Supporting Information for Water mass analysis of the 2018 US GEOTRACES Pacific Meridional Transect (GP15)

R. M. Lawrence [1], A. Shrikumar [1], E. Le Roy [2], J. Swift [3], P. J. Lam [4], G. Cutter [5], and K. L. Casciotti [1]

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA

[2]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

[3]Department of Climate, Atmospheric Science, and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, CA 92037, USA

[4]University of California, Santa Cruz, Department of Ocean Sciences, 1156 High St, Santa Cruz, CA 95064, USA

[5]Department of Ocean and Earth Sciences, Old Dominion University, Norfolk, VA 23529, USA

Contents of this file

Introduction

This document contains supporting text, figures, and tables containing additional information and detail cited in the main text. The GP15 datasets used can be found on the BCO-DMO website (https://www.bco-dmo.org/dataset/778168/

data; https://www.bco-dmo.org/dataset/777951/data; https://www.bco-dmo.org/dataset/824867/data). Only data collected by the Oceanographic Data Facility (ODF, Scripps Institution of Oceanography) group was used. The water mass analysis results have been upload to the Stanford Digital Repository (https://purl.stanford.edu/tv301yr5579) and submitted to BCO-DMO. The water mass analysis was conducted via a modified Optimum Multiparameter (OMP) analysis method using the before-mentioned BCO-DMO GP15 datasets and the newly-developed pyompa package. The pyompa software can be found in Zenodo (https://zenodo.org/record/5733887), and the code to replicate the analysis can be found in Github (https://github.com/nitrogenlab/gp15wmascripts). Any known anomalies in the BCO-DMO datasets were flagged according to the SeaDataNet scheme as is recommended by GEOTRACES (SeaDataNet, 2010; GEOTRACES, n.d.; Cutter et al., 2018). Our analysis excluded data flagged as missing or as a known bad value. Once these data were removed, the thermocline (n=341) and intermediate/deep (n=682) samples with complete data for our parameters of interest (Section 2.2.1) were analyzed.

Text S1. pyompa soft penalty formulation.

The formula for the soft penalties applied to constrain the OMPA is below.

$$penalty(x) = \beta(e^{\alpha max(0, max(lowerbound-x, x-upperbound))} - 1)$$

The 'lowerbound' and 'upperbound' refer to the range over which each penalty is set. In this case, a specific latitude or potential density. The default α and β for latitude and potential density penalties were $\alpha = 0.05$ and $\beta = 100$. These default values were used except in the following cases. In the intermediate and deep water analysis, ENPCW, SPCW, PSUW, ESSW, EqIW, and AAIW, had $\alpha = 0.03$ and $\beta = 50$ for their respective

latitude penalties. In the thermocline analysis, only PSUW and SPCW deviated from the default with $\alpha = 0.03$ and $\beta = 50$ for their respective latitude latitude penalties.

Text S2. Parameter weightings

Our parameter weightings were modified from Peters et al. (2018) as follows. In the thermocline, we started with Peters et al. (2018)'s coded weights of 160, 155, 5, and 10 for temperature, salinity, silicate, and PO (a combination of phosphate and oxygen), respectively. Oxygen was given the same weight as PO, and we weighted phosphate five times the weight of oxygen to give its contribution to the overall cost a similar magnitude. Nitrate was given the same weight as phosphate. We also increased the weight of temperature by 25%. After dividing all the weights by ten to lower the overall weights while maintaining the relative weighting, the weights applied to the thermocline OMP analysis were 20.0, 15.5, 0.5, 5, 5, and 1 for conservative temperature, salinity, $Si(OH)_4$, NO_3^-, PO_4^{3-}, and O_2, respectively.

In the intermediate and deep waters, we started with Peters et al. (2018)'s coded weights of 140, 100, 30, and 10 for temperature, salinity, silicate, and NO (a combination of nitrate and oxygen). Oxygen was given the same weight as NO, and we weighted nitrate five times the weight of oxygen to give its contribution to the overall cost a similar magnitude. Phosphate was given the same weight as nitrate. We also increased the weight of temperature and salinity to account for their conservative nature compared to the other parameters affected by particle dissolution and nutrient regeneration. After dividing all the weights by ten, as described above, intermediate and deep OMP analysis weights were of 56, 80, 3, 5, 5, and 1 for conservative temperature, salinity, $Si(OH)_4$, NO_3^-, PO_4^{3-}, and O_2, respectively.

We compared our parameter weightings to weights cited in (Peters et al., 2018), which are 140 for temperature, 100 for salinity, 30 for silicate, 20 for NO, and 10 for PO. We translated the weights of PO and NO as 20 for oxygen, 20 for nitrate, and 10 for phosphate. These (Peters et al., 2018) weights were kept the same in both the thermocline and intermediate and deep water analyses. The differences in water mass fractions between these GP15 water mass analyses (our water mass fractions minus the water mass fractions resulting from (Peters et al., 2018)'s cited weights) is in figures S2 and S3.

Text S3. Procedure for obtaining thermocline boundaries

Because observations in the thermocline are analyzed using the "thermocline array" technique while observations in the intermediate and deep waters are analyzed with the standard pyompa technique, it is important to separate the observations in the thermocline from those that are in intermediate and deep waters. In prior work (Peters et al., 2018; Jenkins et al., 2015), a single cutoff in potential density was applied across all observations to separate out those in the thermocline. However, in reality, the end of the true thermocline may exist at different potential density cutoffs depending on the station being considered. To account for this, in this work we determined station-specific thermocline thresholds as follows:

First, we downloaded the GP15 ODF CTD data from the BCO-DMO website (`https://www.bco-dmo.org/dataset/778168/data`) and organized the observations by station. In cases where there were multiple CTD casts per station, we focused on data from the cast that had more observations. Then, for each station, we followed these steps:

1. Organize the data for the station into two vectors: the first vector contains the depths of the observations sorted in ascending order, and the second vector contains the corresponding potential density values (set to a reference depth of 0 m).

2. Density should increase monotonically with depth, but sometimes slight noise in the observations can cause density to appear to decrease with increasing depth. To get around this, we use a technique called "isotonic regression", where we fit a non-decreasing function where the input is the depth and the output is the potential density. To fit the isotonic regressor, we used scikit learn's Isotonic Regression class (`https://scikit-learn.org/stable/modules/isotonic.html`).

3. Isotonic regression produces a non-smooth function that increases in sharp steps. To create a smoothed version of the output of the isotonic regressor, we fit the output of the isotonic regressor to a PCHIP interpolator (PCHIP stands for "Piecewise Cubic Hermite Interpolating Polynomial"), as implemented by scipy (`https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.PchipInterpolator.html`). The PCHIP interpolator produces a smooth fit while preserving the monotonicity of the isotonic regressor. We will use σ_0 to denote the function produced by the PCHIP interpolator, such that the potential density output for a depth z is written as $\sigma_0(z)$.

4. Once we have the smooth piecewise polynomial function σ_0, we can calculate its first derivative at any depth. We will use $\sigma_0'(z)$ to denote the first derivative of the potential density at a depth z. We compute this derivative at 1000 evenly-spaced depths between the minimum and maximum depths of the observations at the station.

5. To identify a cutoff for the thermocline (or, more formally, the pycnocline), we make the simplifying assumption that the gradient of the potential density is first monotonically

increasing (i.e. the gradient gets steeper and steeper as we enter the thermocline), and is then monotonically decreasing (i.e. gets less steep as we exit the thermocline and enter the abyssal ocean). Although this assumption is not always true in practice (e.g. the presence of a mode water can create a flatness in the density gradient before the abyssal waters are reached), for our purposes of identifying the upper and lower boundaries of the thermocline, we find that it is sufficient. To put this assumption into practice, we find the maximum of the derivative (over the 1000 depths mentioned in the previous steps), and then split the 1000 depths into two sections around this maximum. For the shallower section, we fit a monotonically increasing isotonic regressor to the values of $\sigma'(z)$ (call the resulting function σ'_{0inc}), and for the deeper section we fit a monotonically decreasing isotonic regressor (call the result function σ'_{0dec}).

6. The upper thermocline boundary is set at which $\sigma'_{0inc}(z)$ is either ≥ 0.01 or is $\geq$ 25% of the maximum derivative (over the 1000 depths), whichever was shallower. We define the lower boundary of the thermocline as the depth in the deeper section at which $\sigma'_{0dec}(z) \leq 0.003$. These cutoffs were chosen based on visual inspection, and will likely need to be adjusted for different datasets.

After performing the procedure above for all stations, we have upper and lower bounds for the thermocline for each station (Figure S3).

Text S4. Thermocline endmember selection

In the thermocline array technique, different end-members are defined for every potential density increment. How do we define the property values for these end-members? To achieve this, we combined the cubic smoothing splines approach used in Jenkins et al. (2015) and Peters et al. (2018) with a novel iterative outlier filtering scheme as follows:

first, we downloaded GLODAPv2 data for the Pacific and Arctic oceans and combined them into one dataset. Next, we used the ranges from Table 1 to filter for datapoints corresponding to a particular end-member. Then, for each property value, we go through the datapoints corresponding to each single end-member and fit a cubic spline as follows:

1. Fit a cubic smoothing spline to predict the value of the property as a function of the potential density (reference depth 0 m). We used the csaps python package (`https://github.com/espdev/csaps`) version 1.1.0, with the smoothing parameter p set to 0.8.

2. Look at all the datapoints corresponding to the end-member and we compute the difference between the true property value and the predicted property value according to the csaps spline from the previous step. We then square these errors, compute the standard deviation, and filter out all datapoints for which the square of the residual is more than two standard deviations greater than zero.

3. Recompute a new smoothing spline on the filtered dataset, and recompute which points would be filtered out according to the new spline. If the set of points getting filtered out is the same as before, end the iteration and designate the newly computed cubic spline as the final cubic spline. Otherwise, repeat this step, but now filtering out points based on the new spline.

Once the spline is fit, we can compute the predicted property values according to the spline at each potential density increment, and use this to define our end-members for the increment.

Figure S1. Comparison of parameter weighting in thermocline water masses. The x-axis is the difference between the GP15 water mass fractions resulting from our thermocline

water mass weighting and from (Peters et al., 2018)'s cited weights. The y-axis is the number of GP15 samples with water mass fractions above zero in at least one set of results from the two weighting schemes. In other words, samples that had a water mass fraction of zero in both of the weighting schemes were excluded from the histogram to focus on the relevant samples from the thermocline. Bins sizes are approximately 0.01 (e.g. the bin including zero is from approximately -0.005 to 0.005).

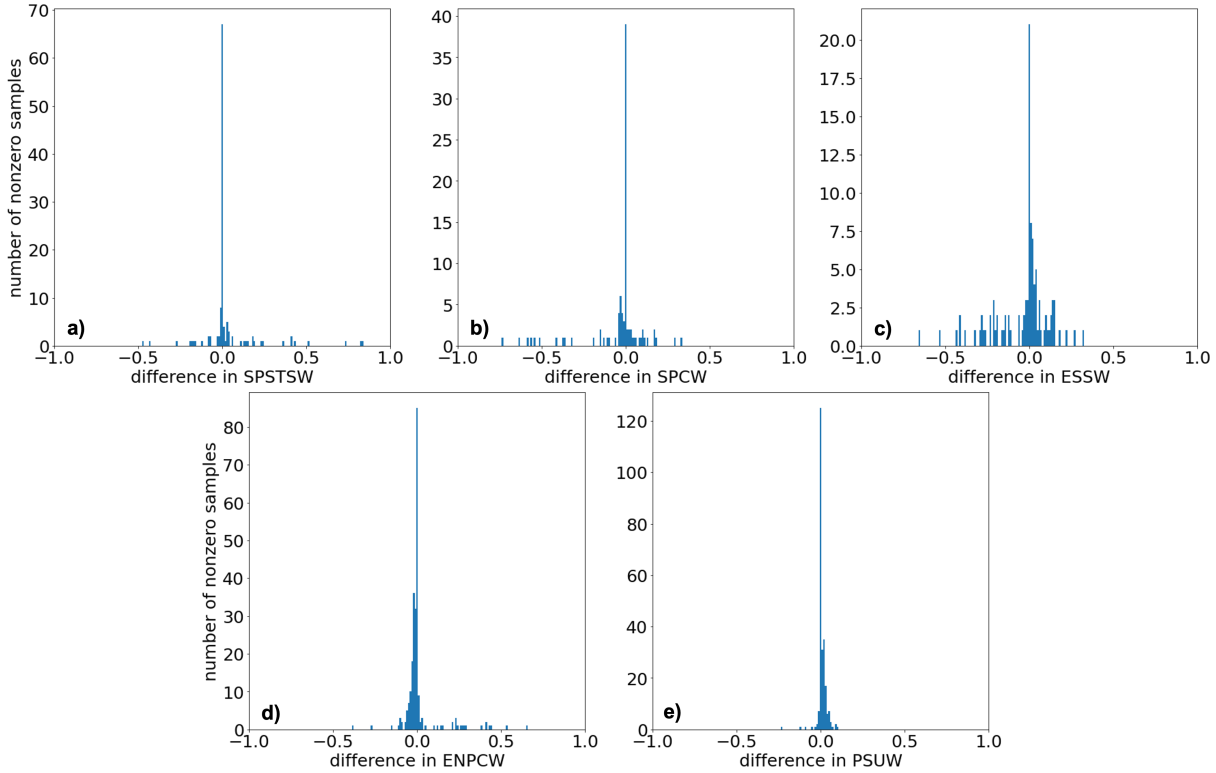


Figure S2. Comparison of parameter weighting in the intermediate and deep water. The x-axis is the difference between the GP15 water mass fractions resulting from our intermediate and deep water mass weighting and from (Peters et al., 2018)'s cited weights. The y-axis is the number of GP15 samples showing a given difference between the two weighting schemes. The y-axis is the number of GP15 samples with water mass fractions above zero in at least one set of results from the two weighting schemes. In other words, samples that has a water mass fraction of zero from both of the weighting schemes were excluded from the histogram to focus on the relevant intermediate and deep water samples. Bins sizes are approximately 0.01 (e.g. the bin including zero is from approximately -0.005 to 0.005).

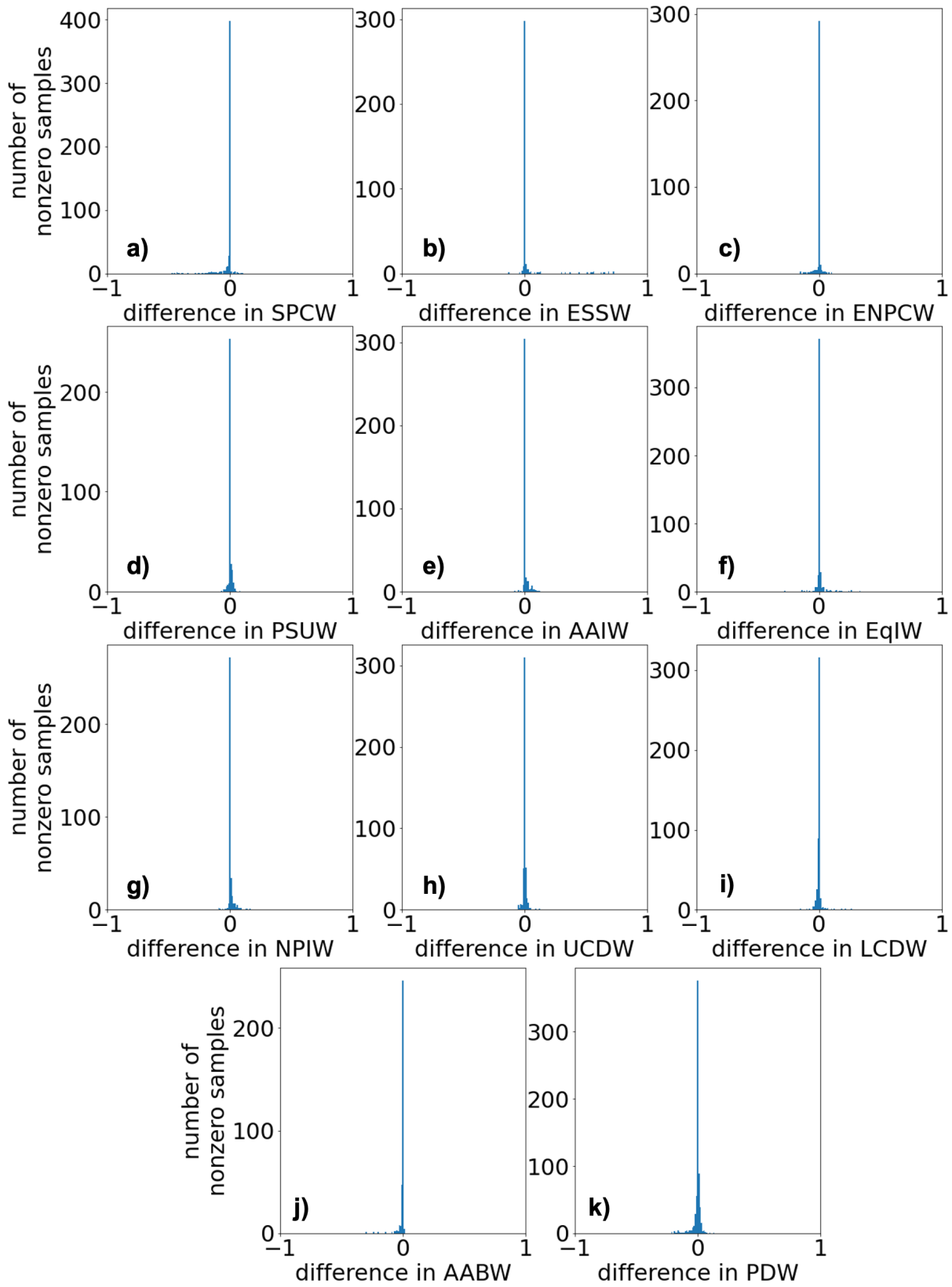

number of nonzero samples
a)
difference in SPCW
b)
difference in ESSW
c)
difference in ENPCW
d)
difference in PSUW
e)
difference in AAIW
f)
difference in EqIW
g)
difference in NPIW
h)
difference in UCDW
i)
difference in LCDW
j)
difference in AABW
k)
difference in PDW

Figure S3. Thermocline a) depth and b) potential density anomaly (σ_0) boundaries for each GP15 station.

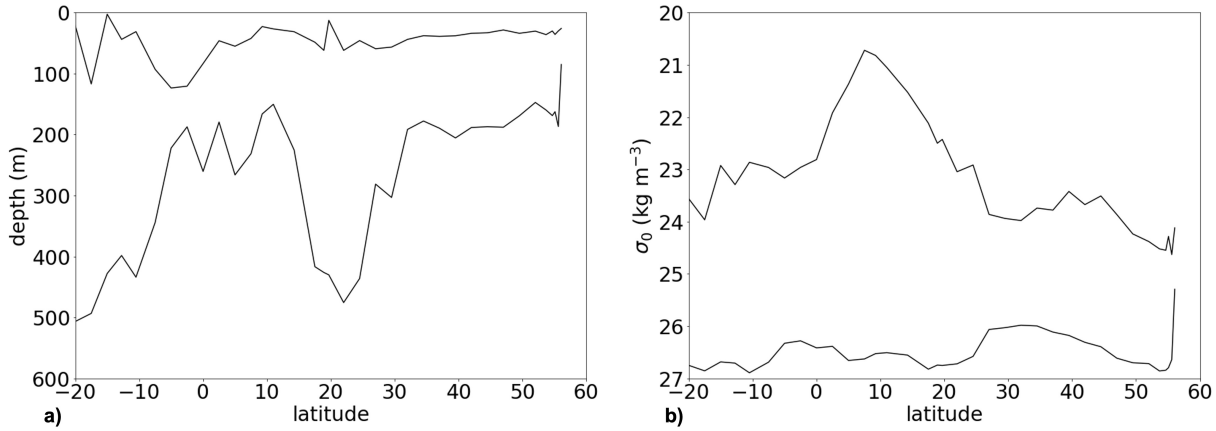


Figure S4. Full water column section plots of a) conservative temperature (℃), b) absolute salinity, c) dissolved oxygen concentrations (μmol kg^{-1}), d) silicate concentrations (μmol kg^{-1}), e) nitrate concentrations (μmol kg^{-1}), and f) phosphate concentrations (μmol kg^{-1}) along the GP15 transect.

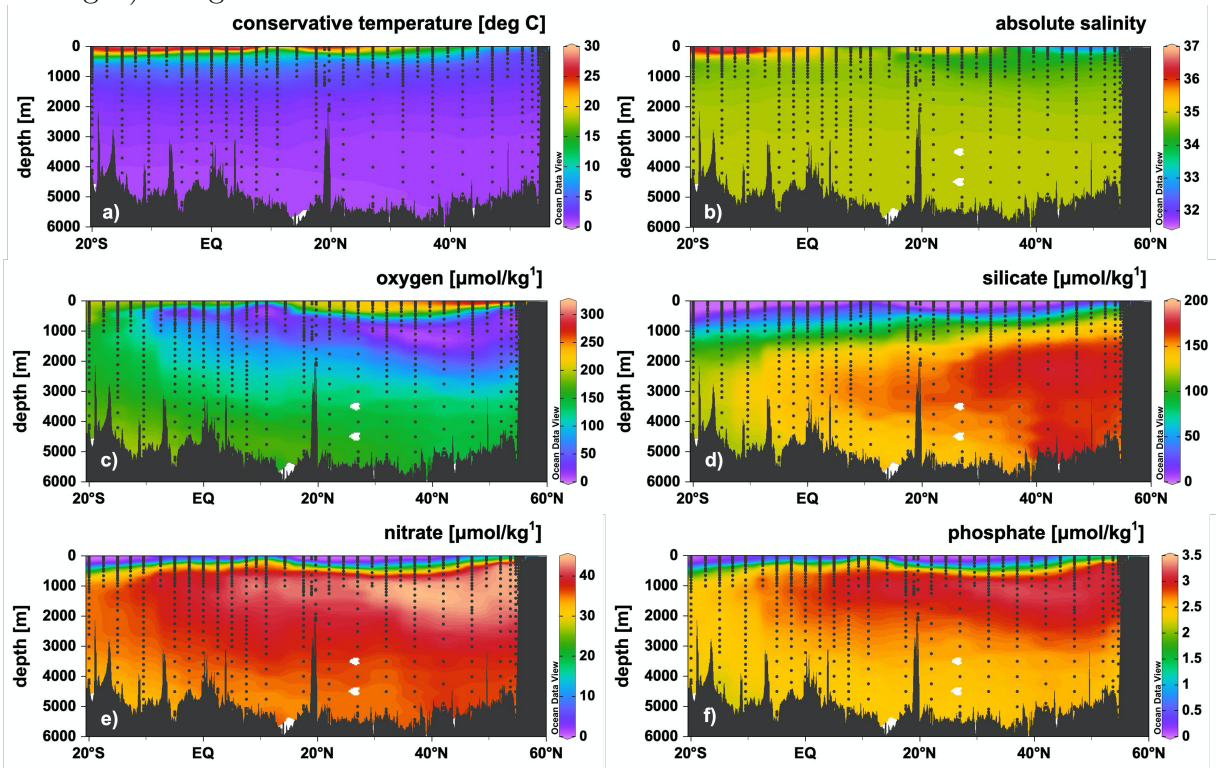


Figure S5. N* (μmol kg^{-1}) with dissolved oxygen (μmol kg^{-1}) contours in the full water column along the GP15 transect.

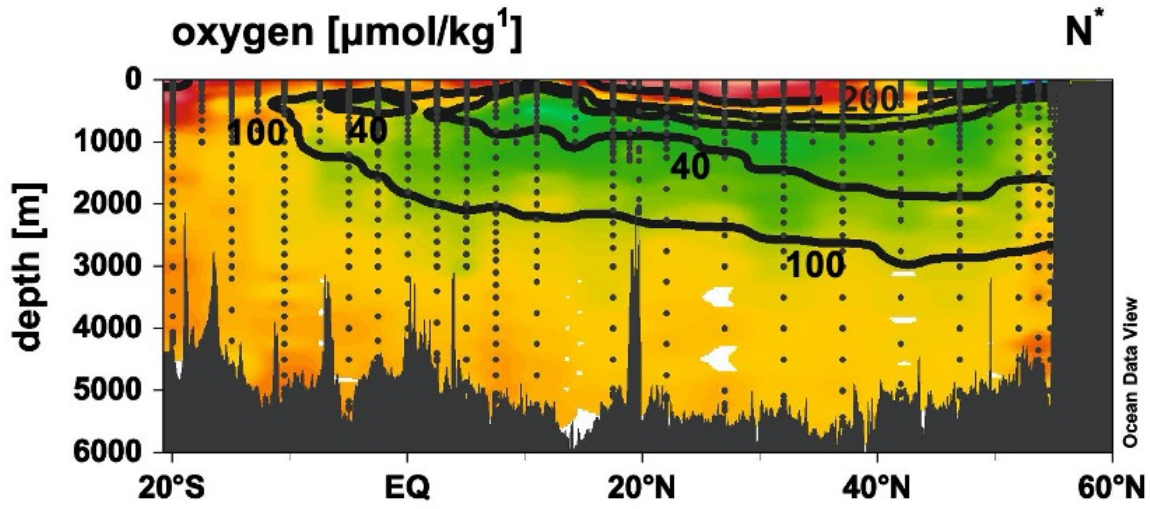


Figure S6. The yellow dots show nitrate and phosphate data flagged as probably bad (3) compared to most of the data in teal flagged as probably good (2). No data flagged as a known bad value (4) was included in the OMP analysis.

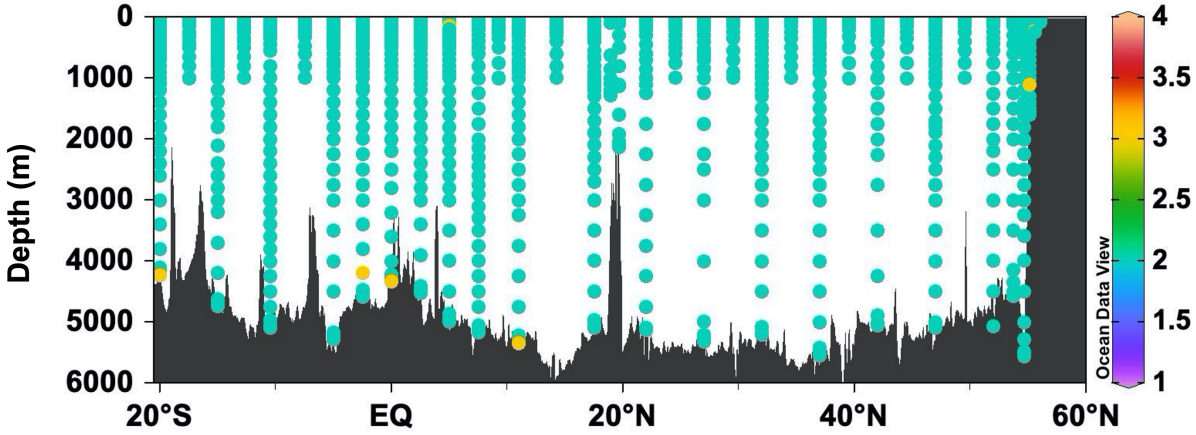


Figure S7. The algorithm-chosen O:P ratio (μmol/μmol) along the GP15 transect where a) phosphate assimilation is more than -0.25 (μmol kg^{-1}) and b) phosphate regeneration is above 0.25 (μmol kg^{-1}).

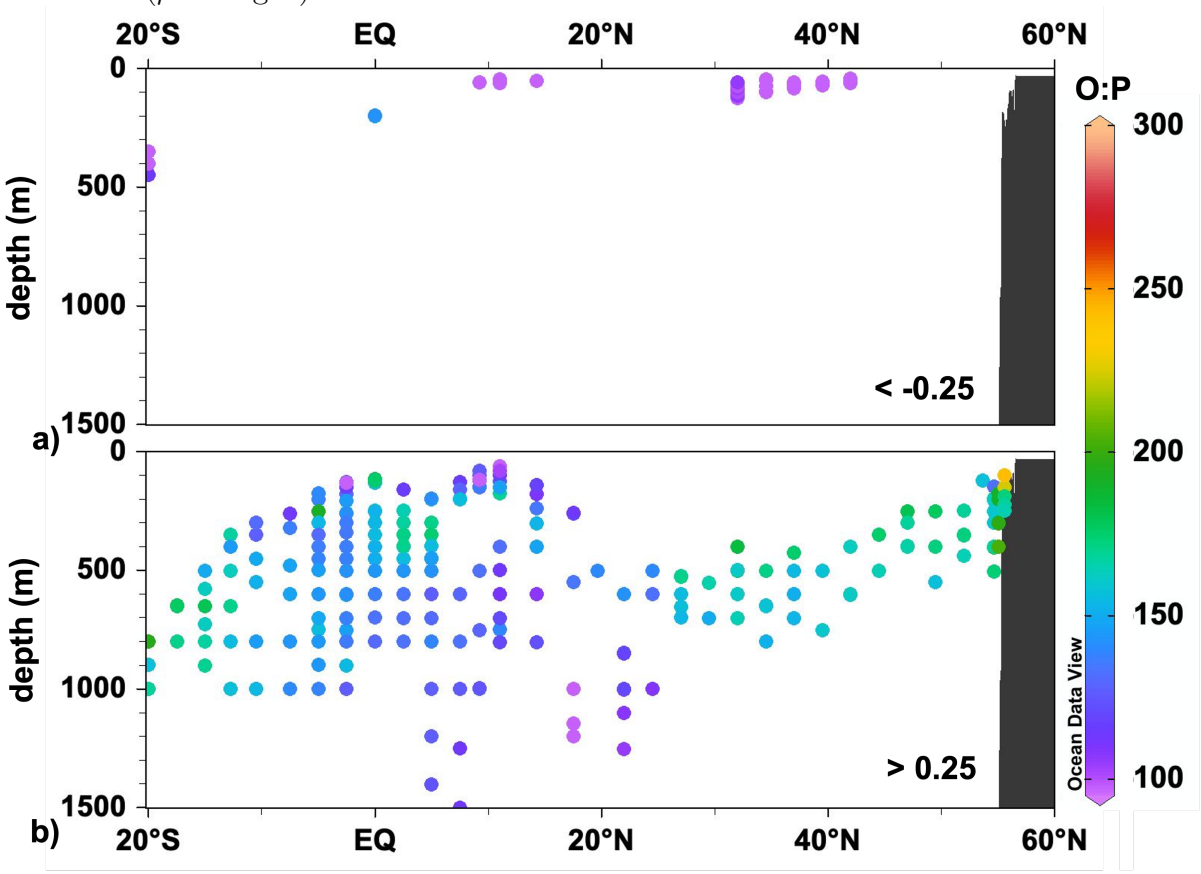
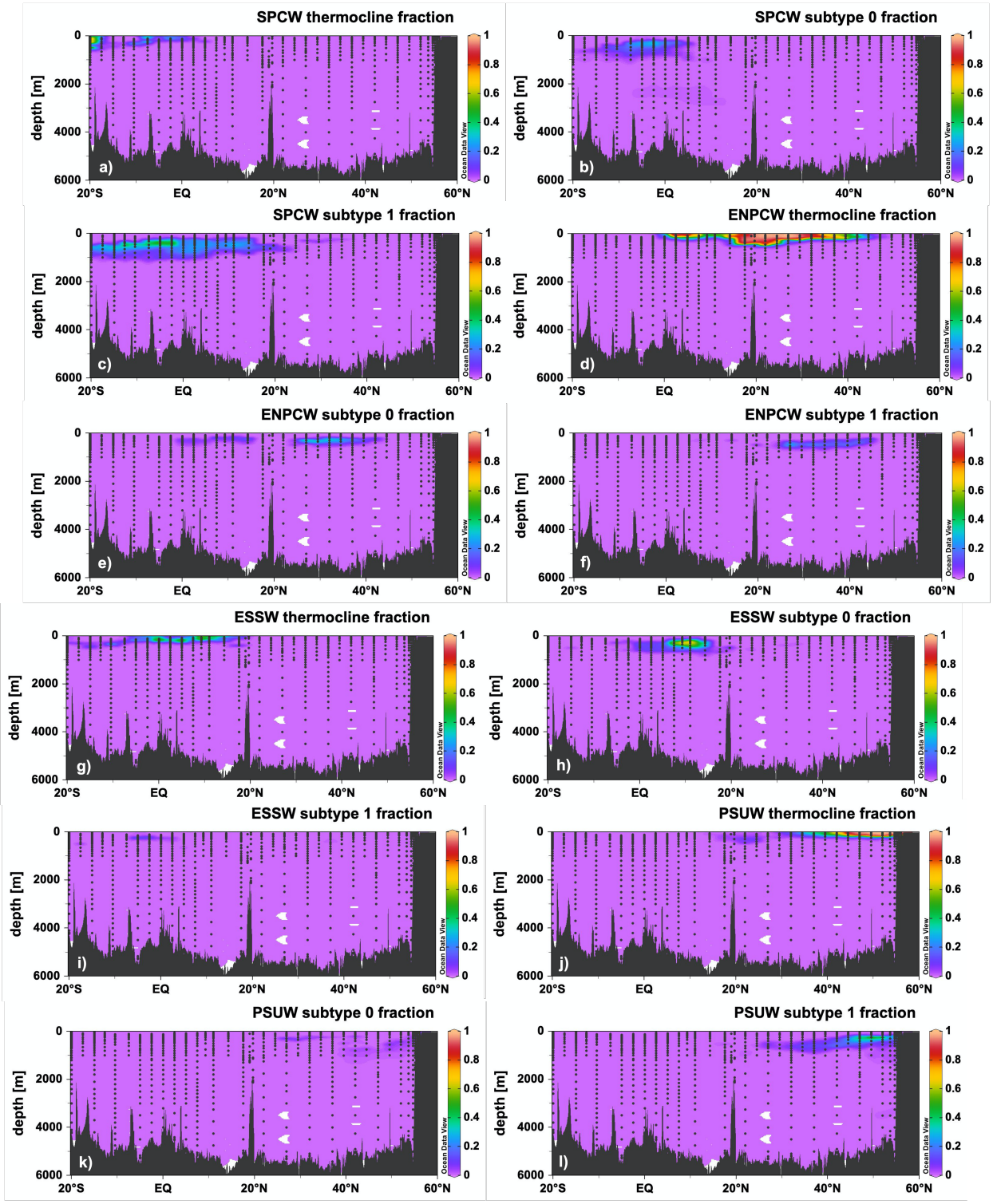


Figure S8. Water mass fractions for water mass subtypes. For the thermocline water masses included in the intermediate and deep analysis, water mass fractions for the thermocline analysis are provided in addition to subtypes. AABW and UCDW are not included as these only have one defined archetype (Table S1). SPSTSW is not included as this water mass was only included in the thermocline analysis. The colorbar represents the water mass fractions for SPCW (a-c), ENPCW (d-f), ESSW (g-i), PSUW (j-l), AAIW (m-n), EqIW (o-p), LCDW (q-r). NPIW (s-u), and PDW (v-x).


SPCW thermocline fraction
SPCW subtype 0 fraction
SPCW subtype 1 fraction
ENPCW thermocline fraction
ENPCW subtype 0 fraction
ENPCW subtype 1 fraction
ESSW thermocline fraction
ESSW subtype 0 fraction
ESSW subtype 1 fraction
PSUW thermocline fraction
PSUW subtype 0 fraction
PSUW subtype 1 fraction
depth [m]
0
2000
4000
6000
20°S
EQ
20°N
40°N
60°N
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
Ocean Data View
1
0.8
0.6
0.4
0.2
0

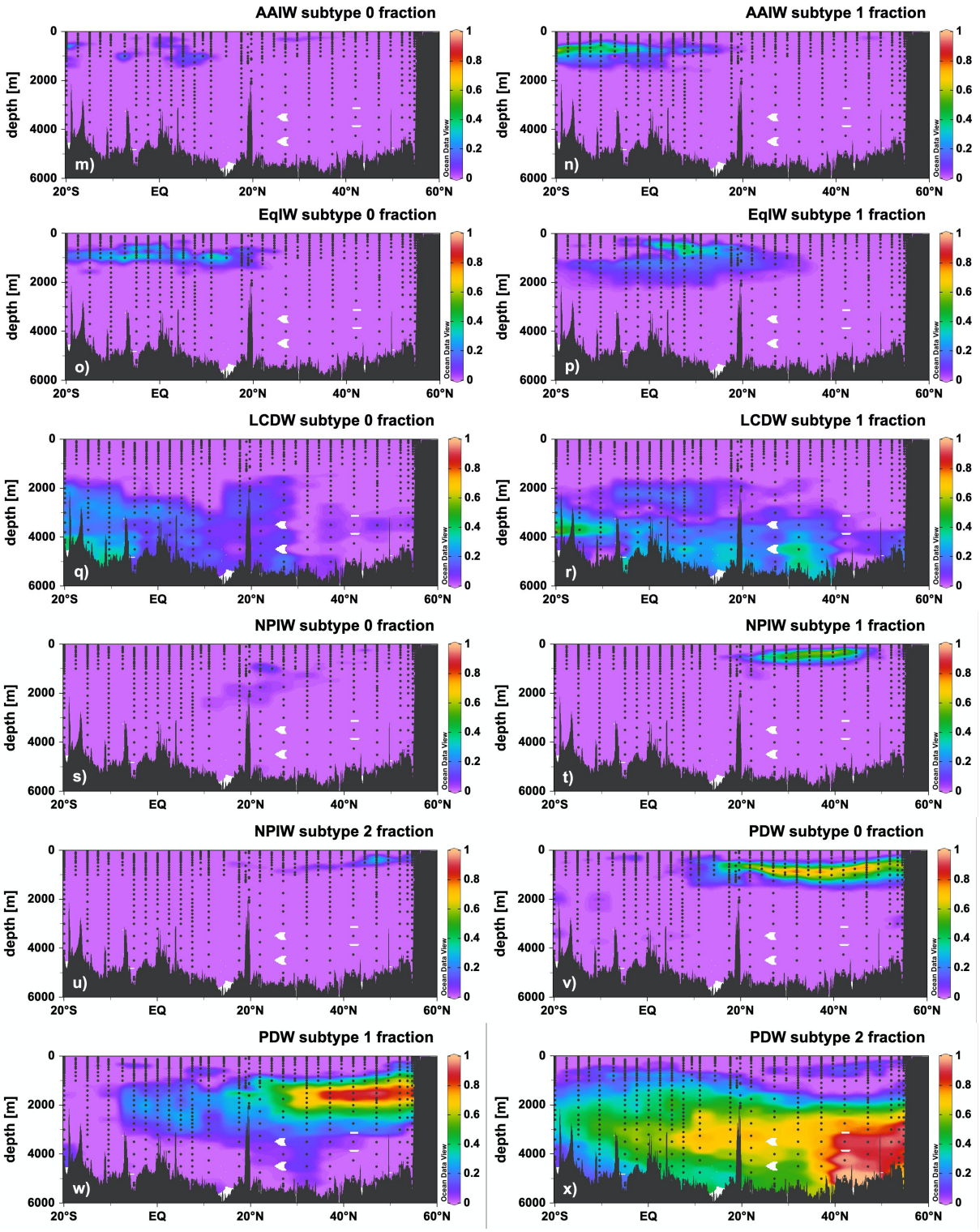

AAIW subtype 0 fraction
AAIW subtype 1 fraction
EqIW subtype 0 fraction
EqIW subtype 1 fraction
LCDW subtype 0 fraction
LCDW subtype 1 fraction
NPIW subtype 0 fraction
NPIW subtype 1 fraction
NPIW subtype 2 fraction
PDW subtype 0 fraction
PDW subtype 1 fraction
PDW subtype 2 fraction
depth [m]
0
2000
4000
6000
20°S
EQ
20°N
40°N
60°N
1
0.8
0.6
0.4
0.2
0
Ocean Data View
m)
n)
o)
p)
q)
r)
s)
t)
u)
v)
w)
x)

Figure S9. Profiles of a) oxygen (μmol kg^{-1}) and b) nitrate (μmol kg^{-1}) for GP15 Station 39, located at 20°S. The oxygen minimum and nitrate maximum characteristic of Upper Circumpolar Deep Water (UCDW) align between 1000 m and 2000 m, where our water mass analysis yielded the highest UCDW mass fractions.

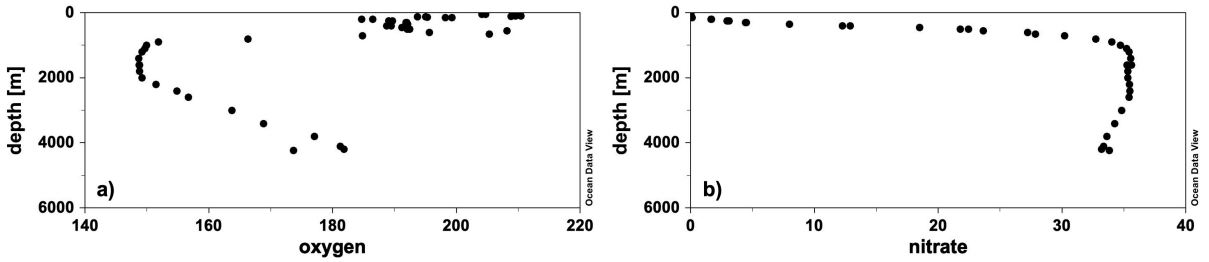


Figure S10. Lower Circumpolar Deep Water (LCDW) fraction with σ_4 contours 45.84 and 45.88 kg m^{-3} overlain as black lines. Figure 10.18 of Talley (2011) places LCDW below 45.84 kg m^{-3} at 28°S and 45.88 kg m^{-3} at 24°N. In our analysis, LCDW water mass fractions were primarily below the depths of these contour lines, in agreement with Talley (2011).

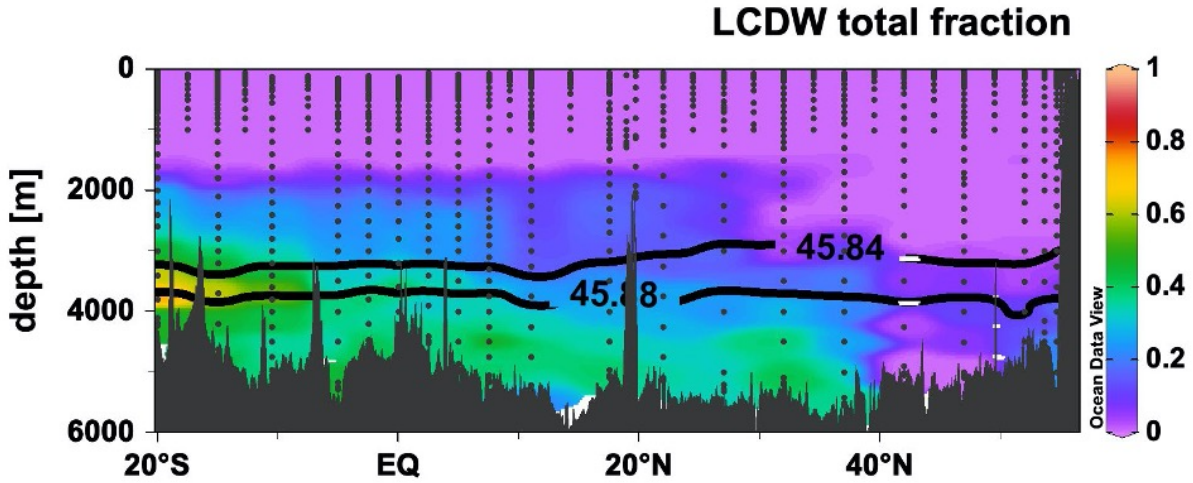


Figure S11. Histograms of a) temperature (℃) b) absolute salinity, c) oxygen (μmol kg^{-1}), d) silicate (μmol kg^{-1}), e) nitrate (μmol kg^{-1}), and f) phosphate (μmol kg^{-1}) residuals for all samples in the thermocline and intermediate and deep water analyses.

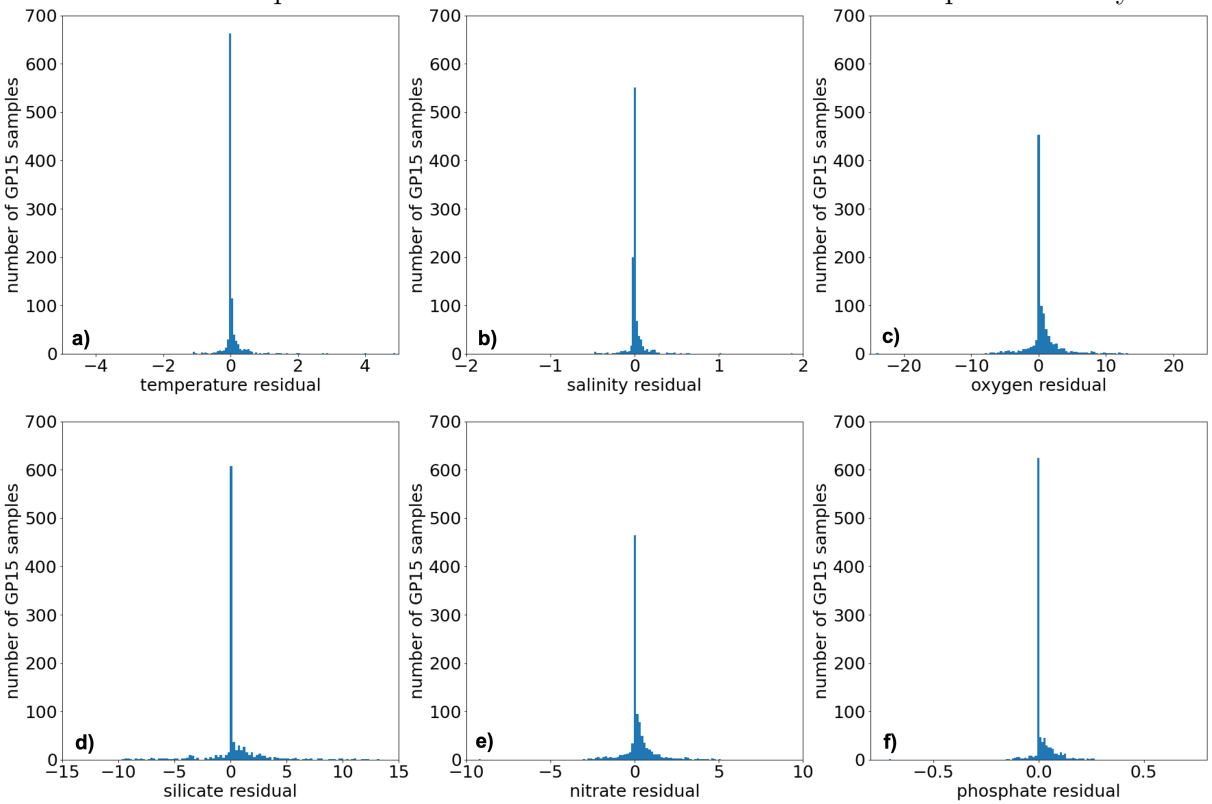


Figure S12. Range in residuals compared with previous studies for a) temperature (°C) b) absolute salinity, c) oxygen (μmol kg^{-1}), d) silicate (μmol kg^{-1}), e) nitrate (μmol kg^{-1}), and f) phosphate (μmol kg^{-1}).

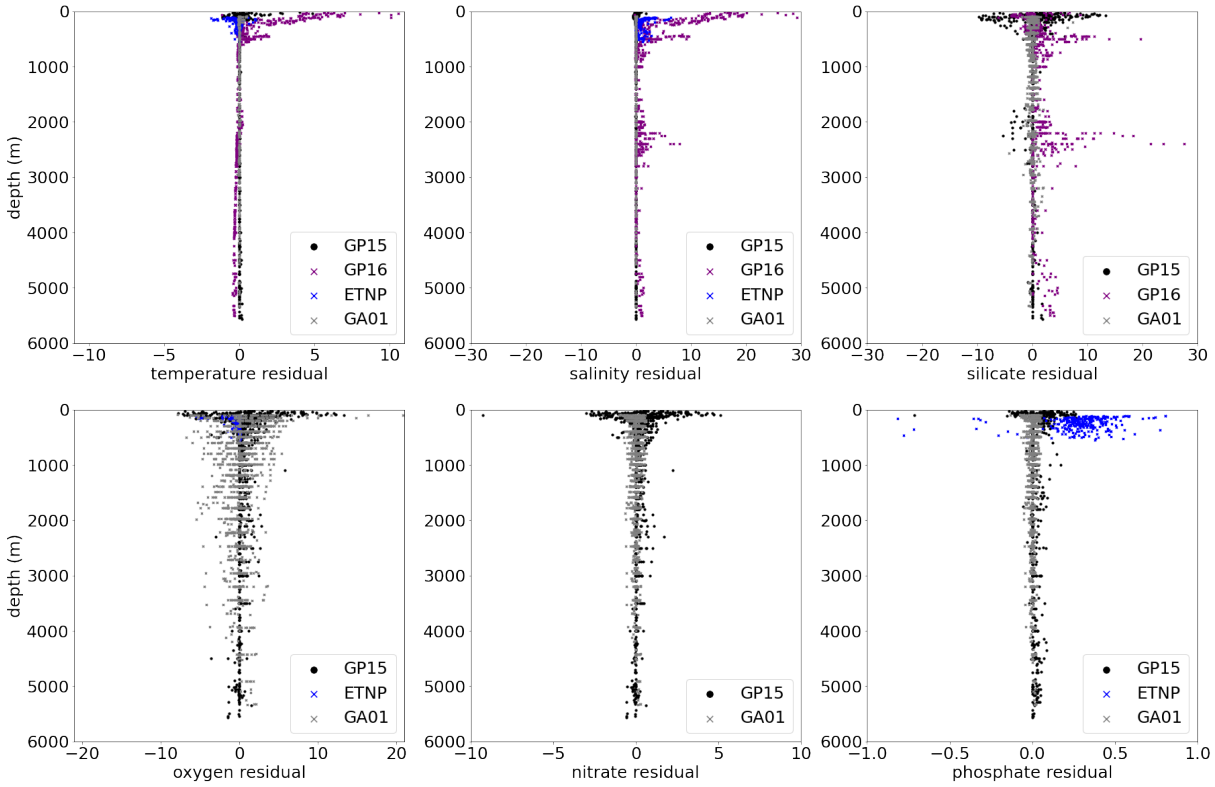


Figure S13. The standard deviation of residuals with altered parameter weightings for a) conservative temperature (℃) b) absolute salinity, c) oxygen (μmol kg^{-1}), d) silicate (μmol kg^{-1}), e) nitrate (μmol kg^{-1}), and f) phosphate (μmol kg^{-1}).

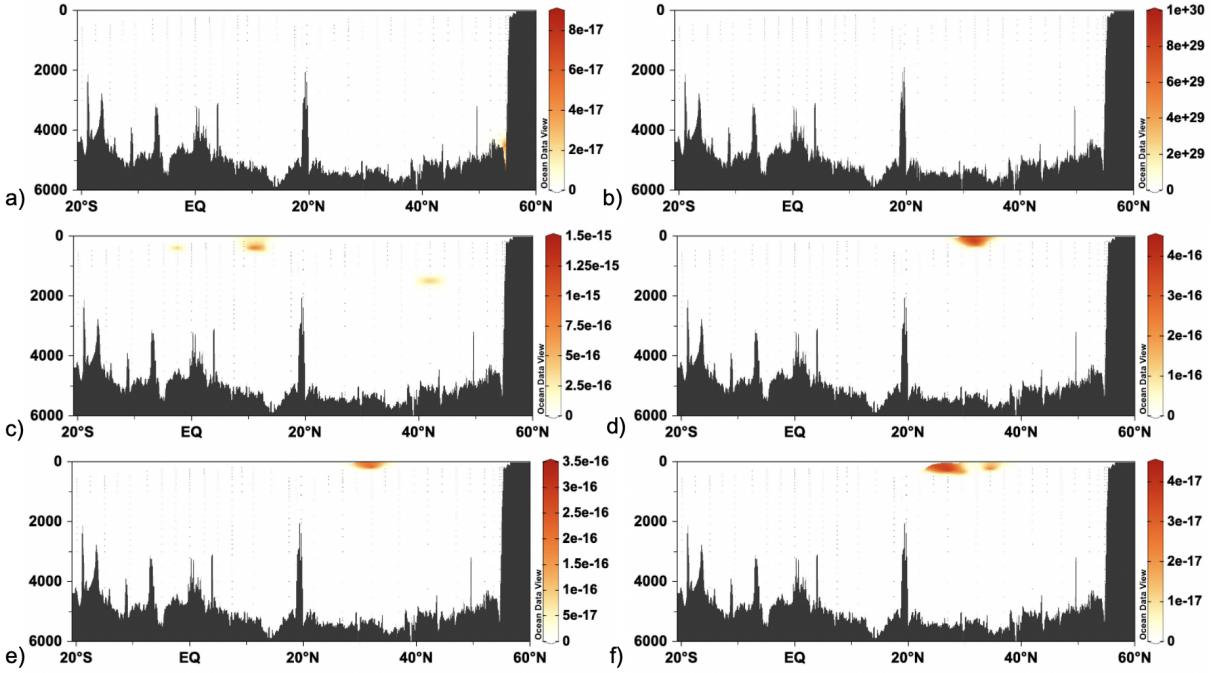


Figure S14. Thermocline endmember properties for a) conservative temperature (℃) b) absolute salinity, c) oxygen (μmol kg^{-1}), d) silicate (μmol kg^{-1}), e) nitrate (μmol kg^{-1}), and f) phosphate (μmol kg^{-1}).

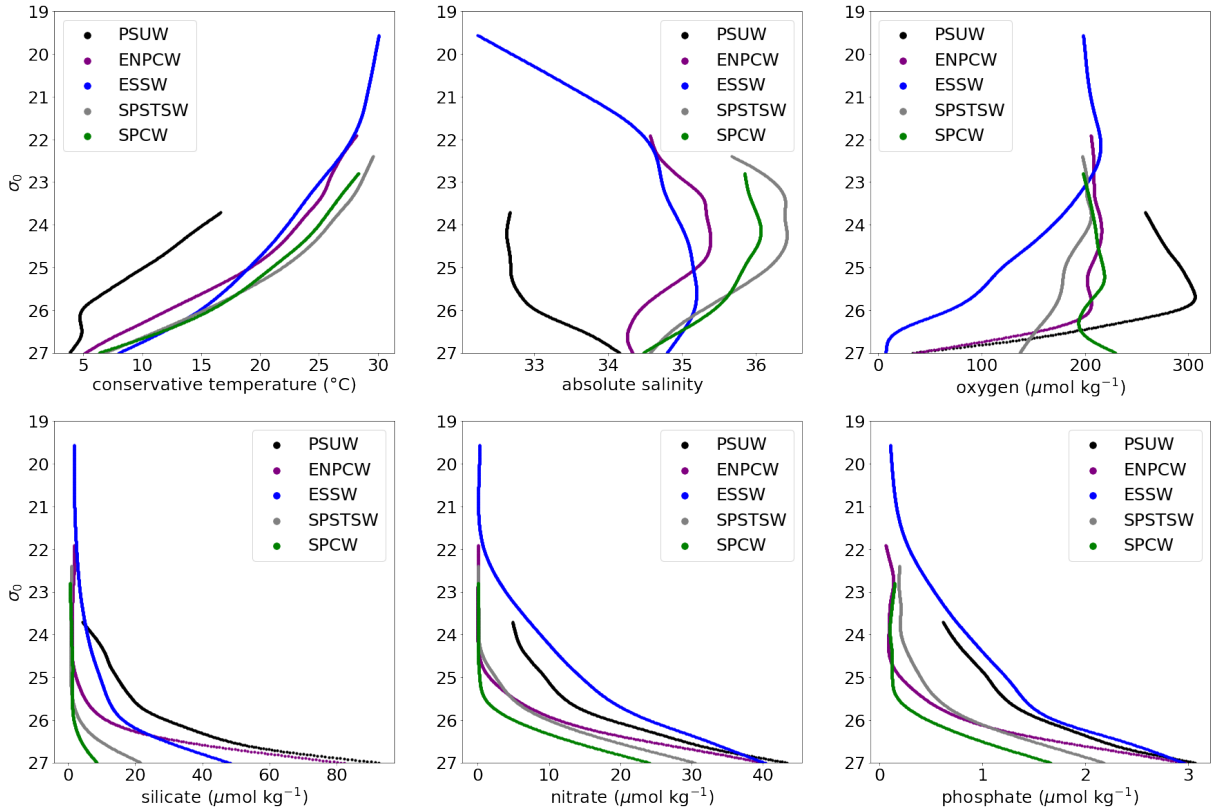


Figure S15. Approximate density range we would expect to find NPCMW. This intersects with several different thermocline and intermediate water types present in our analysis.

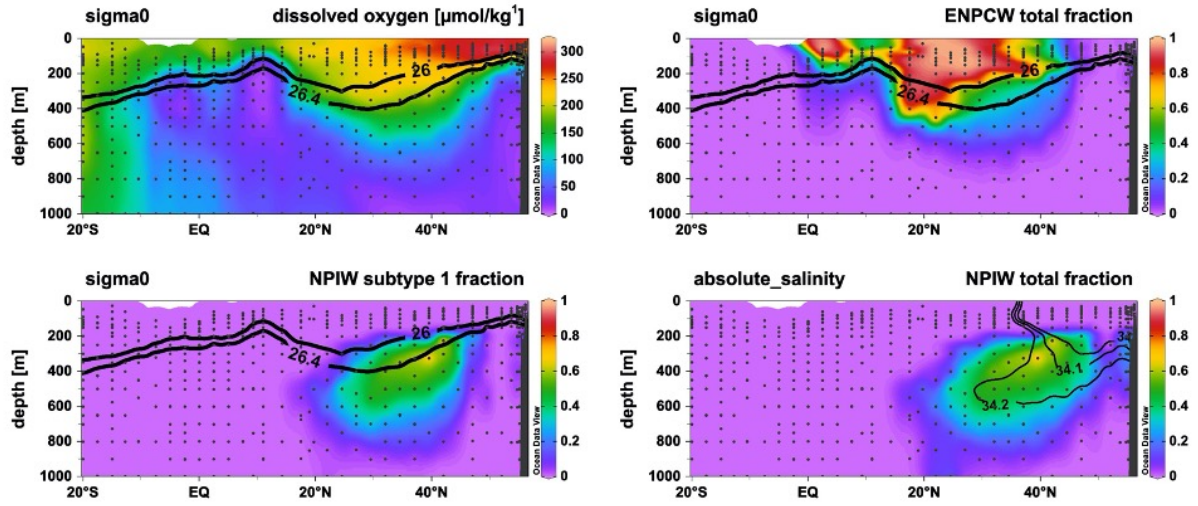


Figure S16. SPCW carries the signal of the Eastern Tropical South Pacific (Peruvian) oxygen deficient zone as shown by the dissolved oxygen contour of 50 μmol kg^{-1}.

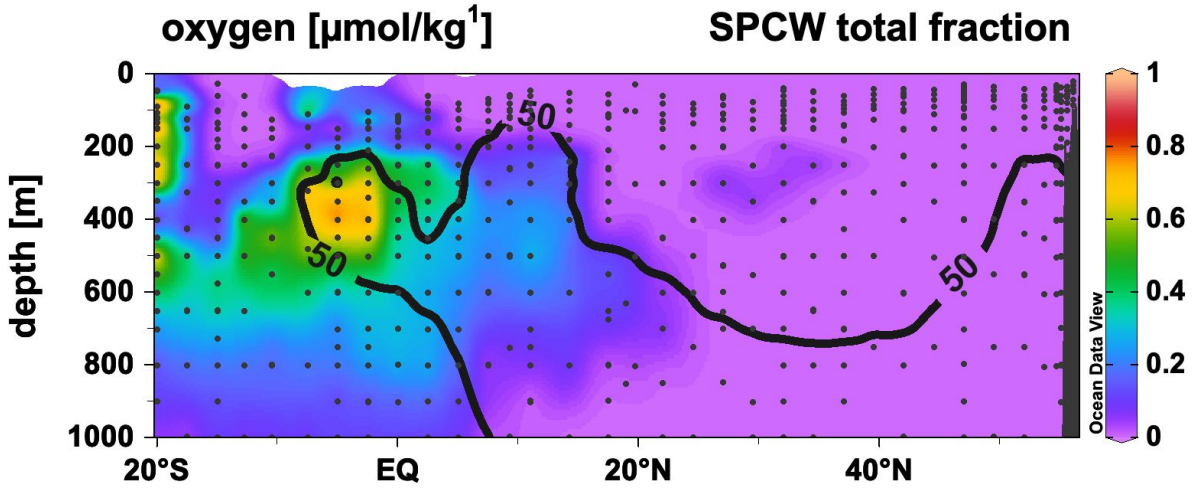


Figure S17. ESSW carries the signal of the Eastern Tropical North Pacific oxygen deficient zone as shown by the dissolved oxygen contour of 50 μmol kg^{-1}.

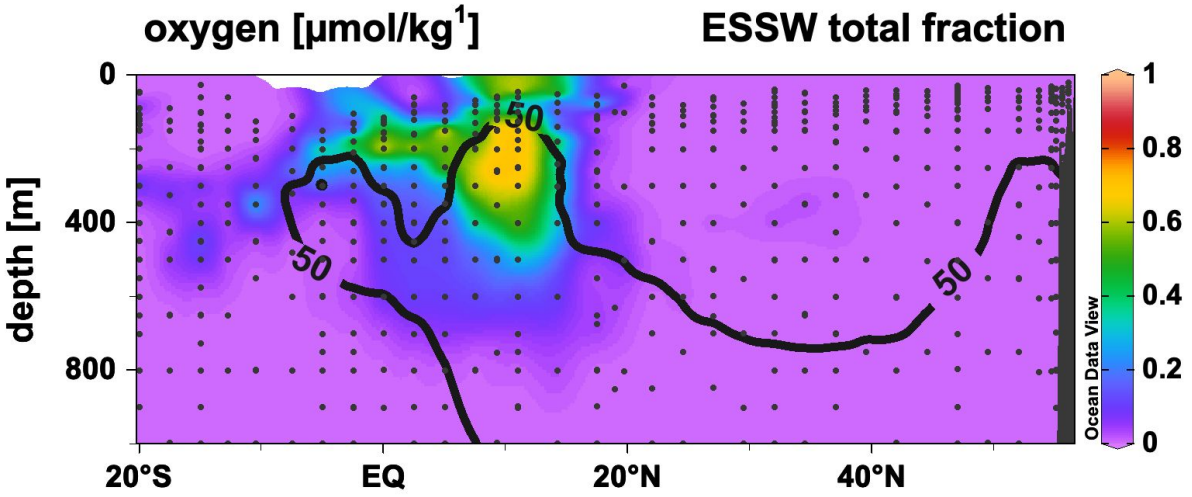


Figure S18. Residuals for the OMP using the Bering Sea endmember a) Conservative temperature (°C) b) absolute salinity, c) oxygen (μmol kg^{-1}), d) silicate (μmol kg^{-1}), e) nitrate (μmol kg^{-1}), and f) phosphate (μmol kg^{-1}).

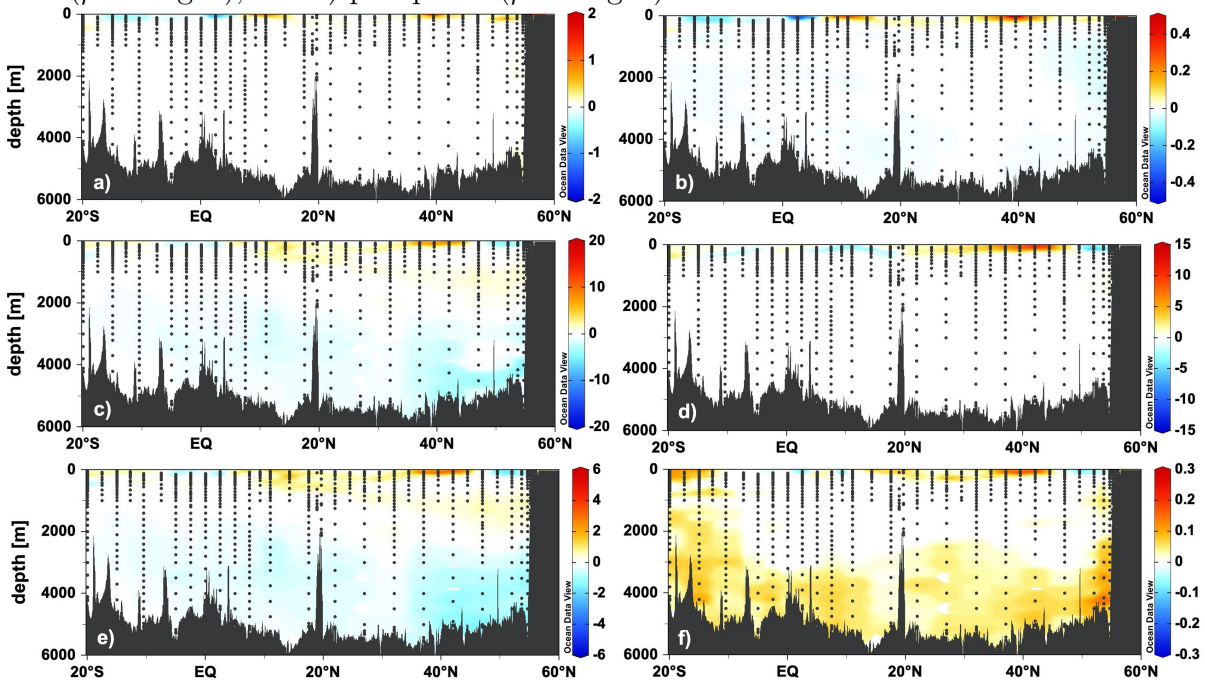


Figure S19. The water mass fractions of a) UCDW b) LCDW, c) AABW, and d) the Bering Sea endmember used in the place of PDW.

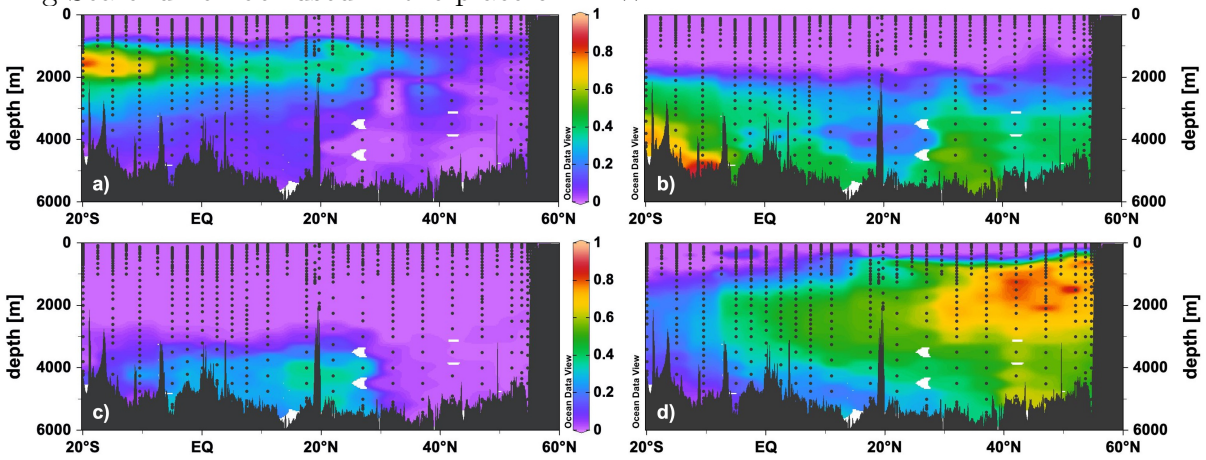


Table S1. Water mass subtype definitions. Each line defines the properties of a water mass subtype (archetype). Subtypes are numbered beginning with '0'.

Water mass	Temperature (°C)	Salinity (S)	Silicate (μmol kg^{-1})	Nitrate (μmol kg^{-1})	Phosphate (μmol kg^{-1})	Oxygen (μmol kg^{-1})
AABW 0	0.07	34.87	128.23	32.46	2.26	216.23
AAIW 0	5.04	34.39	11.51	23.99	1.67	272.31
AAIW 1	4.47	34.40	20.43	28.54	1.99	231.45
ENPCW 0	15.99	34.80	3.54	3.51	0.32	213.72
ENPCW 1	8.55	34.20	28.66	21.55	1.57	177.51
EqIW 0	5.46	34.73	72.17	42.52	3.07	42.82
EqIW 1	8.74	34.84	44.15	38.85	2.86	8.04
LCDW 0	1.55	34.90	93.37	31.50	2.18	190.11
LCDW 1	0.77	34.86	119.83	32.43	2.26	199.11
NPIW 0	9.41	34.34	32.63	19.84	1.41	177.00
NPIW 1	7.15	34.02	25.26	15.05	1.19	268.40
NPIW 2	4.54	34.12	77.91	34.76	2.52	114.14
PDW 0	3.74	34.43	117.45	43.87	3.16	11.38
PDW 1	2.16	34.71	167.51	44.57	3.15	23.94
PDW 2	1.20	34.87	171.60	37.10	2.59	136.80
PSUW 0	6.13	32.71	19.72	11.96	1.21	324.84
PSUW 1	4.43	33.60	50.07	29.99	2.28	178.86
SPCW 0	18.62	35.68	0.76	1.04	0.18	206.65
SPCW 1	8.40	34.62	6.47	19.98	1.43	201.51
UCDW 0	2.50	34.73	83.32	34.11	2.39	168.79

Table S2. Comparison of our AAIW endmember properties with those from SAMW endmembers used by Holte, Talley, Chereskin, and Sloyan (2013). Note Holte et al. (2013) report potential temperature and practical salinity while we report conservative temperature and absolute salinity.

Water mass	temperature (°C)	salinity (S)	Nitrate (μmol kg^{-1})	Phosphate (μmol kg^{-1})	Oxygen (μmol kg^{-1})	Silicate (μmol kg^{-1})
AAIW 0	5.04	34.39	23.99	1.67	272.31	11.51
AAIW 1	4.47	34.40	28.54	1.99	231.45	20.43
SAMW (27.0–27.1 σ_0)	5.30	34.23	25.0	1.7	255	10.7
SAMW (27.1–27.2 σ_0)	4.48	34.26	29.1	2.0	226	21.7

References

Cutter, G. A., Casciotti, K. L., & Lam, P. J. (2018). US GEOTRACES Pacific meridional transect–GP15 cruise report.

GEOTRACES. (n.d.). *GEOTRACES quality flag policy.* https://www.geotraces.org/geotraces-quality-flag-policy/.

Holte, J. W., Talley, L. D., Chereskin, T. K., & Sloyan, B. M. (2013). Subantarctic mode water in the southeast Pacific: Effect of exchange across the Subantarctic Front. *Journal of Geophysical Research: Oceans*, *118*(4), 2052–2066.

Jenkins, W., Smethie Jr, W., Boyle, E., & Cutter, G. (2015). Water mass analysis for the US GEOTRACES (GA03) North Atlantic sections. *Deep Sea Research Part II: Topical Studies in Oceanography*, *116*, 6–20.

Peters, B. D., Jenkins, W. J., Swift, J. H., German, C. R., Moffett, J. W., Cutter, G. A., ... Casciotti, K. L. (2018). Water mass analysis of the 2013 US GEOTRACES eastern Pacific zonal transect (GP16). *Marine Chemistry*, *201*, 6–19.

SeaDataNet. (2010, May). Data quality control procedures (2.0 ed.) [Computer software manual].

Talley, L. D. (2011). *Descriptive physical oceanography: An introduction.* Academic Press.