Manuscript bg-2023-126

"Changes in diazotrophic community structure associated with Kuroshio succession in the northern South China Sea" by Han Zhang et al.

We have taken all the comments of the Reviewers into account in the revision. Our point-by-point responses are provided below in blue fonts. Please note that all the line numbers mentioned in the response refer to those in the Marked-up Manuscript.

General comments:

Reviewer (Comments for the Author):

Zhang et al., 2023 show how the diazotrophic community in the northern South China Sea (nSCS) responds to intrusion by Kuroshio current waters by measuring nifH abundances via qPCR and high throughput sequencing, nitrogen fixation rates, and the degree of Kuroshio intrusion (KI) over two cruises in 2017 and 2018. The big picture question the authors address is an important one, namely how do physical changes in a marine ecosystem affect microbial diversity? The authors show that *Trichodesmium* is more abundant at stations with more KI and that UCYN-B is more abundant at stations with low KI. In addition, the authors show that UCYN-C and gamma-proteobacteria are more prevalent at stations moderately affected by KI. The authors also perform statistical tests to assess the degree environmental factors affect the diazotrophic community structure. Their tests suggests that environmental factors have a bigger affect on diazotrophic community structure in 2017, when KI is strong, than in 2018 when KI is weak.

I have several minor comments but no major revisions and recommend the paper to be accepted with minor revisions. All line numbers refer to the preprint visible on EGUsphere.

Response:

We thank the Reviewer for the comments and consideration of this manuscript for acceptance. We hope that our point-by-point responses below have addressed his/her concern.

General comment

I found it difficult to understand what the Mantel tests vs. the NCM were testing. The abstract and Figure 8 say that the neutral community model is being used to test the effect of environmental variables on diazotrophic community composition. However, in the methods and elsewhere, the Mantel tests are described as the way that relationship is assessed. I think the easiest way to solve this would be to add an explanation in section 2.9 that describes how the NCM was used to assess the variation due to environmental and spatial factors.

Response:

Mantel tests and the Neutral Community Model (NCM) serve distinct purposes and are applied to investigate different aspects of microbial communities. Mantel tests are often used to examine the correlation between distance matrices that reflect the differences in microbial community composition between different samples, such as the Bray-Curtis dissimilarity or UniFrac distances. The Mantel test can also be used to assess how microbial community the correlation composition relates to specific environmental factors or spatial patterns within a given ecosystem. The NCM, on

the other hand, is a theoretical framework used to explore the role of ecological drift and dispersal limitation in shaping microbial community structure. This model assumes that microbial community assembly is primarily governed by random processes, such as birth, death, and migration, rather than deterministic ecological interactions. Together, these approaches contribute to our understanding of the factors driving the structure and dynamics of microbial communities in various environments.

We are grateful for the Reviewer's comment. We agree with the Reviewer that add an explanation in section 2.9 that describes how the NCM was used to assess the variation due to environmental and spatial factors. Therefore, we have added further description about NCM in section 2.9 as below:

Line 213-217: "To assess the role of KI in shaping diazotrophic community assembly, we fit the occurrence frequency of diazotrophic taxa across the wider metacommunity to the neutral community model (NCM), a theoretical framework used to explore the potential impact of stochastic processes (Sloan et al., 2007). Calculation of 95% confidence interval was done using 1,000 bootstrap replicates, and the overall fit to the NCM was indicated by the parameter R^2. All data analyses were processed in R (version 4.1.3) (http://www.r-project.org).".

Line by line comments

Title - I agree with the other reviewers that it should be Changes not Change

Response:

We thank the Reviewer for the suggestion, and we have revised the Title as "Changes in diazotrophic community structure associated with Kuroshio succession in the northern South China Sea".

Line 35 - This phrasing is awkward. I would rephrase "The current state of marine N_2 fixation study ..." to be "The field of marine N_2 fixation is changing dramatically."

Response:

Thanks for the suggestion and we have rephrased it. Please refer to Line 35.

Line 35 : "The field of marine N_2 fixation is changing dramatically.".

Lines 39 - 40 - I would also change the phrasing here. I would replace "Such previously established concept" with "This previously established concept ... "

Response:

Thanks for the suggestion and we have replacedd "such" with "this"..

Line 39-40 : "This previously established concept was challenged by the discovery of unicellular…".

Lines 35-52 - Since UCYN-C is discussed in the abstract, I would explicitly mention it when the new diatom-diazotroph symbioses discovered by Schvarcz, 2022 are discussed. I would also mention another way that the field of marine N_2 fixation is changing, which is the discovery of widespread, high coastal or continental shelf N_2 fixation (Tang et al., 2019).

Response:

We thank the Reviewer for the comment. We have extended the description about new diatom-diazotroph symbioses and the biogeography of marine N_2 fixation as below:

Line 49-56: "Particularly, diazotrophs genomes reconstructed from global metagenomic data have extended the PCR-based amplicon surveys revealing new diatom-diazotroph symbioses (Schvarcz et al., 2022) and new species of non-cyanobacterial diazotrophs (NCDs) (Bombar et al., 2016; Delmont et al., 2018).The Rhopalodiaceae diatoms have been repeatedly isolated from the subtropical North Pacific Ocean, with the endosymbionts having *nifH* gene sequences similar to those of free-living UCYN-C cyanobacteria (Schvarcz et al., 2022). NCDs have been reported to be ubiquitous in ocean ecosystems and contribute to global marine N_2 fixation (Moisander et al., 2014; Chakraborty et al., 2021; Turk-Kubo et al., 2022). Recently, widespread and high N_2 fixation was also discovered in the coastal or continental shelf (Tang et al., 2019).".

Line 94 - I agree with the previous reviewer comment that a 100 micron pre-filter could reduce *Trichodesmium* abundances. I would add in a comment mentioning this if you have not already.

Response:

We completely agree with the reviewer that prefiltration using a 100 μm pore-size nylon mesh might potentially remove colonial trichomes of *Trichodesmium* larger than 100 μm. To be more rigorous, we clarified the rationale of prefiltration and revised Materials and Methods and Discussion sections as below:

Line 100-102: "At each station, 1.5-3 L of seawater was prefiltered through a 100-μm pore-size nylon mesh to remove large zooplankton and fish, and then filtered through a 0.22-μm pore-size 47-mm diameter polycarbonate membrane (Millipore, USA) with low pressure (<100 mm Hg pressure) for subsequent DNA extraction.".

Line 291-293: "It is worth noting that, since we did not observe colonial *Trichodesmium* in our sampling stations, prefilteration using a 100 μm pore-size nylon mesh was not likely to underestimate *Trichodesmium* abundance.".

Line 173 - Please include a reference for your merging criteria.

Response:

We thank the Reviewer for the comment. We have added the reference about the merging criteria as below:

Line 183: "…ii) Sequences having ≥ 10 bp overlap but ≤ 2 bp mismatch were merged (Zhang et al., 2023)…".

Line 177 - Please also include a reference for the RDP classifier algorithm.

Response:

We thank the Reviewer for the comment. We determined relative abundance of OTU using UPARSE with a 97% similarity cutoff according to Quince et al. (2011).

Line 184-189: "Operational taxonomic units (OTUs) were clustered with a 97% similarity cutoff using UPARSE (version 7.1) (http://drive5.com/uparse/) with a novel 'greedy' algorithm that performs chimera filtering and OTU clustering simultaneously (Quince et al., 2011). OTUs were annotated down to genus level using a formatted *nifH* gene database (https://github.com/moyn413/nifHdada2) that was updated in 2023 according to NCBI and Zehr Lab (https://www.jzehrlab.com/*nifH*).".

Line 190 - It is not clear how you define spatial vs. environmental factors. Please state all spatial and environmental factors you test explicitly.

Response:

We thank the Reviewer for the comment. We have added the description about the spatial vs. environmental factors as below:

Line 202-204: "The environmental factors are summarized in Table 1. All environmental parameters, except pH, were log (X+1)-transformed to improve homoscedasticity and normality for multivariate statistical analyses and calculation of the Euclidean distances between samples.".

Line 204-205: A set of spatial variables based on the longitude and latitude coordinates of each sampling station were calculated following the approach of the principal coordinates of neighbor matrices (PCNMs) analysis (Borcard and Legendre, 2002).".

Line 196 - Please also include a citation to justify the choice of a VIF threshold of < 20.

Response:

We thank the Reviewer for the comment. We have added the citation for the choice of a VIF threshold of < 20 as below:

Line 206-207: "To avoid collinearity among factors, explanatory environmental factors with the highest variance inflation factor (VIF) were eliminated until all VIF values were lower than 20 (Blanchet et al., 2008; Chen et al., 2019).".

Line 202 - I do not understand the two phrases used to describe what the NCM is assessing the relationship between "the occurrence frequency of diazotrophic taxa" and "their relative abundance across the wider metacommunity". I would rephrase to more clearly explain the difference between "the occurrence frequency of diazotrophic taxa" and "their relative abundance across the wider metacommunity".

Response:

We thank the Reviewer for the comment. We have rephrased to more clearly explain the difference between "the occurrence frequency of diazotrophic taxa" and "their relative abundance across the wider metacommunity". The frequency of occurrence of diazotrophic taxa refers to how often these organisms are present in a given context, while the relative abundance across the wider metacommunity indicates the proportion or percentage of these taxa in comparison to the entire community. In simpler terms, the occurrence frequency focuses on how often they appear, while relative abundance considers their prevalence within the broader community. we added the description about the NCM in Methods 2.9 as below:

Line 213-217: "To assess the role of KI in shaping diazotrophic community assembly, the occurrence frequency of diazotrophic taxa across the wider metacommunity was fitted to the neutral community model (NCM), a theoretical framework used to explore the potential impact of stochastic processes (Sloan et al., 2007). Calculation of 95% confidence interval was done using 1,000 bootstrap replicates, and the overall fit to the NCM wasindicated by the parameter R^2. All data analyses were processed in R (version 4.1.3) (http://www.r-project.org).".

Lines 218 - 219 - This sentence should be in the methods not in the results.
Response:

We thank the Reviewer for the comment. We have removed the sentence to Methods as below:
Line 163-165: "When the standard clone was diluted 10 times, the corresponding Ct value increased by about 3.3–3.4 units, indicating that the PCR amplification efficiencies among all replicates were between 90% and 100% ($R^2 > 0.99$, Fig. S1).".

Lines 226 - 230 - I didn't see any tables where the relative abundances were clearly stated. I would make a new SI table with the same format as Table S5 that has the relative abundances for the 10 species investigated via qPCR. It is also hard to evaluate the authors' claims about day vs. nighttime abundances since many of the times (19:00, 19:30, 20:00 are on the border of day and night). I would explicitly state with a D or N in Table S5 and the new table which times are day and night. Please also say if times are local time or another timezone.
Response:

We thank the Reviewer for the comment. We apologize for the confusion caused. We have made a new SI table (Table S5) with the same format as Table S4 (Table S5 in the previous version) that has the relative abundances for the 10 species investigated via qPCR. In Table S4 (Table S5 in the previous version) and Table S5, we removed the *nifH* gene copies data of nighttime and focused on surface daytime data in the main text. we would like to clarify that the times mentioned in our study are presented in the local time zone.

Line 232 - This statement is misleading because there is quite a bit of variability across stations, for example stn 1 vs. 12 vs. 4 in 2017 (Table S6) . I would instead describe that there is considerable

station to station variability but that across all stations cyanobacteria and gamma proteobacteria are at approximately equal abundances.

Response:

We thank the Reviewer for the comment. We have revised the statement as below:

Line 246-249: "For DNA-based evaluation of *nifH* amplicons with NGS, the majority (>85%) of the *nifH* amplicons belonged to cyanobacteria and γ-proteobacteria (Table S6). There was considerable station to station variability but across all stations cyanobacteria and γ-proteobacteria were at approximately equal abundances in 2017. In contrast, a higher proportion of cyanobacteria (64%) was detected in 2018 compared to γ-proteobacteria (23%).".

Line 233-Here, I think there also needs to be an acknowledgement that at some stations *Trichodesmium* is much different than 58% of the cyanobacterial *nifH* abundances. For example at stn 1 in 2017 - *Trichodesmium* is ~75%. Like above, I would say that there is considerable variability but that overall *Trichodesmium* is 58% of cyanobacterial abundances.

Overall comment here - describing the overall pattern while acknowledging station to station variability will allow you to transition to the next section - because the station to station variability is consistent with what your correlation analyses show. For example, Stn 1 which is mostly Kuroshio waters has predominantly *Trichodesmium*.

Response:

We thank the Reviewer for the comment. We have revised the statement as below:

Line 249-251: "Overall, *Trichodesmium* and UCYN were approximately 58% and 42%, respectively, of cyanobacterial abundances despite station to station variability. The UCYN-C (22%) and UCYN-B (18%) were the domiant UCYN sublineages in 2017 and 2018, respectively (Table S6)."

Line 242 and/or Figure 5 - In Figure 5, I_{NFR} not S_{NFR} is in the same cluster as SST, SSS, DCM etc. Either there is a typo at line 242 or a typo in Figure 5.

Response:

We thank the Reviewer for the comment. We have changed "S_{NFR}" to "I_{NFR}" in Line 258.

Line 246 and/or Figure 5 - In Figure 5, S_{PP} looks like it is in a separate cluster from I_{DIN}, I_{DIP}, and Nit. If you meant the previous fork in the tree, I would list all six variables that are not in the strong KI cluster.

Response:

We thank the Reviewer for the comment. We have revised the description as below:

Line 259-261: "…, whereas the depth-integrated dissolved inorganic phosphorus (I_{DIP}), dissolved inorganic nitrogen (I_{DIN}), primary production (I_{PP}), surface nitrogen fixation rate (S_{NFR}), surface primary production (S_{PP}) and the depth of nitracline (Nit) with weak KI (Fig. 5).".

Line 257 - Based on figure 8, the variation due to only environmental factors is 0.13 in 2017. The variation due to only environmental factors as well as the variation due to both spatial and environmental factors is 0.33.

Response:

We thank the Reviewer for the comment. We have carefully addressed the comments and made the necessary revisions accordingly. The relative contribution of both components was explained by pure spatial variables, pure environmental variables, spatial variables, environmental variables and the combined effects of both space and environment. In the manuscript, our aim is to illustrate the influence of environmental or spatial factors on the variations in diazotrophic community structure. We regret any confusion arising from the statement "solely by environmental factors." To address this, we have removed the description pertaining to the variation in diazotrophic community structure explained exclusively by environmental factors, as indicated below:

Line 271-274: "The result of VPA further confirmed that the variation in diazotrophic community structure explained by environmental factors was higher in 2017 than in 2018 (0.33 vs. 0.22), whereas the influence of spatial factors was limited and similar between the two years (0.23 vs. 0.26) (Figs. 8a and 8b).".

Lines 275-276 - I would rephrase this sentence, because the 2006 paper does not show what this sentence describes. I would rephrase to "Additionally the relatively lower abundance of UCYN-B in Kuroshio waters may be because *Trichodesmium* has greater genetic resources which allow it to outcompete UCYN-B for P."

Response:

We thank the Reviewer for the comment. We have rephrased the description as below:

Line 289-291: "Additionally, the relatively lower abundance of UCYN B in Kuroshio may be due to the greater genetic resources of *Trichodesmium*, which allows it to outcompete UCYN-B for phosphorus (P) (Dyhrman et al., 2006)."

Line 285 - I think this point should be nuanced more. It looks like UCYN-C was not detected at all at station 10 and in station 9 was detected at lower levels than in 2017.

Response:

We thank the reviewer for the reminder. We have checked our data, and updated the Figure 2, Tables S4 and S5 accordingly.

Line 302-303: "Our data show that UCYN-C was among the least abundant diazotrophs in 2017 when KI was strong, but its relative abundance was 1–2 orders of magnitude higher in 2018 when KI was weak (Fig. 3).".
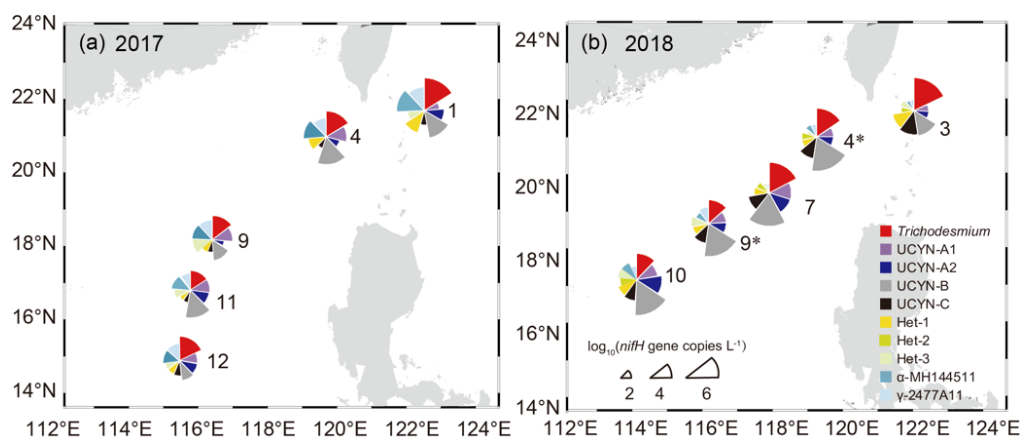

Figure 3. Surface daytime abundance of the ten major diazotrophic groups determined based on TaqMan qPCR assay of the *nifH* gene copies in samples collected in 2017 (a) and 2018 (b). Sector radius represents log-transformed absolute abundance of the *nifH* gene copies, with the angle indicating the relative proportion of each group. For the stations sampled in both years, the station number in 2018 are indicated with an asterisk sign (*).

Lines 288-293 - This paragraph is very confusing. It is difficult to figure out what argument the authors are trying to make.

Response:

We thank the Reviewer for the comment. We regret any confusion arising from the statement. Here we try to explain the differences in the relative abundances of UCYN-C and NCDs between 2017 and 2018, and attribute this primarily to the variations in the adaptive capabilities of different N_2-fixing phylotypes in response to KI. Additionally, we conclude that KI influences the distribution of these diazotrophs mainly as a stochastic factor.

Line 331 - I think predominated is a more accurate word than was distributed.

Response:

We thank the Reviewer for the comments. We have replaced the "distributed" to "predominated".

Line 356-357: "Specifically, *Trichodesmium* was predominated at the stations with high SSS and SST and deep MLD...".

Line 335 - This difference between selective and neutral is not discussed before. It isn't clear to me what makes spatial factors neutral. Please add a section to the methods that defines this.

Response:

We thank the Reviewer for the comment. We have added a section to the methods that defines the difference between selective and neutral processes.

Line 350-356: "Differences in the types and abundances of diazotrophs may arise through selection-driven (deterministic) and/or non-selection driven (stochastic) processes. Deterministic processes may drive differences between communities through species sorting in response to local environmental conditions, while stochastic processes may generate variation through a combination of other assembly processes including dispersal limitation, community drift and speciation (Hughes et al., 2008; Hanson et al., 2012). These stochastic processes—which we define here as 'neutral' processes—are considered in ecological neutral theories and are predicted to produce variation in community structure through space without needing to invoke the actions of selection.".

Line 338 - Another interpretation could be that the rare taxa are not constrained by environmental variables at all (i.e. rare taxa are always rare in this environment)- a really interesting result. I would further discuss this.

Response:

We thank the Reviewer for the comment. We have added the description about the importance of these new groups as below:

Line 363-366: "This result suggests that the AT* and RT subcommunities may response differently to environmental variables, or that RT may not bet constrained by environmental variables possibly due to their low growth rate, low competition potential, and narrow resource range (Pedros-Ali., 2006; Reveillaud et al., 2014).".

Line 339 - I would replace featuring with "correlated with" or "connected to".

Response:

We thank the Reviewer for the comment. We have replaced "featuring" with "correlated with".

Line 365-366: "Among the environmental parameters correlated with KI, temperature is a major factor controlling the distribution of diazotrophs.".

Line 362 - Please explain more - why especially UCYN-A?

Response:

We thank the Reviewer for the comment. To avoid the confusion, we have delete "(especially the UCYN-A)" from sentence.

Line 387-389: "Since N_2 fixation by marine diazotrophs has been proposed as one of the potential negative feedback mechanisms corresponding to ocean warming (Sohm et al., 2011), Kuroshio may transport diazotrophs in the upstream warmer regions including SCS northward to higher latitudes, resulting in a wider distribution of N_2 fixation in the global ocean.".

Lines 379-380 - There is a positive correlation between NFR and KI at just one of the strong KI stations so I don't think this claim is justified.

Response:

We thank the Reviewer for the comment. To avoid the confusion, we have delete "and is responsible for the enhanced N$_2$ fixation" from the sentence.

Line 402-404: "KI has a dominant dilution effect on the nutrient inventory on one hand, it also causes redistribution of the diazotrophic taxa and reallocation of the nutrients on the other hand."

Lines 383-384 - There is already N$_2$ fixation in temperate waters that is high - see Tang et al 2019 paper. I think the bigger picture importance of the paper is better said along the lines of the last sentence in the abstract.

Response:

We thank the Reviewer for the comment. We have revised the sentence as below:

Line 404-406: "As KIs are projected to intensify in a future warming ocean, Kuroshio may potentially cause a wider distribution of diazotrophs at high latitudes".

Figure 2 - Please make the numbers on the contours and axes bigger. They are very very small and hard to read right now. Please also clarify, is the data in panel A the real data used to model the points in panels b-e? I would state this in the legend.

Response:

We thank the Reviewer for the comment. We have made the numbers on the contours and axes bigger. The data in (a) are the real data used to model the points in panels (b-e).
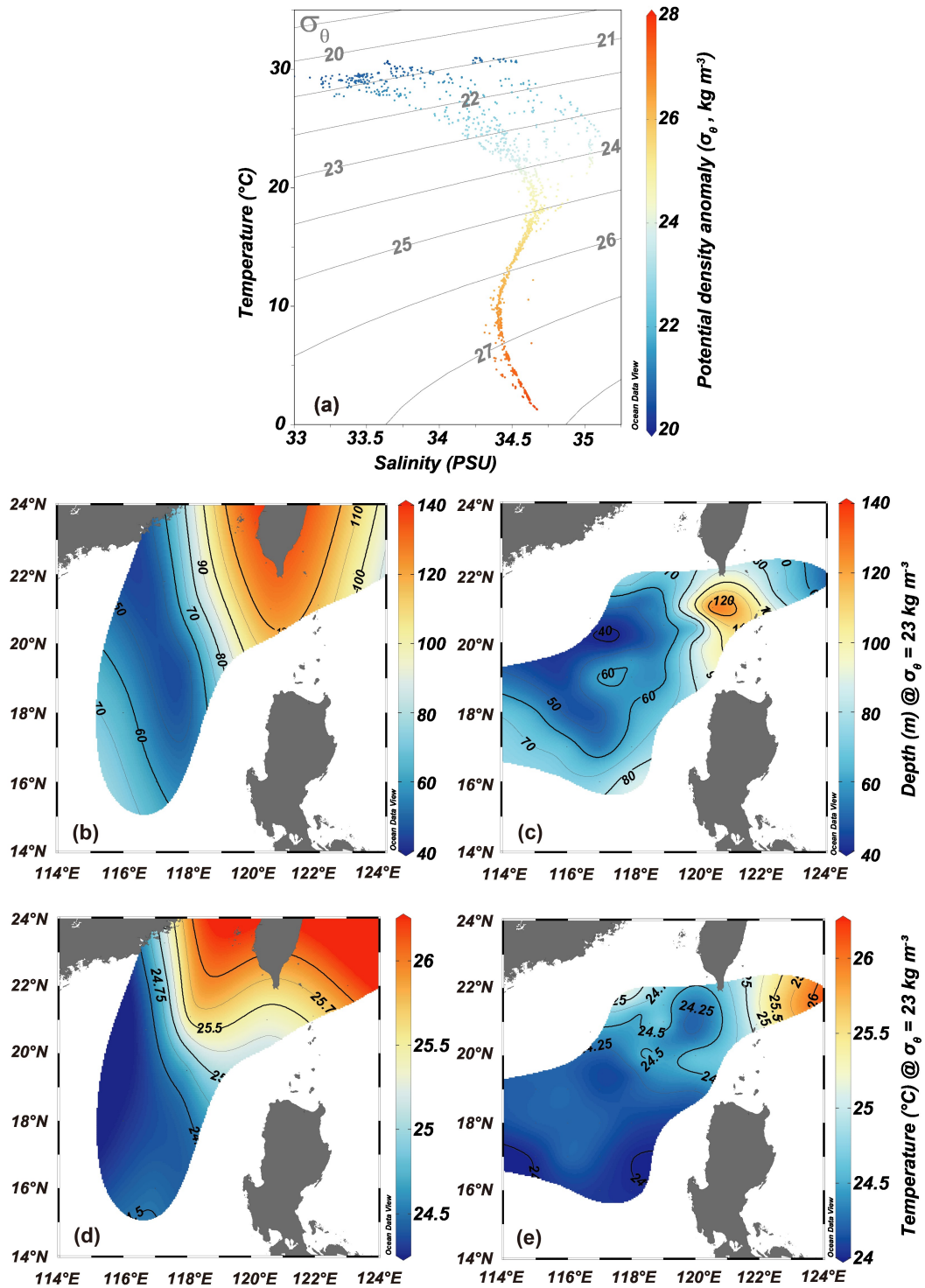
Figure 2. Potential temperatures, salinities and density anomalies modeled for the cruises in 2017 and 2018. (a) Plot of *θ-S* showing potential temperatures (*θ*, °C) and salinities (S, PSU) of water parcels resulting from mixing of the Kuroshio and SCS water masses. Potential density anomalies (σ_θ, kg m^{-3}), shown in grey lines, are imposed on the *θ-S* plot. (b–e) Profiles of depths (b, c) and temperatures (d, e) of isopycnal surface along the σ_θ of 23 kg m^{-3} of water masses depicted for the cruises in 2017 (b, d) and 2018 (c, e). The data points in (a) are the real data used to create the contour plots in panels b–e.

References

Blanchet, F. G., Legendre, P., and Borcard, D.: Forward selection of explanatory variables, Ecology, 89, 2623-2632, https://doi.org/10.1890/07-0986.1, 2008.

Chen, W., Ren, K., Isabwe, A., Chen, H., Liu, M., and Yang, J.: Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons, Microbiome, 7, 138, https://doi.org/10.1186/s40168-019-0749-8, 2019.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H.: Beyond biogeographic patterns: Processes shaping the microbial landscape, Nat Rev Microbiol, 10, 497-506, https://doi.org/10.1038/nrmicro2795, 2012.

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., and Vellend, M.: Ecological consequences of genetic diversity, Ecol Lett, 11, 609-623, https://doi.org/10.1111/j.1461-0248.2008.01179.x, 2008.

Pedrós-Alió, C.: Marine microbial diversity: can it be determined? Trends Microbiol., 14, 257-263, https://doi.org/10.1016/j.tim.2006.04.007, 2006.

Quince, C., Lanzen, A., Davenport, R. J., and Turnbaugh, P. J.: Removing noise from pyrosequenced amplicons, BMC Bioinformatics, 12, 38, https://doi.org/10.1186/1471-2105-12-38, 2011.

Reveillaud, J., Maignien, L., Eren, A. M., Huber, J. A., Apprill, A., Sogin, M. L., and Vanreusel, A.: Host-specificity among abundant and rare taxa in the sponge microbiome, The ISME Journal, 8, 1198-1209, https://doi.org/10.1038/ismej.2013.227, 2014.

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 N-2 fixation and estimating diazotrophic contribution to marine production, Nat. Commun., 10, https://doi.org/10.1038/s41467-019-08640-0, 2019.

Zhang, M., Delgado-Baquerizo, M., Li, G., Isbell, F., Wang, Y., Hautier, Y., Wang, Y., Xiao, Y., Cai, J., Pan, X., and Wang, L.: Experimental impacts of grazing on grassland biodiversity and function are explained by aridity, Nat. Commun., 14, 5040, https://doi.org/10.1038/s41467-023-40809-6, 2023.