Jingwei-Nutrients: A global spatiotemporal reconstruction of ocean nutrients (1965–2023) using multi-task deep learning

Dear Editor,

We sincerely thank you and the reviewers for the careful evaluation of our manuscript and for the constructive comments and suggestions. We have carefully considered all comments raised by the reviewers and revised the manuscript accordingly. We believe that the revised manuscript has greatly benefited from these valuable suggestions.

In the following response document, we provide a detailed point-by-point reply to each reviewer comment, together with the corresponding revisions and additional results. We are now submitting the revised manuscript and the response summary for your consideration.

Thank you for your attention.

Sincerely,

Zhaokun Wang, on behalf of all co-authors

Response to Reviewer #1

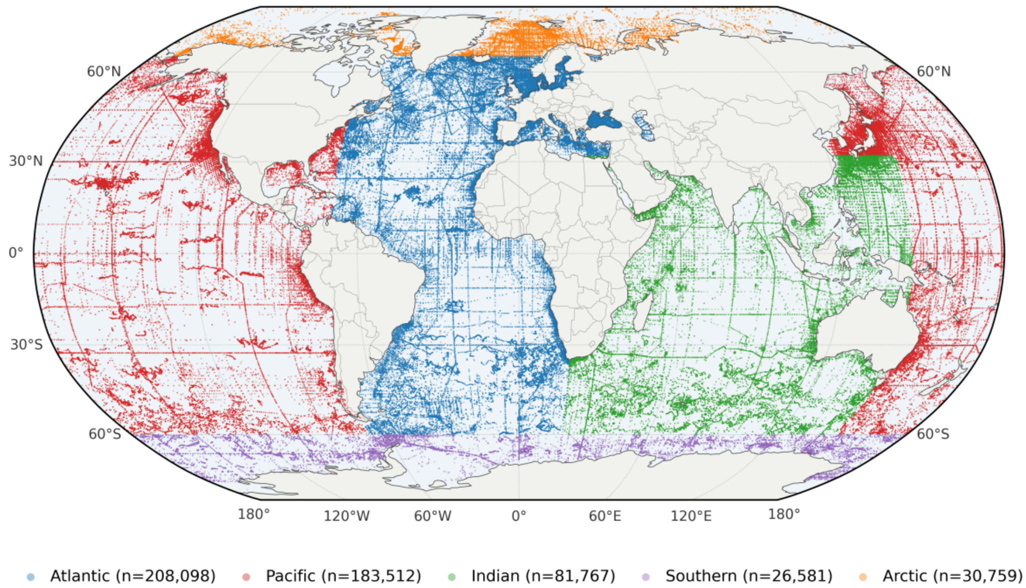
We sincerely thank the reviewer for the positive evaluation of our manuscript and for the constructive suggestions. We are grateful that the reviewer recognized the value of the proposed global nutrient reconstruction dataset and the detailed preprocessing procedures. We have revised the manuscript accordingly to address the reviewer's concerns and improve the overall clarity of the manuscript. Our point-by-point responses are provided below.

Comment 1. The manuscript provides detailed descriptions of the modeling framework and comparison experiments, especially the independent validations conducted for different time periods, which effectively demonstrate the robustness of the proposed workflow. However, I suggest that the authors additionally train and evaluate the model using the complete temporal dataset (i.e., all periods combined) and provide the corresponding results for reference. Furthermore, in the Abstract, are the reported values of 0.980, 0.961, and 0.983 intended to represent Average Accuracy?

Response:

A1: Model Evaluation on Complete Temporal Dataset. We thank the reviewer for this helpful suggestion. We agree that an additional experiment using the complete temporal dataset provides a useful reference for readers.

To address this comment, we conducted an additional all-period experiment using the complete observational dataset from 1965 to 2023. We constructed an independent held-out validation set by sampling according to ocean basin, in which approximately 10% of the observations in each basin were reserved for validation; the spatial distribution of these validation observations is shown in Response Fig. R1. The remaining observations were used for model training. Based on the same held-out validation set, we evaluated two reconstructed products: (1) a full-period product trained using all remaining training samples, and (2) a 6-fold ensemble product obtained by averaging the predictions from six fold-specific models. The reconstructed nutrient fields were then sampled at the corresponding observation time, depth, latitude, and longitude using nearest-neighbor matching, and compared with the reserved observations.



Response Fig. R1. Spatial distribution of the validation observations used in the all-period experiment.

The overall validation results are summarized in Response Table R1.

Response Table R1. Validation performance of the K-fold ensemble product and the full-period product on the independent held-out validation set used in the all-period experiment.

		K-Fold Product		Full-Period Product	
Nutrient	Number	RMSE	R^2	RMSE	R^2
Nitrate	314,758	1.993	0.982	2.047	0.979
Phosphate	272,824	0.207	0.963	0.217	0.959
Silicate	223,758	4.797	0.981	5.135	0.977

The comparison shows that both the K-fold product and the full-period product achieve strong agreement with the hold-out observations. The K-fold product gives slightly better performance for all three nutrients, with lower RMSE and higher R^2 than the full-period single-model product.

This improvement is likely attributable to the ensemble effect of the K-fold product. Although the full-period product is trained as a single model using all available training samples, its predictions may still be influenced by local observational noise, random initialization, training stochasticity, and specific characteristics of the training distribution. In contrast, the K-fold product averages predictions from multiple independently trained models, each learned from a different subset of the training data. This model averaging can reduce prediction variance and smooth out fold-specific or model-specific random errors. As a result, the K-fold mean product provides more stable reconstructions and slightly lower RMSE values than the full-period single-model product. Therefore, the additional experiment supports the robustness of the proposed workflow and suggests that the K-fold ensemble strategy is beneficial for reducing random reconstruction errors in the global nutrient fields.

A2: Reported Value in the Abstract. We also clarified the meaning of the values reported in the Abstract. The values 0.980, 0.961, and 0.983 are the mean R^2 values across *the chronological K-fold validation folds* for nitrate, phosphate, and silicate, respectively. Specifically, they correspond to the fold-averaged R^2 values of the MTL model reported in the "Avg" row of Table 3. To avoid ambiguity, we revised the Abstract to explicitly state "mean validation R^2 across chronological K-fold cross-validation".

Manuscript revision:

We revised the Abstract to clarify that the reported values are mean R^2 values across the chronological K-fold validation folds.

"Evaluation on chronological K-fold cross-validation yields mean validation R^2 values across folds of 0.980, 0.961, and 0.983, with RMSEs of 2.21, 0.23, and 6.35 μmol kg^{-1} for nitrate, phosphate, and silicate, respectively."

Comment 2. Lines 45–46: The study by Deutsch indeed suggests strong relationships among multiple nutrients, and similar concepts are reflected by Prof. Redfield. However, it remains unclear why the reconstructed dataset itself should inherently preserve such nutrient interrelationships. Please provide a more in-depth explanation and justification.

Response:

Thanks for your comment. In our work, we adopted the Multi-Task Learning (MTL) framework as a flexible, data-adaptive mechanism to encourage consistency with observed nutrient interrelationships. In the shared Transformer backbone, gradients from the three nutrient-specific prediction heads jointly update the shared representation. Therefore, information from one nutrient can help shape the latent hydrographic-biogeochemical representation used by the other nutrients, without forcing the model to follow a fixed stoichiometric ratio. This mechanism allows the model to capture flexible, regionally varying nutrient relationships rather than imposing a globally uniform relationship.

Besides, we would like to emphasize that the reconstructed dataset does not preserve nutrient interrelationships through a hard-coded Redfield ratio or any explicit stoichiometric constraint. As discussed in the Introduction, although nitrate, phosphate, and silicate are biogeochemically linked through biological uptake, remineralization, and water-mass processes, their ratios are not globally fixed. Instead, nutrient relationships vary across regions, depths, seasons, and long-term temporal scales. Therefore, imposing a rigid empirical ratio could introduce unrealistic constraints and potentially degrade the reconstruction in regions where nutrient stoichiometry deviates from the canonical relationship.

Comment 3. Related to the previous comment, Lines 324–332 Figure 3 indicates that the multi-task learning (MTL) framework does not appear to provide substantial improvements over the single-task approach for nitrogen (N) and silicate (Si). What are the underlying reasons for this behavior? Please analyze the potential sources of error and discuss why the benefits of MTL differ among nutrients. In addition, please specify the value of K used in the K-fold cross-validation.

Response:

We thank the reviewer for this important comment. We agree that the visual difference between MTL and STL in Fig. 3 may appear modest, especially because both models already achieve high overall R^2 values. However, the quantitative metrics in Table 3 indicate that MTL consistently improves performance across all three nutrients. The fold-averaged RMSE reductions are significant, approximately 24.2% for nitrate, 24.4% for phosphate, and 8.9% for silicate.

For nitrate, the relatively modest visual improvement does not indicate that MTL provides little benefit. Rather, the STL nitrate model already achieves a high R^2, leaving limited room for further improvement in the R^2 metric and in the density scatter plot. Nevertheless, nitrate shows a substantial RMSE reduction under MTL, comparable to that of phosphate. Therefore, the improvement for nitrate is more clearly reflected by RMSE than by the visual difference in Fig. 3.

For silicate, the smaller improvement is likely related to its more distinct biogeochemical controls. Unlike nitrate and phosphate, which are more directly coupled through biological uptake and remineralization, silicate is strongly affected by additional processes such as diatom uptake, opal dissolution, and deep-water accumulation. These processes are not always tightly synchronized with nitrate and phosphate, reducing the amount of transferable information from the other nutrient tasks and leading to a smaller MTL gain for silicate.

We also clarified the K value used in the chronological K-fold cross-validation. In this study, K = 6, corresponding to the six chronological folds for nearly per decade: 1965–1975, 1976–1985, 1986–1995, 1996–2005, 2006–2015, and 2016–2023.

Manuscript revision:

We revised Section 2.5.4 to explicitly specify the value of K used in the chronological K-fold cross-validation.

"We adopt a strict time-based K-fold cross-validation strategy to prevent temporal data leakage. In this study, K = 6, and the 59-year dataset is divided into six chronological folds: 1965–1975, 1976–1985, 1986–1995, 1996–2005, 2006–2015, and 2016–2023. By partitioning the dataset into chronological blocks rather than random subsets, distinct time periods are rigorously isolated as test sets, providing an objective assessment of the model's generalization."

Comment 4. Line 103: The manuscript states that "the MTL architecture maximizes the utility of fragmented historical observations." What evidence supports this statement? In particular, what periods are referred to as "data-sparse eras"? Please provide a clearer definition and supporting analysis.

Response:

We appreciate this suggestion. In this study, "data-sparse eras" do not refer to a single fixed period that is identical for all three nutrients. Rather, they refer to periods in which the available observations for a given nutrient are limited in number or unevenly distributed in space, as shown by the spatiotemporal observation distribution in Fig. 1. For example, nitrate observations are relatively sparse in the earliest chronological fold, especially 1965-1975. For phosphate and silicate, observational sparsity remains evident in the recent period after 2006, because the modern BGC-Argo network provides a large number of nitrate observations but does not provide corresponding phosphate or silicate measurements. Therefore, the fragmented nature of the historical archive is both time-dependent and nutrient-dependent.

As described in the original manuscript, our MTL framework is designed to make better use of this fragmented observation record by jointly learning multiple nutrient-related tasks. Unlike a training strategy that requires nitrate, phosphate, and silicate to be simultaneously available, the MTL framework can use samples that contain valid temperature/salinity predictors and at least one available nutrient target. During training, samples with partial nutrient labels still contribute to the shared feature extractor through their available task-specific losses. In this way, observations of one nutrient can help improve the shared representation used by the other nutrient tasks, allowing different nutrient records to mutually complement each other within the same model. The effectiveness of this design is first supported by the STL-MTL comparison in Table 3, where the MTL framework consistently reduces the fold-averaged RMSE for nitrate, phosphate, and silicate compared with the corresponding single-task models.

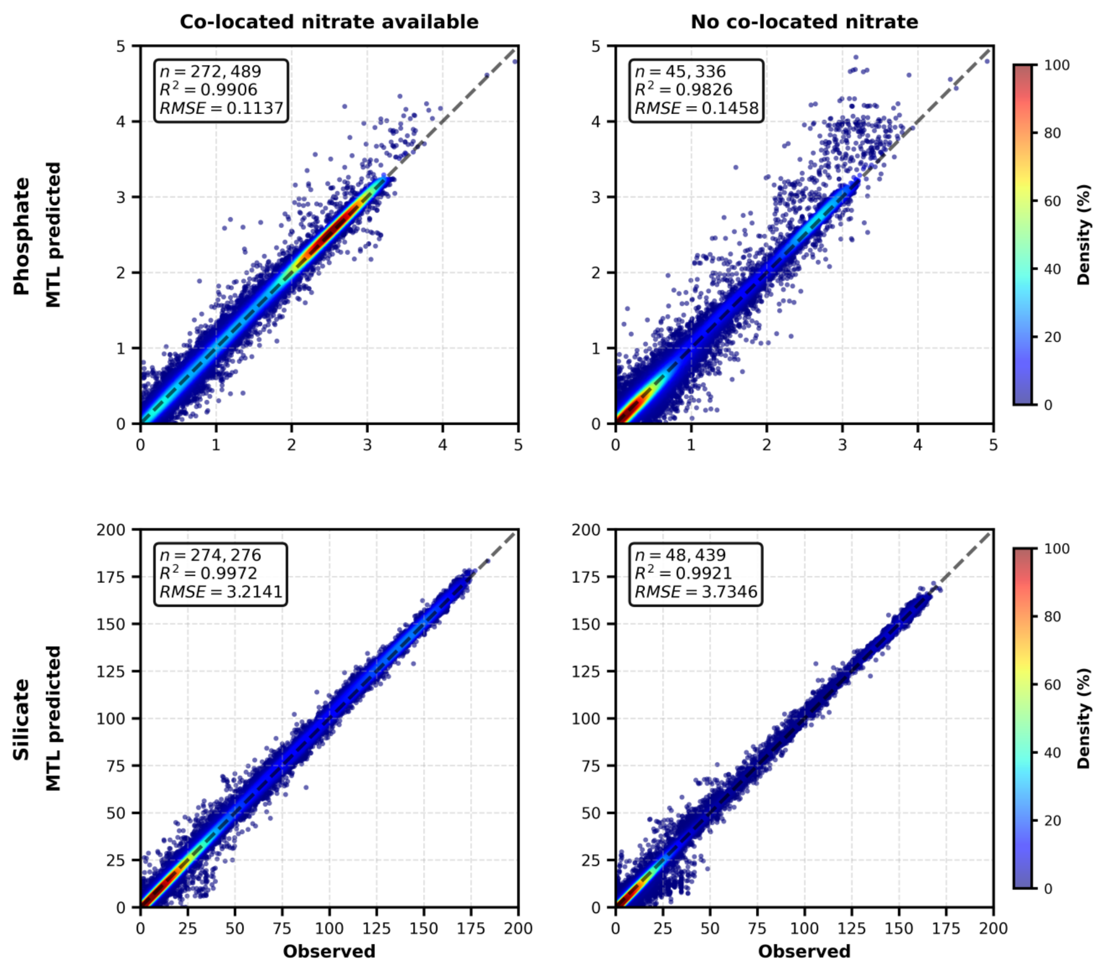
To further examine whether such fragmented observations can provide useful information for nutrient reconstruction, we conducted an additional last-fold withholding experiment. The final chronological fold, 2016-2023, was used as a controlled inference period. In this experiment, all phosphate and silicate observations from 2016-2023 were excluded from model training and retained only as independent observational references. The MTL model was trained using nitrate, phosphate, and silicate observations from the earlier folds, together with nitrate observations from 2016-2023. In contrast, the STL phosphate and silicate models were trained only with phosphate or silicate observations from the earlier folds, respectively. After training, both approaches were applied to infer phosphate and silicate concentrations for 2016-2023, and the inferred results were compared with the withheld phosphate and silicate observations from the same period. To reduce the influence of random initialization, the experiment was repeated multiple times, and the mean performance metrics are reported in Response Table R2.

Response Table R2. Performance of the last-fold withholding experiment.

	Multi-Task Learning		Single-Task Learning	
Nutrient	R^2	RMSE	R^2	RMSE
Phosphate	0.9892	0.1189	0.9679	0.1401
Silicate	0.9968	3.2998	0.9884	3.8713

The results show that the MTL model outperforms the STL models for both phosphate and silicate in the controlled inference period. For phosphate, R^2 increases from 0.9679 to 0.9892, while RMSE decreases from 0.1401 to 0.1189 μmol kg^{-1}. For silicate, R^2 increases from 0.9884 to 0.9968, and RMSE decreases from 3.8713 to 3.2998 μmol kg^{-1}. These consistent RMSE reductions indicate that nitrate observations available in the target period provide useful shared information for reconstructing phosphate and silicate, supporting the ability of the MTL framework to utilize fragmented historical observations.

To further visualize the cross-nutrient auxiliary effect of nitrate information, we additionally grouped the withheld phosphate and silicate samples in the final fold according to whether valid collocated nitrate observations were available at the same sampling point. The observed-versus-predicted scatter plots for the MTL model are shown in Response Fig. R2.



Response Fig. R2. Observed-versus-predicted scatter plots used to examine the cross-nutrient auxiliary effect of nitrate information in the MTL framework.

The withheld phosphate and silicate samples from 2016–2023 were divided into two groups according to whether valid co-located nitrate observations were available at the same sampling point. Scatter plots compare observations and MTL predictions, with colors indicating point density and dashed lines showing the 1:1 reference. Samples with co-located nitrate observations show higher R^2 and lower RMSE for both phosphate and silicate, indicating that nitrate information provides useful cross-nutrient support for reconstructing phosphate and silicate in the MTL framework.

Comment 5. Line 115: The term "physically consistent" is used multiple times throughout the manuscript. Its meaning should be explicitly defined and supported by appropriate references. Expressions such as "considering physical correlations" or "integrating biogeochemical information" may be more precise. A similar concern applies to the term "physically interpretable" in Line 389.

Response:

We agree with the reviewer that these terms should be used more carefully. In our manuscript, "physically consistent" does not mean that the model explicitly solves physical governing equations. Rather, it refers to the agreement of the reconstructed nutrient fields with known hydrographic and biogeochemical features, including global spatial patterns, vertical nutrient structures, seasonal cycles, station-scale temporal variations, and cruise-section distributions. These aspects are evaluated through the WOA comparison, seasonal-cycle analysis, HOT/KERFIX station validation, and GO-SHIP cruise verification. To avoid overstatement, we clarified the meaning of this term and replaced some occurrences with more precise expressions.

Manuscript revision:

We revised the two expressions to make their meanings more specific and avoid overstatement.

For Line 115, we replaced "physically consistent foundation" with a more precise expression:

"Ultimately, the Jingwei-Nutrients dataset offers a validated and biogeochemically coherent baseline for investigating multi-decadal ocean biogeochemical dynamics and ecosystem responses to global climate change."

For Line 389, we replaced "physically interpretable" with "consistent with known oceanographic processes":

"Overall, these regional variations are consistent with known oceanographic processes, and the comparison indicates that the Jingwei-Nutrients product provides a consistent representation of global biogeochemical distributions."

Comment 6. Table 1: The numbers of observations for nitrogen and phosphorus are relatively similar, which may weaken the effectiveness of the multi-task learning framework. However, phosphorus exhibits a good performance improvement under MTL. How can we explain this result? Please provide further discussion.

Response:

We thank the reviewer for this insightful comment. Although the total numbers of nitrate and phosphate observations are relatively similar, their data-source compositions and spatiotemporal distributions are different. In particular, BGC-Argo provides a large number of nitrate observations but does not provide phosphate observations. Therefore, nitrate is more strongly constrained by direct observations, especially in recent decades, whereas phosphate can benefit from nitrate-rich information through the shared MTL representation.

We also note that nitrate itself benefits substantially from MTL. As shown in Table 3, the fold-averaged RMSE of nitrate is reduced from 2.9216 μmol kg^{-1} in the STL model to 2.2138 μmol kg^{-1} in the MTL model, corresponding to a reduction of approximately 24.2%. This reduction is comparable to the RMSE reduction for phosphate, which is approximately 24.4%. However, the STL model already achieves a high R^2 for nitrate, leaving limited room for further R^2 improvement. In contrast, phosphate has a lower STL baseline R^2, so the improvement introduced by MTL is more clearly reflected in the R^2 metric.

This asymmetric behavior is also consistent with the biogeochemical relationships among nutrients. Nitrate and phosphate are tightly coupled through biological uptake and remineralization over large parts of the ocean. Thus, phosphate can benefit from nitrate-rich observations through the shared MTL backbone, particularly in regions or periods where phosphate observations are relatively sparse. By contrast, nitrate already has stronger direct observational constraints, so the additional information transferred from phosphate produces a smaller marginal gain in R^2, even though the RMSE improvement remains substantial.

Comment 7. Line 243: The term "natural language" appears unrelated to the scope of this study and should be reconsidered. Line 263: I do not fully agree with the description involving "biogeochemical constraints." Please consider revising the terminology.

Response:

Thank you for pointing out this concern. The phrase "natural language" was used only to introduce the original development context of the Transformer architecture, but it is not necessary for this study and may distract from the oceanographic focus. We revised the sentence to state that the Transformer was "*originally developed for sequence representation learning*" and has since been adapted to Earth system and environmental modeling.

Regarding the phrase "biogeochemical constraints," our intention was to describe the implicit coupling among nutrients introduced by the MTL architecture. In our framework, this coupling is not imposed through fixed empirical ratios. Instead, it is implemented through the shared Transformer backbone, which is jointly optimized by the nitrate, phosphate, and silicate prediction tasks. During training, the learning signals from different nutrients are propagated back to the shared feature extractor, allowing the model to learn a common hydrographic-biogeochemical representation. In this way, each nutrient task can benefit from information provided by the others while still retaining its own task-specific prediction head. Because nutrient relationships vary across regions, depths, seasons, and long-term temporal scales, this shared-representation strategy provides a more flexible way to incorporate multi-nutrient relationships than imposing a fixed stoichiometric constraint. To make this point clearer, we revised the wording from "biogeochemical constraints" to more precise expressions.

Manuscript revision:

We replaced "natural language processing" with "sequence representation learning."

"This framework, originally developed for sequence representation learning (Vaswani et al., 2017), has been recently adapted for high-dimensional Earth system modeling (e.g., Bi et al., 2023; Lam et al., 2023) and complex oceanographic field reconstructions (Reichstein et al., 2019; Sonnewald et al., 2021)."

We replaced "specific biogeochemical constraints" with "information learned by each nutrient decoding head."

“Crucially, during training, the information learned by each nutrient decoding head is continuously propagated back to the shared layers. By jointly shaping the shared representation space, the MTL framework enables the joint learning of coherent biogeochemical patterns.”

Comment 8. Line 298: The word “constructive” is enclosed in quotation marks, which is somewhat confusing. Please provide a more direct and precise description rather than a figurative expression. Similarly, the phrase “autonomously infer” in Line 274 sounds unusual in this context and could be revised for clarity.

Response:
We agree with the reviewer and have revised these expressions to make the description more direct and precise. The phrase “autonomously infer” was replaced with “learn from observations,” which better describes the data-driven learning process. In addition, the figurative expression involving a “constructive” manifold was replaced with a clearer description of the role of PCGrad. Specifically, we now state that PCGrad reduces destructive gradient interference by modifying conflicting task gradients, thereby helping the shared parameters update in directions that are less detrimental to other tasks.

Manuscript revision:
We revised the wording in Section 2.5.3.

“Consequently, we employ a data-driven multi-task learning approach, enabling the neural network to learn and represent these dynamic elemental relationships from observations within its shared latent space.”

We also revised the PCGrad description in Section 2.5.3.

“By iteratively applying this projection, PCGrad mitigates inter-task gradient conflicts by suppressing destructive gradient components, thereby facilitating more balanced optimization of the shared parameters and reducing negative transfer among nutrient predictions.”

Comment 9. Figure 10: Regarding the HOT station, please clarify whether the comparison is based on dedicated long-term HOT station observations or on data extracted from the corresponding latitude and longitude location.

Response:

The observational reference used in Fig. 10 is the dedicated long-term HOT station dataset, rather than a subset of all merged observations located near the HOT coordinates. Specifically, we used the official HOT time-series nutrient measurements as the independent observational benchmark. The reconstructed values were extracted from the nearest Jingwei-Nutrients grid cell and matched to the HOT records according to the corresponding month and depth level. We have revised Section 3.3 and the caption of Fig. 10 to make this procedure explicit.

Comment 10. Several figures would benefit from labeling subpanels with letters (e.g., a, b, c, d) to facilitate clearer descriptions and discussions in the text.

Response:

Thanks for your kind advice. We have added subpanel labels to the relevant figures and revised the text and captions to refer to the labeled subpanels where appropriate.

Response to the Second Round Comments from Reviewer #1

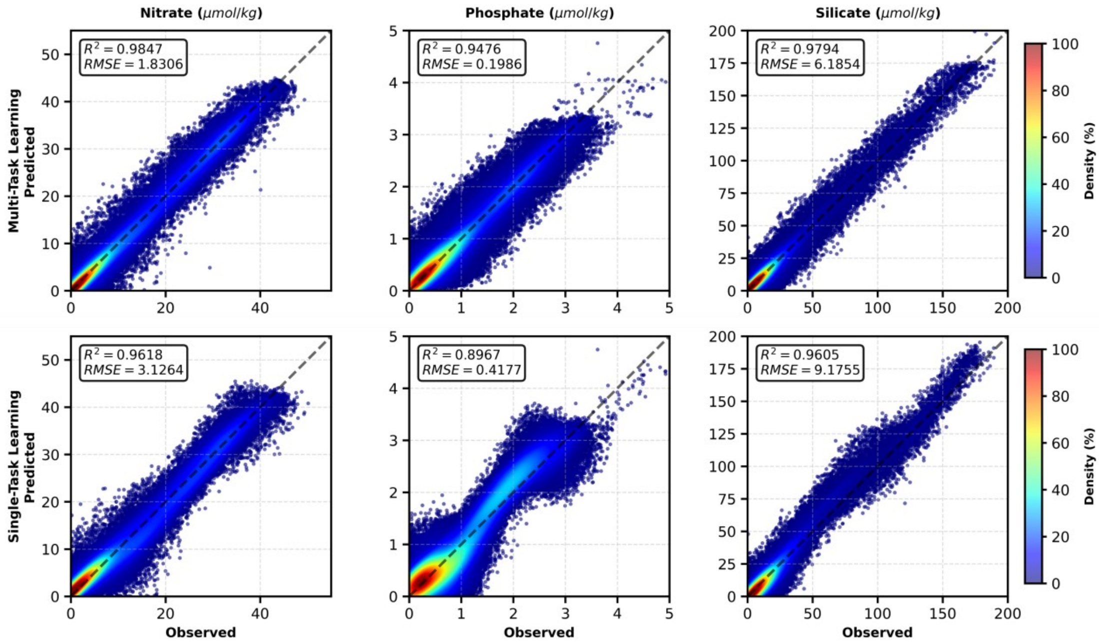
We sincerely thank the reviewer for the positive evaluation of our revised manuscript and for the constructive additional suggestions. We are grateful that the reviewer recognized the adequacy of our previous responses, the newly added validation analyses, and the improvements made to the manuscript. We have carefully considered the remaining comments and further revised the manuscript accordingly. In particular, we added supplementary endpoint extrapolation results and improved the presentation of the data quality control workflow and related methodological descriptions. Our point-by-point responses are provided below.

Comment 1. Using 50 years of data to predict the other 10 years need more data supporting. Averaging results can reduce part of the errors, but two extrapolation tests still need supplementary verification: reconstructing the earliest 10 years with the latter 50 years, and predicting the latest 10 years using the former 50 years. Unlike interpolation for other periods, these two settings are typical extrapolation tasks. I suggest the authors give detailed results of the two tests, including scatter plots and corresponding evaluation data.

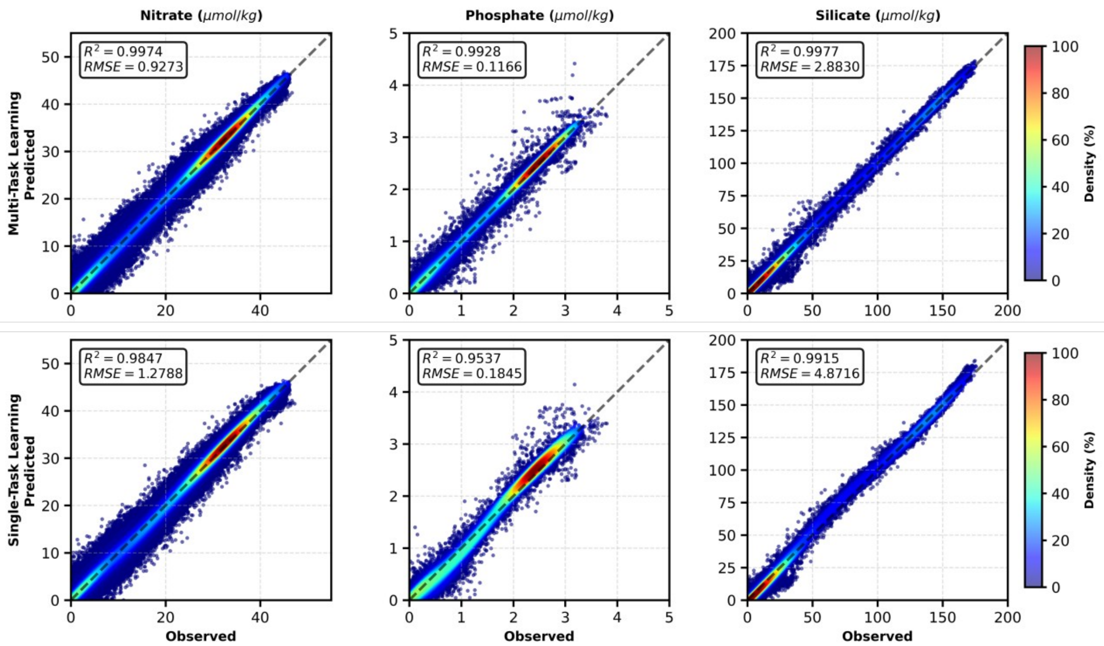
Response：

We thank the reviewer for this helpful suggestion. We agree that the first and last chronological folds represent more stringent extrapolation tests than the intermediate folds. To address this point, we explicitly examined the two endpoint extrapolation settings: reconstructing the earliest fold using the later folds and reconstructing the latest fold using the earlier folds.

The detailed scatter plots and corresponding R^2 and RMSE values are provided in Response Fig. R1 and Response Fig. R2. Overall, the predicted values remain generally aligned with the 1:1 reference line for nitrate, phosphate, and silicate, providing additional support for the temporal generalization capability of the proposed reconstruction framework under the two endpoint extrapolation settings.



Response Fig. R1. Density scatter plots for the earliest-fold extrapolation test.

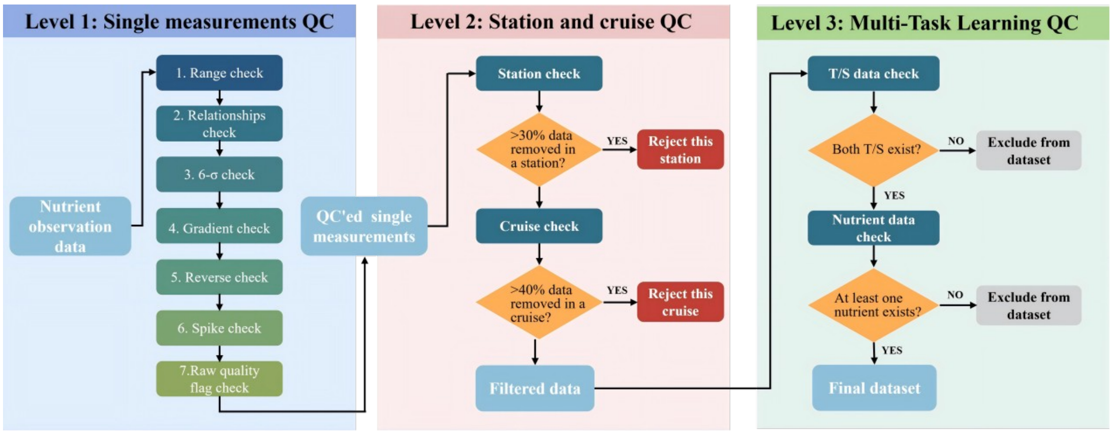


Response Fig. R2. Density scatter plots for the latest-fold extrapolation test.

Comment 2. Besides, please optimize the presentation of key data quality control steps, descriptions of comparative experiments and main model figures, and polish the language to improve readability.

Response：

We thank the reviewer for this valuable suggestion. We have further optimized the presentation of the manuscript to improve readability. First, we added a schematic workflow figure to summarize the hierarchical data quality control procedure, which consists of three levels: single-measurement QC, station and cruise QC, and the final data-selection step for multi-task learning. The first level removes abnormal individual measurements through range, relationship, statistical, gradient, reversal, spike, and raw-quality-flag checks. The second level excludes unreliable stations and cruises with high rejection rates. The final level retains samples with valid temperature and salinity and at least one available nutrient target for model training. This figure provides a clearer overview of how raw nutrient observations are filtered and prepared before being used in the multi-task learning framework.



Response Fig. R3. Schematic workflow of the hierarchical data quality control procedure.

In addition, we revised the descriptions of the comparative experiments to make the experimental settings, validation logic, and differences between MTL and STL models clearer. We also refined the presentation of the main model-related figures and polished the relevant text throughout the manuscript to improve readability.

Response to Reviewer #2

We sincerely thank the reviewer for the careful evaluation of our manuscript and for the constructive and insightful comments. We greatly appreciate the reviewer's recognition of the scientific value of the Jingwei-Nutrients data product, the motivation of using a multi-task learning framework for joint nutrient reconstruction, and the importance of comprehensive validation for a global four-dimensional biogeochemical product. The comments raised by the reviewer are particularly valuable because they point to several key aspects that determine the reliability and usability of the data product, including the independence of external validation datasets, the physical consistency of reconstructed nutrient ratios, the quality control of BGC-Argo nitrate observations, and the need for more direct comparisons with existing nutrient products.

In response to these comments, we have carefully revised the manuscript and conducted a series of additional analyses and experiments. *First, we clarified the validation-data withholding strategy and explicitly stated that the GO-SHIP cruise sections and the HOT/KERFIX station observations used for independent validation were excluded from the training dataset before model development. Second, to more directly evaluate the biogeochemical consistency of the reconstruction, we added new analyses of nutrient stoichiometric ratios. Third, in response to the concern about BGC-Argo nitrate quality control, we recollected and used adjusted BGC-Argo nitrate profiles with calibration information, reconstructed the training dataset, and reran the full set of model training, validation, and product-generation experiments. Fourth, to better position Jingwei-Nutrients relative to currently available data products, we selected several existing short-term or regional nutrient products for additional comparison and provided quantitative evaluations and discussion of their similarities and differences.*

These revisions and additional experiments have substantially improved the methodological transparency, physical interpretability, and external comparability of the manuscript. We believe that the revised version more clearly demonstrates the independence of our validation framework, the reliability of the reconstructed nutrient fields, and the added value of Jingwei-Nutrients as a long-term global nutrient data product. Our detailed point-by-point responses are provided below.

Major comments

Comments 1 and 2.

Table 1 shows that CCHDO data contribute 69,355 nitrate, 98,652 phosphate, and 73,165 silicate profiles to the training dataset. I think that the GO-SHIP cruises used for independent spatial validation in Section 3.4 (P16N 2015, P16S 2014, P06E 2017) are disseminated through CCHDO. So, it looks like these cruise data were included in the training set. If it is the case, the spatial validation in Figs. 13, 14 and 15 is not independent. The authors must clarify whether these specific cruises were withheld from training, and if not, please provide alternative independent spatial validation (e.g., GEOTRACES sections or GO-SHIP lines not present in CCHDO).

Both HOT and KERFIX are long-running time-series stations whose data are archived in WOD, which contributes the majority of training observations. The authors do not state whether profiles from these two stations were explicitly withheld from training. If HOT data (running from the late 1980s to the present) were included in WOD training profiles, the apparent high R^2 values at HOT (≥0.994 for all three nutrients) may partly reflect data memorisation rather than generalisation. Please clarify whether HOT and KERFIX were excluded from training and, if not, recompute station-level metrics using withheld data only.

Response:

We thank the reviewer for raising this important concern. We fully agree that the independence of external validation datasets is essential for demonstrating the generalization capability of the reconstruction model. ***We would like to clarify that the GO-SHIP cruise sections and the HOT/KERFIX station observations used for validation were excluded from the training dataset before model training.*** Specifically, all observations from the GO-SHIP P16N (2015), P16S (2014), and P06E (2017) cruises, as well as all nutrient observations from the HOT and KERFIX time-series stations, were removed from the training dataset. Because these records may be distributed through public archives such as CCHDO and WOD, we identified them using cruise identifiers, station information, and spatiotemporal coordinates during data harmonization and deduplication. Any duplicate or overlapping records associated with these validation datasets were also excluded from the training samples. Therefore, the validation results reported for these stations and cruise sections are based on withheld observations rather than on samples used during model development. The withheld observations were used only as independent references for the station and cruise-section validations. We will revise the relevant descriptions in the manuscript to make this withholding procedure clearer.

Manuscript revision:

We also recognize that the original manuscript did not describe this withholding procedure clearly enough. Although we used the term "***independent validation***" in the manuscript, the validation-specific data exclusion strategy was not stated explicitly, which may have led to the misunderstanding that the validation records were included in the training dataset because Table 1 summarizes the quality-controlled multi-source archive. To avoid this ambiguity, ***we will revise the relevant descriptions in the Methods and Results sections to clearly state that the HOT, KERFIX, P16N, P16S, and P06E observations were withheld before model training and used only for independent validation.*** We will also clarify that Table 1 summarizes the quality-controlled observational archive before validation-specific withholding, whereas the actual training subset excludes the station and cruise observations used for external validation.

Comment 3. The primary motivation for the MTL approach is that jointly learning the three nutrients encourages physically consistent elemental ratios. However, the paper contains no direct evaluation of whether the reconstructed N:P or N:Si ratios are consistent, and whether they are more consistent in the MTL product than in the STL baseline. Given that this is the central scientific claim, the authors should compute and map Redfield ratios for example N:P across depth levels and regions and compare them with observations.

Response:

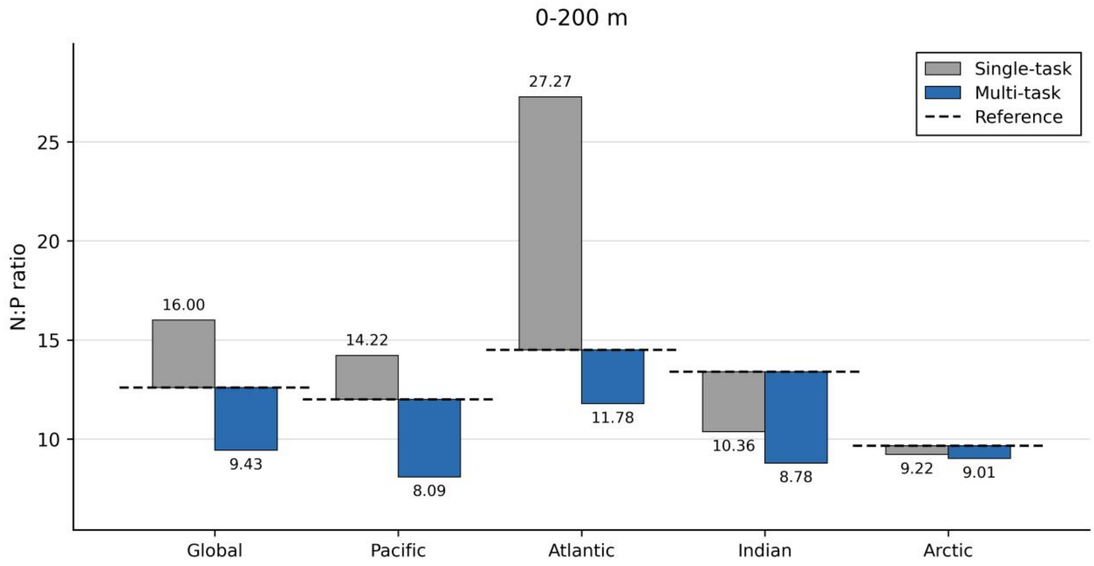
We thank the reviewer for this important suggestion. We agree that evaluating each nutrient separately using R^2 and RMSE is not sufficient to directly demonstrate whether the reconstructed product preserves the coupled biogeochemical relationships among nutrients. To address this comment, ***we added a new stoichiometric-ratio evaluation experiment, in which the N:P ratios derived from the MTL reconstruction were compared with those from the corresponding STL reconstruction*** under the same spatial grid, temporal coverage, and depth range.

We used the recent study by Liu et al. (2025), "***Global-scale shifts in marine ecological stoichiometry over the past 50 years***", published in Nature Geoscience, as the reference for ocean nutrient stoichiometry. Based on this study, we calculated the N:P ratios of the MTL and STL reconstructions across different ocean basins and depth layers, and compared them with the reported reference ratios. The detailed results are summarized in Response Table 1. Response Table 1. Comparison of reconstructed N ratios from the MTL and STL products across different ocean basins and depth layers. The reference N ratios are taken from Liu et al. (2025). ***Bold values indicate the reconstruction result with the smaller deviation from the reference ratio for each basin and depth layer. A dash indicates that the corresponding reference value was not available in Liu et al. (2025).***

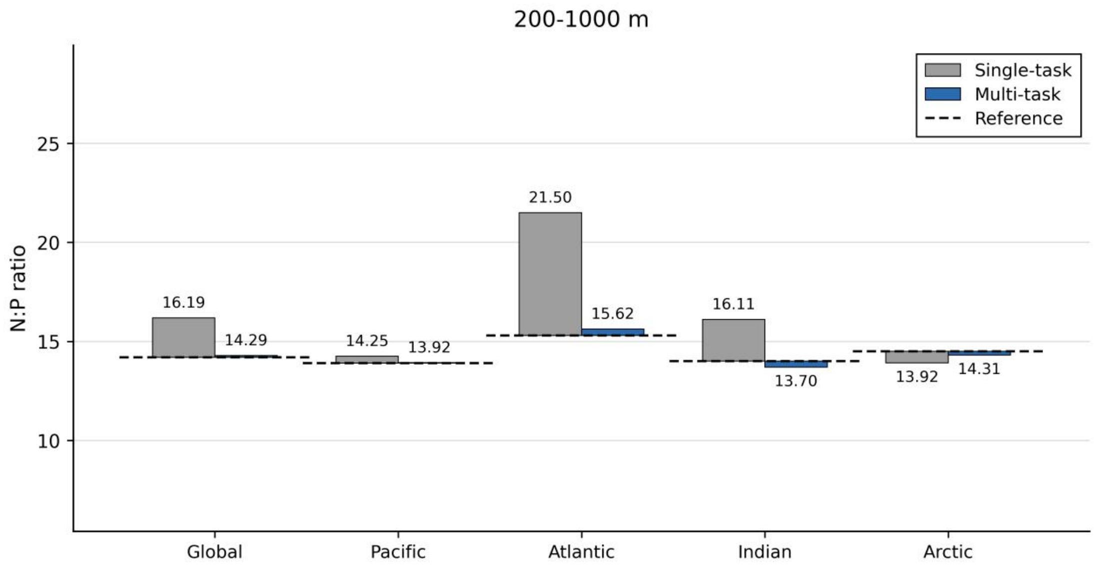
Response Table R1. Comparison of reconstructed N ratios from the MTL and STL products across different ocean basins and depth layers.The reference N ratios are taken from Liu et al. (2025). Bold values indicate the reconstruction result with the smaller deviation from the reference ratio for each basin and depth layer. A dash indicates that the corresponding reference value was not available in Liu et al. (2025).

Depth (m)	Basin	MTL N:P	STL N:P	Reference N:P
0–200	Global	**9.43**	16.00	12.60
	Pacific	8.09	**14.22**	12.00
	Atlantic	**11.78**	27.27	14.50
	Indian	8.78	**10.36**	13.40
	Southern Ocean	14.41	14.20	—
	Arctic Ocean	9.01	**9.22**	9.66
200–1000	Global	**14.29**	16.19	14.20
	Pacific	**13.92**	14.25	13.90
	Atlantic	**15.62**	21.50	15.30
	Indian	**13.70**	16.11	14.00
	Southern Ocean	14.43	14.06	—
	Arctic Ocean	**14.31**	13.92	14.50
1000–2000	Global	14.37	15.05	—
	Pacific	14.18	14.19	—
	Atlantic	15.08	18.04	—
	Indian	14.04	14.28	—
	Southern Ocean	14.43	14.10	—
	Arctic Ocean	14.55	14.01	—

The results show that the MTL reconstruction provides a more consistent nitrate–phosphate stoichiometric relationship than the STL reconstruction in most basin-depth comparisons. This improvement is especially clear in the 200–1000 m layer, where the MTL-derived N ratios are very close to the reference values in the global ocean, Pacific Ocean, Atlantic Ocean, Indian Ocean, and Arctic Ocean. For example, in the Atlantic Ocean at 200–1000 m, the reference N ratio is 15.30, while the MTL reconstruction gives 15.62, compared with 21.50 from the STL reconstruction. In the upper 0–200 m layer, the advantage of MTL is less uniform, likely because surface nutrient ratios are more strongly affected by regional biological uptake, seasonal variability, and nutrient limitation. Nevertheless, MTL substantially reduces the unrealistic high N ratio produced by STL in the Atlantic surface layer.



Response Fig. R1. Comparison of N:P ratios derived from the MTL and STL reconstructions with the reference ratios from Liu et al. (2025) in the upper ocean layer of 0–200 m. Grey bars indicate the STL reconstruction, blue bars indicate the MTL reconstruction, and dashed horizontal lines indicate the reference N ratios. Values are shown for major ocean basins where reference ratios are available.



Response Fig. R2. Comparison of N:P ratios derived from the MTL and STL reconstructions with the reference ratios from Liu et al. (2025) in the intermediate ocean layer of 200–1000 m. Grey bars indicate the STL reconstruction, blue bars indicate the MTL reconstruction, and dashed horizontal lines indicate the reference N ratios. Values are shown for major ocean basins where reference ratios are available.

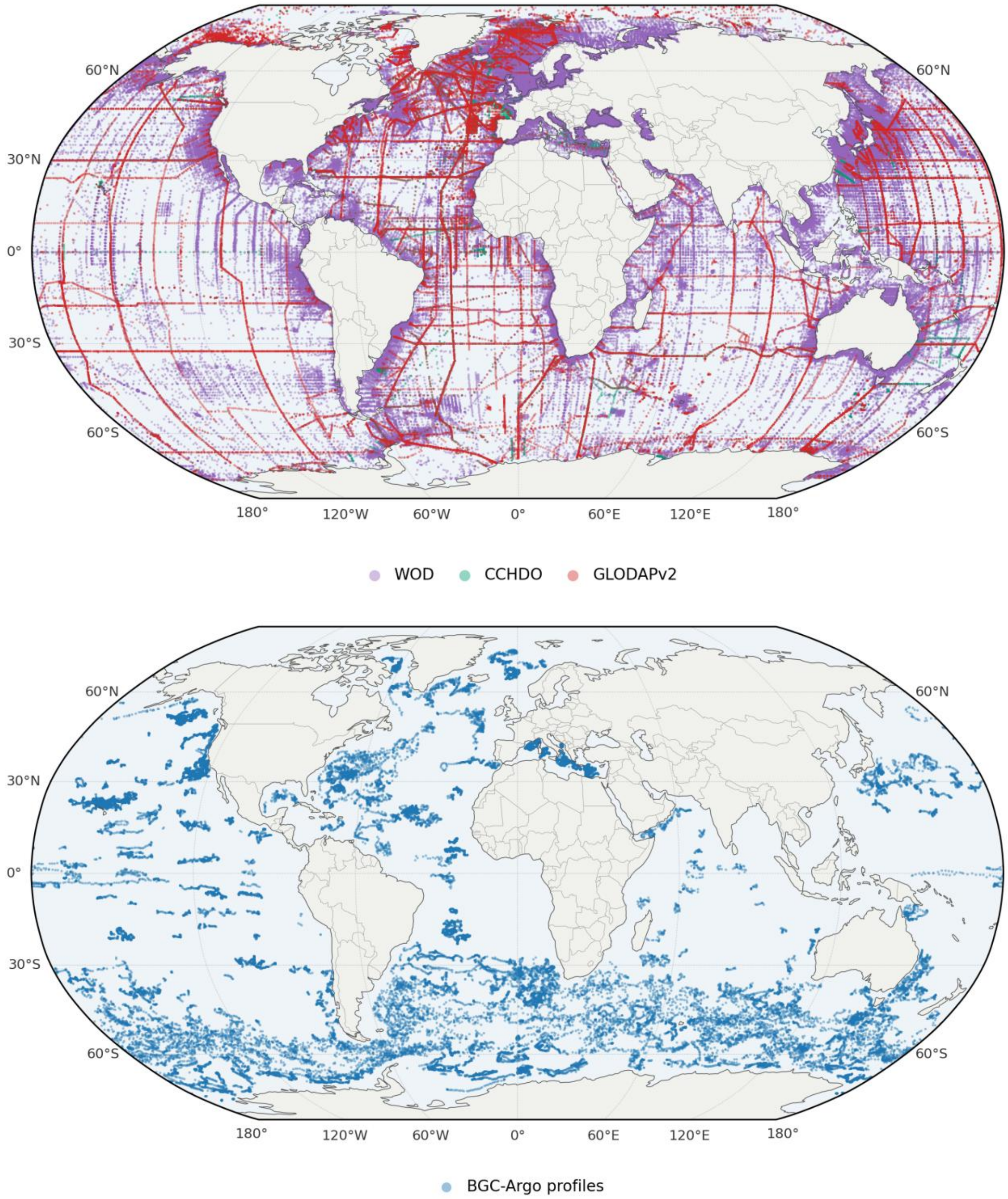
These results provide direct evidence that the benefit of the MTL framework is not limited to improving the prediction accuracy of individual nutrients. By jointly learning nitrate and phosphate within a shared representation space, the MTL model also better preserves their coupled stoichiometric relationship, especially below the surface layer where nutrient fields are more strongly controlled by water-mass structure and remineralization processes. Therefore, the added analysis supports the role of multi-task learning in improving the biogeochemical consistency of the Jingwei-Nutrients reconstruction.

Comment 4. BGC-Argo contributes 1,367,531 nitrate measurements. BGC-Argo nitrate is measured by ISUS/SUNA sensors that require careful calibration and are known to drift, particularly in deep-water deployments. The QC section (2.4) applies generic boundary and statistical checks but does not mention float-specific calibration procedures (e.g., cross-referencing with GLODAP climatological values at depth, as recommended by the BGC-Argo QC manual). How was BGC-Argo nitrate specifically validated, and were adjusted (DOXYcorrected) profiles used?

Response:

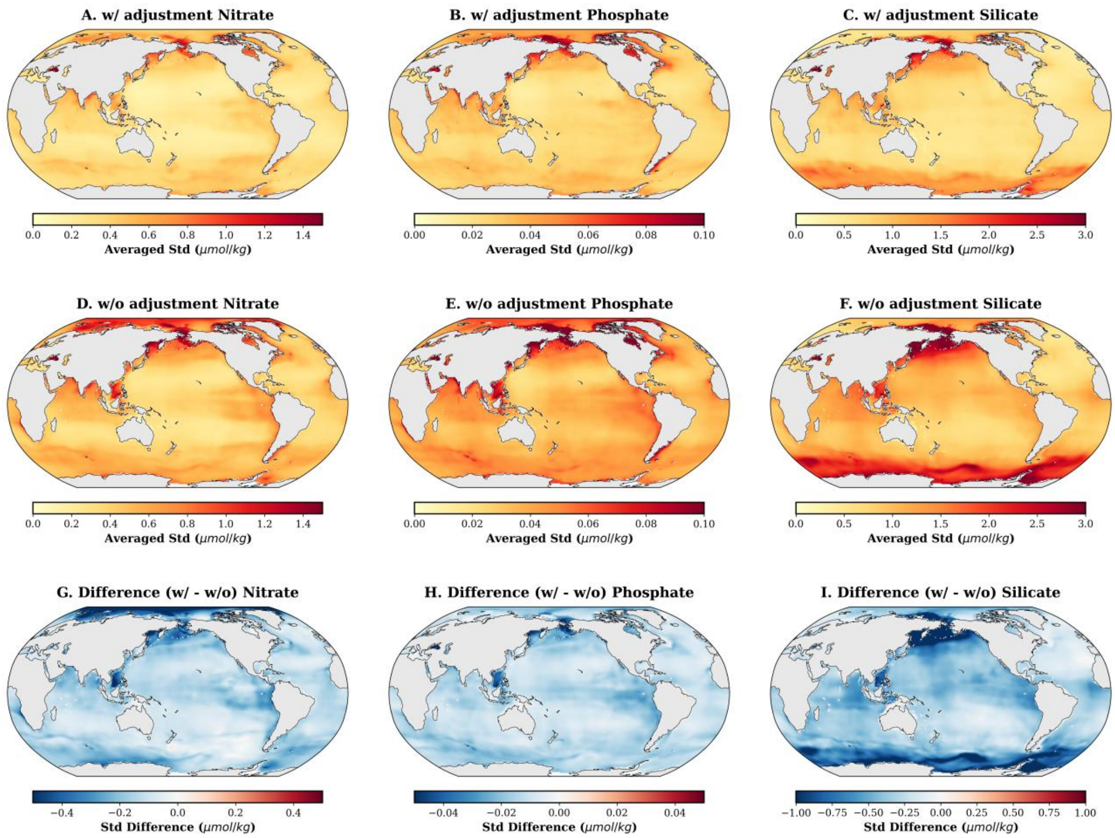
We thank the reviewer for this important and valuable suggestion. We agree that BGC-Argo nitrate observations require specific calibration treatment, because nitrate measurements from ISUS/SUNA sensors may be affected by sensor drift and calibration uncertainty, especially during long-term float deployments. We found that officially adjusted BGC-Argo nitrate profiles with calibration information are available. Therefore, following the reviewer's suggestion, ***we replaced the original BGC-Argo nitrate records in our previous workflow with these adjusted profiles. Based on this updated observational archive, we reconstructed the training dataset and reran the full model training, validation, global reconstruction, and uncertainty estimation experiments.***

The BGC-Argo nitrate data provide an important complement to the three ship-based data sources, especially in regions where historical bottle observations are sparse. As shown in Response Fig. R3, the ship-based observations from WOD, CCHDO, and GLODAPv2 are mainly distributed along cruise tracks and are relatively dense in historically well-sampled regions, whereas the BGC-Argo profiles provide additional open-ocean nitrate observations, with a particularly important contribution in the Southern Ocean. These profiles provide valuable high-latitude nitrate constraints under open-ocean hydrographic conditions, helping the model better learn nitrate variability in regions that are difficult to fully sample using ship-based cruises alone, and this improved nitrate constraint can further benefit the reconstruction of phosphate and silicate through the shared representation in the MTL framework.



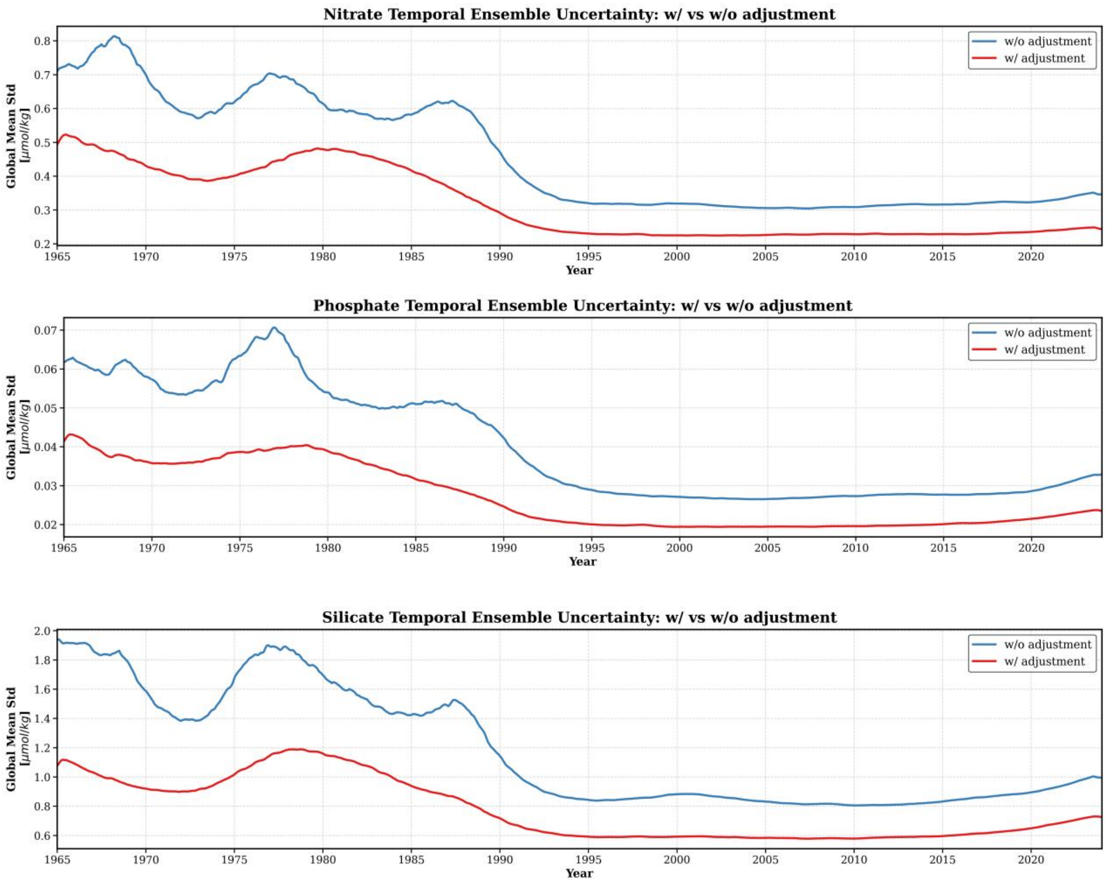
Response Fig. R3. Spatial distribution of nitrate observations from ship-based datasets and BGC-Argo profiles. The upper panel shows nitrate observations from WOD, CCHDO, and GLODAPv2_2022, and the lower panel shows nitrate observations from BGC-Argo profiles. Compared with the ship-based observations, which are mainly distributed along cruise tracks, BGC-Argo provides additional open-ocean coverage, particularly in the high-latitude Southern Hemisphere.

We then examined how replacing the original BGC-Argo nitrate records with the adjusted BGC-Argo nitrate profiles affects reconstruction uncertainty. As shown in Response Fig. R4, the adjusted-BGC-Argo-based reconstruction shows lower spatial ensemble uncertainty than the reconstruction based on the original BGC-Argo nitrate records. Because both experiments include BGC-Argo data, this reduction should not be interpreted as the effect of adding new spatial coverage. Instead, it reflects the benefit of using adjusted nitrate profiles with improved calibration consistency. After adjustment, the BGC-Argo nitrate observations are more consistent with the other quality-controlled data sources, reducing calibration-related discrepancies in the merged training dataset. This provides more reliable nitrate constraints for model training, especially under high-latitude open-ocean conditions where ship-based observations are sparse and the model relies more strongly on learned hydrographic-biogeochemical relationships. With more reliable nitrate constraints in these data-sparse regimes, the model can learn more stable nitrate variability patterns and reduce reconstruction uncertainty in high-latitude regions. Although the adjustment is applied directly to nitrate, uncertainty reductions are also observed for phosphate and silicate. This is reasonable because, in the MTL framework, nitrate, phosphate, and silicate share a common representation space; therefore, improving the reliability of one data-rich task can also help stabilize the shared features used by the other nutrient tasks. Consequently, the reconstructed fields show lower spatial uncertainty for all three nutrients after using the adjusted BGC-Argo nitrate profiles.



Response Fig. R4. Comparison of spatial uncertainty distributions before and after using the adjusted BGC-Argo nitrate profiles. The first row shows the averaged standard deviation of the reconstructed nitrate, phosphate, and silicate fields after using the adjusted BGC-Argo nitrate data. The second row shows the corresponding uncertainty distributions using the unadjusted BGC-Argo nitrate data. The third row shows the difference between the two experiments, calculated as the uncertainty with adjusted BGC-Argo data minus that without adjusted BGC-Argo data. Negative values indicate reduced reconstruction uncertainty after using the adjusted BGC-Argo nitrate profiles. “w/o adjustment” denotes the reconstruction based on the original BGC-Argo nitrate records, and “w/ adjustment” denotes the reconstruction based on adjusted BGC-Argo nitrate profiles.

In addition to the spatial uncertainty analysis, we further examined the temporal evolution of ensemble uncertainty. As shown in Response Fig. R5, the reconstruction based on adjusted BGC-Argo nitrate profiles shows lower global mean uncertainty throughout the reconstruction period for all three nutrients. This further supports that the adjusted BGC-Argo nitrate profiles provide more stable observational constraints for model training. The uncertainty reduction is observed not only for nitrate, but also for phosphate and silicate, which is consistent with the cross-nutrient information sharing in the MTL framework.



Response Fig. R5. Temporal evolution of global mean ensemble uncertainty before and after using the adjusted BGC-Argo nitrate profiles. The blue lines indicate the reconstruction based on the original BGC-Argo nitrate records, and the red lines indicate the reconstruction based on the adjusted BGC-Argo nitrate profiles. Results are shown for phosphate, nitrate, and silicate from 1965 to 2023. "w/o adjustment" denotes the reconstruction based on the original BGC-Argo nitrate records, and "w/ adjustment" denotes the reconstruction based on adjusted BGC-Argo nitrate profiles.

We further evaluated the reconstruction performance using the same independent test set before and after replacing the original BGC-Argo nitrate records with the adjusted BGC-Argo nitrate profiles. As summarized in Response Table R2, the reconstruction based on adjusted BGC-Argo nitrate data shows improved performance for all three nutrients. For nitrate, R^2 increases from 0.9803 to 0.9911, while RMSE decreases from 2.2138 to 1.2951 μmol kg^{-1}, corresponding to a 41.5% reduction in RMSE. For phosphate, R^2 increases from 0.9610 to 0.9691, and RMSE decreases from 0.2324 to 0.1627 μmol kg^{-1}, corresponding to a 30.0% reduction. For silicate, R^2 increases from 0.9834 to 0.9902, and RMSE decreases from 6.3485 to 4.1085 μmol kg^{-1}, corresponding to a 35.3% reduction. These improvements indicate that the float-specific adjustment of BGC-Argo nitrate has a measurable positive effect on the downstream reconstruction, supporting our decision to use the adjusted BGC-Argo nitrate profiles in the revised Jingwei-Nutrients product.

Response Table R2. Comparison of MTL reconstruction performance before and after using the adjusted BGC-Argo nitrate profiles. Both experiments were evaluated using the same independent test set. “w/o adjustment” denotes the reconstruction based on the original BGC-Argo nitrate records, and “w/ adjustment” denotes the reconstruction based on adjusted BGC-Argo nitrate profiles. RMSE values are reported in μmol kg⁻¹.

Nutrient	w/o adjustment R^2	w/o adjustment RMSE	w/ adjustment R^2	w/ adjustment RMSE	R^2 increase	RMSE reduction
Nitrate	0.9803	2.2138	0.9911	1.2951	+0.0108	0.9187, 41.5%
Phosphate	0.9610	0.2324	0.9691	0.1627	+0.0081	0.0697, 30.0%
Silicate	0.9834	6.3485	0.9902	4.1085	+0.0068	2.2400, 35.3%

Overall, using adjusted BGC-Argo nitrate profiles improves the consistency between BGC-Argo and the other quality-controlled observational sources, thereby reducing calibration-related discrepancies in the training dataset. This improvement is reflected in lower spatial and temporal ensemble uncertainty and better reconstruction performance on the same independent test set. ***Based on these results, we adopted the adjusted-BGC-Argo-based reconstruction as the revised Jingwei-Nutrients product and updated all corresponding figures, tables, validation statistics, uncertainty analyses, and data products in the manuscript. The updated results are provided at the end of this response.***

Comment 5. I can't find this reference: Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olafsson, J., and Ríos, A. F.: A global monthly climatology of macronutrients in the surface ocean, Earth Syst. Sci. Data, 13, 2593–2612, https://doi.org/10.5194/essd-13-2593-2021, 2021. A comparison with relevant products is expected. At minimum, provide surface R^2, RMSE, and bias, and discuss where Jingwei-Nutrients improves and where it does not.

Response:

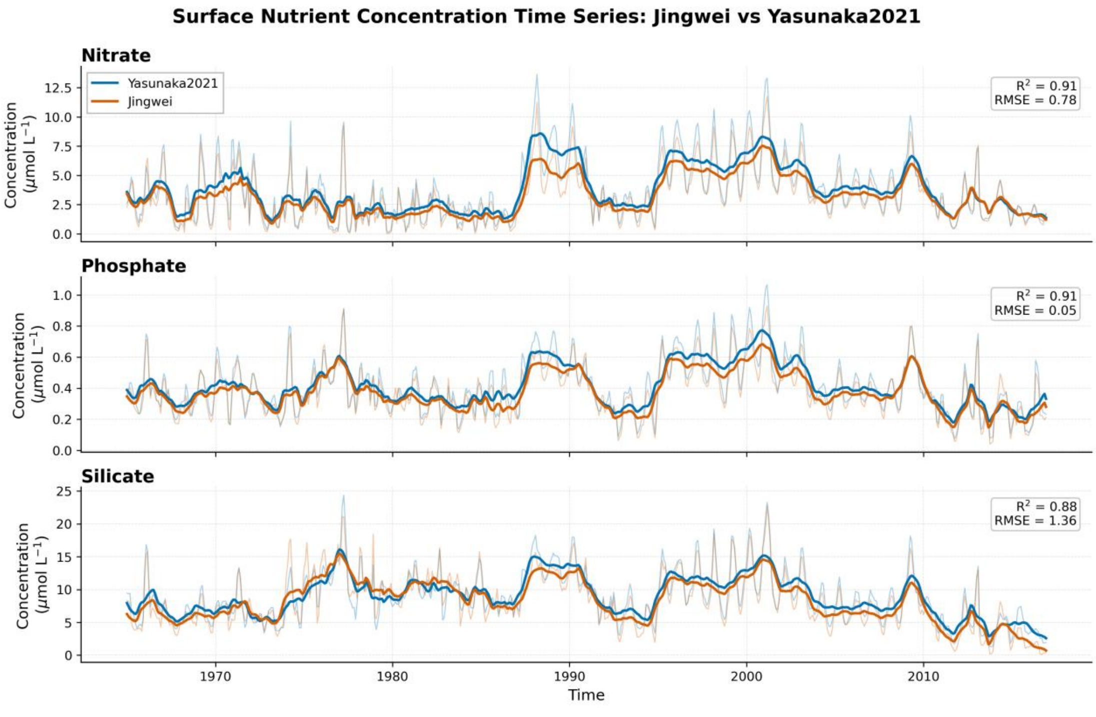
We thank the reviewer for pointing this out. We apologize for the incorrect reference citation in the original manuscript, and we have carefully checked and corrected the relevant reference.

<u>We agree with the reviewer that comparison with existing nutrient products is important for evaluating the characteristics and added value of Jingwei-Nutrients.</u> After further investigation, we found that directly comparable global four-dimensional nutrient products that simultaneously cover nitrate, phosphate, and silicate over a multi-decadal period remain very limited. Nevertheless, we identified several existing products that are suitable for partial comparison in terms of variables, spatial domain, depth coverage, or temporal coverage. ***In particular, we added comparisons with two available nutrient products: (1) the North Pacific surface nutrient product associated with Yasunaka et al. (2021), "Nutrient and dissolved inorganic carbon variability in the North Pacific", which provides monthly surface nitrate, phosphate, and silicate concentrations in the North Pacific from January 1961 to December 2012, covering 10°N–60°N and 120°E–90°W; and (2) the three-dimensional nitrate reconstruction product associated with Yu et al. (2025), "A continual-learning-based multilayer perceptron for improved reconstruction of three-dimensional nitrate concentrations", which provides monthly nitrate concentrations over the pan-European coastal region from 2010 to 2023, covering the upper 2000 m.***

(1) <u>Comparison with the North Pacific surface product of Yasunaka2021</u>

We first compared Jingwei with the North Pacific surface nutrient product of Yasunaka et al. (2021). This product provides monthly surface nitrate, phosphate, and silicate concentrations in the North Pacific, allowing a direct comparison with Jingwei over their overlapping surface domain and temporal period. Since Yasunaka2021 was generated using an interpolation-based mapping approach, whereas Jingwei was reconstructed using a Transformer-based multi-task learning framework, the agreement between the two products provides a useful cross-method check of the reconstructed surface nutrient variability and spatial structures.

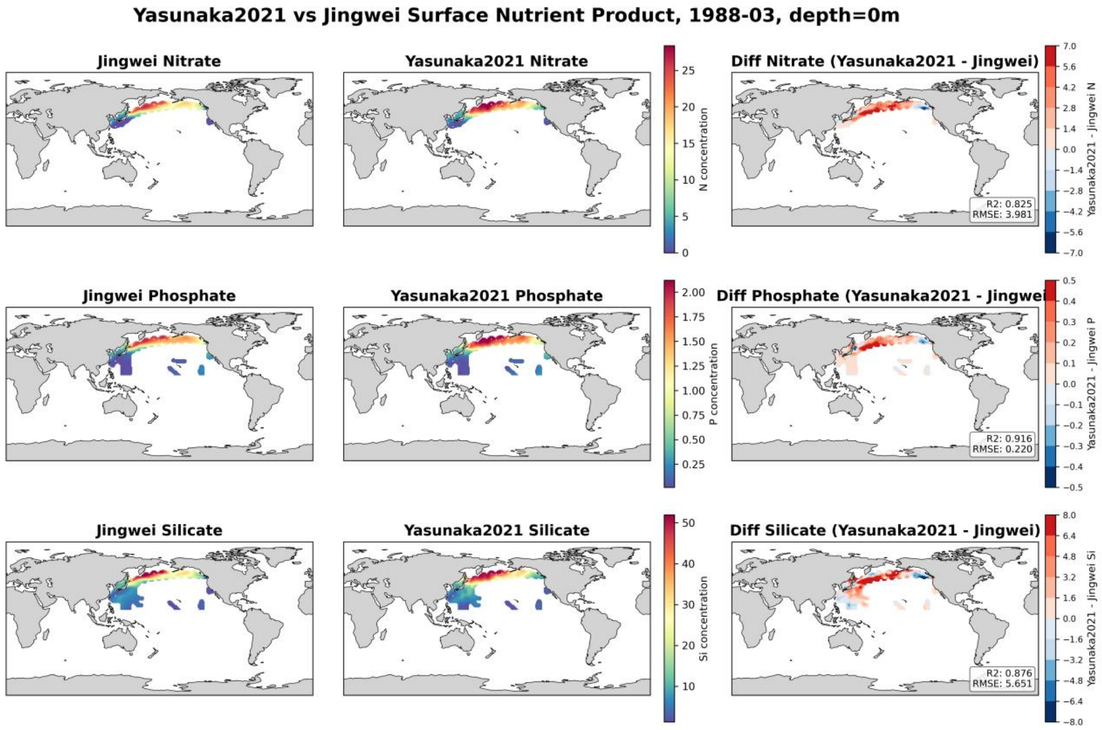
Response Fig. R6 shows the monthly surface concentration time series of nitrate, phosphate, and silicate averaged over the overlapping North Pacific domain. Jingwei and Yasunaka2021 show highly consistent temporal variations for all three nutrients. The two products capture similar seasonal-to-interannual variability and comparable long-term fluctuations. The agreement is also supported by the quantitative metrics, with R^2 values of 0.91, 0.91, and 0.88 for nitrate, phosphate, and silicate, respectively. This strong consistency suggests that Jingwei can reliably reproduce surface nutrient variability in the North Pacific, and the agreement between the two independently generated products provides mutual support for the robustness of the reconstructed surface signals.



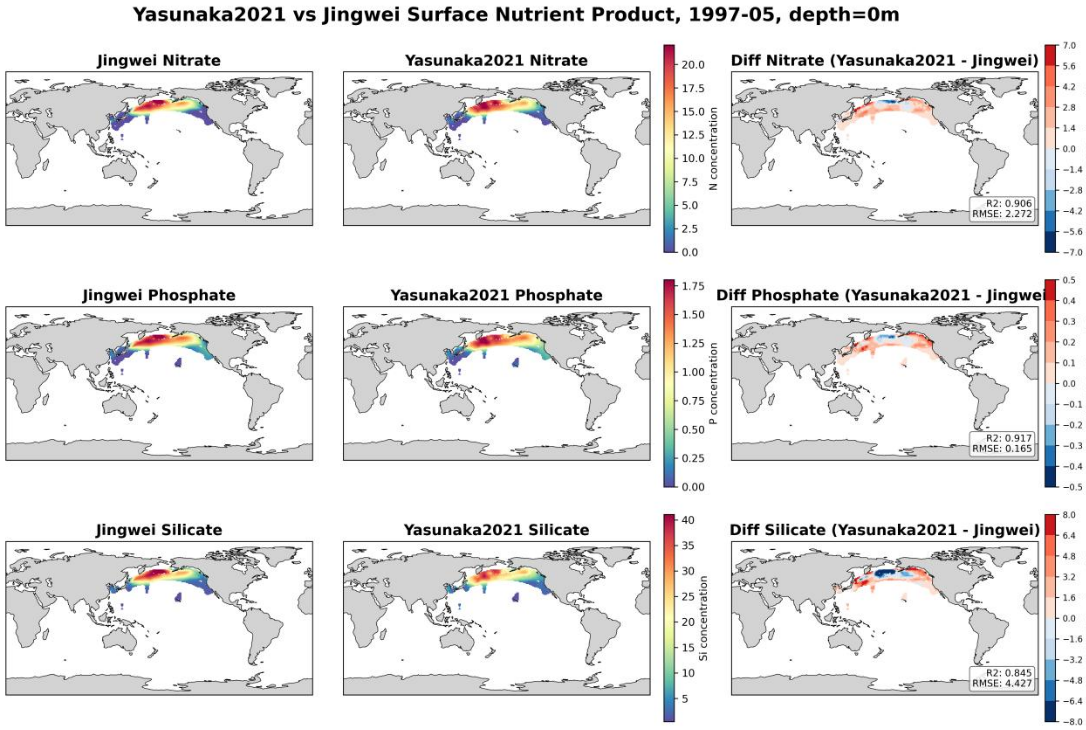
Response Fig. R6. Comparison of monthly surface nutrient concentration time series between Jingwei and Yasunaka2021 over the overlapping North Pacific domain. Results are shown for nitrate, phosphate, and silicate. The corresponding R^2 and RMSE values are shown in each panel.

We further compared the spatial distributions of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021 for selected months. As shown in

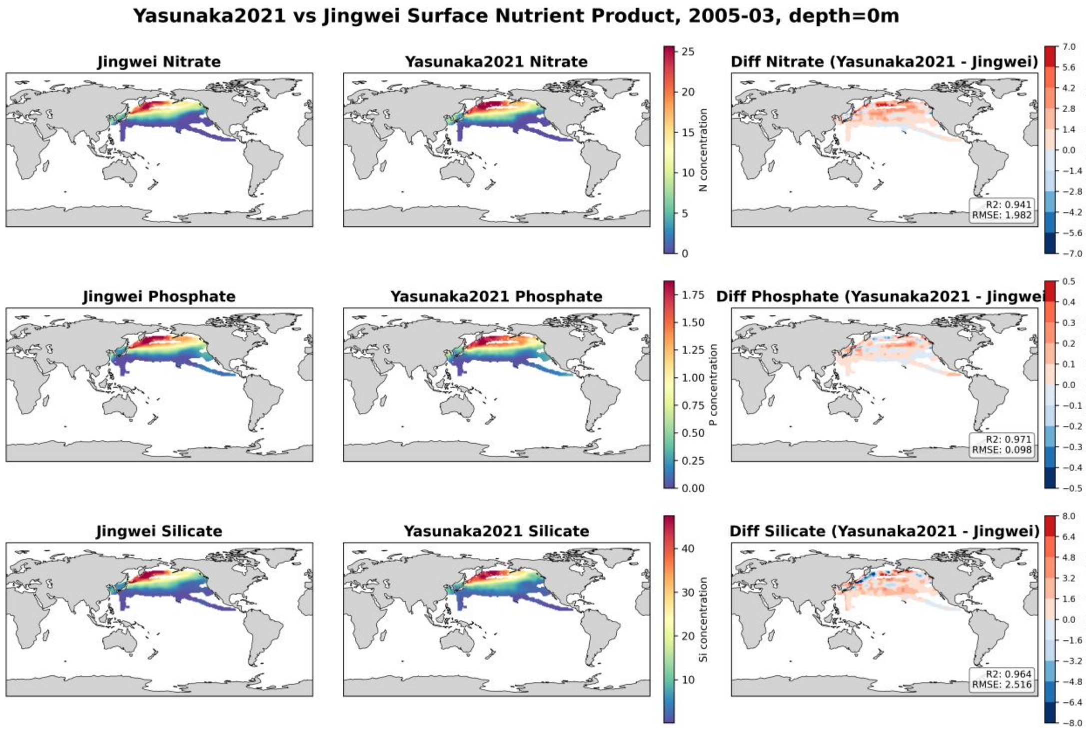
Response Figs. R7–R10, the two products exhibit similar large-scale surface nutrient patterns in the North Pacific. Both products reproduce the major meridional nutrient gradients, including higher nutrient concentrations in the northern North Pacific and lower concentrations in the subtropical region. Although local differences exist, especially near sharp frontal zones and marginal regions, the overall spatial structures are highly consistent. These results further demonstrate that Jingwei and Yasunaka2021, despite being developed using different reconstruction approaches, produce consistent large-scale surface nutrient structures in the North Pacific, providing a useful cross-method verification of the Jingwei reconstruction.



Response Fig. R7. Spatial comparison of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021 in March 1988. The first column shows Jingwei, the second column shows Yasunaka2021, and the third column shows their difference, calculated as Yasunaka2021 minus Jingwei.

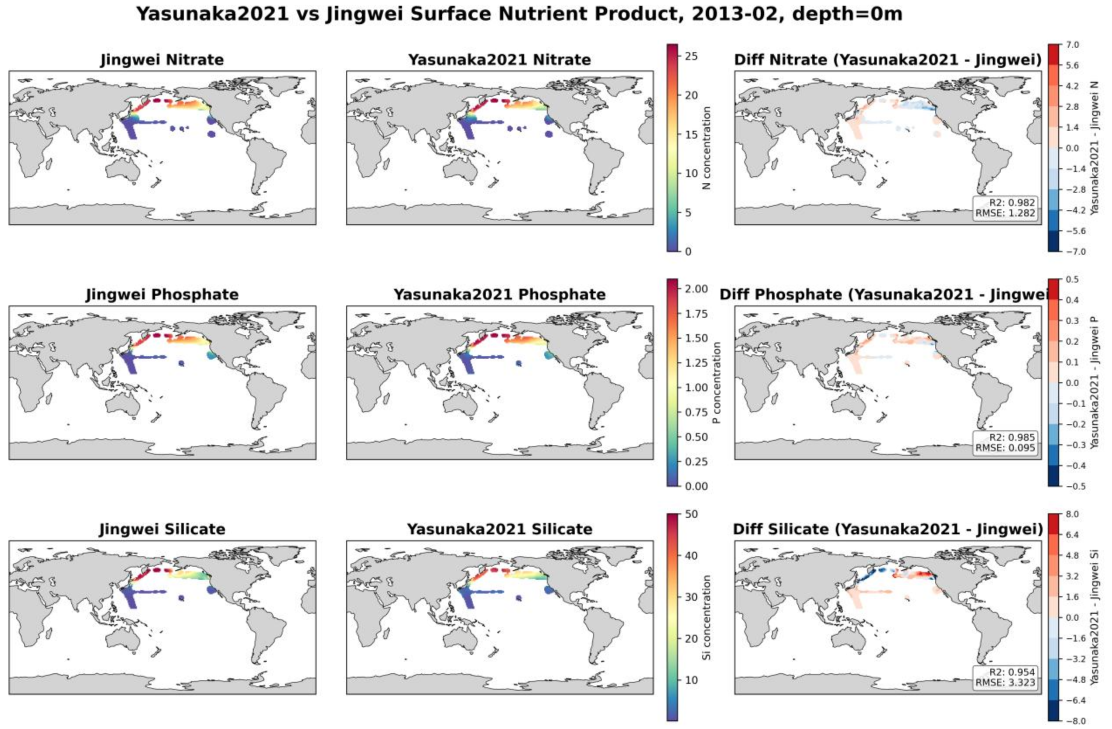


Response Fig. R8. Spatial comparison of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021 in May 1997. The first column shows Jingwei, the second column shows Yasunaka2021, and the third column shows their difference, calculated as Yasunaka2021 minus Jingwei.



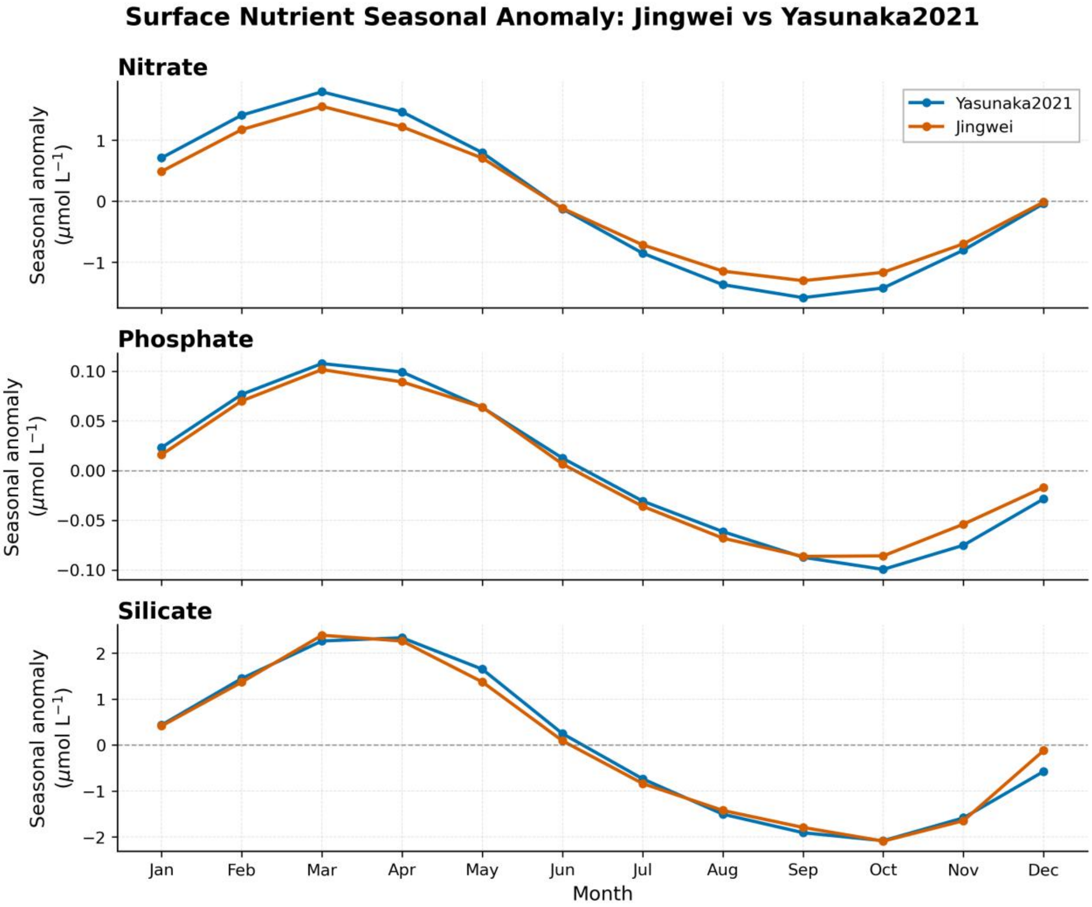
Response Fig. R9. Spatial comparison of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021 in March 2005. The first column shows Jingwei, the second column shows

Yasunaka2021, and the third column shows their difference, calculated as Yasunaka2021 minus Jingwei.



Response Fig. R10. Spatial comparison of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021 in February 2013. The first column shows Jingwei, the second column shows Yasunaka2021, and the third column shows their difference, calculated as Yasunaka2021 minus Jingwei.

We also compared the seasonal anomaly cycles of surface nitrate, phosphate, and silicate between Jingwei and Yasunaka2021. As shown in Response Fig. R11, the two products show highly consistent seasonal phases and amplitudes for all three nutrients. Both products reproduce positive anomalies in late winter to spring and negative anomalies in summer to autumn, reflecting the dominant seasonal nutrient cycle in the North Pacific surface ocean. The close agreement in seasonal anomalies further indicates that Jingwei is not only consistent with Yasunaka2021 in long-term temporal variability and spatial distribution, but also reproduces the main seasonal evolution of surface nutrients in the overlapping region. This provides an additional cross-method verification of the Jingwei reconstruction in the North Pacific surface ocean.



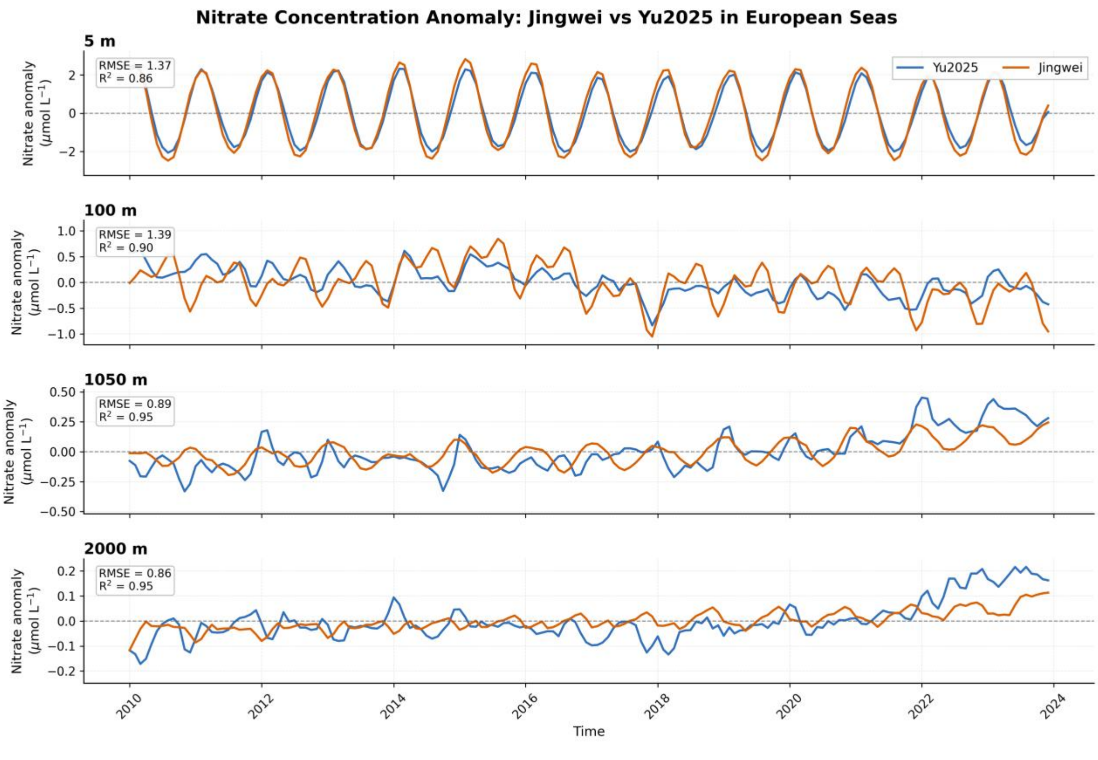
Response Fig. R11. Comparison of surface nutrient seasonal anomalies between Jingwei and Yasunaka2021 over the overlapping North Pacific domain. Results are shown for nitrate, phosphate, and silicate. The seasonal anomalies were calculated from the monthly climatological cycle after removing the annual mean.

Overall, the comparison with Yasunaka2021 shows strong consistency between Jingwei and an independently developed interpolation-based surface product in the North Pacific. This agreement provides a useful cross-method verification of the Jingwei reconstruction in the overlapping surface domain. However, Yasunaka2021 is a regional surface product, and its interpolation-based framework is more suitable for regions with relatively dense observational constraints. In comparison, Jingwei reconstructs nitrate, phosphate, and silicate over the global ocean from the surface to 2000 m depth. ***Therefore, while Jingwei maintains strong consistency with Yasunaka2021 in the North Pacific surface ocean, it further extends nutrient reconstruction to a global four-dimensional product that is more suitable for broad-scale biogeochemical analyses.***

(2) ***Comparison with the pan-European nitrate product of Yu2025***

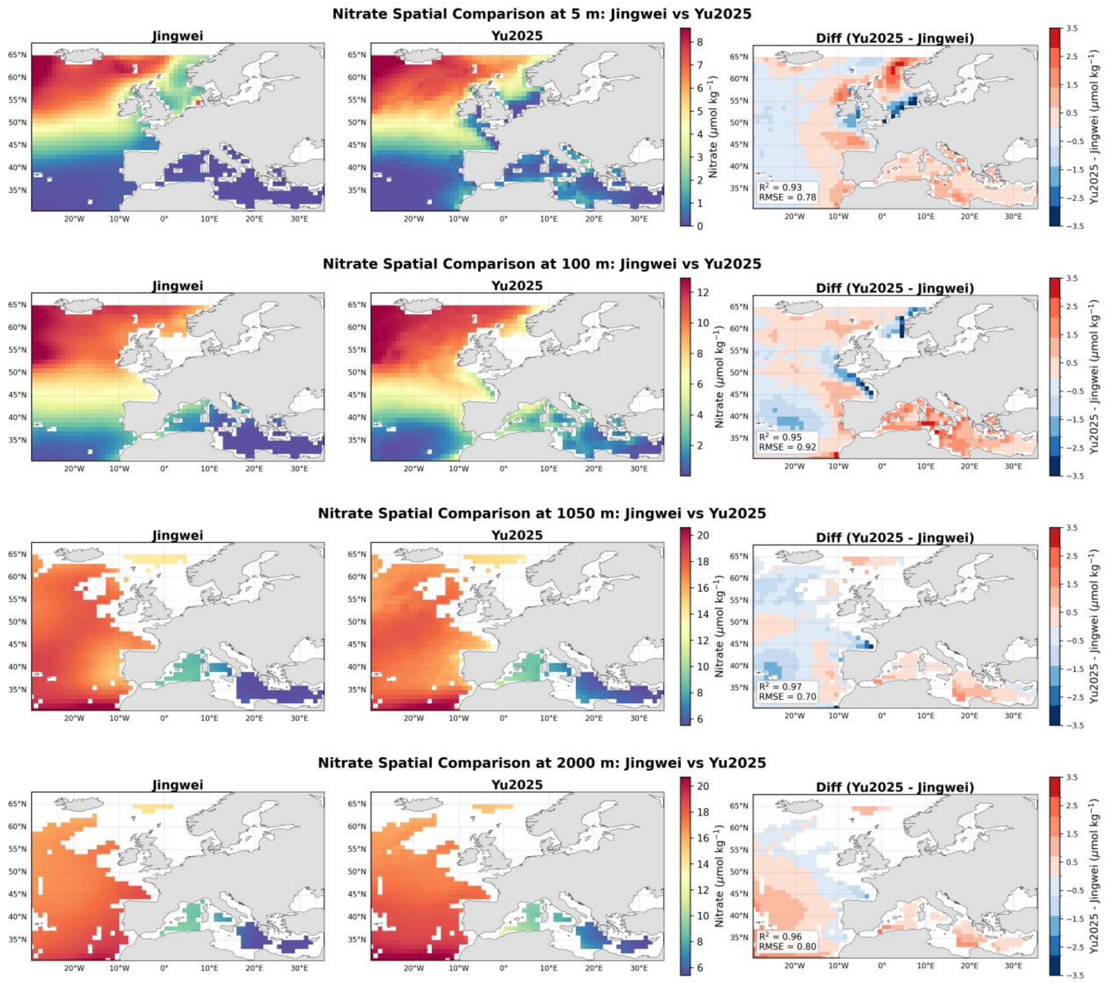
We next compared Jingwei with the four-dimensional nitrate reconstruction product of Yu et al. (2025). Yu2025 provides monthly nitrate concentrations over the pan-European coastal region from 2010 to 2023, covering the upper 2000 m. Therefore, we extracted Jingwei nitrate over the same spatial domain, temporal period, and representative depth levels for comparison. Since Yu2025 and Jingwei were developed using different reconstruction strategies, their comparison provides an additional cross-product assessment of nitrate variability in the overlapping domain.

Response Fig. R12 compares the nitrate concentration anomalies from Jingwei and Yu2025 at four representative depths: 5, 100, 1050, and 2000 m. Overall, the two products show consistent temporal variability across these depth levels. The agreement is particularly strong in the deeper layers, with R^2 values of 0.95 at both 1050 and 2000 m. In the upper ocean, both products capture the dominant seasonal variability, although local differences in amplitude and short-term fluctuations are visible, especially at 100 m. These differences may reflect the stronger influence of regional biological processes, vertical mixing, and product-specific reconstruction strategies in the upper ocean. Nevertheless, the overall consistency indicates that Jingwei captures nitrate temporal variability comparable to an independently developed regional 3D nitrate product.



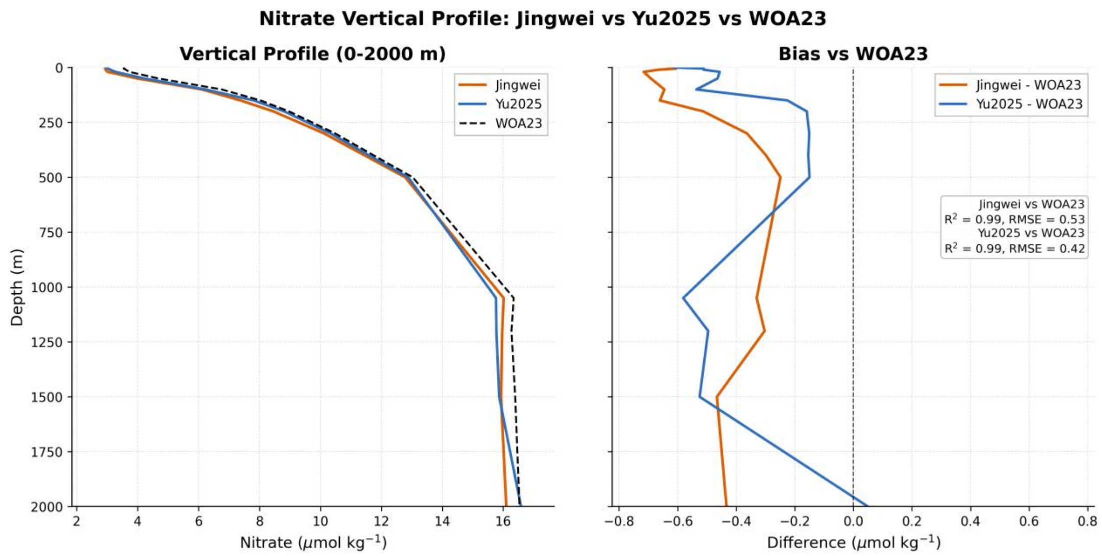
Response Fig. R12. Comparison of nitrate concentration anomalies between Jingwei and Yu2025 over the pan-European coastal region at representative depths of 5, 100, 1050, and 2000 m. The blue lines indicate Yu2025, and the orange lines indicate Jingwei. The corresponding R^2 and RMSE values between the two products are shown in each panel.

We also compared the spatial nitrate distributions at the same representative depth levels. As shown in Response Fig. R13, Jingwei and Yu2025 exhibit broadly consistent spatial patterns from the near-surface layer to 2000 m. Both products reproduce the relatively high nitrate concentrations in the northeastern Atlantic and lower nitrate concentrations in the Mediterranean region. The spatial agreement remains strong across depths, with high spatial R^2 values in the selected layers. Some regional differences are still observed, especially near coastal margins, semi-enclosed seas, and regions with sharp gradients. These differences are expected because regional nitrate reconstruction is sensitive to the treatment of boundary conditions, sparse observations, and the use of different auxiliary information.



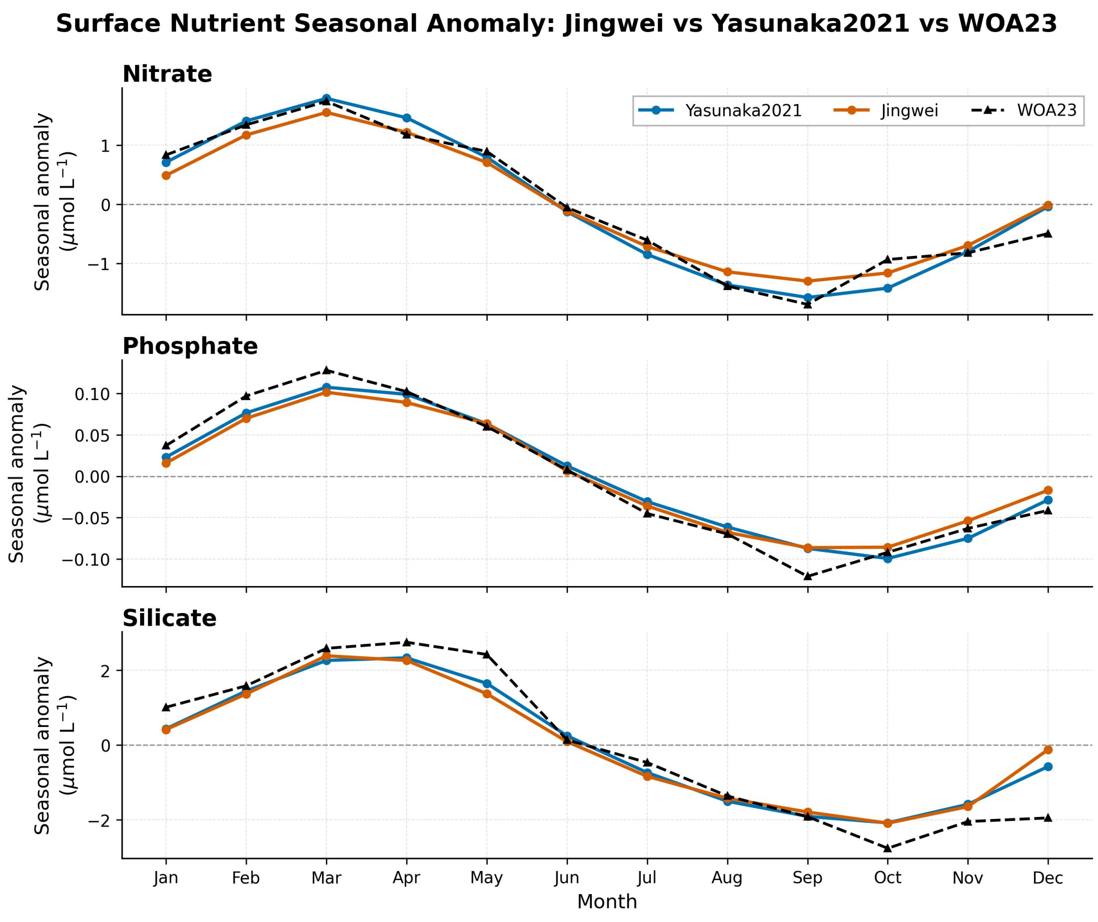
Response Fig. R13. Spatial comparison of nitrate concentrations between Jingwei and Yu2025 over the pan-European coastal region at 5, 100, 1050, and 2000 m. For each depth, the first column shows Jingwei, the second column shows Yu2025, and the third column shows their difference, calculated as Yu2025 minus Jingwei.

To further assess the vertical consistency, we compared the regional mean nitrate vertical profiles from Jingwei, Yu2025, and WOA23. As shown in Response Fig. R14, Jingwei and Yu2025 both reproduce the major vertical structure of nitrate concentrations, including low concentrations in the surface layer and increasing concentrations with depth. The two products also show similar agreement with the WOA23 climatological profile. This result suggests that Jingwei is consistent with existing regional nitrate reconstructions in representing the large-scale vertical nitrate structure. The small differences among the products should not be interpreted as a simple error relative to a single reference, but rather as product-dependent differences arising from different data sources, reconstruction strategies, and regional constraints.



Response Fig. R14. Comparison of regional mean nitrate vertical profiles among Jingwei, Yu2025, and WOA23 over the pan-European coastal region. The left panel shows the nitrate vertical profiles from 0 to 2000 m, and the right panel shows the corresponding differences relative to WOA23.

We further compared the seasonal nitrate cycles at selected upper-ocean and intermediate depths. As shown in Response Fig. R15, Jingwei and Yu2025 show highly consistent seasonal phases at 0 and 50 m, with high nitrate concentrations in winter–spring and low concentrations in summer–autumn. This agreement indicates that the two products reproduce similar seasonal nitrate variability in the upper ocean. At 150 and 500 m, the seasonal signal becomes weaker and the differences among Jingwei, Yu2025, and WOA23 become more apparent. Nevertheless, Jingwei and Yu2025 remain broadly comparable in magnitude and seasonal range, especially considering that these deeper layers are less directly constrained by seasonal surface processes. The comparison with WOA23 provides an additional climatological reference, showing that Jingwei preserves the main vertical and seasonal structure of nitrate while maintaining consistency with an independently developed regional nitrate product. Overall, the seasonal-cycle comparison further supports the reliability of Jingwei in representing nitrate variability across different depth layers in the overlapping European seas.



Response Fig. R15. Comparison of nitrate seasonal cycles among Jingwei, Yu2025, and WOA23 over the pan-European coastal region at 0, 50, 150, and 500 m. The seasonal cycles were calculated from monthly climatological nitrate concentrations over the overlapping domain.

Overall, the comparison with Yu2025 provides another useful cross-product verification of Jingwei in a regional four-dimensional setting. ***Jingwei and Yu2025 show consistent nitrate temporal variability, spatial distributions, vertical structures, and seasonal cycles in their overlapping domain.*** At the same time, the two products have different design goals and data foundations. Yu2025 focuses on nitrate reconstruction over the pan-European coastal region and uses a continual-learning strategy that incorporates information from CMEMS simulated nitrate fields. Because the CMEMS biogeochemical hindcast is a numerical-model product without data assimilation, model-dependent biases may be introduced into downstream reconstructions. In contrast, Jingwei is designed as a global four-dimensional reconstruction of nitrate, phosphate, and silicate based on quality-controlled multi-source observations and a multi-task learning framework. ***Therefore, while the agreement with Yu2025 supports the reliability of Jingwei in the overlapping European seas, Jingwei further extends the reconstruction to the global ocean and to three major macronutrients, making it more suitable for global-scale biogeochemical analyses.***

Minor comments

Comments 1. Please make the model code (training scripts, architecture, inference pipeline) publicly available (e.g., GitHub with a Zenodo DOI) and cite the repository.

Response:

We thank the reviewer for this suggestion. We agree that open access to the model code is important for improving the transparency, reproducibility, and reusability of the Jingwei-Nutrients product. We have organized the relevant code, including the model architecture, training scripts, data preprocessing workflow, inference pipeline, and post-processing scripts, and made them publicly available through a GitHub repository. We will also cite this repository in the revised manuscript.

The code repository is available at: https://github.com/jingwei-sjtu/Jingwei-Nutrients

---

Comments 2. Section 5 (Data availability) is brief. Can you adapt it to include technical description about the data product: variable names, units, fill values, coordinate, depth..., and the structure of the ensemble uncertainty files? Please add a

Data Format subsection or expand Section 5.

Response:

We thank the reviewer for this helpful suggestion. We agree that the original Data availability section was too brief and did not provide sufficient technical information for users to directly understand and use the data product. We have expanded Section 5 by adding a new Data Format subsection, including the file format, variable names, units, fill values, coordinate dimensions, depth levels, time encoding, data-variable dimension order, compression information, metadata, and the structure of the ensemble uncertainty files.

Manuscript revision:

We added the following description to Section 5:

"The Jingwei-Nutrients product is provided as a CF-1.8-compliant NetCDF4/HDF5 file, named Jingwei_Nutrients_1965_2023_0-2000m.nc. The file contains monthly global reconstructed fields from January 1965 to December 2023. All nutrient variables are stored with dimensions (time, depth, lat, lon). The time dimension contains 708 monthly time steps, the depth dimension contains 67 standard levels from 0 to 2000 m, the latitude coordinate ranges from 82.5°S to 89.5°N, and the longitude coordinate ranges from 179.5°W to 179.5°E. The three main variables are nitrate, phosphate, and silicate, with units of μmol kg^{-1}. Missing or invalid ocean values are stored as NaN. The corresponding ensemble uncertainty files use the same coordinates, dimensions, depth levels, and missing-value convention as the main product. These files provide ensemble-based uncertainty estimates for nitrate, phosphate, and silicate, expressed as the standard deviation among ensemble members, with the same units as the reconstructed nutrient concentrations."

Comments 3. The WOD23 dataset is cited as Garcia et al. (2026). I can't find it in the list of references, I recommend a full revisions of reference list.

Response:

We thank the reviewer for pointing this out. We have carefully checked the reference list and corrected the citation related to WOD23. In addition, we conducted a full revision of the reference list to ensure that all cited datasets, data products, and methodological references are correctly included and consistently formatted in the revised manuscript.

Manuscript revision:

We added the missing reference for the WOD23 dataset to the reference list:

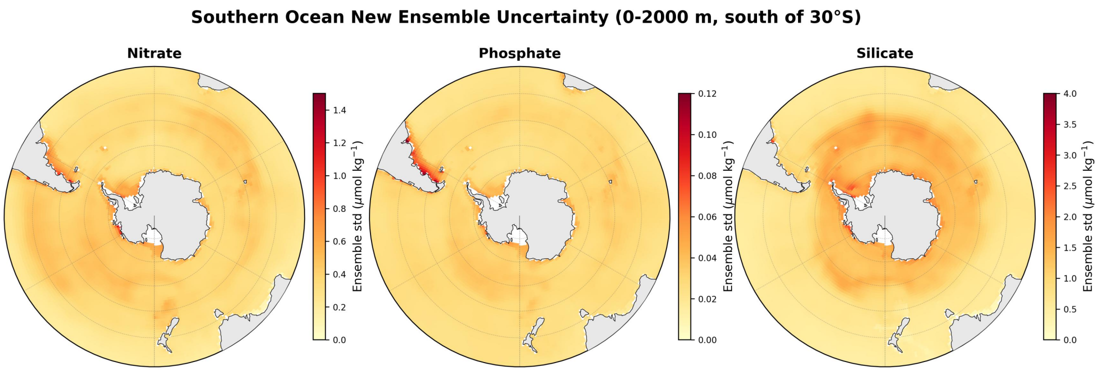
Garcia, H., Boyer, T., Levitus, S., Reagan, J., Mishonov, A., Jiang, L.-Q., Wang, Z., Paver, C., Nyadjro, E., Cross, S., Bouchard, C., Hogan, P., Baranova, O., and Locarnini, R.: World Ocean Database 2023: A Foundational Data Resource for and by the Global Ocean and Coastal Communities, Scientific Data, 13, 613, https://doi.org/10.1038/s41597-026-06957-2, 2026.

Comments 4. The R^2 for surface phosphate at KERFIX is 0.758, attributed to iron-driven N:P decoupling. Given that the Southern Ocean is both undersampled and biogeochemically important, users would benefit from a regional uncertainty map or bias analysis for the Southern Ocean surface layer, beyond the depth-averaged uncertainty shown in Fig. 16.

Response:

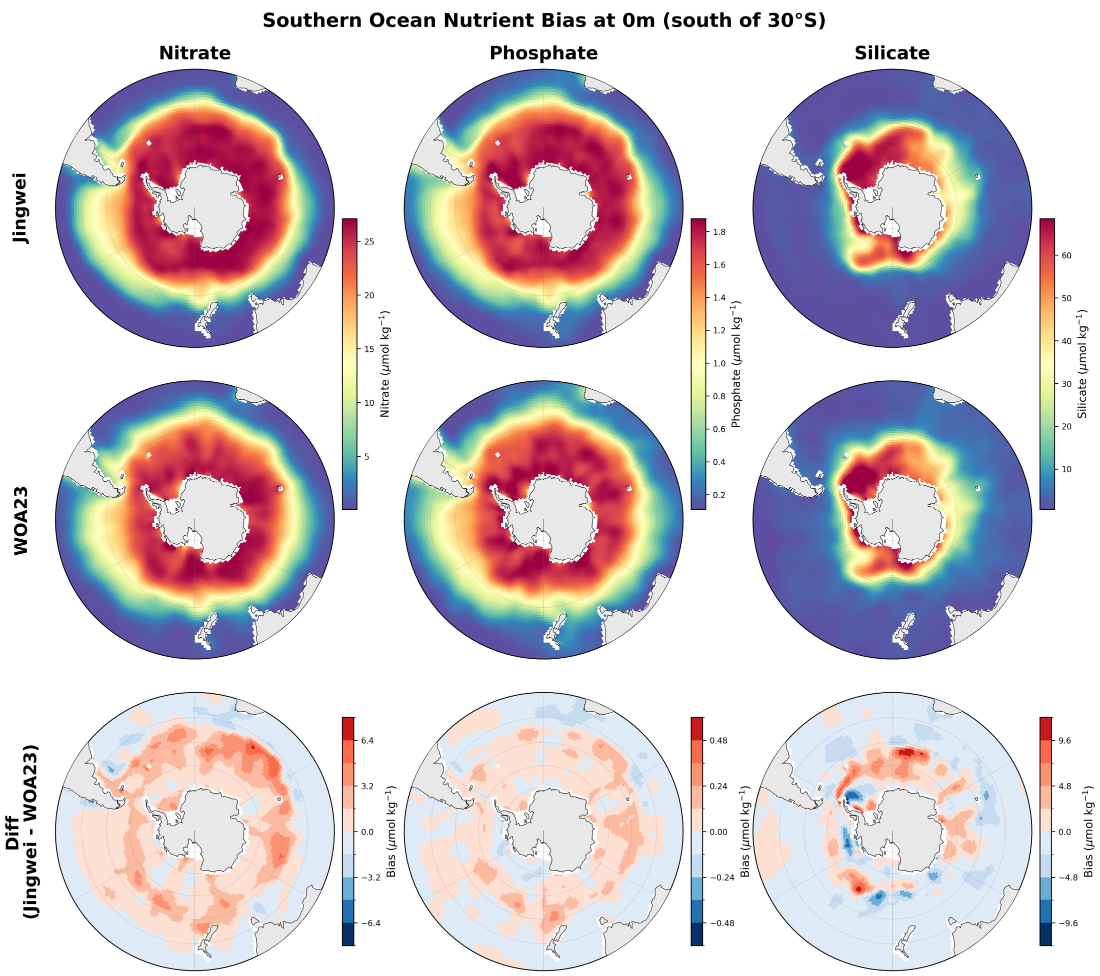
We thank the reviewer for this helpful suggestion. We agree that the Southern Ocean is a particularly important region for nutrient reconstruction because of its sparse observational coverage, strong biogeochemical gradients, and complex nutrient limitation processes. To provide users with a clearer regional assessment, we added two additional Southern Ocean analyses: an ensemble uncertainty map for nitrate, phosphate, and silicate south of 30°S, and a surface comparison between Jingwei and WOA23 for the same three nutrients.

As shown in Response Fig. R16, the Southern Ocean uncertainty map provides a more focused view of the regional reconstruction uncertainty than the global depth-averaged figure. The uncertainty is spatially heterogeneous, with relatively larger values near strong frontal zones and regions close to the Antarctic margin, especially for nitrate and silicate. This pattern is consistent with the sparse observations and strong physical-biogeochemical gradients in the Southern Ocean.



Response Fig. R16. Regional ensemble uncertainty of Jingwei-Nutrients in the Southern Ocean south of 30°S. The panels show the vertically averaged ensemble standard deviation from 0 to 2000 m for nitrate, phosphate, and silicate. Units are μmol kg^{-1}.

We further added a surface nutrient comparison with WOA23 to evaluate the regional spatial structure of Jingwei in the Southern Ocean surface layer. As shown in Response Fig. R17, Jingwei reproduces the broad circumpolar nutrient distribution and the large-scale gradients shown by WOA23 for nitrate, phosphate, and silicate. Local differences remain, particularly around frontal regions and near the Antarctic coastal zone, where strong mesoscale variability, seasonal biological uptake, iron limitation, and sparse observations can all contribute to regional discrepancies. These additional analyses provide useful context for the relatively lower surface phosphate R^2 at KERFIX and help users better interpret the uncertainty and potential bias of Jingwei in the Southern Ocean surface layer.



Response Fig. R17. Surface nutrient comparison between Jingwei and WOA23 in the Southern Ocean south of 30°S. The first row shows Jingwei surface concentrations, the second row shows WOA23 surface climatological concentrations, and the third row shows their difference, calculated as Jingwei minus WOA23. Results are shown for nitrate, phosphate, and silicate. Units are μmol kg^{-1}.

Minor language and typographical issues

Comments:

- Line 75: "they lost temporal variance" should read "they lose temporal variance".
- Line 105: "four-dimensional" is hyphenated here but not at line 19; please be consistent, and change to data product not a dataset. Data products are generally derived from datasets;
- Line 383: "anomalies are generally low" consider "anomalies are generally small"

(concentrations, not anomalies, can be low).

- Figure captions for Figs. 13–15 do not align with the order of discussion in the text (Fig. 13 caption references P16N, but P16S is discussed first).

Response:

We thank the reviewer for carefully pointing out these language and typographical issues. We have accepted all of these suggestions and revised the corresponding expressions in the manuscript. Specifically, we changed “they lost temporal variance” to “they lose temporal variance”, standardized the use of “four-dimensional” throughout the manuscript, and replaced “dataset” with “data product” where appropriate. We also revised “anomalies are generally low” to “anomalies are generally small” to improve the accuracy of the expression. In addition, we adjusted the order and captions of the GO-SHIP section figures to ensure that Figs. 13–15 are consistent with the order of discussion in the text.

Updated Results Based on Adjusted BGC-Argo Nitrate Profiles

Following the reviewer’s suggestion, we recompiled the observational dataset using adjusted BGC-Argo nitrate profiles, retrained the model, reran all validation and reconstruction experiments, and updated the corresponding statistics, figures, tables, and uncertainty analyses. The updated results are provided below.

Table 1. Summary of nutrients and hydrographic data compiled from four major databases after quality control.

	WOD	CCHDO	GLODAPV2	BGC-Argo	Total
T/S	5,735,819	0	0	0	5,735,819
Nitrate	1,738,637	69,355	609,016	1,808,366	4.225.374
Phosphate	2,902,300	98,652	577,388	0	3,578,340
Silicate	2,252,702	73,165	606,131	0	2,931,998

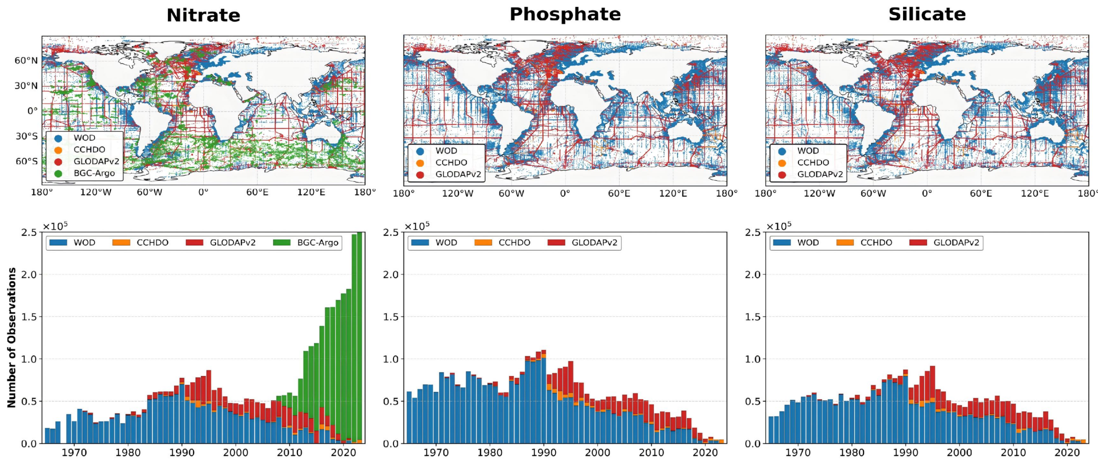


Figure 1. Spatial and temporal distribution of the nutrients data after quality control.

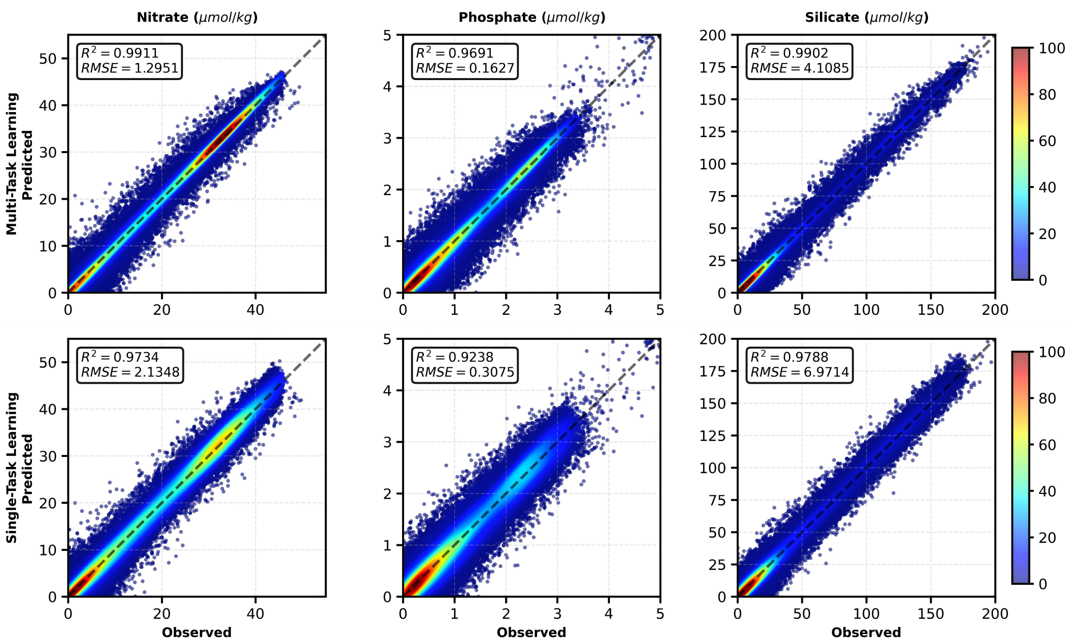


Figure 3. Density scatter plots comparing the Jingwei-Nutrients reconstructed concentrations against in situ observations for nitrate, phosphate, and silicate. The plots aggregate all matched test samples from the strict time-based K-fold cross-validation over the entire study period (1965–2023). Colors indicate the density of data points, and the dashed line represents the 1:1 ideal agreement.

Table 3. Statistical evaluation metrics ($R^2$and RMSE) of the chronological K-fold cross-validation. The results highlight the decadal performance evolution and compare the predictive accuracy of the MTL architecture against independent STL models for nitrate, phosphate, and silicate reconstructions. The unit for all RMSE values is ***$\mu mol\ kg^{-1}$***.

	Multi-Task Learning						Single-Task Learning					
Fold	N (R^2/RMSE)		P (R^2/RMSE)		Si (R^2/RMSE)		N (R^2/RMSE)		P (R^2/RMSE)		Si (R^2/RMSE)	
1965-1975	0.9847	1.8306	0.9476	0.1986	0.9794	6.1854	0.9618	3.1264	0.8967	0.4177	0.9605	9.1755
1976-1985	0.9865	1.5452	0.9430	0.2094	0.9834	4.7658	0.9651	2.6278	0.8908	0.4080	0.9672	8.8442
1986-1995	0.9889	1.3298	0.9618	0.1688	0.9881	4.1327	0.9694	2.2483	0.9146	0.3368	0.9753	7.7289
1996-2005	0.9937	1.1494	0.9831	0.1501	0.9958	3.5784	0.9781	1.9364	0.9415	0.2586	0.9883	5.8910
2006-2015	0.9955	0.9883	0.9863	0.1327	0.9968	3.1057	0.9813	1.5911	0.9455	0.2394	0.9900	5.3172
2016-2023	0.9974	0.9273	0.9928	0.1166	0.9977	2.8830	0.9847	1.2788	0.9537	0.1845	0.9915	4.8716
Avg	0.9911	1.2951	0.9691	0.1627	0.9902	4.1085	0.9734	2.1348	0.9238	0.3075	0.9788	6.9714

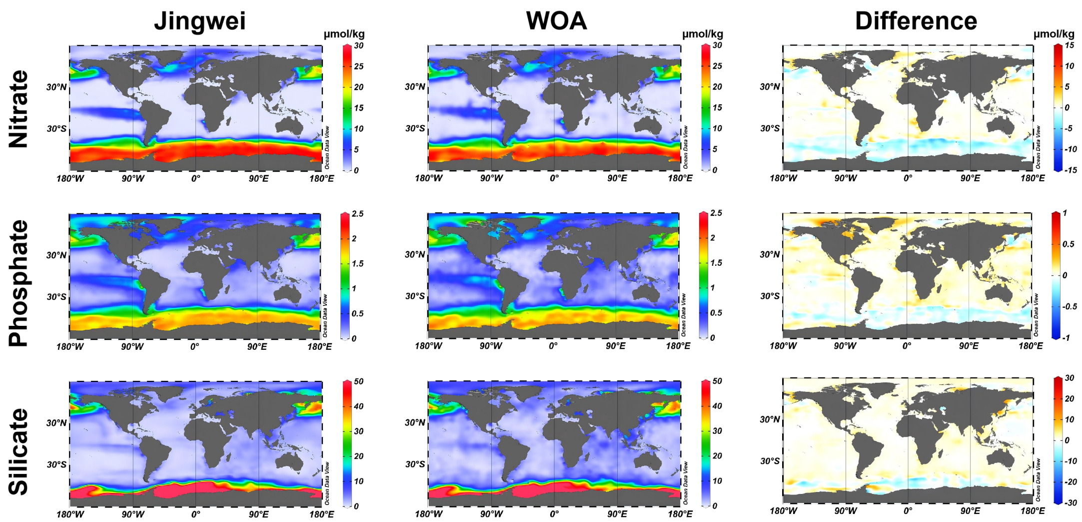


Figure 4. Spatial comparison of the climatological mean distributions of nitrate, phosphate, and silicate at the sea surface (0 m). The left column displays the *Jingwei-Nutrients* reconstruction, the middle column shows the WOA23 standard reference, and the right column presents the absolute difference (*Jingwei* minus WOA23). The surface fields highlight the model's precise capture of nutrient-depleted subtropical gyres and nutrient-rich outcropping zones.

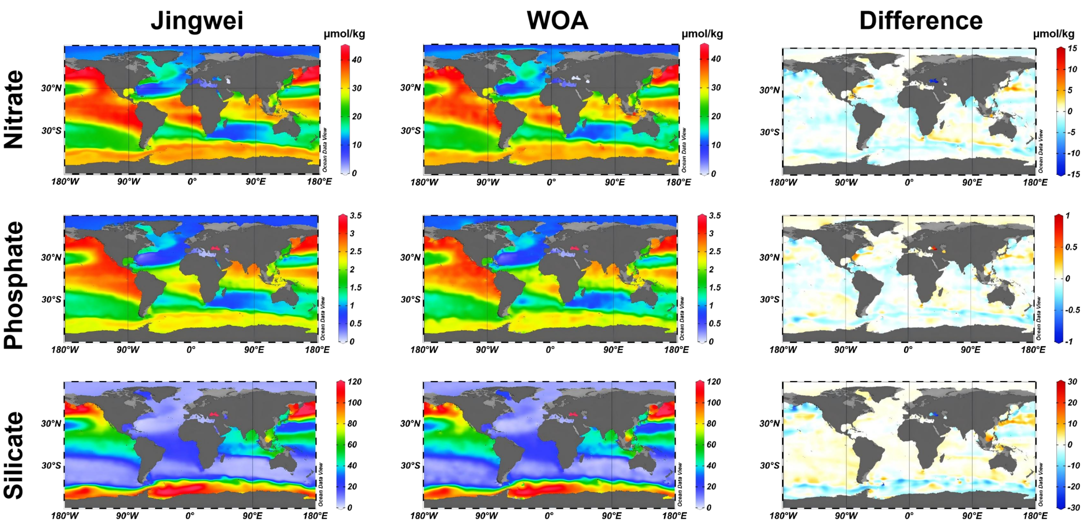


Figure 5. The Global climatological distributions of nitrate, phosphate, and silicate at the intermediate layer (500 m). The column and row layout is identical to Fig. 4.

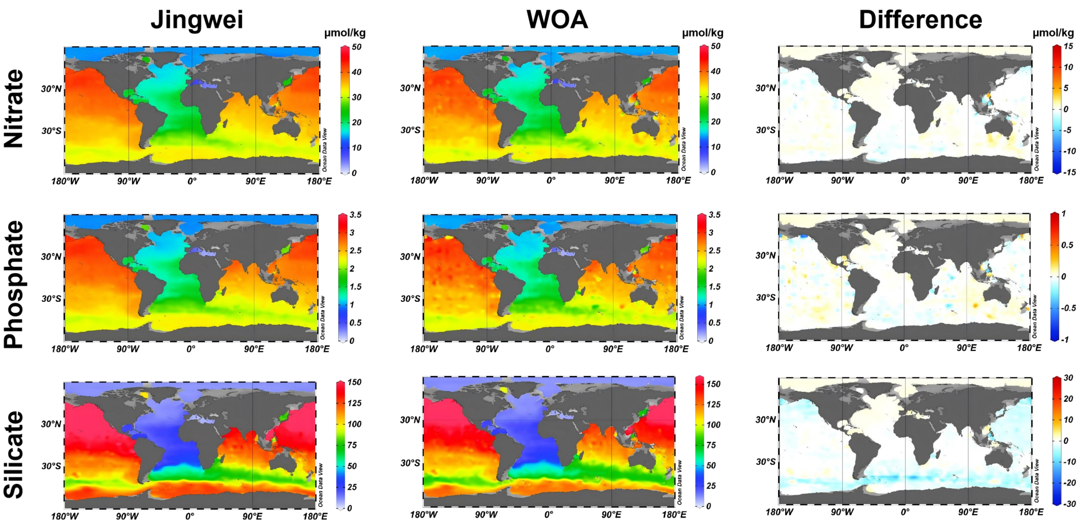


Figure 6. Global climatological distributions of nitrate, phosphate, and silicate at the deep layer (2000 m). The column and row layout is identical to Fig. 4.

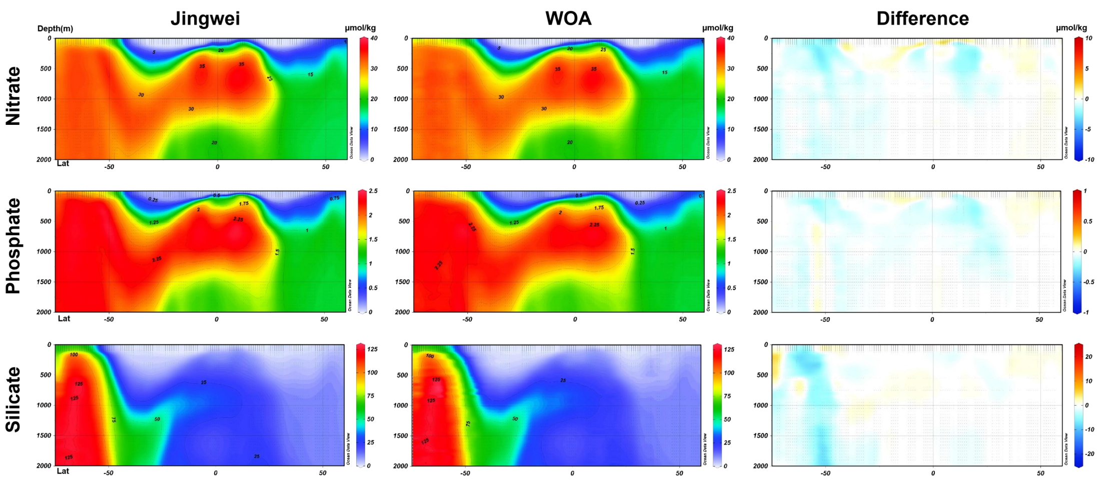


Figure 7. Vertical distribution of climatological mean nutrients along the 25°W meridional section in the Atlantic Ocean.The left column presents the *Jingwei-Nutrients* reconstruction, the middle column displays the WOA23 standard reference, and the right column shows the absolute difference (*Jingwei* minus WOA23). Rows from top to bottom correspond to nitrate, phosphate, and silicate, respectively. Units for all concentrations and differences are $\mu mol\ kg^{-1}$.

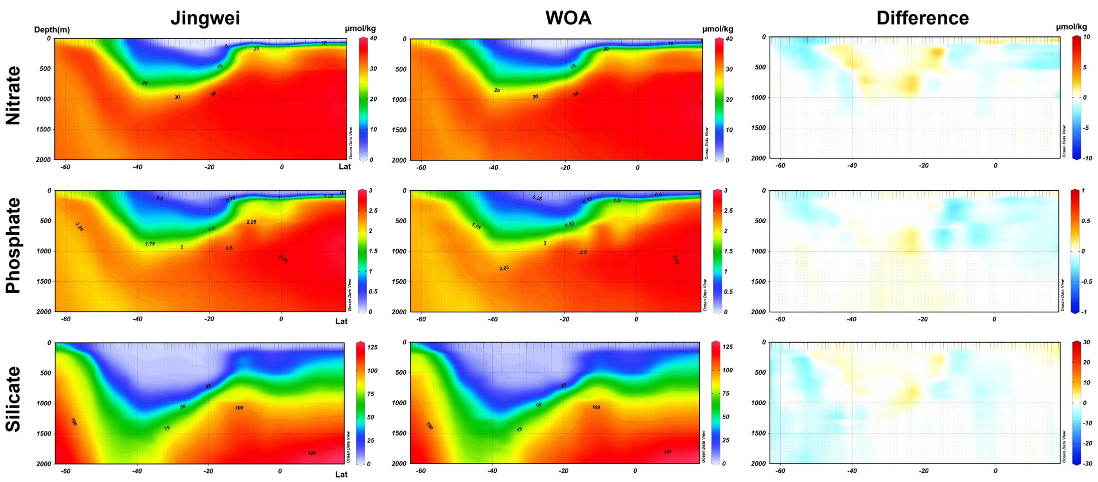


Figure 8. Vertical distribution of climatological mean nutrients along the 90°E meridional section in the Indian Ocean. The panel layout and represented variables are identical to those in Figure 7. Units for all concentrations and differences are $\mu mol\ kg^{-1}$.

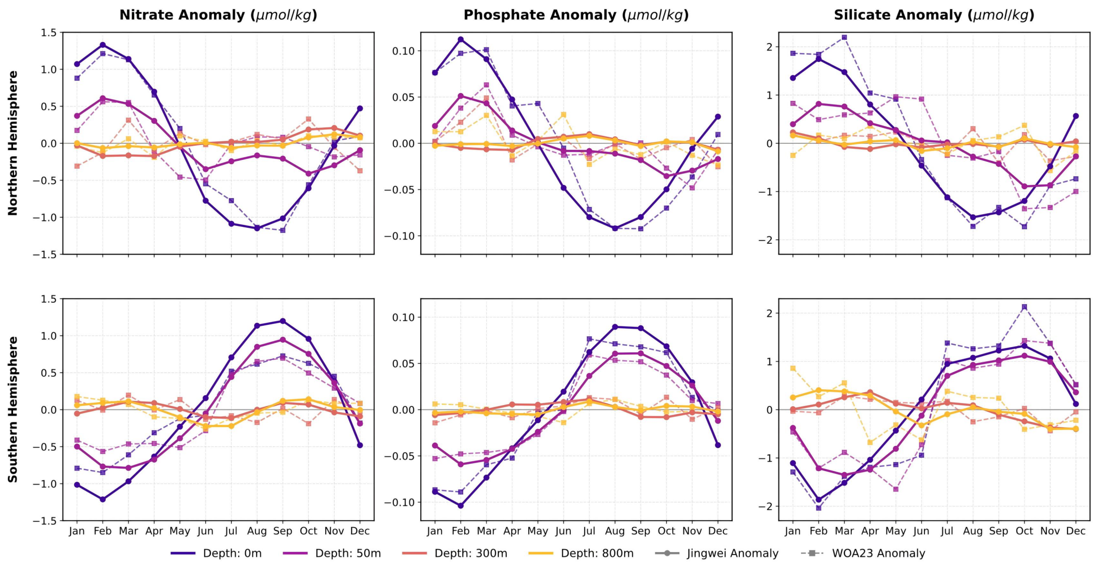


Figure 9. Seasonal cycles of nutrient anomalies across different hemispheres and depth layers. Solid lines indicate the *Jingwei-Nutrients* reconstruction, and dashed lines represent the WOA23 reference. The top row shows the Northern Hemisphere, and the bottom row shows the Southern Hemisphere, encompassing nitrate, phosphate, and silicate (from left to right) at 0, 50, 300, and 800 m depths. Units are $\boldsymbol{\mu mol\ kg^{-1}}$.

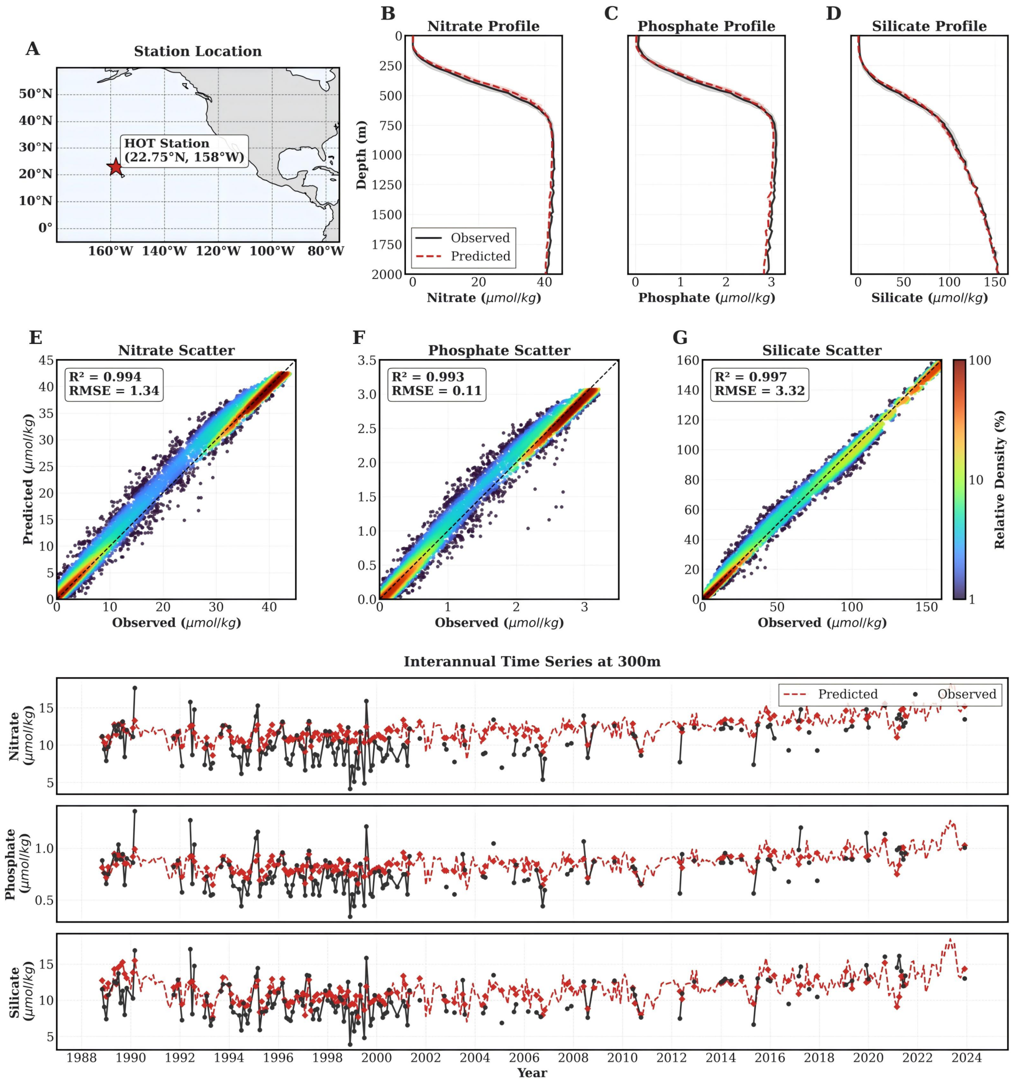


Figure 10. Multi-scale temporal and vertical validation at the Hawaii Ocean Time-series (HOT) station in the North Pacific. (A) Geographical location of the HOT station. (B–D) Climatological vertical profiles comparing in situ observations (black solid lines) and *Jingwei-Nutrients* predictions (red dashed lines) for nitrate, phosphate, and silicate. (E–G) Density scatter plots evaluating the overall predictive accuracy, with R^2 and RMSE metrics embedded. The bottom panels display the interannual time series of nutrient concentrations evaluated at the 300 m depth layer, demonstrating the model's capability to maintain stable, long-term baselines.

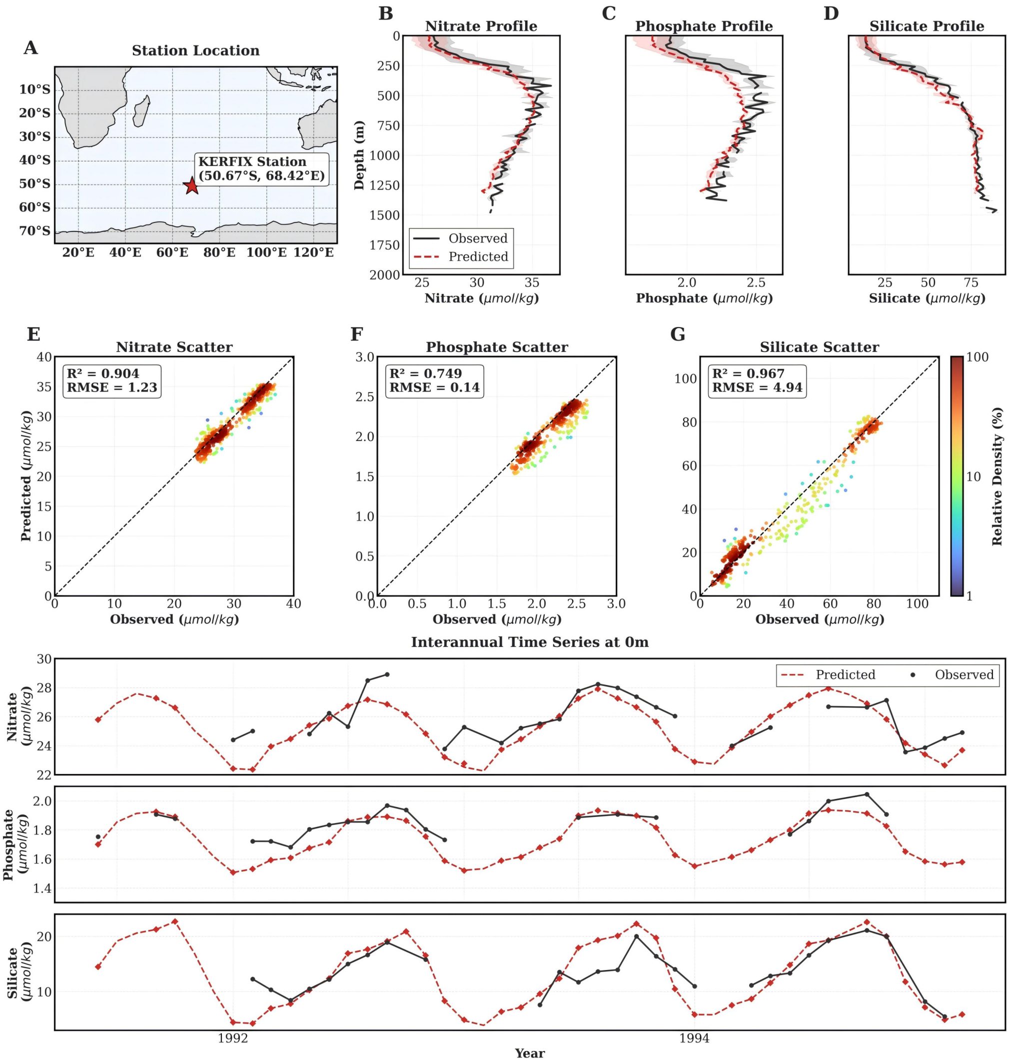


Figure 11. Multi-scale temporal and vertical validation at the KERFIX station in the Southern Ocean. The panel layout is identical to Fig. 10. This figure highlights the model's performance in a highly dynamic, high-latitude regime. The interannual time series (bottom panels) are evaluated at the surface layer (0 m) to explicitly track the massive seasonal amplitudes driven by deep convective winter mixing and summer biological drawdowns.

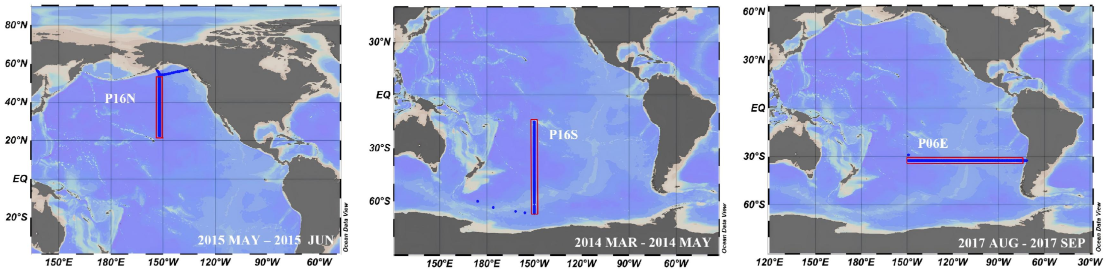


Figure 12. Map of the selected U.S. GO-SHIP cruise trajectories used for independent spatial validation. The solid lines indicate the continuous high-resolution shipboard measurement sections: the P16N (2015) and P16S (2014) meridional cruises in the Pacific Ocean, and the P06E (2017) zonal cruise in the South Pacific.

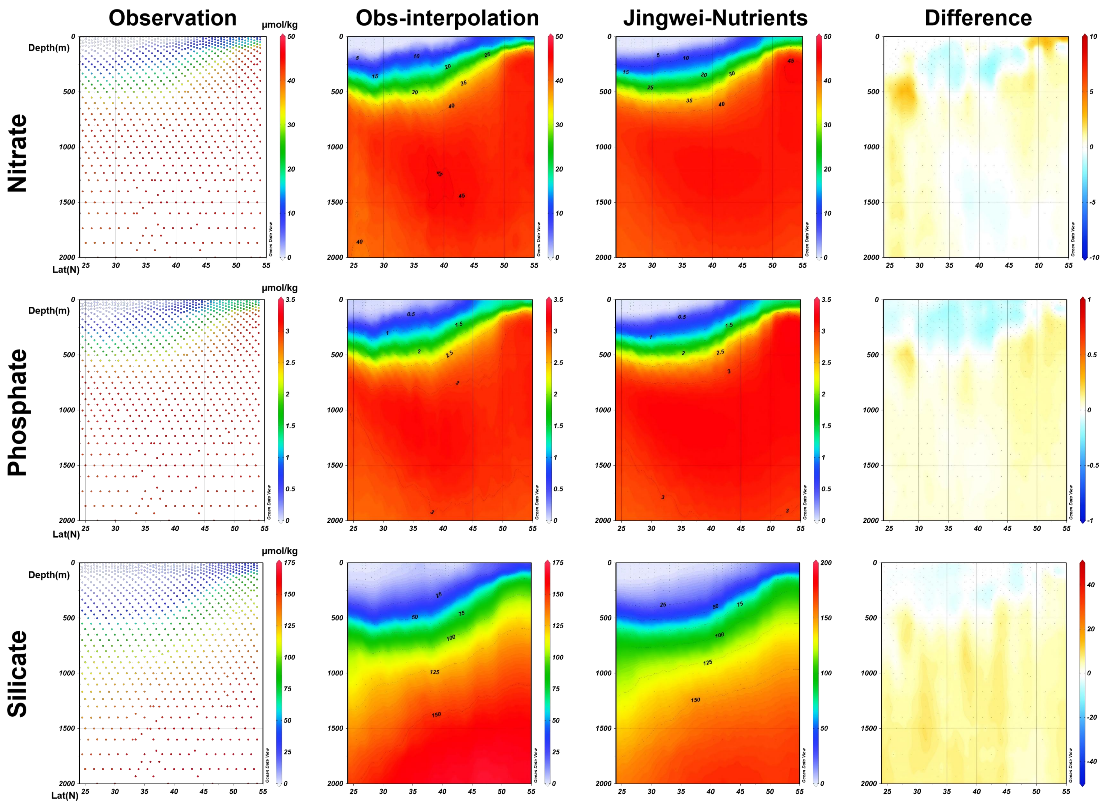


Figure 13. Spatial validation of nutrients along the P16N meridional section in the North Pacific. Columns from left to right display the in situ discrete shipboard measurements, the objectively interpolated observational field, the coincident cross-section extracted from the *Jingwei-Nutrients* product, and the absolute difference (*Jingwei* minus Observation). Rows from top to bottom correspond to nitrate, phosphate, and silicate, respectively.

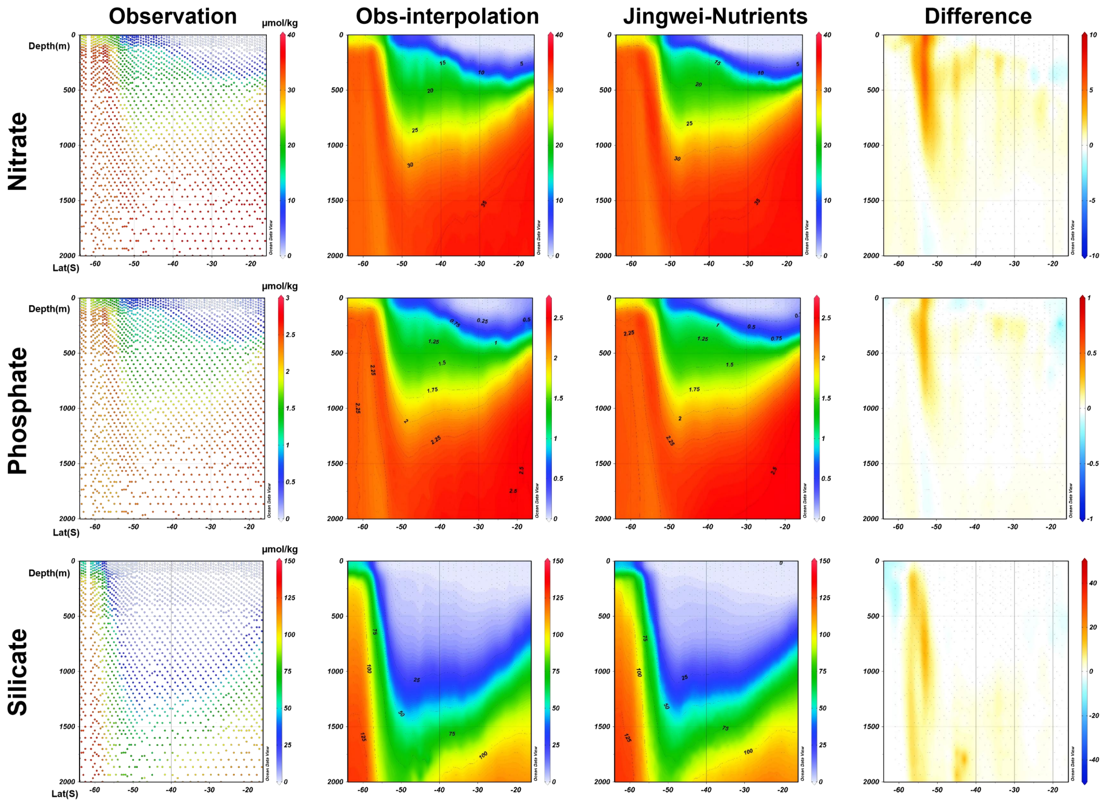


Figure 14. Spatial validation of nutrients along the P16S meridional section in the South Pacific and Southern Ocean. The panel layout and represented variables are identical to those in Fig. 13.

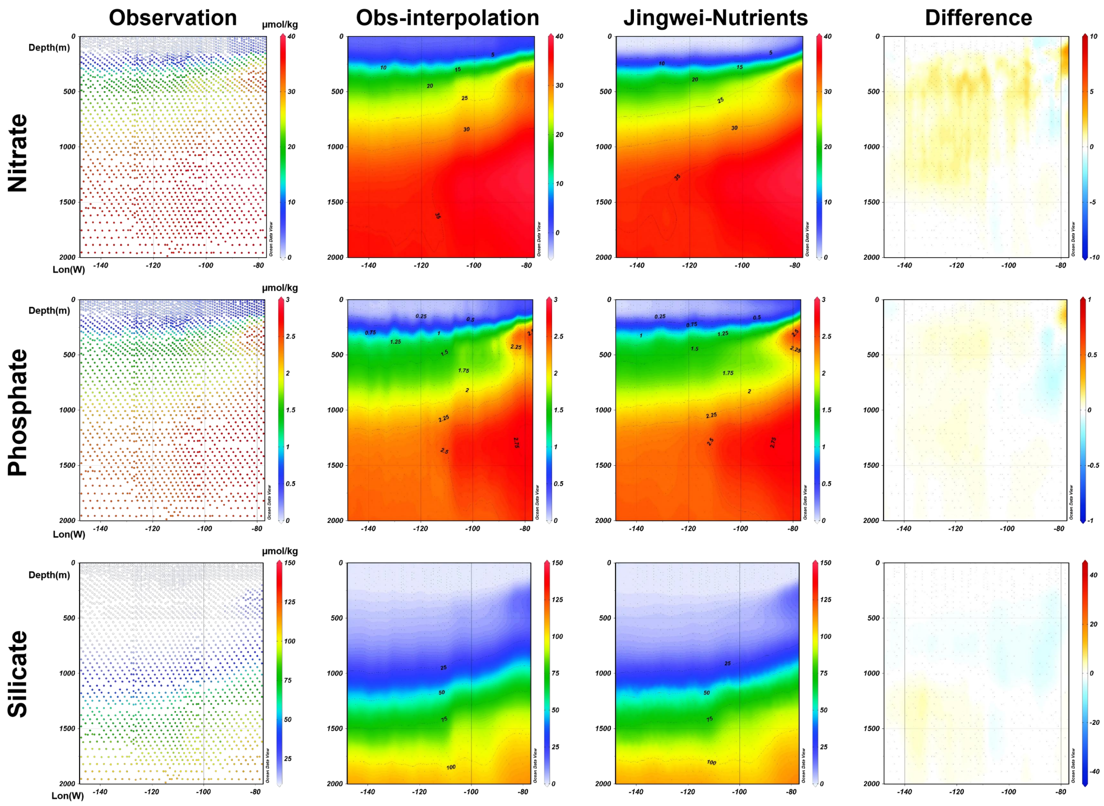


Figure 15. Spatial validation of nutrients along the P06E zonal section in the South Pacific.The panel layout and represented variables are identical to those in Fig. 13.

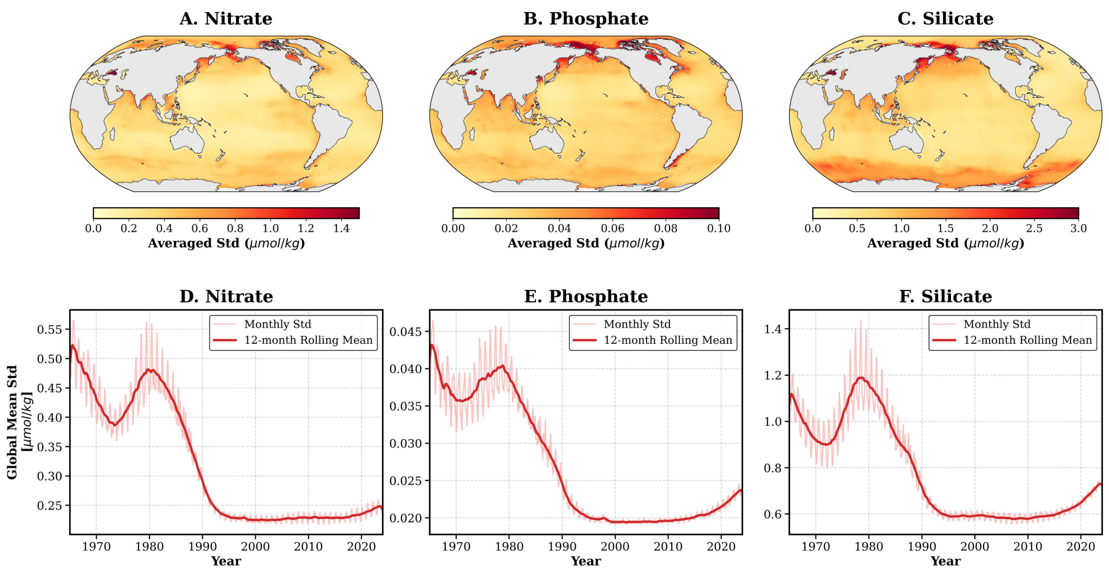


Figure 16. Spatiotemporal uncertainty analysis of the *Jingwei-Nutrients* reconstruction based on the ensemble standard deviation across the 6-fold cross-validation models. The top row displays the spatial distribution of the depth-averaged standard deviation for nitrate, phosphate, and silicate. The bottom row presents the temporal evolution of the global mean standard deviation from 1965 to 2023. Light red lines represent the raw monthly uncertainty, while the thick red lines indicate the 12-month rolling mean.