J., Chen, Z., Shang, Y., Liu, X., Cao, Z., Achterberg, E. P., Dai, M., and Shi, D.: Significant contribution of the unicellular cyanobacterium UCYN-B to oceanic nitrogen fixation, Natl. Sci. Rev., 12, nwaf337, https://doi.org/10.1093/nsr/nwaf337, 2025.

Kantor, E. J. H., Robicheau, B. M., Tolman, J., Archibald, J. M., and LaRoche, J.: Metagenomics reveals the genetic diversity between sublineages of UCYN-A and their algal host plastids, ISME Commun., 4, ycae150, https://doi.org/10.1093/ismeco/ycae150, 2024.

Lai, J., Tang, J., Li, T., Zhang, A., and Mao, L.: Evaluating the relative importance of predictors in generalized additive models using the gam.hp R package, Plant Divers., 46, 542–546, https://doi.org/10.1016/j.pld.2024.06.002, 2024.

Marra, G., and Wood, S. N.: Practical variable selection for generalized additive models, Comput. Stat. Data Anal., 55, 2372–2387, https://doi.org/10.1016/j.csda.2011.02.004, 2011.

Montoya, J. P., Voss, M., Hler, P. K., and Capone, D. G.: A simple, high-precision, high-sensitivity tracer assay for N_2 fixation, Appl. Environ. Microbiol., 62, 986–993, https://doi.org/10.1128/aem.62.3.986-993.1996, 1996.

Nguyen, A., Ustick, L. J., Larkin, A. A., and Martiny, A. C.: Global phylogeography and microdiversity of the marine diazotrophic cyanobacteria *Trichodesmium* and UCYN-A, mSphere, 10, e00245-25, https://doi.org/10.1128/msphere.00245-25, 2025.

Phillips, S. J., Anderson, R. P., and Schapire, R. E.: Maximum entropy modeling of species geographic distributions, Ecol. Model., 190, 231–259, https://doi.org/10.1016/j.ecolmodel.2005.03.026, 2006.

Schvarcz, C. R., Wilson, S. T., Caffin, M., Stancheva, R., Li, Q., Turk-Kubo, K. A., White, A. E., Karl, D. M., Zehr, J. P., and Steward, G. F.: Overlooked and widespread pennate diatom-diazotroph symbioses in the sea, Nat. Commun., 13, 799, https://doi.org/10.1038/s41467-022-28065-6, 2022.

Schvarcz, C. R., Stancheva, R., Turk-Kubo, K. A., Wilson, S. T., Zehr, J. P., Edwards, K. F., Steward, G. F., Archibald, J. M., Oatley, G., Sinclair, E., Santos, C., Paulini, M., Aunin, E., Gettle, N., Niu, H., McKenna, V., O'Brien, R., Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics Team, EBI Aquatic Symbiosis Genomics Data Portal Team, and Aquatic Symbiosis Genomics Project Leadership: The genome sequences of the marine diatom *Epithemia pelagica* strain UHM3201 (Schvarcz, Stancheva & Steward, 2022) and its nitrogen-fixing, endosymbiotic cyanobacterium, Wellcome Open Res., 9, 232, https://doi.org/10.12688/wellcomeopenres.21534.1, 2024.

Tang, W., and Cassar, N.: Data-driven modeling of the distribution of diazotrophs in the global ocean, Geophys. Res. Lett., 46, 12258–12269, https://doi.org/10.1029/2019GL084376, 2019.

Tschitschko, B., Esti, M., Philippi, M., Kidane, A. T., Littmann, S., Kitzinger, K., Speth, D. R., Li, S., Kraberg, A., Tienken, D., Marchant, H. K., Kartal, B., Milucka, J., Mohr, W., and Kuypers, M. M. M.: Rhizobia–diatom symbiosis fixes missing nitrogen in the ocean, Nature, 630, 899–904, https://doi.org/10.1038/s41586-024-07495-w, 2024.

White, A. E., Granger, J., Selden, C., Gradoville, M. R., Potts, L., Bourbonnais, A., Fulweiler, R. W., Knapp, A. N., Mohr, W., Moisander, 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 $^{15}N_2$ tracer method to measure diazotrophic production in pelagic ecosystems, Limnol. Oceanogr. Methods, 18, 129–147, https://doi.org/10.1002/lom3.10353, 2020.

Xiao, W., Wang, L., Laws, E., Xie, Y., Chen, J., Liu, X., Chen, B., and Huang, B.: Realized niches explain spatial gradients in seasonal abundance of phytoplankton groups in the South China Sea, Prog. Oceanogr., 162, 223–239, https://doi.org/10.1016/j.pocean.2018.03.008, 2018.