Response to Reviewer #1

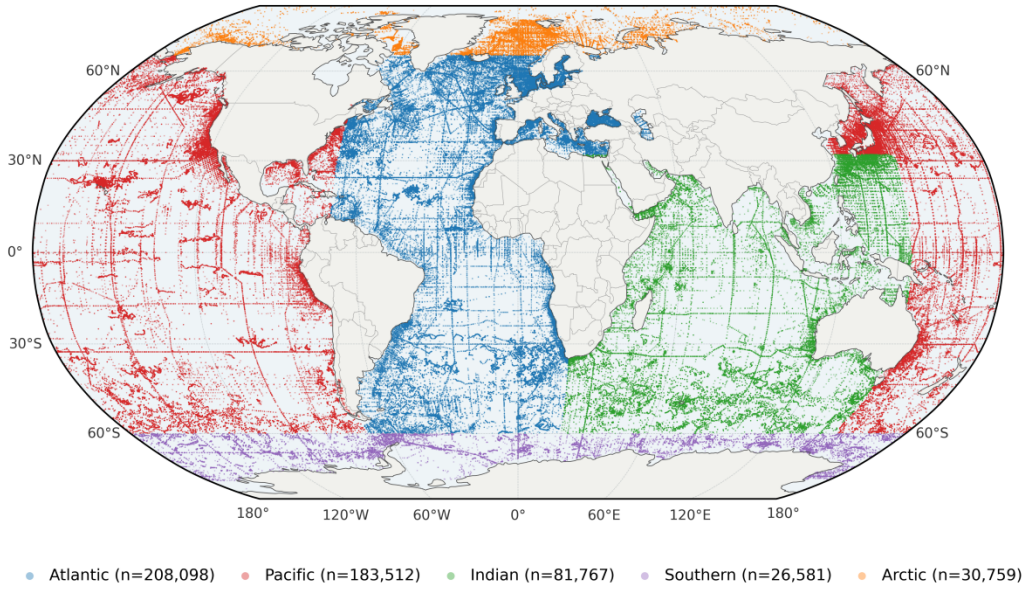
We sincerely thank the reviewer for the positive evaluation of our manuscript and for the constructive suggestions. We are grateful that the reviewer recognized the value of the proposed global nutrient reconstruction dataset and the detailed preprocessing procedures. We have revised the manuscript accordingly 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.

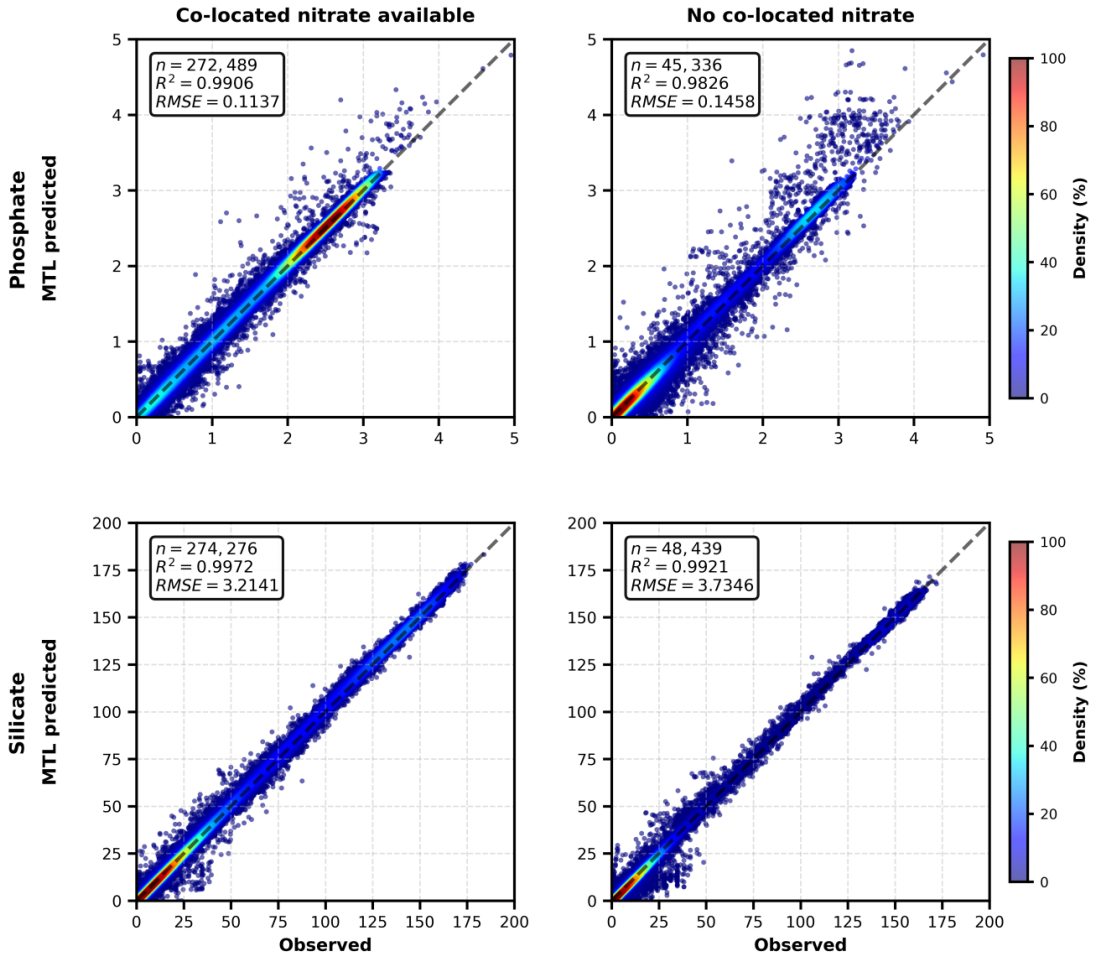
To further examine whether such fragmented observations can provide useful information for nutrient reconstruction, we conducted an additional last-fold withholding experiment. The final chronological fold, 2016-2023, was used as a controlled inference period. In this experiment, all phosphate and silicate 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.