

Supporting Information for:

**A historical nutrient dataset (1895–2024) for the North Pacific:
reconstructed from machine learning and hydrographic observations**

Chuanjun Du^{1*}, Naiwen Zheng¹, Shuh-Ji Kao¹, Minhan Dai², Zhimian Cao², Dalin
Shi², Qiancheng Li¹, Hao Wang¹, and Xiaolin Li^{2*}

¹School of Marine Sciences, Hainan University, Haikou 570228, China

²State Key Laboratory of Marine Environmental Science, College of Ocean and Earth
Sciences, Xiamen University, Xiamen 361102, China

***Corresponding Authors:** Chuanjun Du, cjdu@hainanu.edu.cn; Xiaolin Li,
xlli@xmu.edu.cn

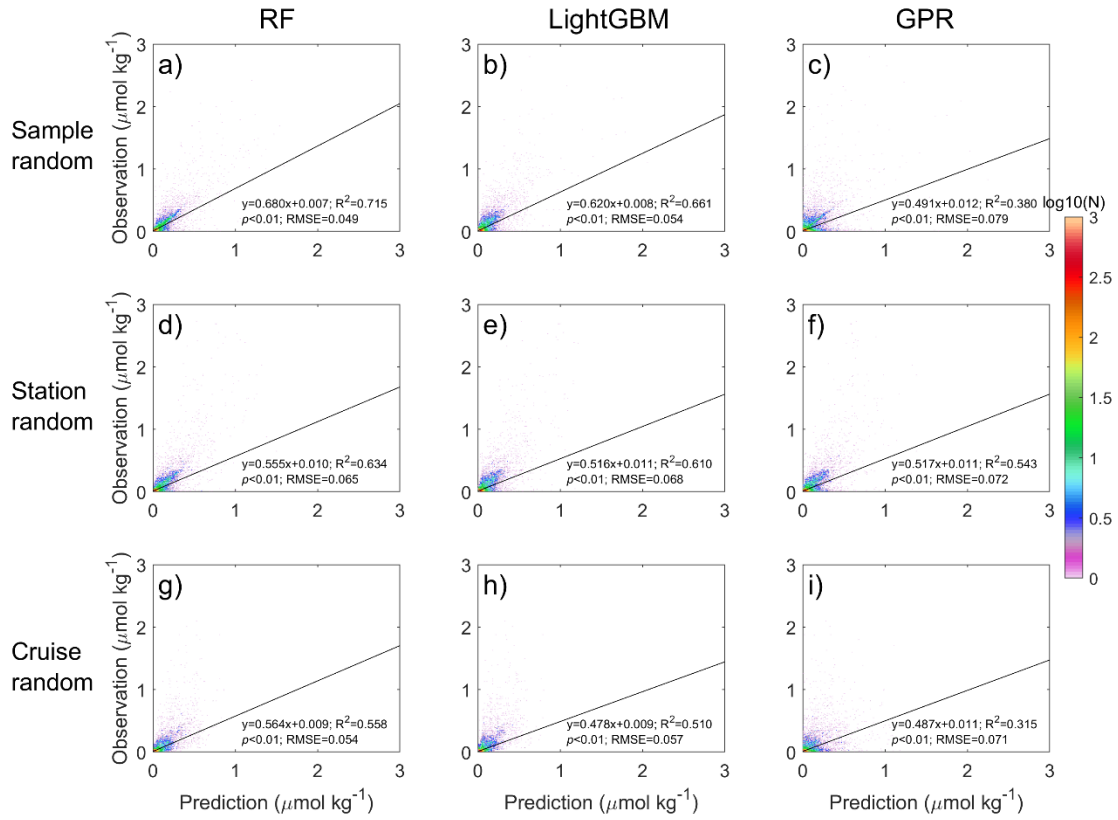


Figure S1. Validating the reconstructed NO_2^- concentrations using leave-one-out cross-validation with different data selection strategies and machine learning methods. Plots shown in row 1 correspond to the sample random strategy (a-c), row 2 correspond to the station random strategy (d-e), and row 3 correspond to the cruise random strategy (g-i). Plots shown in column 1 correspond to the Random Forest (RF; a, d, and g), column 2 correspond to the LightGBM (b, e, and h), and column 3 correspond to the Gaussian Process Regression (GPR; c, f, and i). The black lines and text show the fitted linear regressions, regression equations, coefficient of determination (R^2), p values, and Root Mean Squared Errors (RMSE). The color represents the data density (N , number of observations). Note that the logarithmic scale of N is applied.

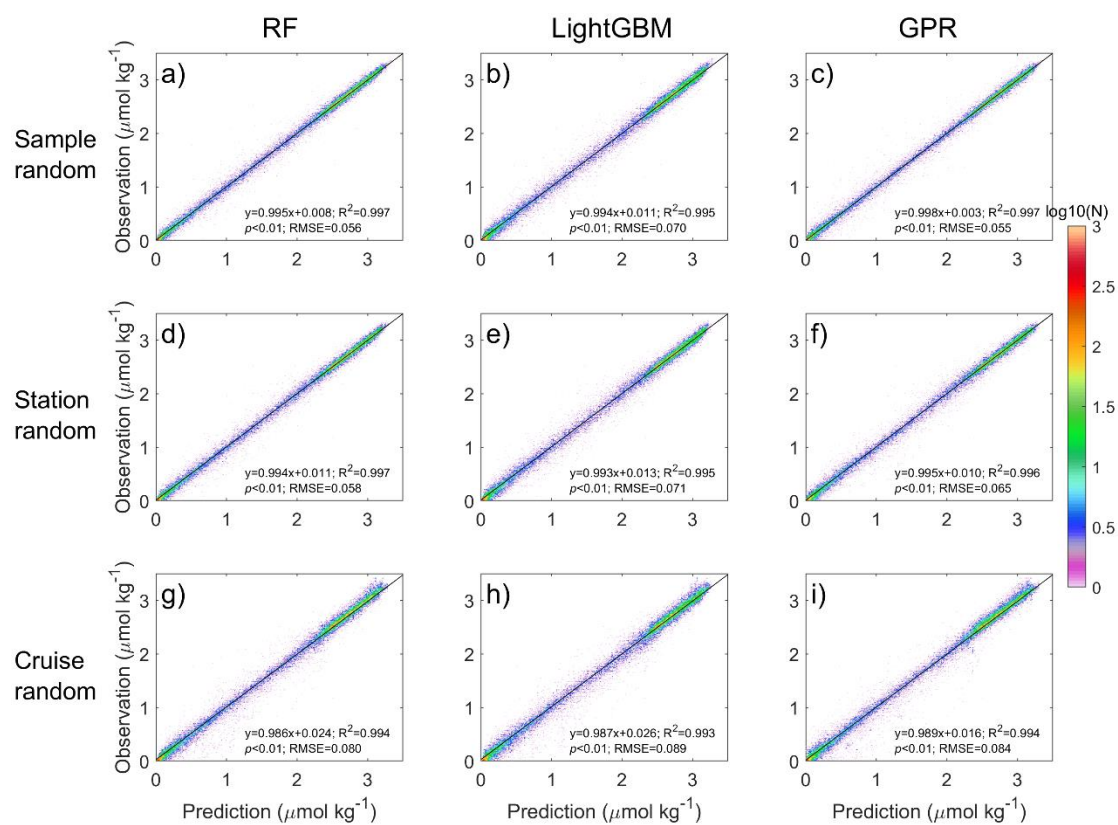


Figure S2. Similar to Fig. S1, but for DIP.

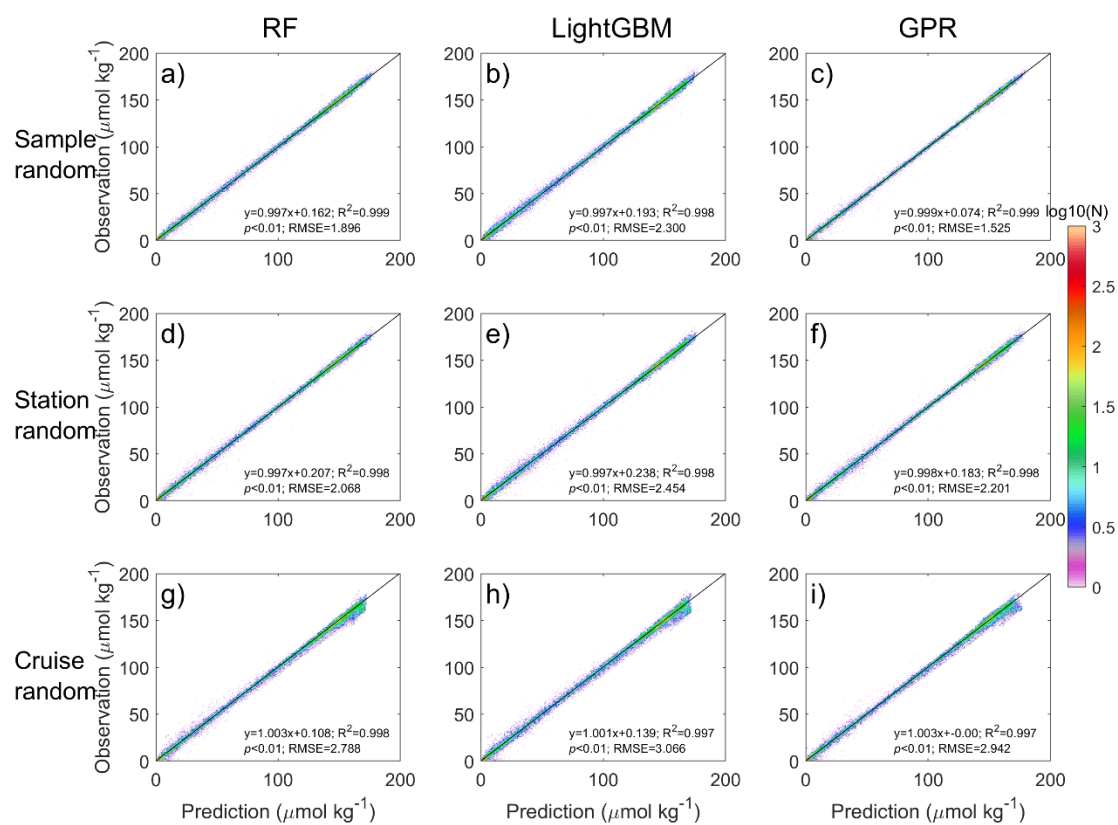


Figure S3. Similar to Fig. S1, but for Si(OH)_4 .

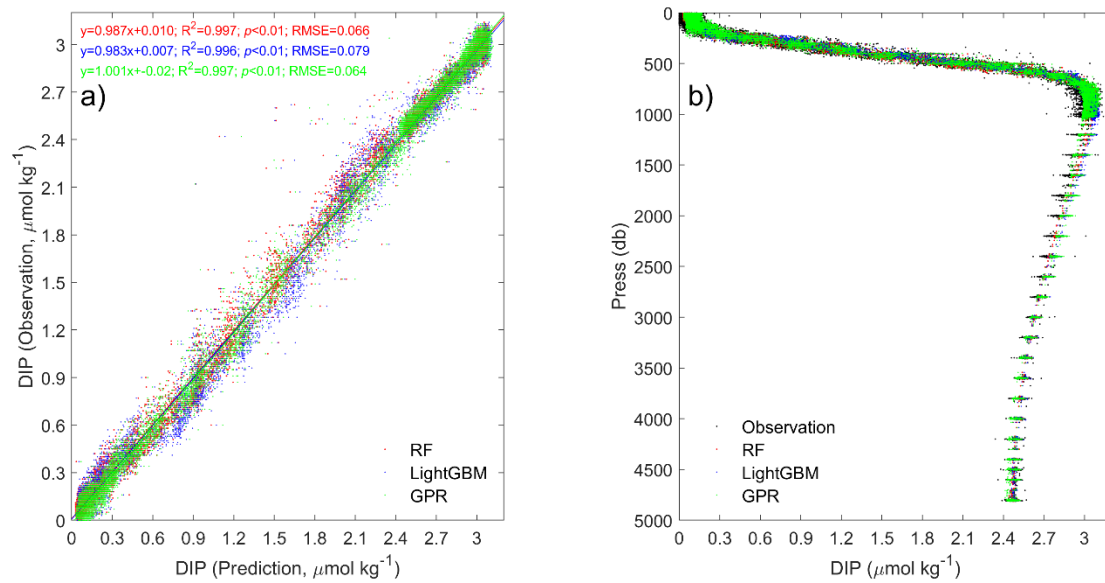


Figure S4. Validating the reconstructed DIP concentrations at Station ALOHA. a) Reconstructed DIP vs. observations: Random Forest (RF; red dots), LightGBM (blue dots), and Gaussian Process Regression (GPR; green dots); b) Profiles of observed (black dots) and reconstructed DIP from RF (red dots), LightGBM (blue dots), and GPR (green dots).

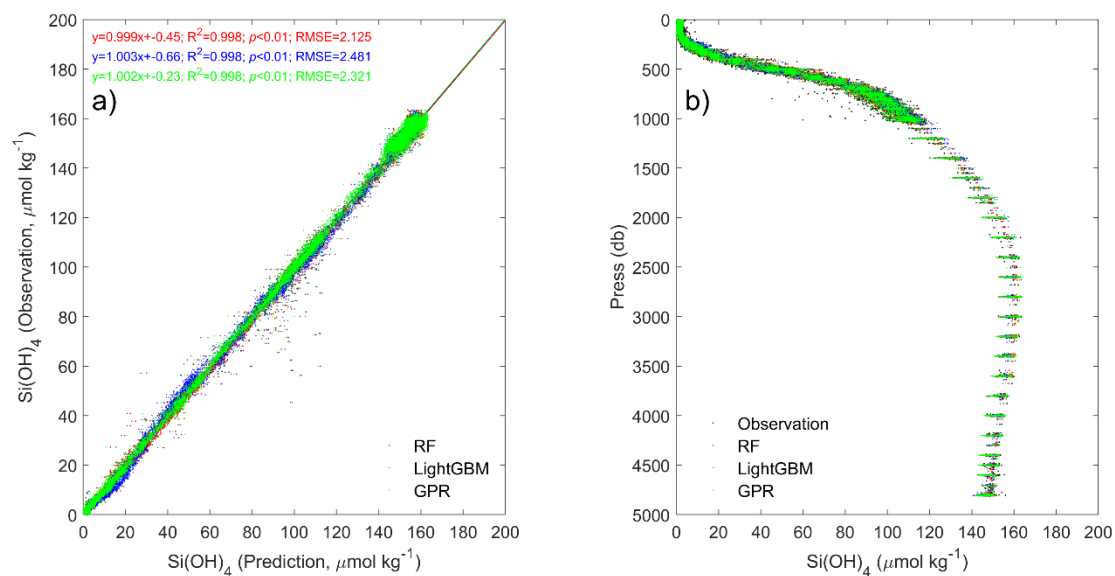


Figure S5. Similar to Fig. S4, but for Si(OH)_4 .

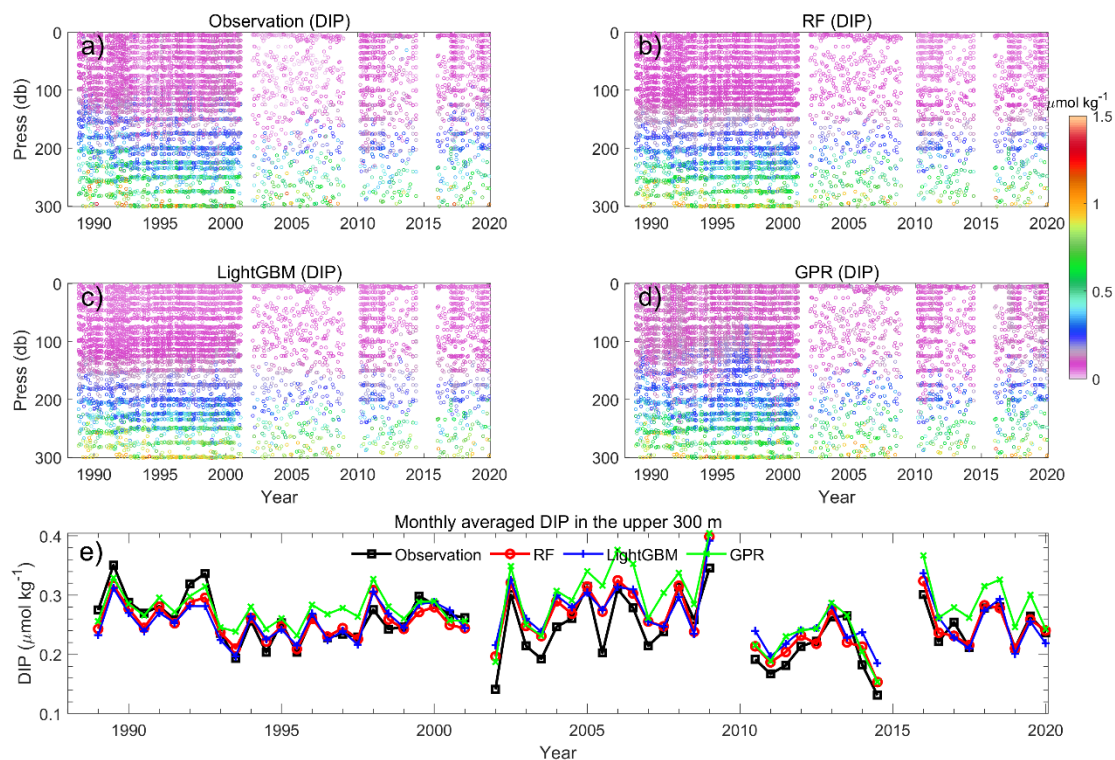


Figure S6. Temporal variations of DIP concentrations in the upper 300 m at Station ALOHA from 1988 to 2021 for observed (a) and reconstructed DIP by Random Forest (RF; b), LightGBM (c), and Gaussian Process Regression (GPR; d). (e) Time series of monthly averaged NO_x^- concentrations in the upper 300 m from observations, and reconstructions by RF, LightGBM, and GPR, respectively.

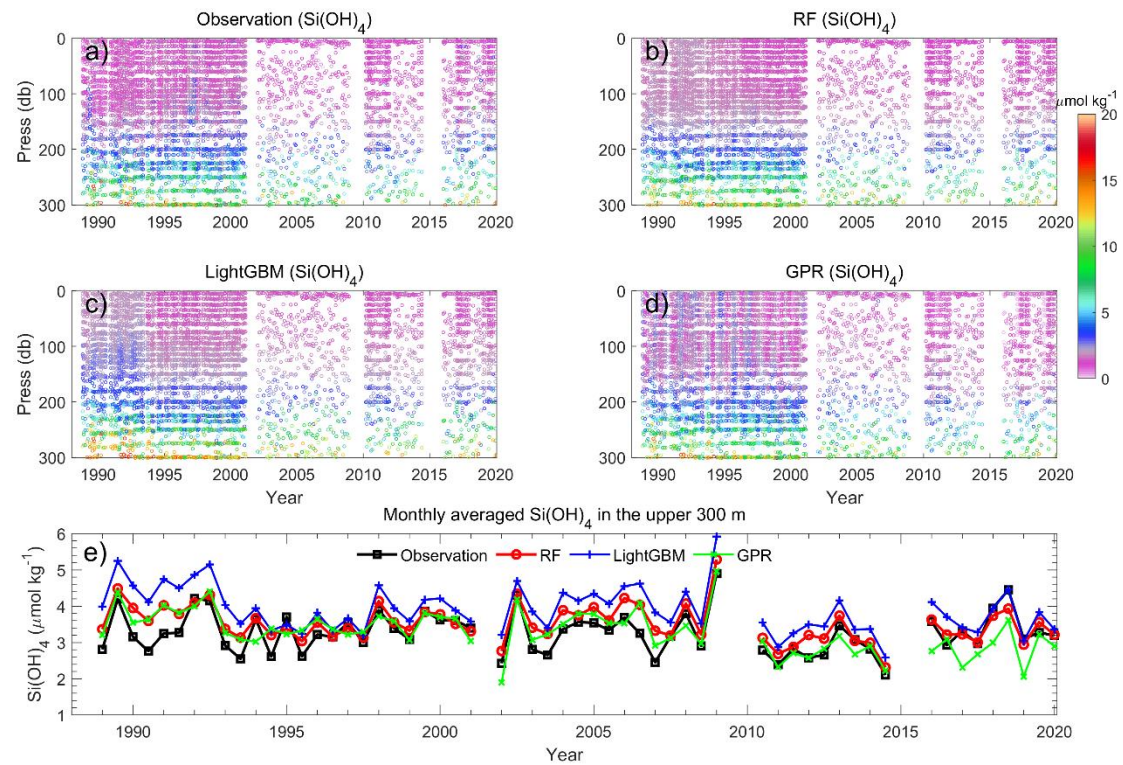


Figure S7. Similar to Fig. S6, but for Si(OH)_4 .

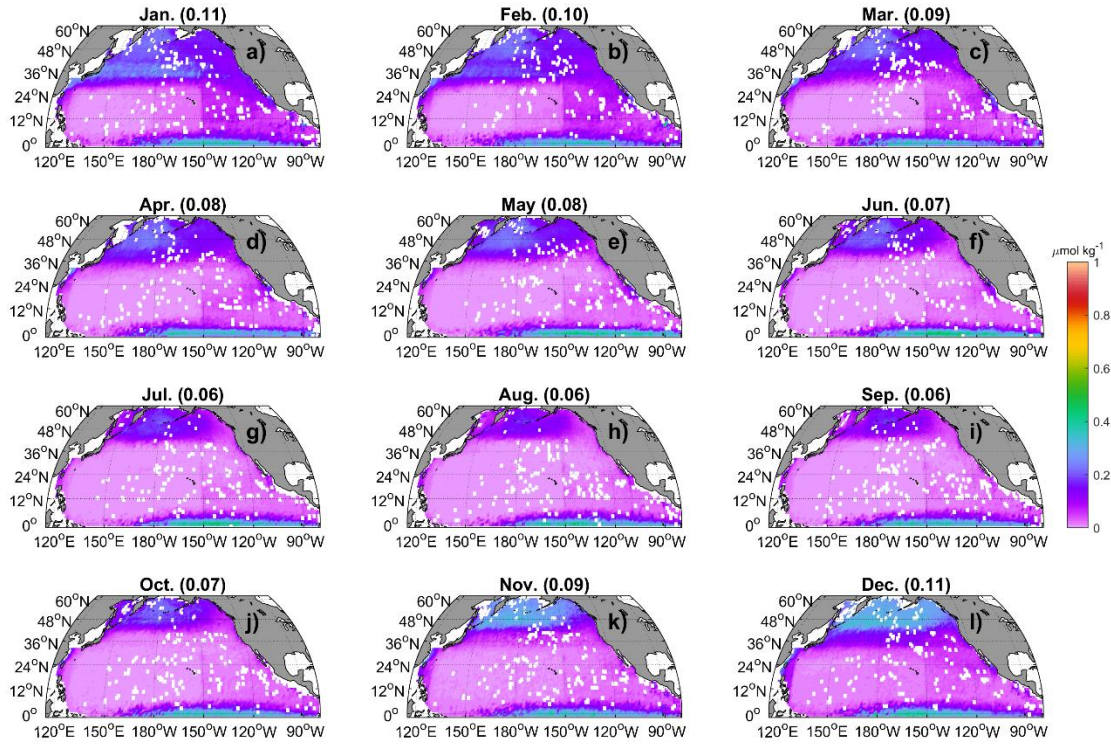


Figure S8. The monthly climatology of NO_2^- at 5 m in the North Pacific. Data are binned and averaged within $1 \times 1^\circ$ grid cell. The values in the title represent the spatial mean values.

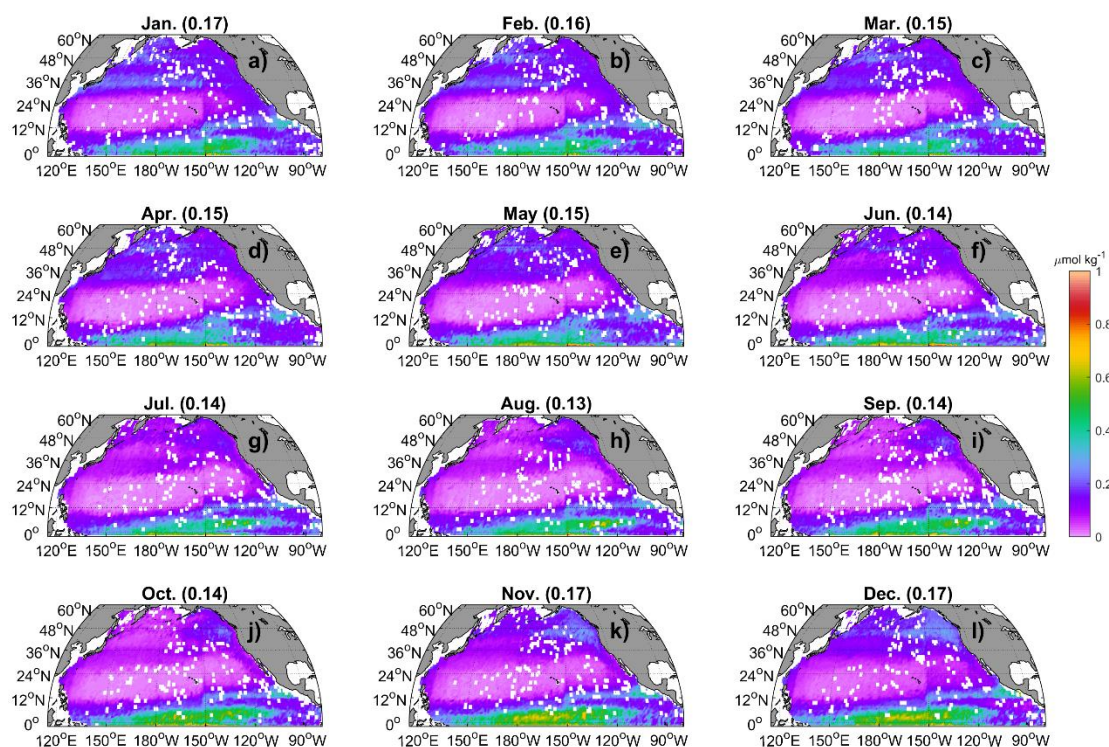


Figure S9. Similar to Fig. S8, but for a depth of 100 m.

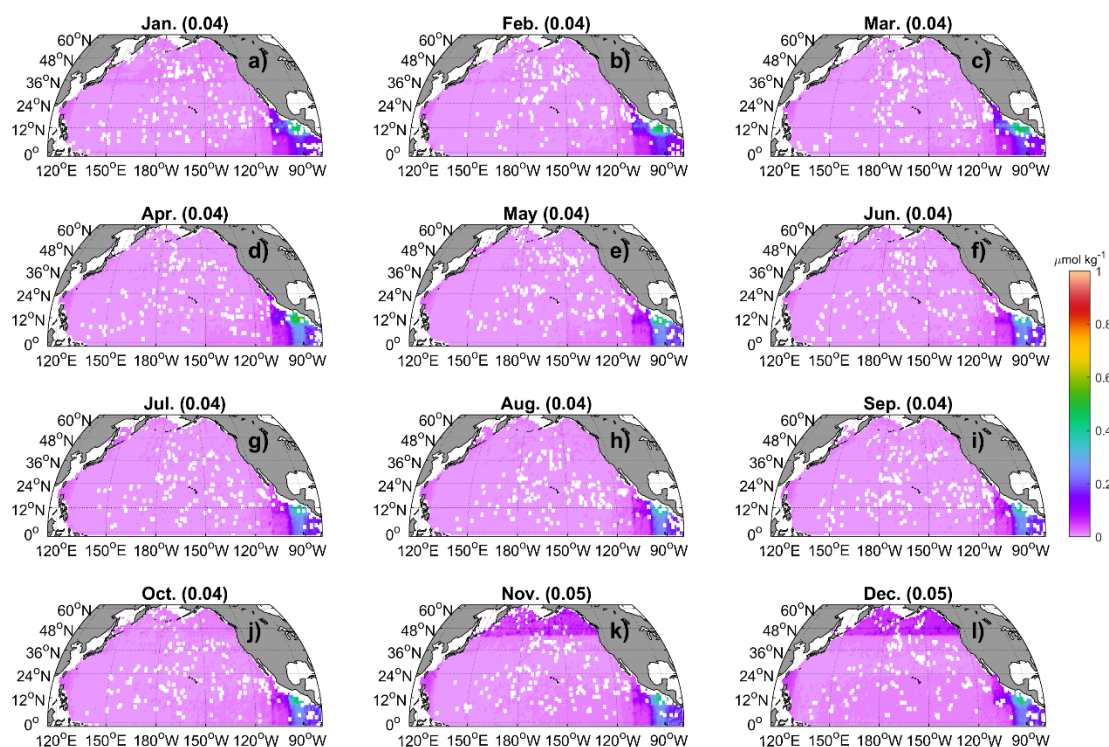


Figure S10. Similar to Fig. S8, but for a depth of 500 m.

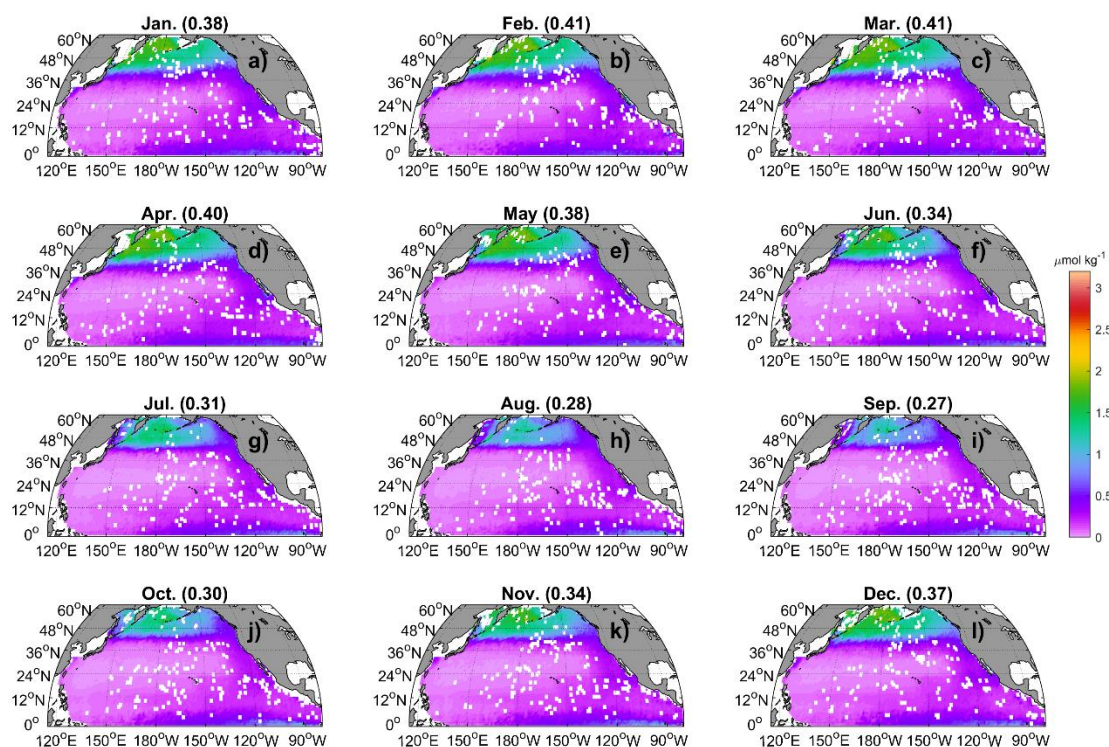


Figure S11. Similar to Fig. S8, but for DIP and at a depth of 5 m.

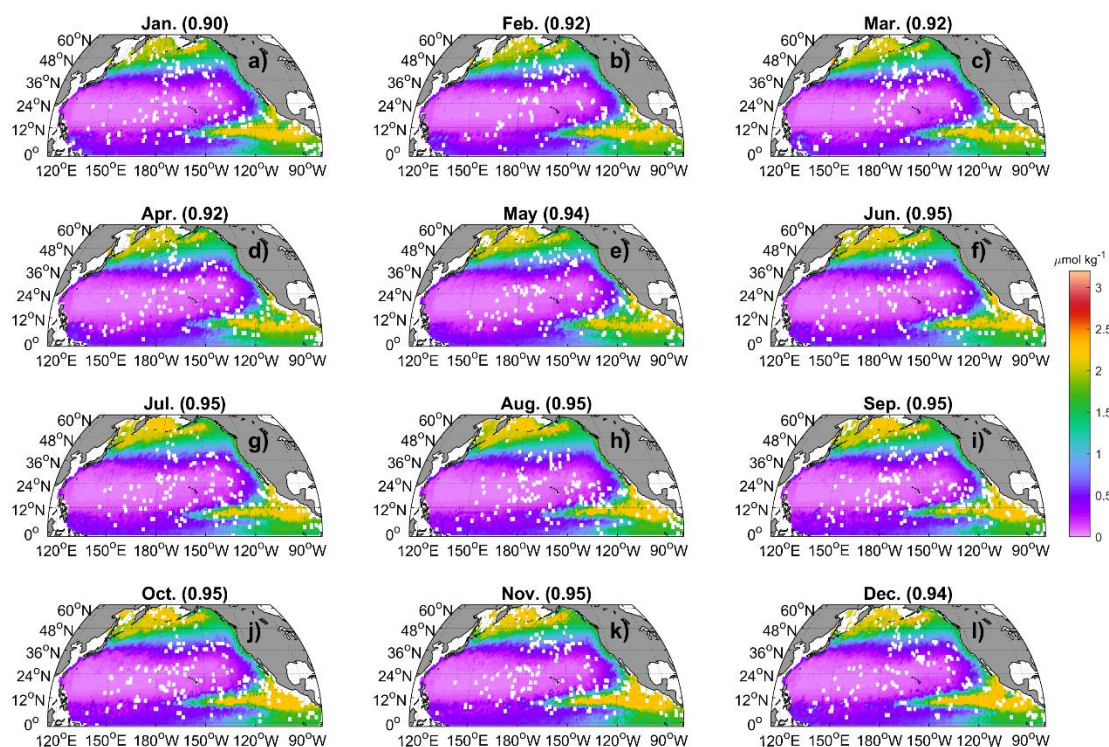


Figure S12. Similar to Fig. S8, but for DIP and at a depth of 100 m.

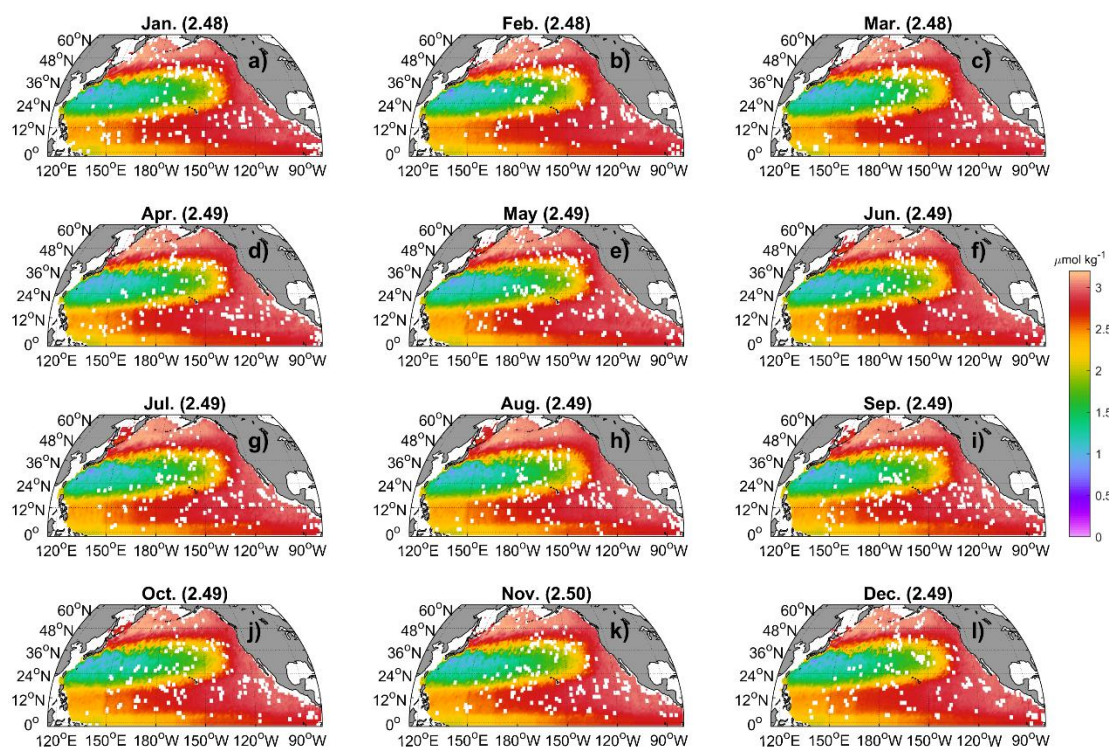


Figure S13. Similar to Fig. S8, but for DIP and at a depth of 500 m.

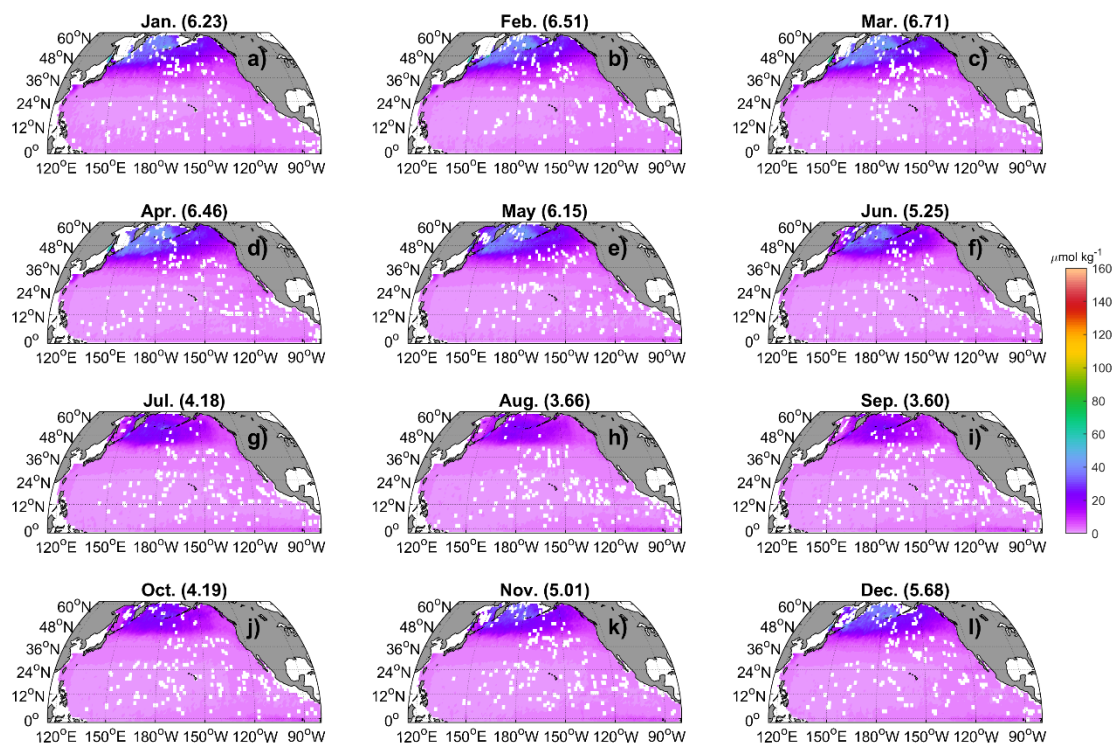


Figure S14. Similar to Fig. S8, but for Si(OH)_4 and at a depth of 5 m.

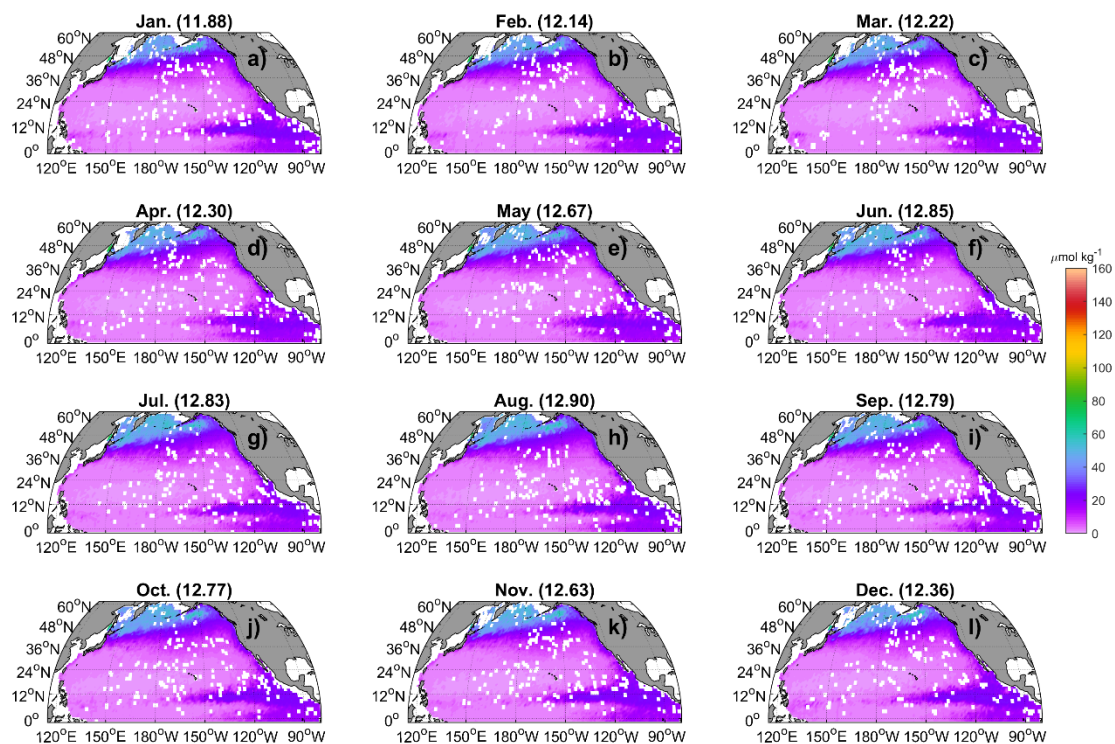


Figure S15. Similar to Fig. S8, but for Si(OH)_4 and at a depth of 100 m.

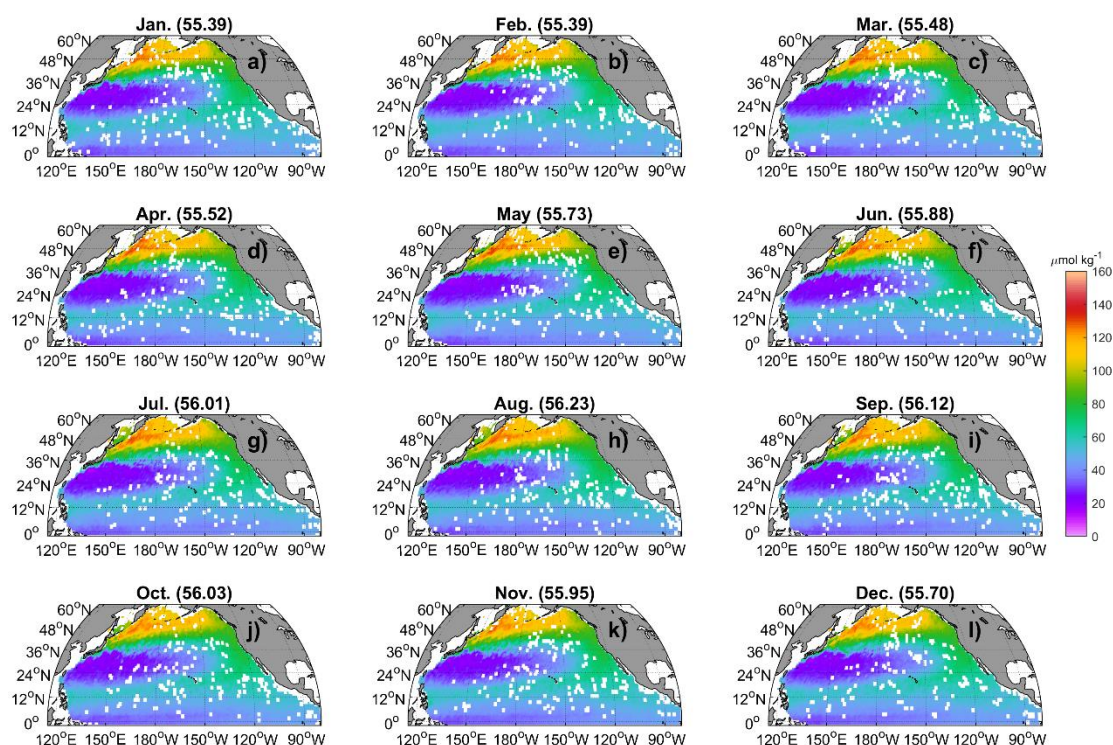


Figure S16. Similar to Fig. S8, but for Si(OH)_4 and at a depth of 500 m.

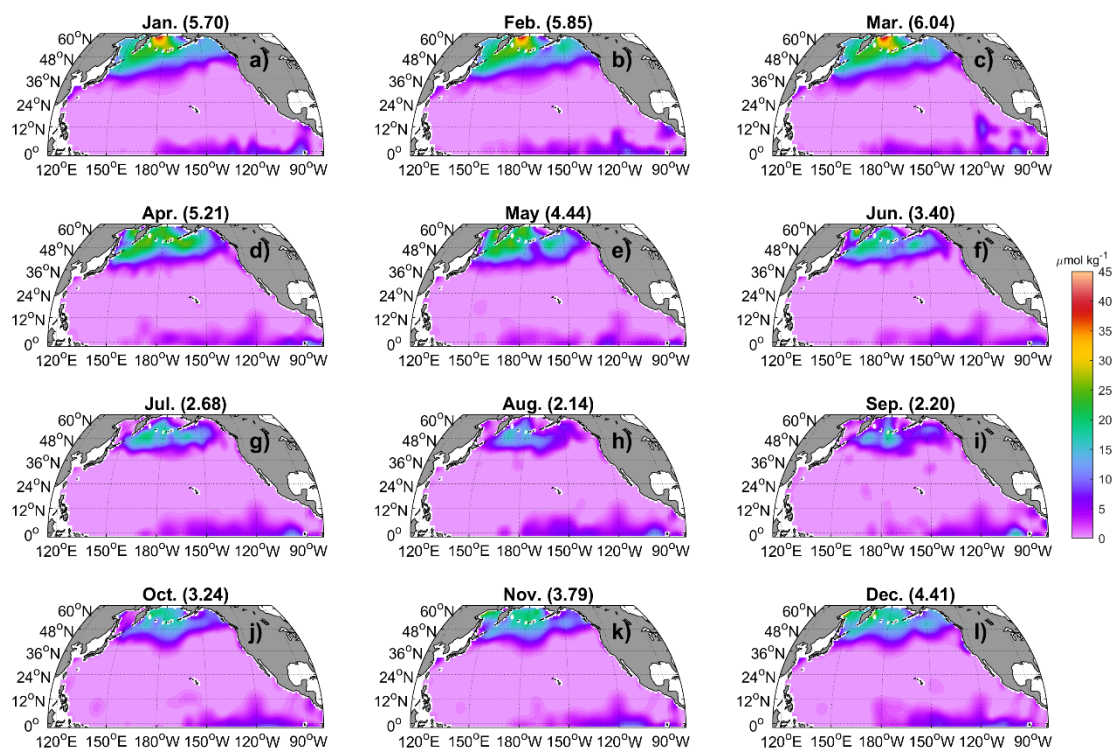


Figure S17. The WOA23 monthly climatology of NO_3^- at 5 m in the North Pacific.

The values in the title represent the spatial mean values.

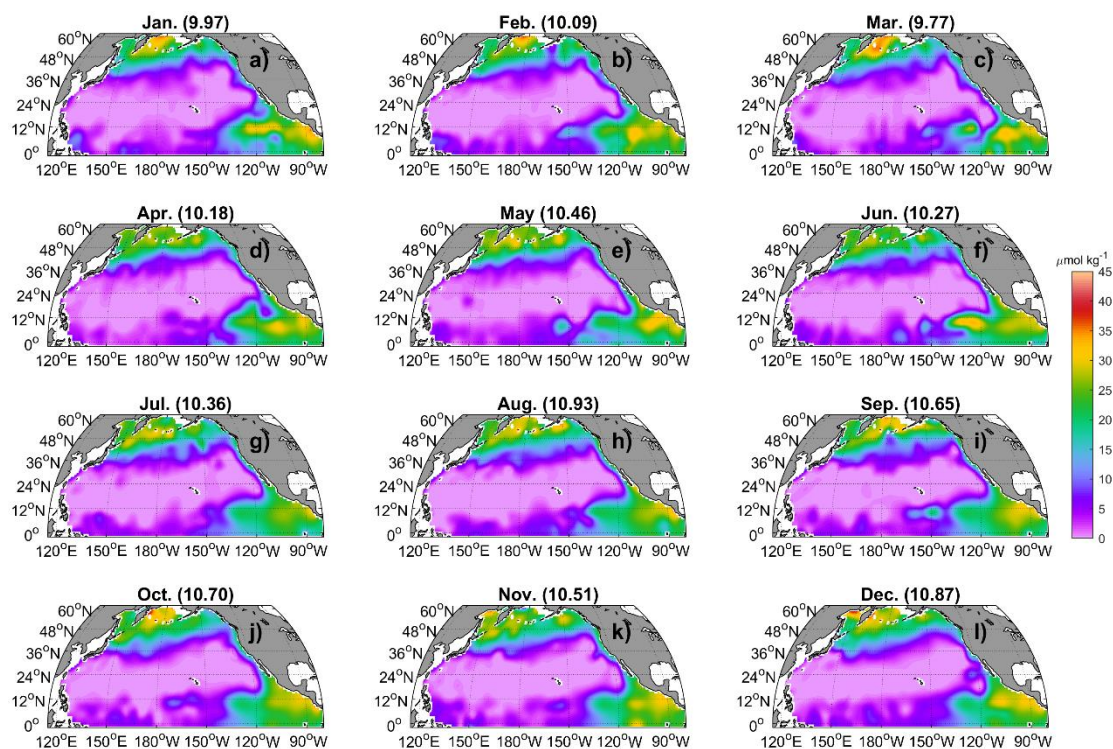


Figure S18. Similar to Fig. S17, but for a depth of 100 m.

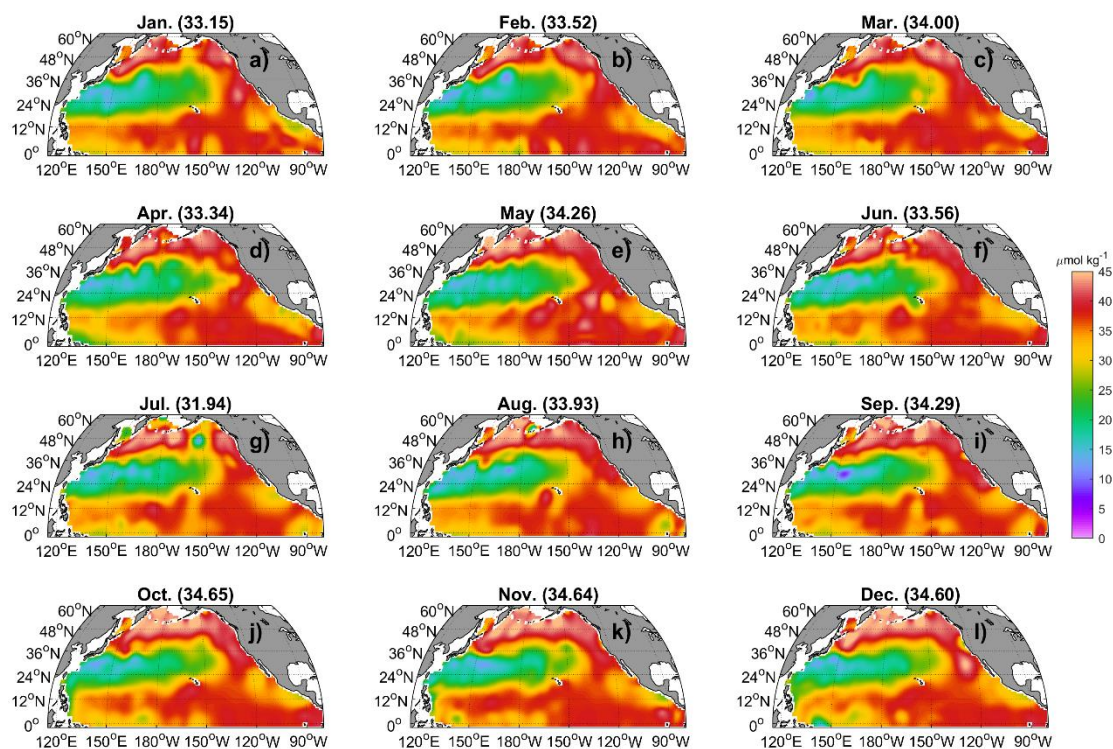


Figure S19. Similar to Fig. S17, but for a depth of 500 m.

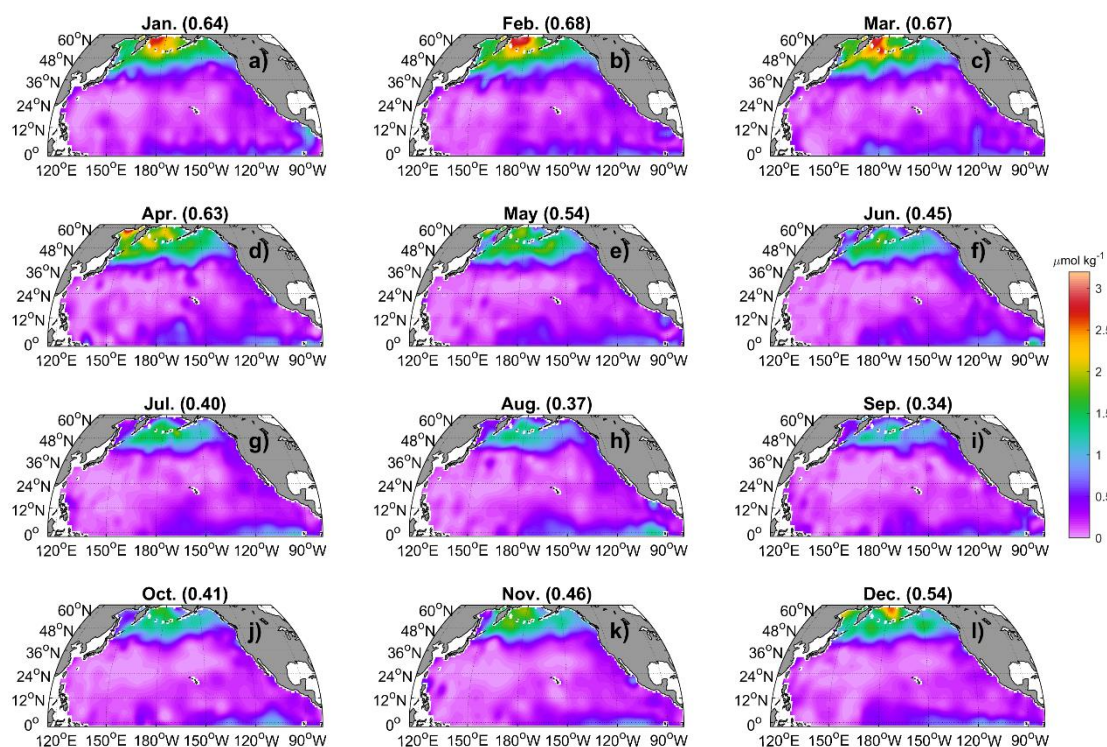


Figure S20. Similar to Fig. S17, but for DIP and at a depth of 5 m.

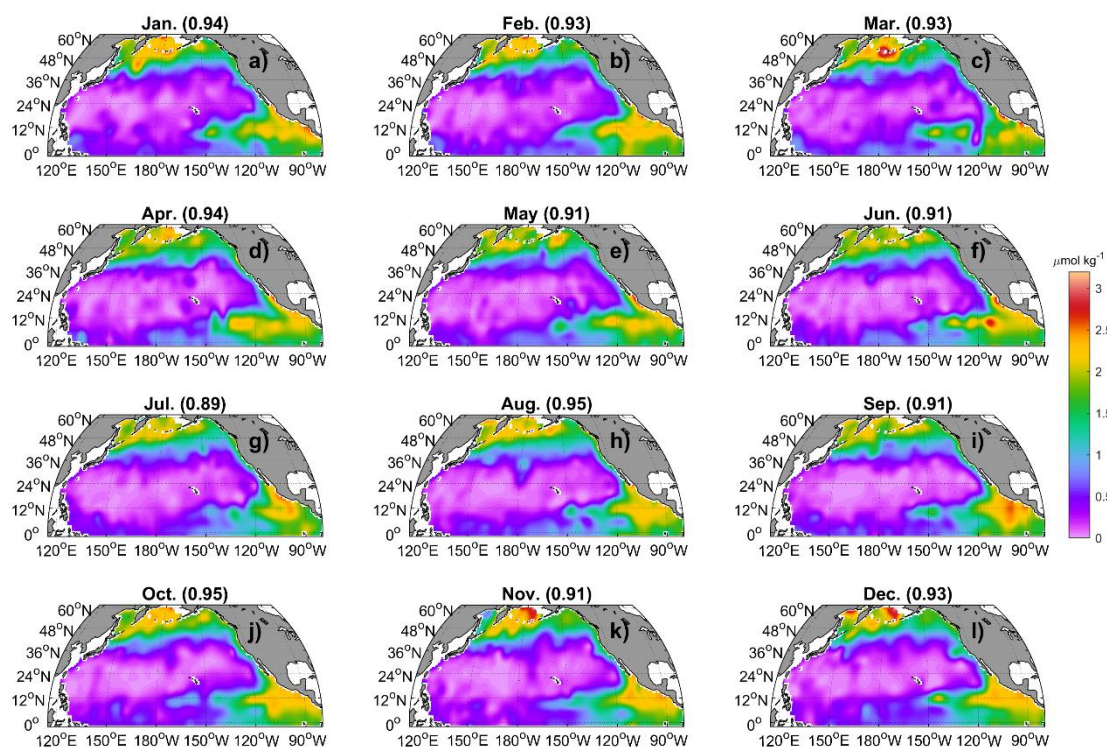


Figure S21. Similar to Fig. S17, but for DIP and at a depth of 100 m.

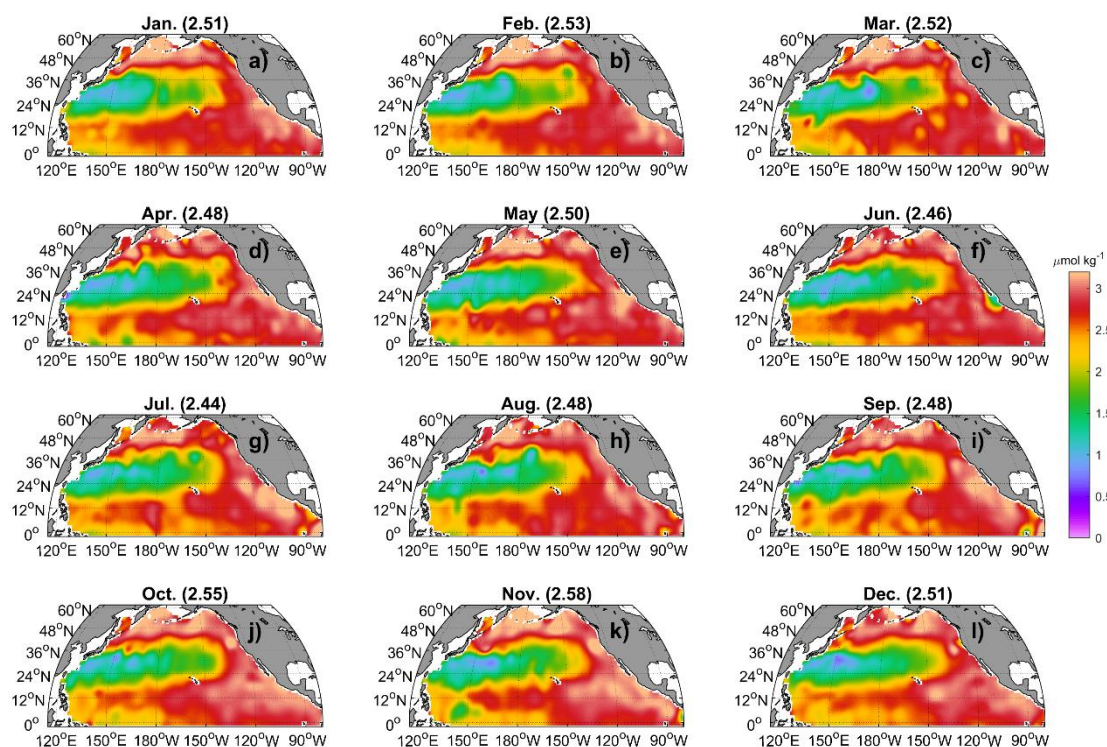


Figure S22. Similar to Fig. S17, but for DIP and at a depth of 500 m.

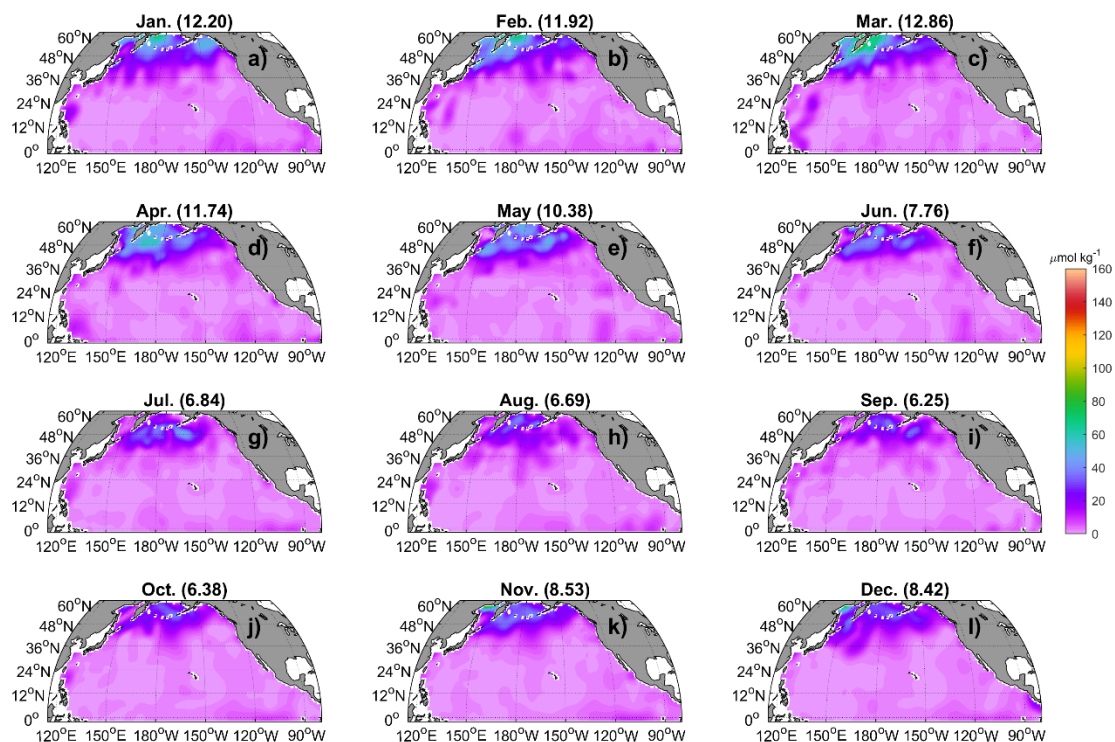


Figure S23. Similar to Fig. S17, but for Si(OH)_4 and at a depth of 5 m.

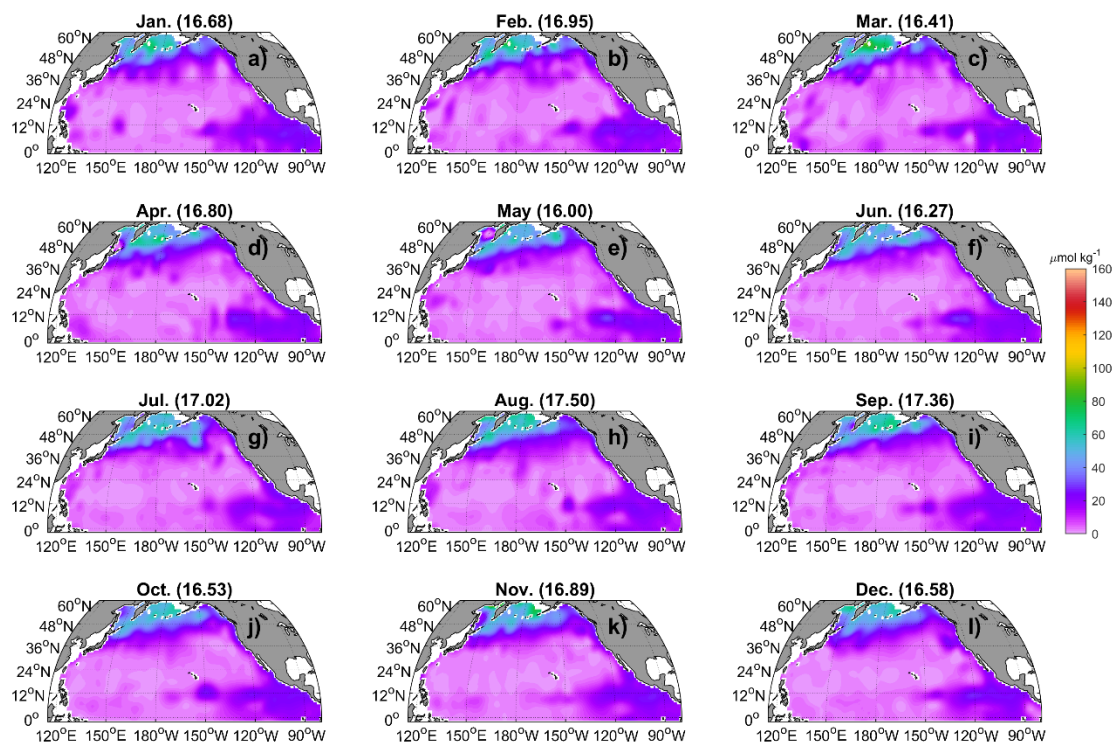


Figure S24. Similar to Fig. S17, but for Si(OH)_4 and at a depth of 100 m.

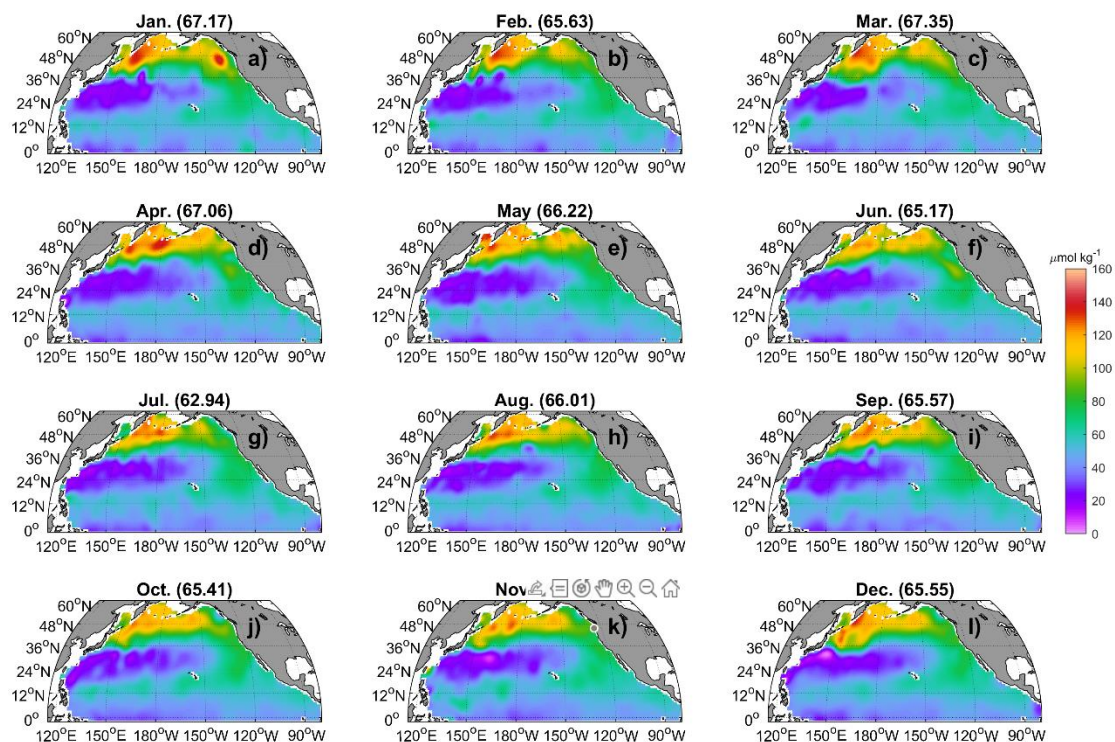


Figure S25. Similar to Fig. S17, but for Si(OH)_4 and at a depth of 500 m.