

We appreciate the valuable feedback from the reviewers. Their constructive comments have significantly improved the manuscript. We have carefully revised the paper in response to their suggestions to meet the high standards of Earth System Science Data.

Reviewer 1#

This manuscript presents a valuable contribution to the field of chemical oceanography. The authors have reconstructed a massive database of historical nutrient data points for the North Pacific, greatly expanding original observations. The rigorous four-level quality control and the use of multiple machine learning (ML) architectures make this a strong candidate for the journal Earth System Science Data.

Overall, this is a good paper. However, I have some concerns regarding the temporal extrapolation, which must be addressed to ensure the dataset's reliability for historical hindcasts. Besides, the paper would benefit from a more in-depth discussion on the long-term trend of nutrients, which can help strengthen the utility of such a historical dataset.

I recommend minor revisions to strengthen the methodology and discussion before publication further.

[Response] We thank the reviewer for the positive evaluation and thoughtful suggestions. As advised, we have added new analyses and discussions on temporal extrapolation and long-term nutrient trends to improve the reliability and utility of the dataset. The corresponding revisions can be found in Sections 3.1 and 3.4 of the revised manuscript.

Major comments:

1. *Temporal extrapolation robustness*: The use of three validation strategies (sample-random, station-random, and cruise-random) provides a transparent view of error, and the cruise-random approach should be most convincing in validating spatial extrapolation. However, in terms of temporal extrapolation, the training dataset (CCHDO) only spans from 1973 to 2022, then the model is applied to reconstruct nutrients going back to 1895. The inclusion of year as a predictor might be biased if any trend learned from 1973–2022 does not map onto the 1895–1972 era. The authors need to justify that their approach can be extrapolated not only spatially but also temporally, maybe through some discussion about whether the water mass-nutrient relationships remained relatively stationary over the last century (but is this really true given the acceleration of anthropogenic forcing?), or validate with some "time-slices" to prove the temporal predictor is robust.

[Response] We thank the reviewer for raising this critical point regarding the robustness of temporal extrapolation. The reviewer is correct to highlight that our core training data (CCHDO) spans 1973–2022, while the reconstruction extends back to 1895. We address this concern through dedicated validation analyses and a physical rationale, as detailed below and in the revised manuscript.

Our decision not to incorporate pre-1973 nutrient data from sources like the Ocean Station Data (OSD) from World Ocean Database (WOD) for model training was

primarily due to data quality concerns. Prior to the standardization of modern oceanographic methods, nutrient measurements—particularly from earlier decades—were subject to greater analytical errors, inconsistent sampling protocols, and varied determination techniques. This is evident in the sporadic and sometimes physically implausible deep nutrient profiles found in WOD for that era (e.g., discrete values at depths >1000 m; see Fig. R1 in our response, now included as Fig. S22 in the revised Supporting Information).

To directly assess the temporal extrapolation capability of our model, we performed a targeted validation using the relatively higher-quality nutrient observations available from WOD for five specific years with more abundant data: 1929, 1947, 1953, 1958, and 1966. After applying the same quality-control criteria outlined in our Methods, we used the historical hydrography (temperature/salinity) from those years to reconstruct nutrient concentrations. The comparisons between reconstructed and observed values for NO_x^- , DIP, and $\text{Si}(\text{OH})_4$ are shown in Figs. R2–R5 (now Figs. S23–S26 in the revised Supporting Information).

The results show that the reconstruction errors for these historical periods are, as expected, larger than those from the cruise-random validation (which uses modern, high-quality data). The root mean square errors (RMSEs) reach upper limits of approximately $5.7 \mu\text{mol kg}^{-1}$ for NO_x^- , $0.40 \mu\text{mol kg}^{-1}$ for DIP, and $22.9 \mu\text{mol kg}^{-1}$ for $\text{Si}(\text{OH})_4$. We interpret these as conservative error estimates for the 1895–1970 period, acknowledging that they encompass both model prediction uncertainty and the larger observational errors inherent in early measurements.

We argue that this extrapolation should be reasonable because the variations of temperature-salinity-nutrient relationships in the ocean's interior might be small over the past century, providing a basis for temporal extrapolation. First, the residence time of nitrogen in deep and intermediate waters can be up to 2000 years in the North Pacific. Consequently, the imprint of centennial-scale change on nutrient inventories is attenuated. Second, the long-term variations of nutrient concentrations are not evident within our core training period (1973–2022; Figs. 9e and 17 in the revised manuscript). Finally, the mean nutrient profiles derived from the 1920–1970 and 1973–2022 periods are not evidently different in the central North Pacific (Fig. R6 and Fig. S21 in the revised Supporting Information). Therefore, while the North Pacific may experience long-term variability, it might be masked by the reconstruction error, and the use of hydrographic properties as predictors for nutrients is justified for historical reconstructions. In addition, the analysis of anthropogenic influence requires further and more detailed investigation in future studies.

In summary, we have added discussions at L386–410 and L451–488 in the revised manuscript that presents this temporal validation exercise and discusses the associated uncertainties. We acknowledge that reconstruction errors are likely higher for the pre-1973 period, and the error estimated here should be considered as a "best estimate" with quantified uncertainties, and encourage users to consider these error bounds when applying the dataset to early twentieth-century conditions.

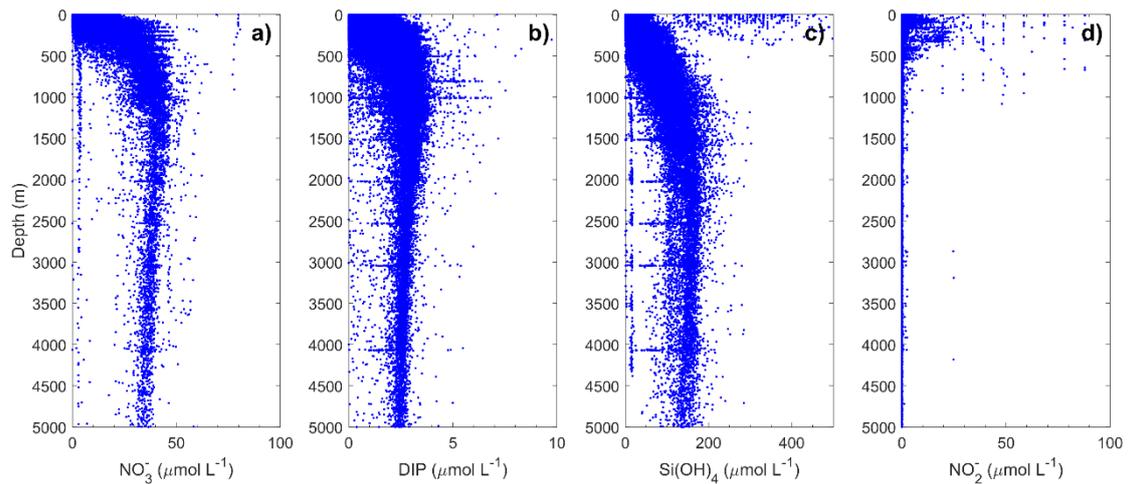


Figure R1. Profiles of nutrients collected from the Ocean Station Data in the World Ocean Database (before 1970) in the North Pacific basin: (a) NO_3^- ; (b) DIP; (c) Si(OH)_4 ; (d) NO_2^- , illustrating the scattered vertical distribution.

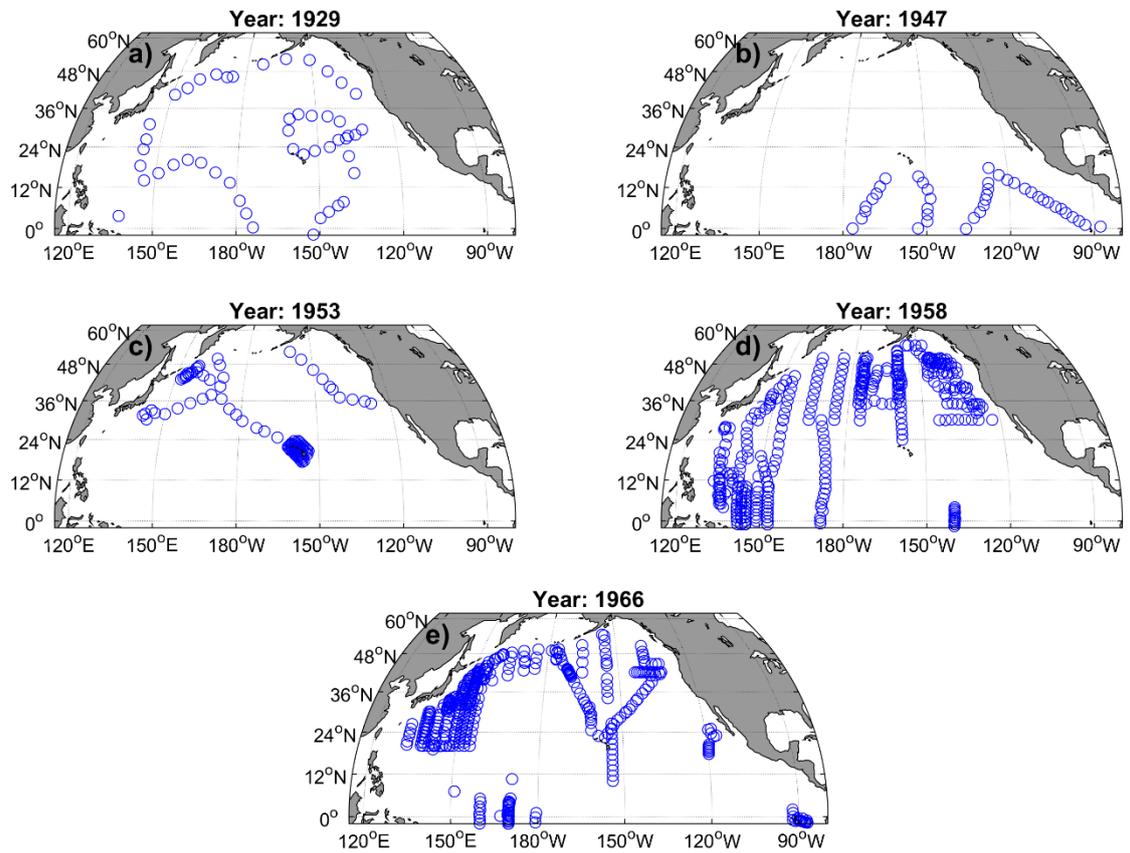


Figure R2. Hydrographic and nutrient sampling locations for selected years prior to 1970 (collected from Ocean Station Data in the World Ocean Database): (a) 1929, (b) 1947, (c) 1953, (d) 1958, and (e) 1966.

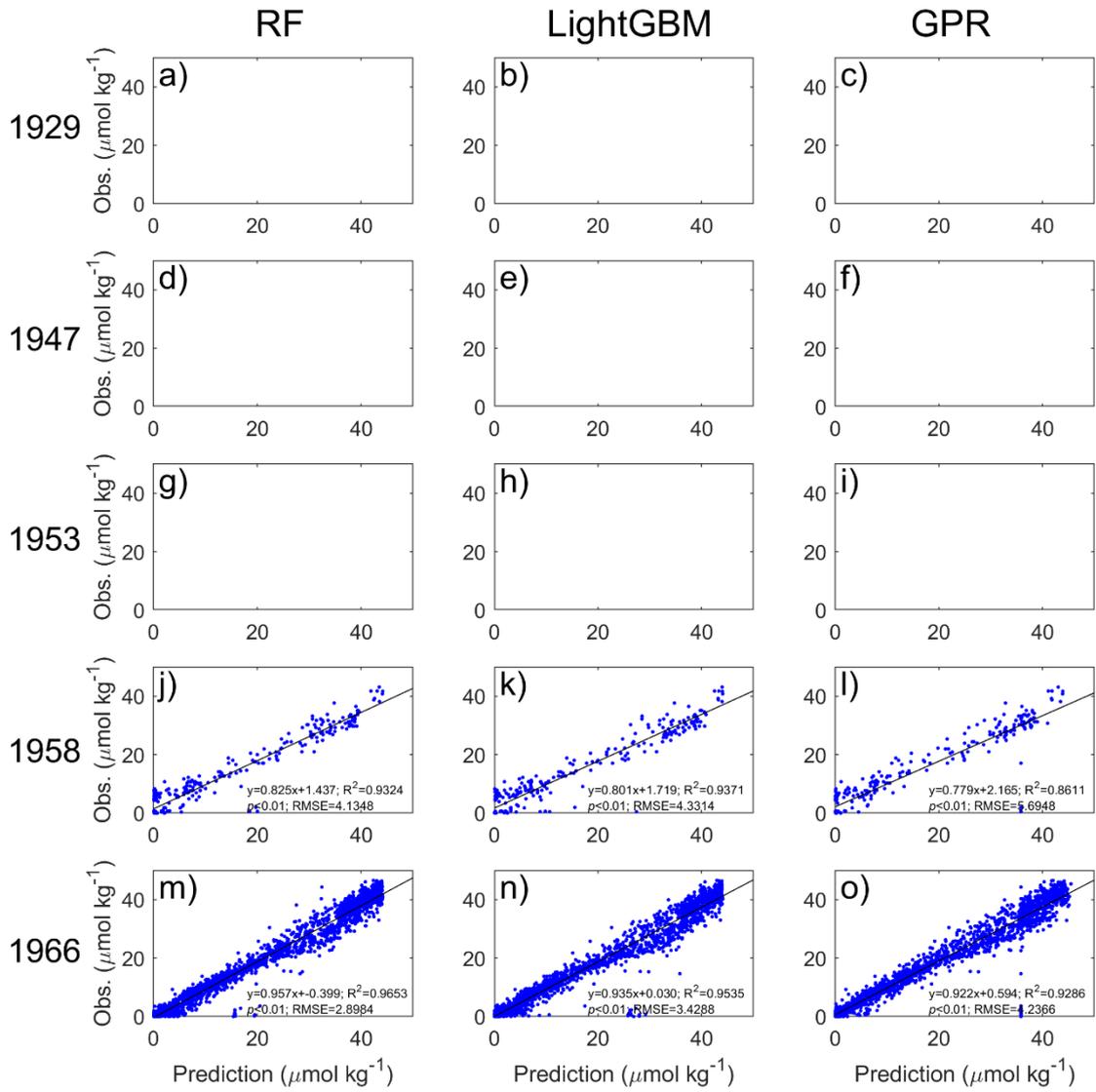


Figure R3. Validation of reconstructed NO_3^- concentrations from hydrography-based reconstructions for 1929 (row 1), 1947 (row 2), 1953 (row 3), 1958 (row 4), and 1966 (row 5). Station locations are shown in Fig. S23. Columns correspond to different reconstruction methods: Random Forest (RF; a, d, g, j, m), LightGBM (b, e, h, k, n), and Gaussian Process Regression (GPR; c, f, i, l, o). Each panel displays the linear regression fit (black line), regression equation, coefficient of determination (R^2), p-value, and root mean square error (RMSE). Note: no NO_3^- observations were available for 1929, 1947, or 1953.

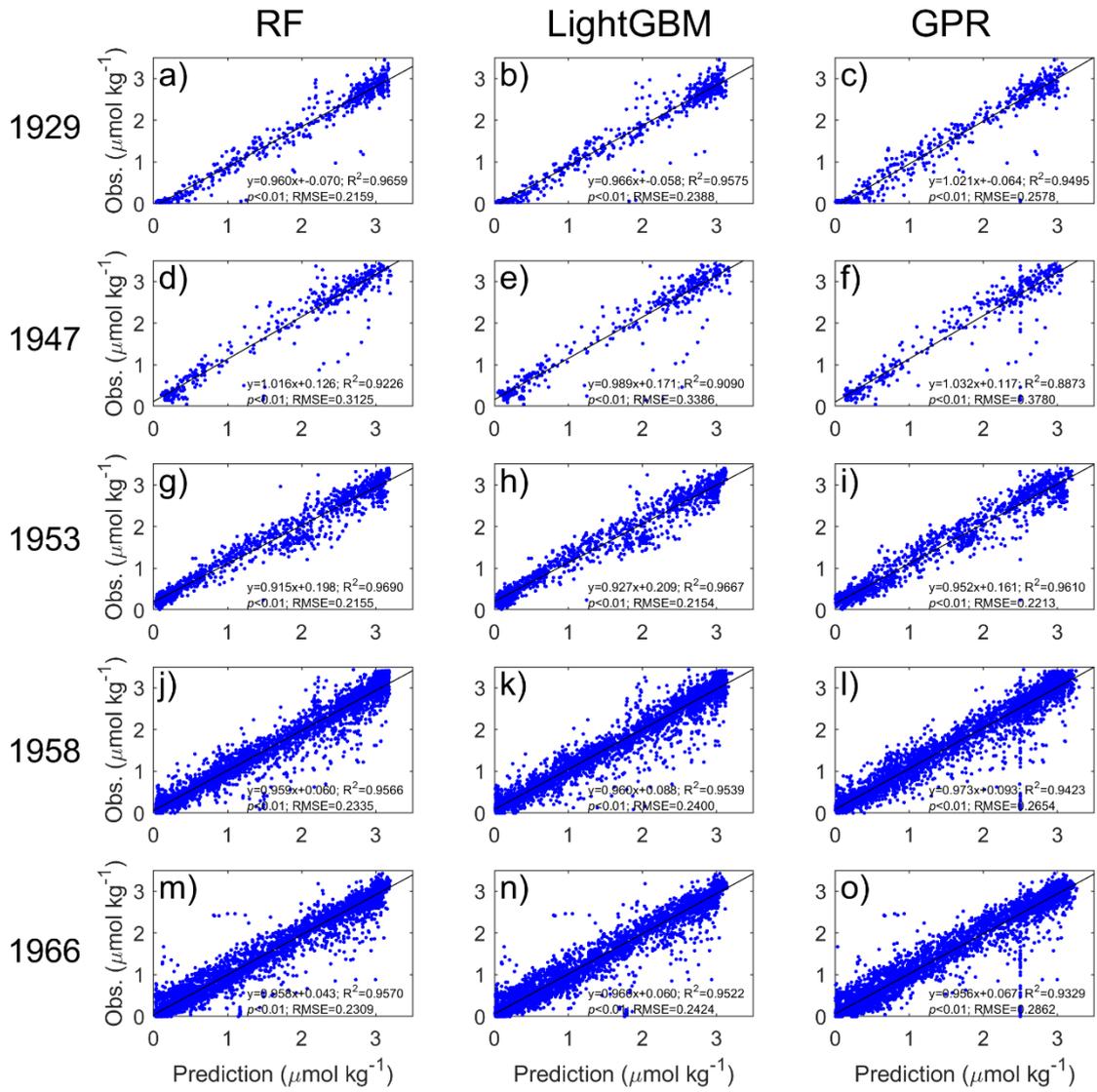


Figure R4. Same as Fig. R3, but for DIP.

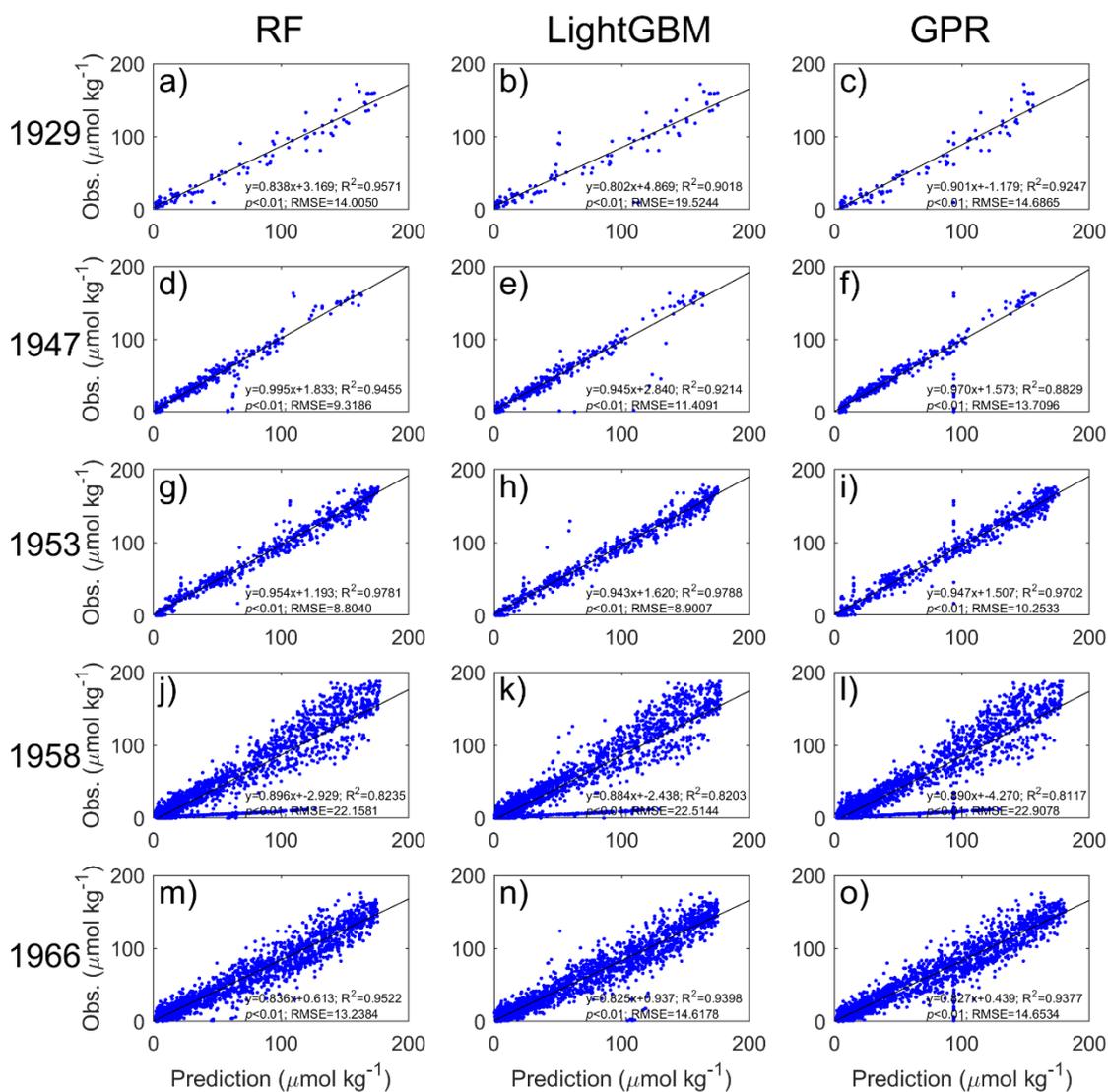


Figure R5. Same as Fig. R3, but for Si(OH)_4 .

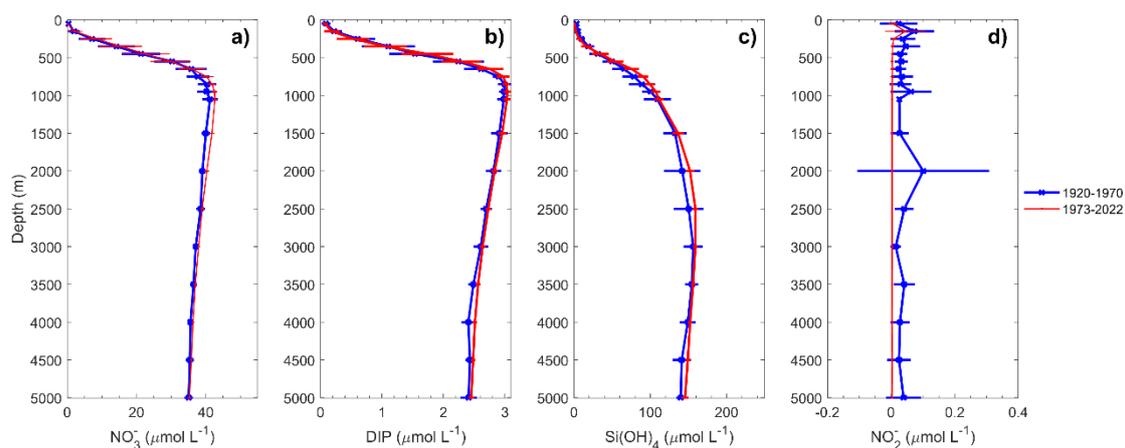


Figure R6. Mean profiles (vertical lines) with standard deviations (horizontal lines) of nutrients between 1920–1970 (Ocean Station Data in the World Ocean Database after quality control) and 1973–2022 (CCHDO data) in the North Pacific basin region (180°E – 150°W , 15°N – 30°N): (a) NO_3^- ; (b) DIP; (c) Si(OH)_4 ; (d) NO_2^- .

2. *Missing long-term analyses*: A major selling point of this paper is the temporal extent of the reconstruction (1895–2024). However, the Results section is dominated by climatological maps (Figs. 10–13), which effectively collapse the temporal dimension that the authors worked so hard to reconstruct. Providing 130 years of data without showing a single long-term trend analysis (e.g., decadal shifts in the nutricline depth, or basin-scale nutrient inventory changes) undermines the claim that this dataset is ‘historical.’ I suggest the authors add a section analyzing a long-term trend or any regime shift using their reconstructed nutrient data. This would serve as a proof of concept that the reconstruction captures low-frequency climate variability and is not just a high-resolution climatology.

[Response] We thank the reviewer for this important suggestion. We agree that demonstrating the dataset's ability to capture long-term variability is crucial. Accordingly, we have added a new Section 3.4 ("Long-term variations of nutrients") in the revised main manuscript. In this section, we present an initial analysis of long-term nutrient changes by examining five representative regions in the North Pacific (covering the subarctic gyre, subtropical gyre, and equatorial areas).

As shown in Figs. R7–R12, we plot the reconstructed monthly and spatially averaged concentrations of NO_x^- ($\text{NO}_3^- + \text{NO}_2^-$), NO_2^- , DIP, and $\text{Si}(\text{OH})_4$ at several standard depths (10 m to 1000 m) for these regions from 1895 to 2024. These time series reveal notable interannual to decadal-scale fluctuations, providing a first-order view of low-frequency variability captured by the reconstruction.

We acknowledge the reviewer's point regarding more integrated metrics like nutricline depth or basin-scale inventory changes. However, performing such analyses is challenging given the current dataset structure. The reconstructed nutrient fields are inherently aligned with the irregular spatiotemporal distribution of the underlying hydrographic observations (see examples in Fig. R12). Calculating physically meaningful trends therefore requires careful gridding and interpolation, which would introduce substantial computational overhead and could amplify uncertainties, especially in the data-sparse early decades. As the primary focus of this study is on the reconstruction methodology and the resulting dataset, we have chosen to present these regional time series as a foundation for demonstrating the dataset's utility in long-term studies.

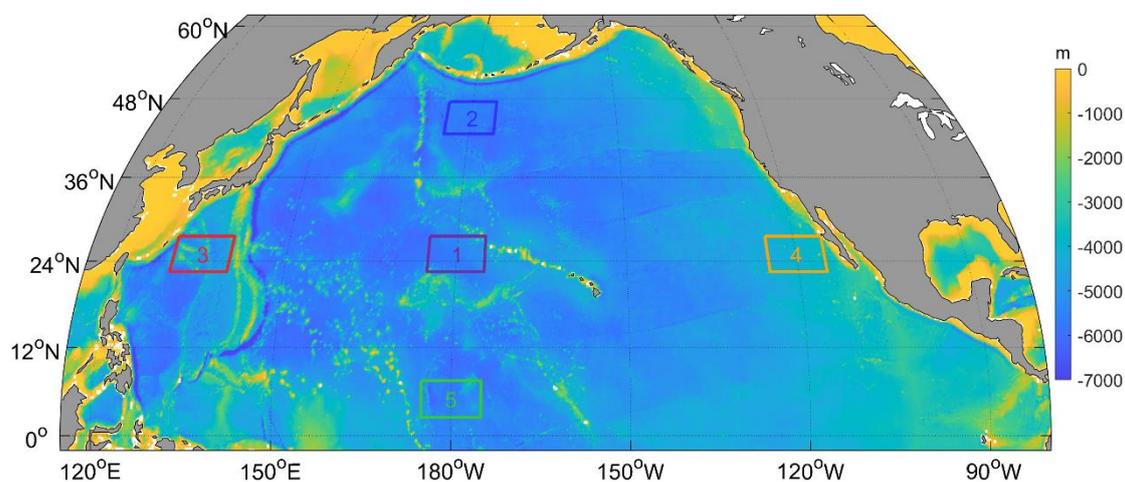


Figure R7. Locations of five representative regions for analyzing long-term nutrient variations.

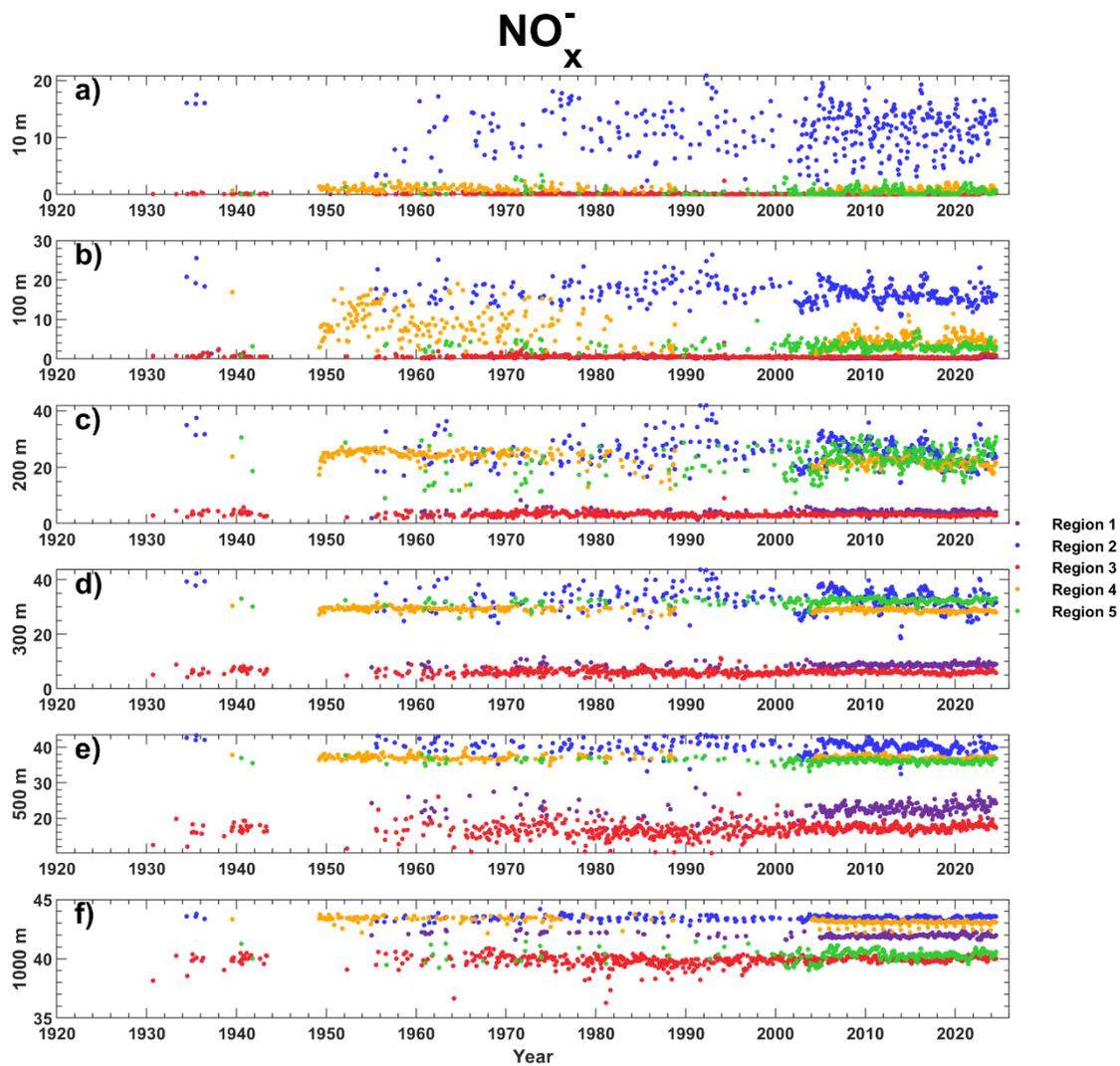


Figure R8. Time series of reconstructed NO_3^- concentrations at 10 m (a), 100 m (b), 200 m (c), 300 m (d), 500 m (e), and 1000 m (f) for regions 1–5 (see Fig. 16). Data were binned by depth and region and then averaged by month.

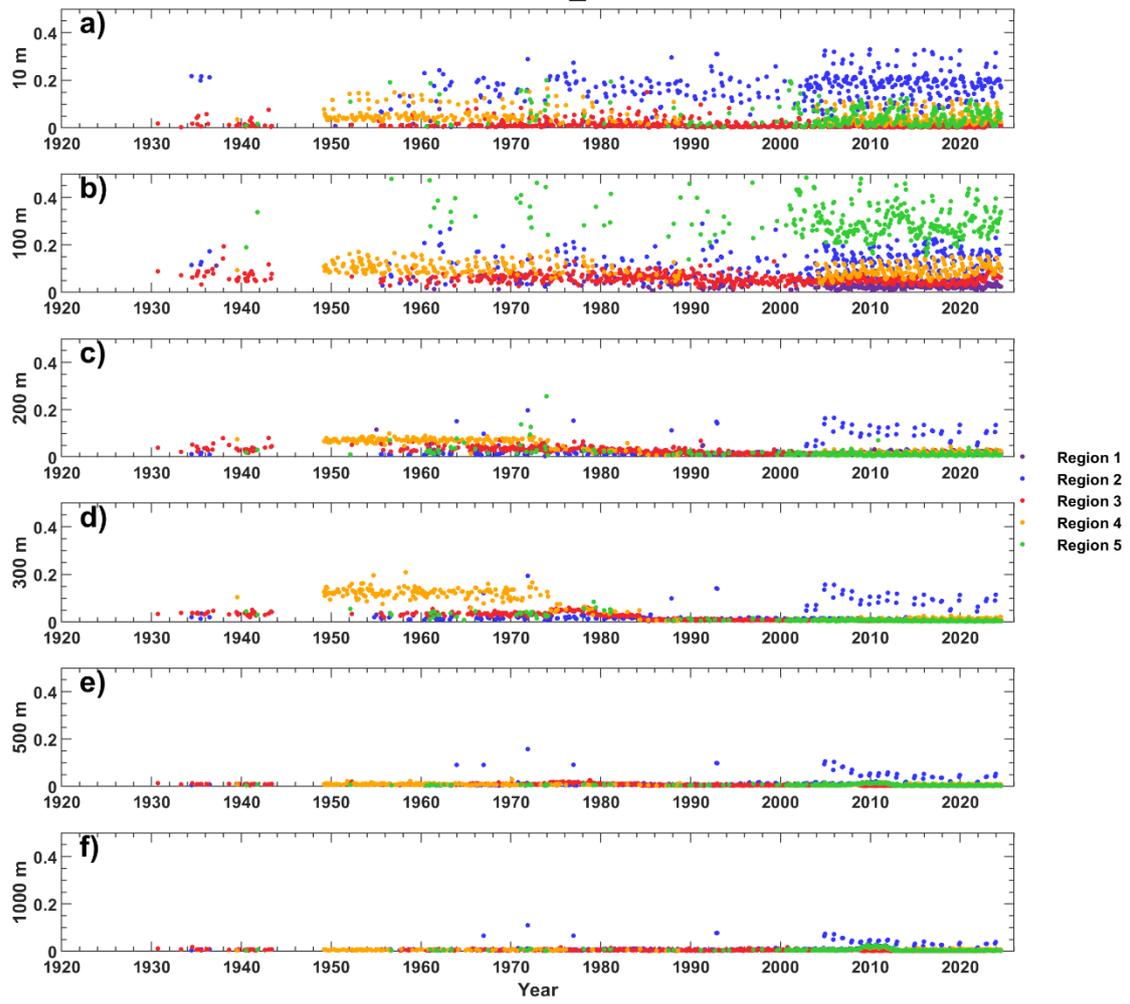


Figure R9. Same as Figure R8, but for NO_2^- .

DIP

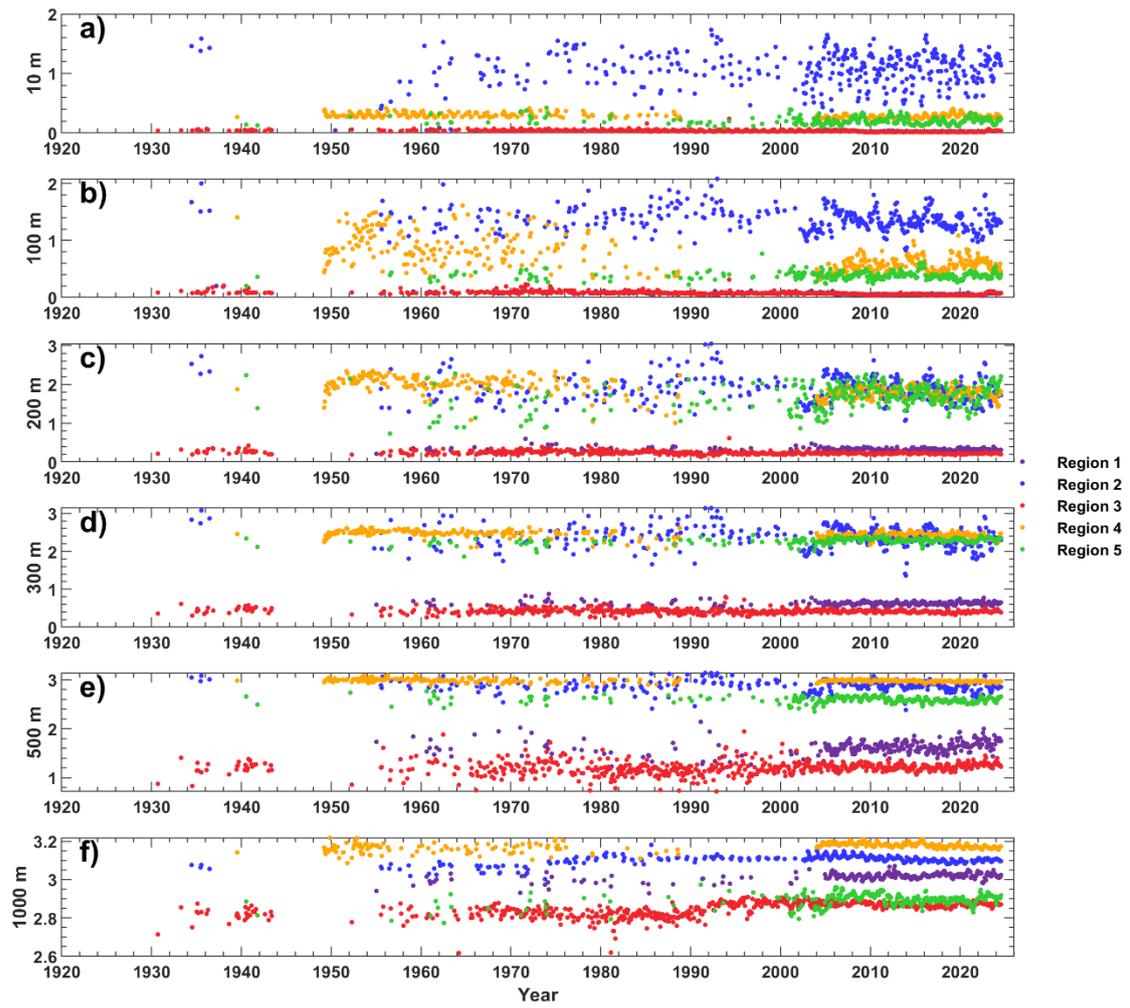


Figure R10. Same as Figure R8, but for DIP.

Si(OH)₄

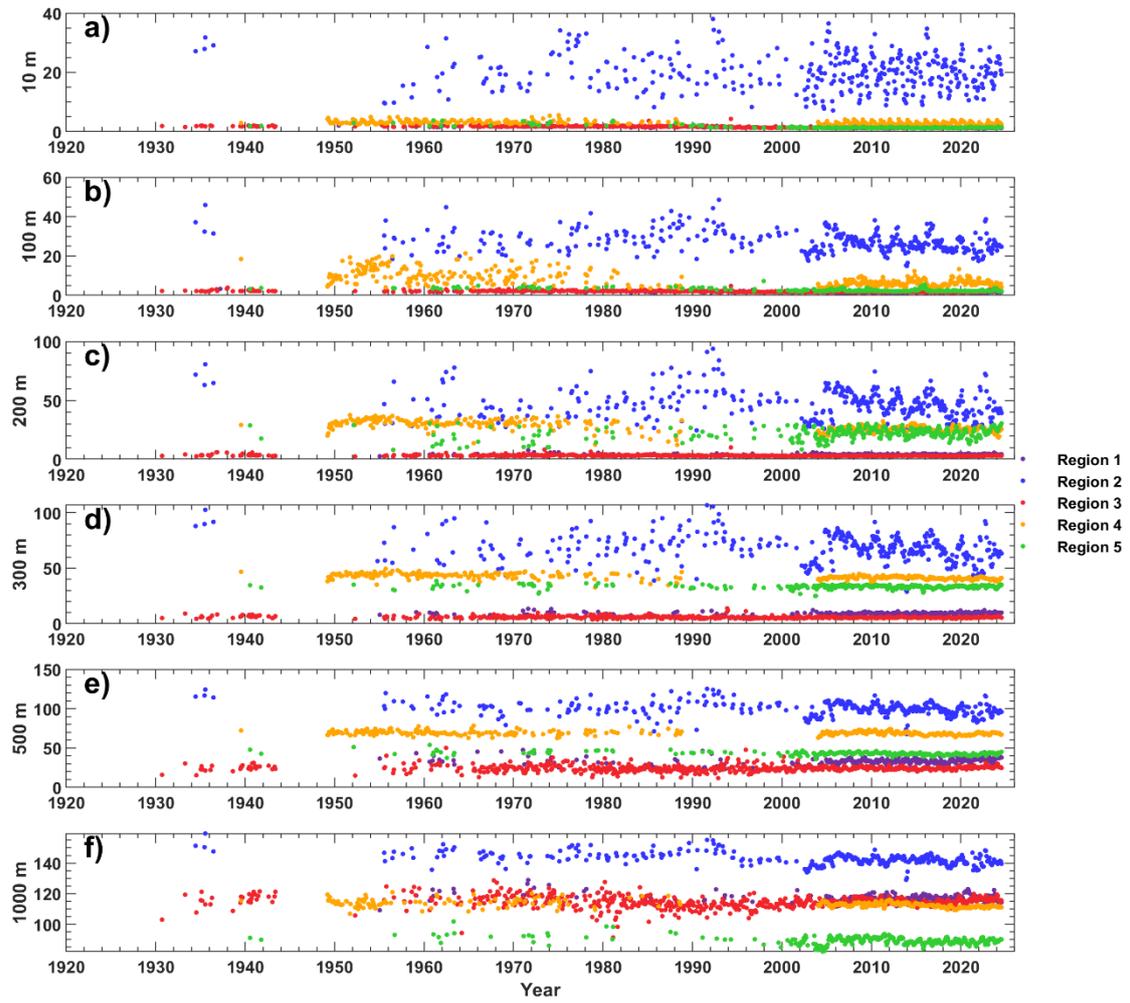


Figure R11. Same as Figure R8, but for Si(OH)₄.

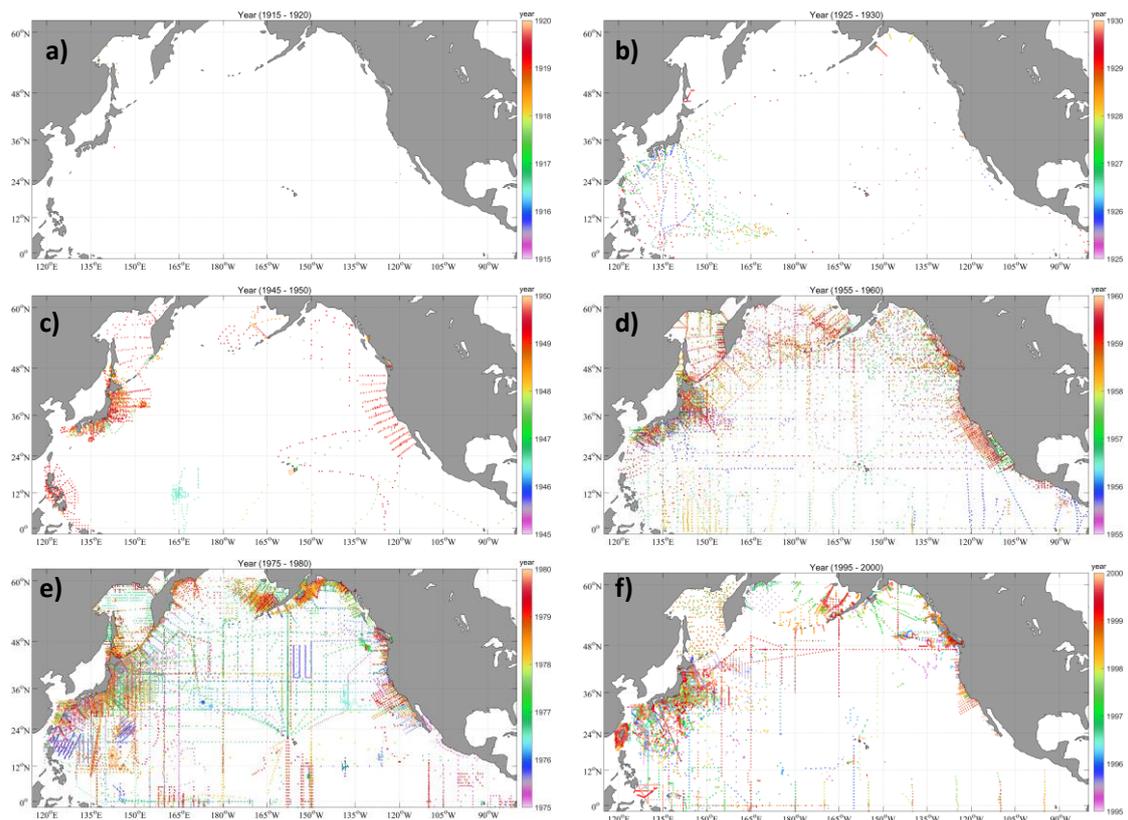


Figure R12. Spatial distribution of hydrographic locations (which correspond to the reconstructed nutrient data locations) during six selected periods: (a) 1915–1920, (b) 1925–1930, (c) 1945–1950, (d) 1955–1960, (e) 1975–1980, and (f) 1995–2000. The colorbar indicates the year of sampling.

3. *Elaboration on potential future applications:* I think the reconstructed datasets would be impactful and have broad utility, but their applications are written in a generic way. To increase the impact of this paper, I recommend expanding the discussion to explicitly list potential future applications of this dataset. Specific examples could be to use this 4D dataset to spin up ocean biogeochemical models, or investigate nutrient stoichiometric changes, etc.

[Response] We thank the reviewer for the constructive suggestion. In response, we have expanded the discussion in the manuscript to explicitly elaborate upon specific potential applications of the reconstructed dataset. The revised text now highlights its use for: (1) examining basin-scale nutrient transport and budgets; (2) spinning up and validating ocean biogeochemical models; and (3) investigating long-term nutrient trends and stoichiometric changes in response to climate variability and anthropogenic forcing. These additions have been incorporated in the Abstract (L36-39 in the revised manuscript) and the Discussion section (Section 5 and L627-633 in the revised manuscript) of the revised manuscript.

Minor comments:

- Section 2.1: Oxygen is a fundamental tracer for remineralization and is physically coupled with nutrients via the Redfield ratio and AOU, but is not included in the predictors. Can the authors explain why it is not included? Is it because many datasets lack this property?

[Response] We thank the reviewer for raising the point regarding the role of oxygen as a predictor. Indeed, oxygen (or apparent oxygen utilization, AOU) is a key tracer of remineralization and is stoichiometrically linked to nutrients via the Redfield ratio. The decision not to include oxygen in our predictor set was based on the following practical and methodological considerations:

1) Data co-occurrence and completeness: Nutrient observations in our primary training dataset (CCHDO) are not always accompanied by concurrent, high-quality dissolved oxygen (DO) measurements. Including DO as a predictor would remove the nutrient data lacking a corresponding DO value, drastically reducing the size and spatial coverage of the training dataset, particularly for the Gaussian Process Regression model.

2) Model generality and applicability: Our objective was to develop a reconstruction model that could be applied wherever historical temperature and salinity profiles exist. A model dependent on DO would be inapplicable to the vast number of historical hydrographic casts that measured only hydrographic parameters. Technically, our chosen Gaussian Process Regression framework cannot natively handle missing values (NaN) in predictors, and imputing DO for missing cases would introduce significant uncertainty.

3) Data quality and consistency: Ensuring a uniformly high standard of quality control for historical DO measurements—which are affected by sensor drift, calibration differences, and methodological evolution—is a complex task. Our ongoing work is focused on creating a consistent, quality-controlled DO product, but it was not available for this study.

In summary, while we agree that incorporating oxygen could theoretically improve the reconstruction, the practical constraints of data availability, model design, and current data quality led us to prioritize a robust, widely applicable model based on hydrographic data.

- Table 1: The salinity data count increases after quality control. Typo?

[Response] We thank the reviewer for spotting this discrepancy. The text in the original Table 1 contained a writing error regarding the salinity data count after quality control. This has been corrected in the revised manuscript.

- Figure 3a: Hard to visualize the low station counts in the open ocean. Consider plotting the colorbar in log scale?

[Response] Accepted. The color scale in Figure 3a has been changed to a logarithmic scale for better visualization.

- Line 373: The model performance for NO₂ is notably lower ($R^2 = 0.32-0.72$) compared to other nutrients. Given that NO₂ is biologically dynamic, the utility of a

T/S-based reconstruction is questionable. Consider removing NO₂⁻ from the primary dataset or flagging it with a high-uncertainty warning?

[Response] We appreciate the reviewer's critical observation on the relatively lower model performance for NO₂⁻. Upon re-examination, we discovered an error in the axis assignment of the initial 1:1 regression plot for all parameters, although it has no influence on the RMSE and R² (Fig. R13, and see updated Figs 7 and S1–S3), the performance metrics (mainly the slope) for NO₂⁻ improved, though they remain lower than those for other nutrients—consistent with its high biological reactivity.

The lowest reported R² of 0.32 corresponded to the GPR model. When using Random Forest (RF), the lowest R² is 0.56. While we acknowledge that the prediction uncertainty for NO₂⁻ is higher, we believe the reconstructed data can still provide value at climatological or spatially averaged scales, where errors are reduced through averaging. Therefore, we have opted to retain NO₂⁻ in the dataset. In response to the reviewer's concern, we have now added a high-uncertainty warning for NO₂⁻ in the revised manuscript (e.g., Lines 379–382) and in the Zenodo description, advising users to exercise caution in applications requiring high precision.

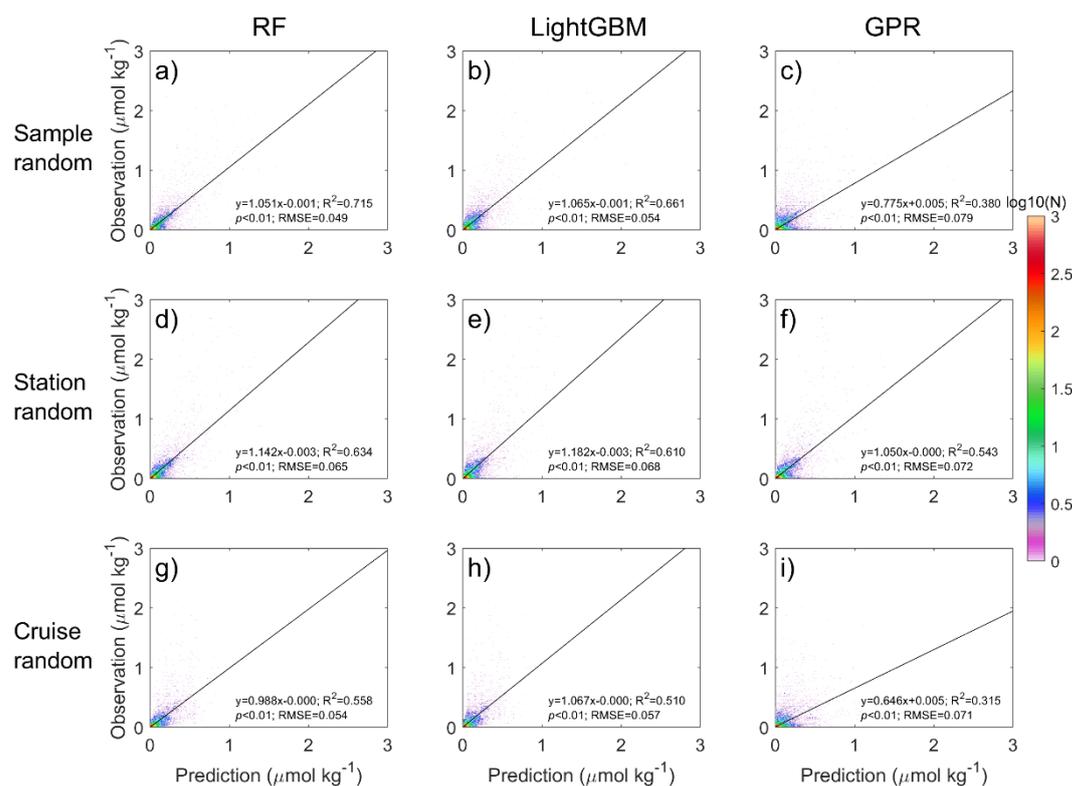


Figure R13. Validating the reconstructed NO₂⁻ concentrations using leave-one-out cross-validation with different data selection strategies and machine learning methods.

- Line 417: The manuscript notes that "most data points are located above 2,000 m." How should we interpret the deep data then? Do they have larger RMSE? If so, to what extent can the reconstructed deep nutrient fields be considered reliable for full-depth modeling applications?

[Response] The statement that "most data points are located above 2,000 m" reflects the depth distribution of the underlying hydrographic observations to which our nutrient reconstructions are tied. Specifically, a large portion of the source data—particularly from Argo floats—is concentrated in the upper ocean, which naturally limits the vertical coverage of the final product.

Importantly, this observational bias does not imply that reconstruction errors increase with depth. In fact, as requested by the reviewer (and also noted by Reviewer 2), we have added a new analysis of RMSE as a function of depth in the revised manuscript (see new Fig. S4-S7 and L386-391 in the revised manuscript). The results show that reconstruction errors in deep layers are generally smaller than those in the upper ocean, owing to lower natural variability and stronger physical–nutrient correlations at depth. Therefore, the reconstructed nutrient fields can be considered reliable for full-depth modeling applications within the quantified uncertainty bounds.

Reviewer 2#

General comments

The authors' goal is to reconcile the current imbalance in data availability between hydrographic tracers (T and S), which are widely measured, and nutrients, which are much sparser. They (1) compile nutrient data from available datasets, (2) develop a pipeline for quality control and filtering, and (3) use the "cleaned" nutrient compilation with an ensemble of machine-learning methods to build a predictive model that infers nutrient concentrations as a function of T, S, and time. I think this is valid, well-written work and it is worthy of publication.

[Response] We sincerely thank the reviewer for their positive assessment of our work and for highlighting the importance of clearly describing the final data product.

One criticism is that the current version of the manuscript does not clearly state the nature of the final product generated using the trained model. Initially, I thought the authors were aiming to produce a fully time-resolved climatology (e.g., monthly means for each year spanning 1895–2024 or 1973–2022). Then, I interpreted the product as being analogous to WOA23 (i.e., monthly climatological means that are not explicitly resolved through time). I then downloaded the data from Zenodo using the link included in the abstract. However, the Zenodo download page is also somewhat unclear, with acronyms that are not explained (e.g., ABP, GLD, PFL, UOR), each associated with different files. Based on a preliminary inspection of some of these files, it appears that the authors provide time-resolved predictions (potentially at ~2-day resolution from 2004 to 2024). This makes the dataset more interesting, but the authors should do a better job—both in the manuscript and on Zenodo—of clearly describing the characteristics of the product they are publishing.

[Response] The reviewer correctly notes the initial ambiguity regarding the nature of the reconstructed dataset. We apologize for any confusion caused. To clarify: the final product is not a gridded product. Instead, it provides hydrography-attached nutrient concentrations—that is, nutrient values reconstructed specifically at the locations and

depths of the original hydrographic observations (from WOD and other sources) where direct nutrient measurements might be unavailable or did not pass quality control.

In response to this comment, we have revised the manuscript (e.g., L491–503 in the revised manuscript) to explicitly state the characteristics of the dataset and define all acronyms used for platform types (e.g., APB, CTD, GLD, PFL, UOR). Furthermore, we have updated the dataset description on the Zenodo repository page (<https://zenodo.org/records/17451417>) to include a clear explanation of the data structure, temporal resolution, and file organization. This ensures users can accurately interpret and utilize the resource.

In addition, I think the comparison with WOA23 should be more quantitative. The main manuscript and supplementary information include many maps from WOA23 and from the authors' reconstructed dataset, but the comparison would be more effective if it also included maps of the differences between the two products. The text describing the advantages of the new dataset relative to WOA23 is also somewhat generic (e.g., “seasonal patterns are similar but concentration is lower”; how much lower? see lines 441–449).

[Response] We appreciate the reviewer's suggestion for a more quantitative comparison with WOA23. In response, we have added difference maps showing spatial deviations between reconstructed nutrient fields and WOA23 climatologies, and explicitly state their differences (Figs. R14–R16; see L544–545 and Figs. S45–S47 in the revised manuscript).

However, we want to clarify a fundamental distinction in dataset design: our product is not a gridded dataset. Instead, it provides hydrographic-attached nutrient reconstructions at the original observation locations/depths. We have stated this at L496–503 in the revised manuscript.

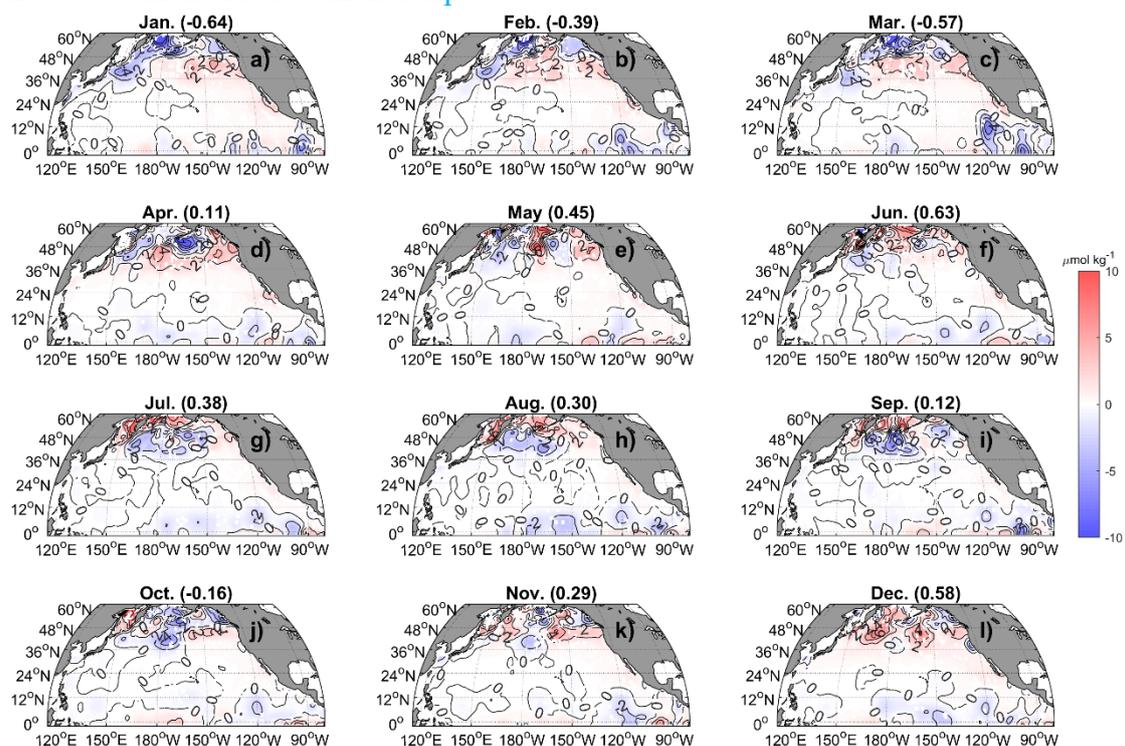


Figure R14. Difference between the predicted (Random Forest, RF) and World Ocean Atlas (WOA) climatologies of NO_x^- ($\text{NO}_3^- + \text{NO}_2^-$) at 5 m depth (RF minus WOA).

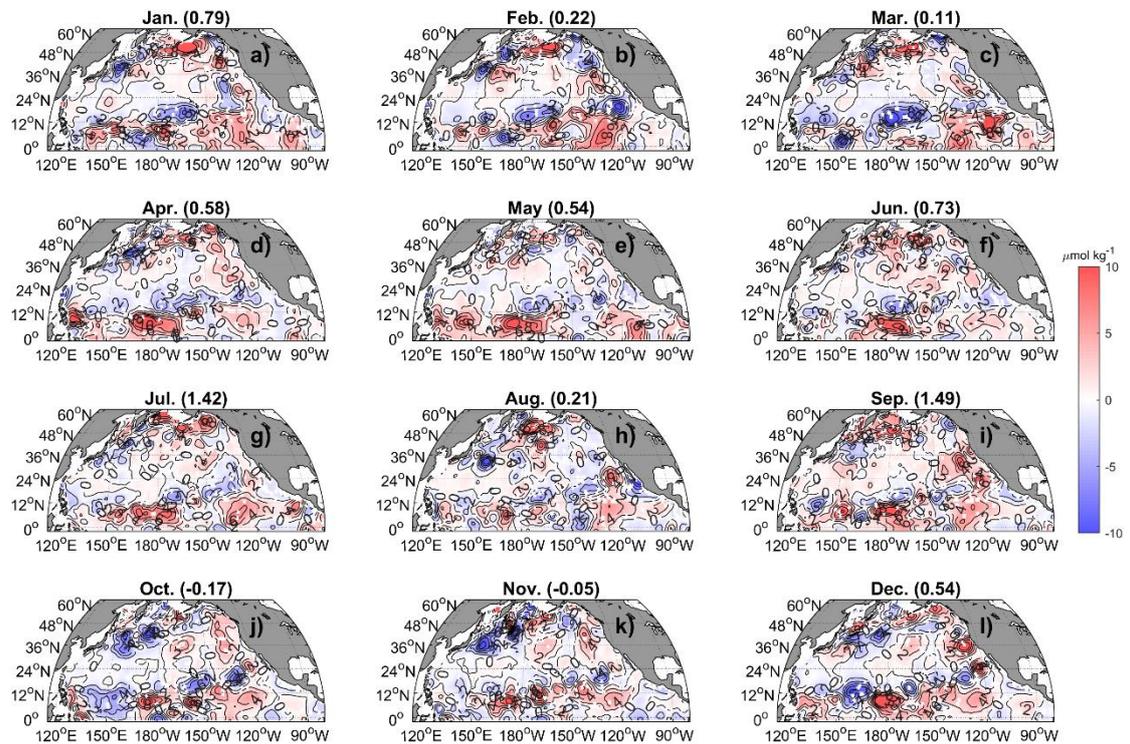


Figure R15. Same as Figure R14, but for 200 m.

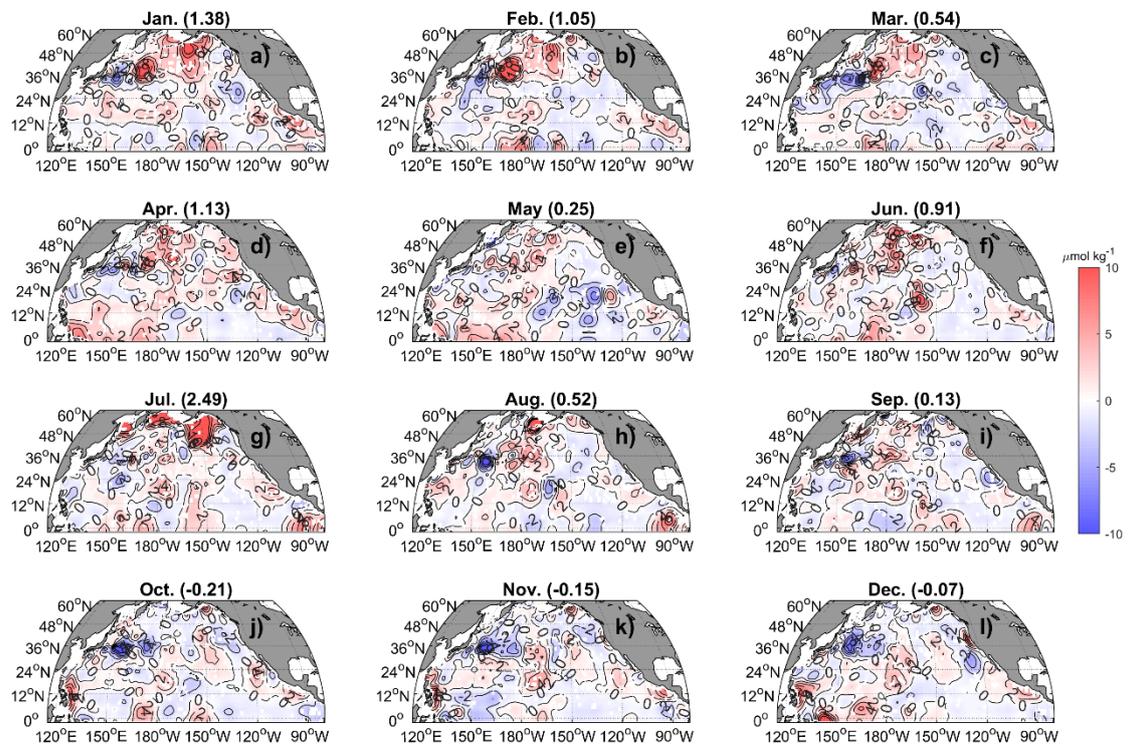


Figure R16. Same as Figure R14, but for 500 m.

Specific technical comments

The reported RMSE values are low when compared with the concentration ranges of these nutrients, but the authors should provide additional detail on how the error varies with depth. For instance, an error of $\sim 1.5 \mu\text{mol kg}^{-1}$ for NO_3^- may be acceptable in deep waters, but it would be a very large error across much of the surface ocean. The same issue applies across different biogeochemical provinces (e.g., nutrient-rich upwelling regions versus nutrient-poor subtropical gyres). The applicability of the dataset will depend strongly on the vertical and lateral structure of the reconstruction error.

[Response] We thank the reviewer for this insightful suggestion. We agree that understanding the spatial structure of reconstruction errors is critical for assessing dataset applicability. In response:

1) Depth-resolved error analysis: we have added new depth profiles of RMSE for all nutrients in the revised manuscript (L386–391, Figs. S4–S7, Fig. R17), demonstrating that errors in deep waters are consistently lower.

2) Spatial error mapping: error distribution maps (L391-399 and Figs. S8-S11 in the revised manuscript; Fig. R18) highlight how errors change with biogeographic regimes.

3) We have paid particular attention to the oligotrophic regimes and the nutrient reconstruction errors (Table R1, Table 4 in the revised manuscript). This confirms that absolute errors decrease in oligotrophic regimes.

These additions directly address the reviewer’s concern by demonstrating that the dataset remains viable across diverse provinces when used with appropriate error context.

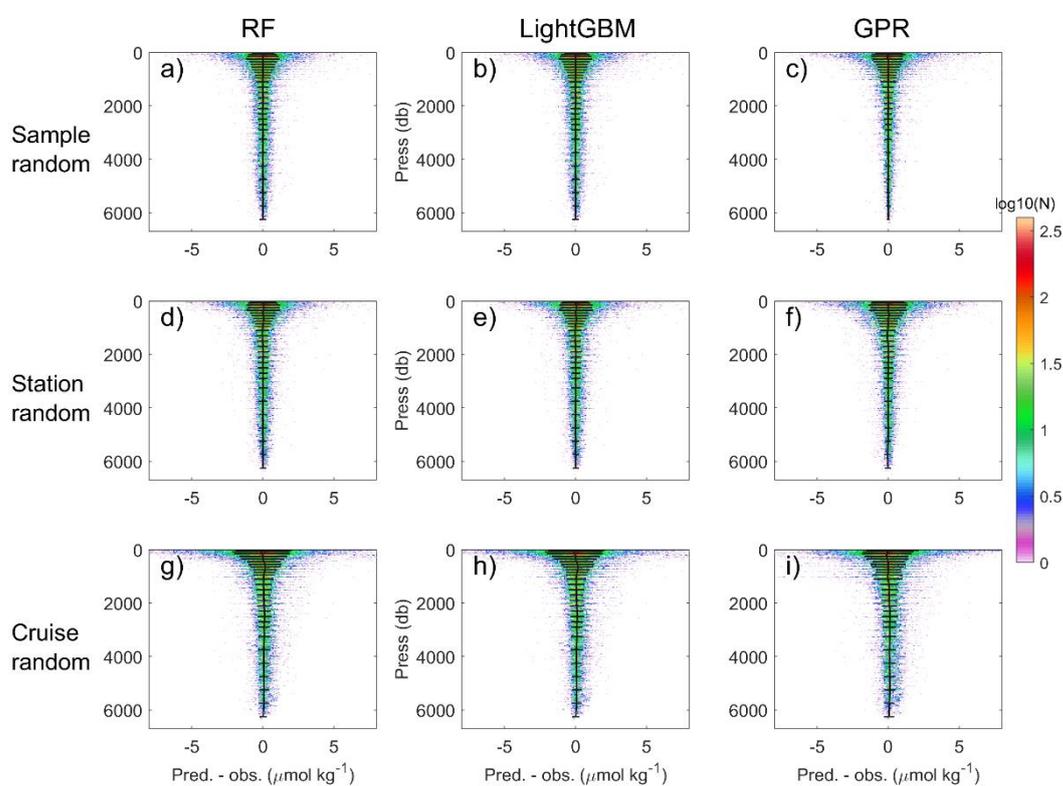


Figure R17. Vertical profiles of NO_x^- ($\text{NO}_3^- + \text{NO}_2^-$) reconstruction errors with different data selection strategies and machine learning methods. Colors indicate data

density; the solid black vertical line represents the mean absolute error ($\mu\text{mol kg}^{-1}$); and the horizontal error bars denote the ± 1 standard deviation range. Plots in row 1 correspond to the sample random strategy (a-c); plots in row 2 correspond to the station random strategy (d-f); and plots in row 3 correspond to the cruise random strategy (g-i). Plots in column 1 correspond to the Random Forest (RF; a, d, g); plots in column 2 correspond to LightGBM (b, e, h); and plots in column 3 correspond to Gaussian Process Regression (GPR; c, f, i).

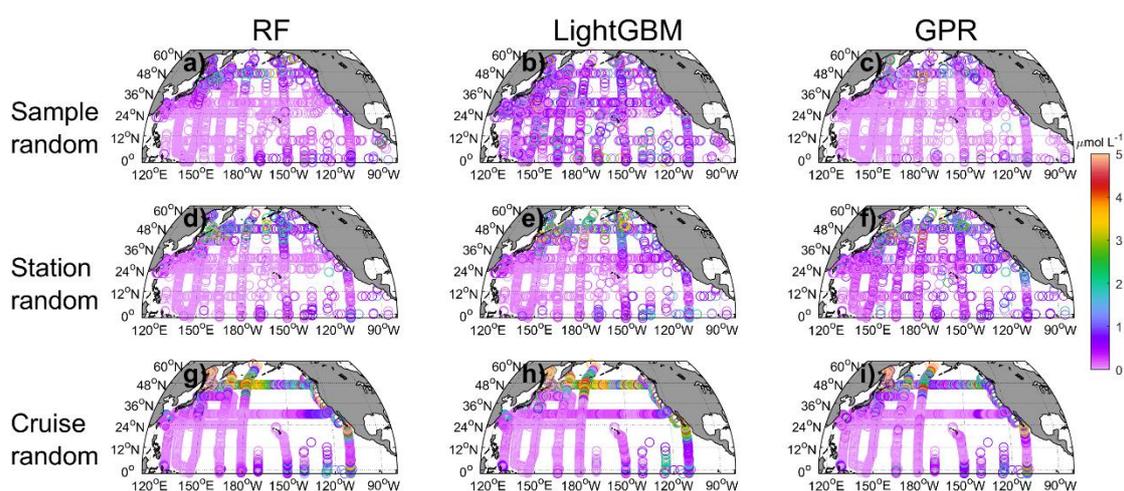


Figure R18. Spatial distribution of absolute errors (difference between predicted and observed values) for surface nutrient concentrations during a random selection of data.

Table R1 The Root Mean Squared Errors of nutrient reconstruction from different error evaluation strategies in surface oligotrophic regimes (unit: $\mu\text{mol kg}^{-1}$). The oligotrophic regimes are defined as the region where NO_x^- , NO_2^- , DIP, and $\text{Si}(\text{OH})_4$ concentrations < 0.2 , < 0.2 , < 0.2 , and $< 5.0 \mu\text{mol kg}^{-1}$, respectively.

Data selection strategy	NO_x^-			NO_2^-			DIP			$\text{Si}(\text{OH})_4$		
	RF	Light GBM	GPR	RF	Light GBM	GPR	RF	Light GBM	GPR	RF	Light GBM	GPR
Sample random	0.290	0.567	0.444	0.018	0.035	0.048	0.028	0.042	0.039	1.19	0.90	1.30
Station random	0.303	0.457	0.474	0.030	0.030	0.043	0.036	0.045	0.043	1.24	1.51	1.51
Cruise random	0.378	0.457	0.574	0.030	0.029	0.056	0.075	0.077	0.084	1.85	1.88	1.75

A potential concern is that nutrient data quality and sampling density have changed substantially over time. Also, summer observations are three times greater than in winter. I think the authors should comment on whether this could introduce time-dependent biases in a fully time-resolved reconstruction. Are uncertainty and model skills explicitly evaluated by era, depth, and region?

[Response] The reviewer raises a valid concern regarding potential biases arising from uneven temporal and seasonal sampling distributions. We address these points as follows:

1) Seasonal bias (summer vs. winter): We acknowledge that summer observations significantly outnumber winter data. In response, we have conducted a seasonal error analysis. The results indicate that for NO_x^- under cruise-random validation, the RMSE is indeed higher in winter (Fig. R19). However, this seasonal discrepancy is not consistently evident for other nutrients or validation strategies. We have added a discussion of this seasonal performance variation in the revised manuscript (L403-407 and Fig. S12 in the revised manuscript).

2) Spatial and depth-dependent model skill: The concern regarding performance variation by depth and region aligns with the previous comment on error structure. As noted in our response to that comment, we have now explicitly evaluated and presented the reconstruction errors as a function of depth and across oligotrophic regimes in the revised manuscript (see L386-403 in the revised manuscript).

3) Time-dependent bias (by era): The concern about model skill over different time periods is directly related to the major comment by Reviewer #1 on "Temporal extrapolation robustness..." Please refer to our detailed response to that comment, where we validated the model's performance using historical data from pre-1973 eras (1929, 1947, etc.) and discussed the associated uncertainties. Please refer to L408-410 and L451-488 in the revised manuscript.

In summary, we have incorporated new analyses to explicitly address how uncertainties vary with season, region, and depth. The issue of performance across historical eras is addressed in our response to Reviewer #1, Major Comment #1.

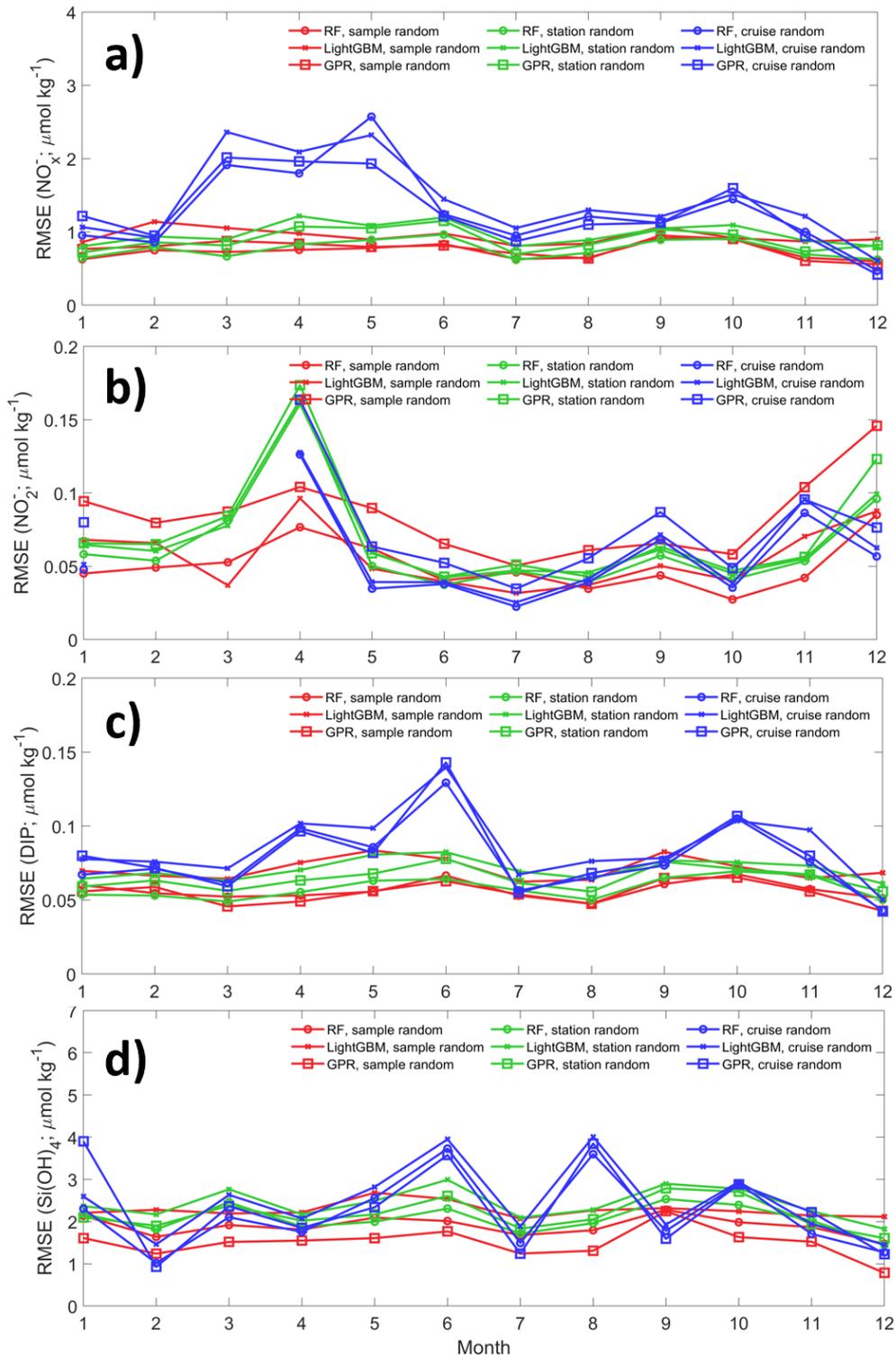


Figure R19. Monthly variations in reconstruction errors (root mean square error) for different parameters under various data selection strategies and machine learning methods.

My understanding is that the authors train the models on nutrient data from 1973 to 2022 and then apply the trained model to salinity and temperature data from 1895–2024. I guess that the assumption is that the relationship between predictors (T and S, physical

tracers) and targets (nutrients, biological) remains the same between 1895-1973 and 1973-present. If so, the authors should explain why they think this is a strong assumption.

[Response] The reviewer identifies a key assumption underlying our methodological approach: that the fundamental relationships between the hydrographic predictors and nutrient concentrations remain statistically stationary from the historical period (1895–1972) through the modern training era (1973–2022). We appreciate the opportunity to clarify the rationale behind this assumption, which we have now explicitly stated in the revised manuscript (L451-488 in the revised manuscript). Our position is supported by several lines of evidence:

1) Lack of strong long-term trends: As shown in the multi-decadal time series from Station ALOHA (Fig. 12 in the manuscript), nutrient concentrations in the open ocean do not exhibit pronounced long-term trends over the 30-year observational period, suggesting relative stability on centennial timescales, at least in the oligotrophic gyres.

2) Long ocean nutrient residence time: The biogeochemical cycling timescale for nitrogen is even longer (~2,000 years). Given these long timescales, significant alteration of the large-scale, depth-dependent nutrient-versus-hydrography relationships over a mere 100-year window is physically small, especially for waters beneath the main thermocline.

3) Consistency between historical and modern oceanographic profiles: Perhaps the most compelling indirect evidence comes from comparing mean oceanic nutrient profiles derived from pre-1970s data with those from the modern era. As illustrated in Fig. R19, these profiles are strikingly similar despite major differences in sampling methodologies, analytical precision, and spatiotemporal coverage. This result implies that the North Pacific Ocean has not undergone a fundamental shift within the averaging errors. We also acknowledge while the North Pacific may experience T/S-nutrient relationships variability, it might be masked by the reconstruction error, and the use of hydrographic properties as predictors for nutrients is justified for historical reconstructions.

However, the collective evidence provides reasonable confidence in applying the model for historical reconstruction, with the understanding that uncertainties are likely higher for the earliest decades, as discussed in our response regarding temporal extrapolation (comment by Reviewer 1#: “1. Temporal extrapolation robustness: The use of three validation strategies...”).

Line 99: what’s striking of the Pacific (for example, relative to the Atlantic) is longitude not the latitude range

[Response] We thank the reviewer for this additional point. We have revised the text to emphasize its exceptional longitudinal span, which distinguishes it from other ocean basins (L101 in the revised manuscript).

Line 99-118 One aspect of the Pacific that is unique and maybe should be highlighted here is that it hosts, unlike the Atlantic, all major N fluxes (including water column denitrification and N₂ fixation)

[Response] We agree with the reviewer’s suggestion to highlight the Pacific’s unique role in hosting all major nitrogen fluxes. The revised text now notes that the Pacific basin encompasses key processes, including water-column denitrification zones (e.g., Eastern Tropical Pacific oxygen minimum zones) and major N_2 fixation hotspots (e.g., North Pacific Subtropical Gyre). Please refer to L106-112 in the revised manuscript.

319-337 I think they have a good error-estimation strategy; however, I don’t think that the validation splits explicitly test for time-dependent changes in data quality (e.g., through era-based validation) or quantify how reconstruction error varies systematically with depth and biogeochemical province, where model skill and sampling density can differ substantially. Time validation is performed in Aloha but it’s for a “short” time range relative to the time range of the hydrographic properties archive (1988-2021 vs 1895-2024) and for one specific biogeochemical province.

[Response] The reviewer rightly notes that our initial validation did not fully address era-dependent errors or province-specific skill variations. As detailed in our responses to Reviewer #1 (Major Comment 1) and Reviewer #2 (Comment on temporal extrapolation assumptions), we have now: 1) added era-specific validation using pre-1970s data (1929–1966) to quantify errors in historical periods; 2) explicitly evaluated depth- and region-dependent errors.

Figure 9 I think this figure would benefit from including a plot of the residuals (predicted minus observed)

[Response] Accepted. We have now included residual plots (predicted minus observed values) for all nutrients (Fig. R20 as an example; Figs. S45–S47 in the revised Supporting Information).

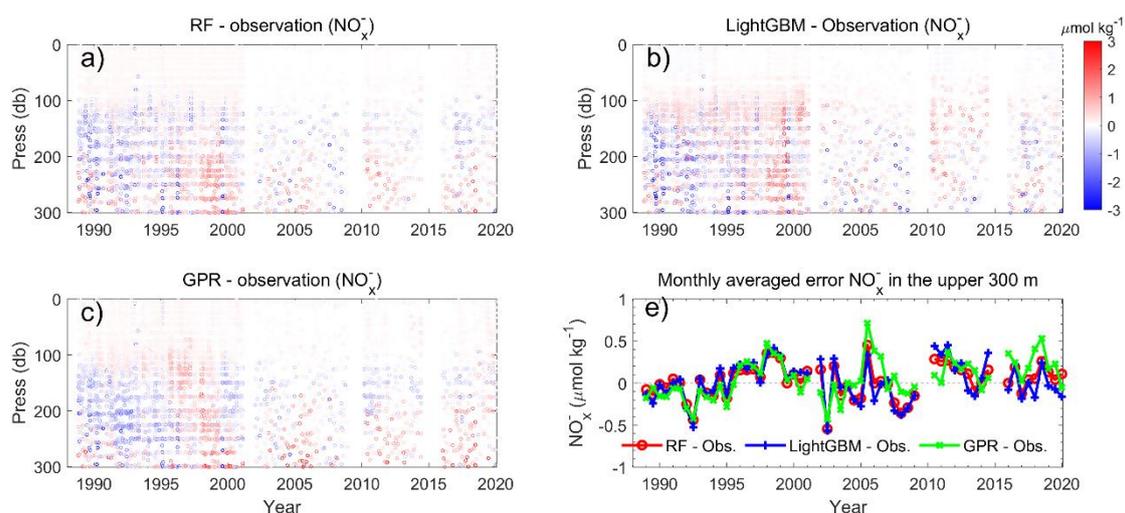


Figure R20. Temporal variations of reconstruction error profiles of NO_x^- ($NO_3^- + NO_2^-$) (prediction minus observations) in the upper 300 m at Station ALOHA from 1988 to 2021 with different machine learning models: Random Forest (RF; a), LightGBM (b), and Gaussian Process Regression (GPR; c). Note that the colorbar plotted in (b) is also

applicable to (a) and (c). (e) Temporal variations of monthly averaged NO_x^- prediction errors (prediction minus observations) in the upper 300 m from RF, LightGBM, and GPR.