Supplementary Material

Here we present the extra analyses showing that our results are robust to various potential biases.

Seasonal and Diel Cycles Effects Many – mostly pluricellular – plankton taxa conduct diel vertical migrations (DVM) (Lampert, 1989). Yet, our analysis showed no significant effect of this migration on plankton community composition (i.e., relative concentrations) (Table S3), in line with previous findings (Soviadan et al., 2022). To further investigate the circadian effect separately from the seasonal effect, we conducted an additional test on pairs of profiles performed both during day and night at the same location (geographic distance < 2 km, time distance < 24h). For the 172 pairs of such profiles existing in our dataset, we compared raw concentrations in the epipelagic layer at day and night for the four most abundant taxa (*Trichodesmium*, Copepoda, Phaeodaria and Acantharea), using a paired a Wilcoxon-Mann-Whitney test after removing double zeros. This revealed statistical differences for all four tested taxa (Figure S6).

Both phytoplankton and zooplankton concentrations also vary seasonally: spring (and possibly autumn) blooms cause an increase in productivity and plankton concentration (Behrenfeld & Boss, 2014). But plankton may also bloom outside seasonal blooms, due to favourable conditions following water mass displacements (McGillicuddy et al., 2007). These sudden events, restricted spatially and temporally, are called intermittent blooms. For example, *Trichodesmium* can bloom locally in tropical and subtropical oceans (Westberry & Siegel, 2006). Colonies formed during these events can be detected by the UVP5. However, although seasonality affects absolute concentrations, our results suggested a negligible effect of season on community composition.

Briefly, both diel and seasonal effects were detected on absolute concentrations. Their non-significance in our analysis was therefore due to Hellinger's transformation, focusing our analyses on relative rather than absolute concentrations (Legendre & Legendre, 2012). With such focus on community composition and at the broad taxonomic level studied, the large-scale geographical effect dominated over seasonal and diel cycles.

Sampling Effort Heterogeneity UVP5 profiles were distributed unevenly: some areas were sampled intensively (California Current, Peruvian upwelling, Mediterranean Sea), others were rarely visited (Indian Ocean, Southern Ocean, Figure S1). Moreover, sampling was heterogeneous in time too: high latitudes were not visited during winter months (Figure S7).

To make sure that our results are not solely representative of oversampled areas, we conducted our analyses on a subsample of our data. Focusing on the epipelagic layer, variograms computed on the concentrations of Copepoda, *Trichodesmium* and Collodaria showed a scale of autocorrelation around 1000 km. Thus, a maximum of 20 profiles were selected in squares of 10° by 10° (~1000 km × 1000 km), for a total of 1388 selected profiles. These 1388 profiles were used to perform a subset PCA and build a factorial space in which all 2517 profiles from the epipelagic layer were projected. Projections on PC1 and PC2 of all 2517 profiles were extracted and compared to projections obtained from the PCA performed with all profiles. This resulted in good correlations (PC1: $R^2 = 0.97$, $p < 0.001$; PC2: $R^2 = 0.90$, $p < 0.001$), showing that our analyses are robust to down-sampling.

Furthermore, our analysis does not explicitly consider location or date, only each sample's community and environmental conditions; so the relevant question is: does UVP sampling cover environmental conditions representative of global scale variance? For this, we compared conditions distribution at UVP samples' locations to the same variables distribution at global scale. Of course, simultaneous worldwide *in situ* observations are not available. Instead, we used annual climatologies on a 1° grid from World Ocean Atlas (WOA) (Boyer et al., 2018) for important water characteristics: temperature, salinity, and oxygen. We first checked that those climatologies were representative of the *in situ* conditions at locations sampled by UVP5, over the epipelagic and mesopelagic layers previously defined; this was the case since correlations were good (all $R^2 > 0.84$, except for AOU in the epipelagic: $R^2 = 0.35$, Figure S8). Then, we compared each variable distributions from the WOA data at UVP5 profiles' locations vs. worldwide (Figure S9), for two depth layers (0 - 200 m; 200 - 500 m), since the above dynamic boundary could not be computed from WOA data. Distributions were similar, showing UVP samples covered diverse enough environmental conditions, representative of worldwide oceans.

Figures

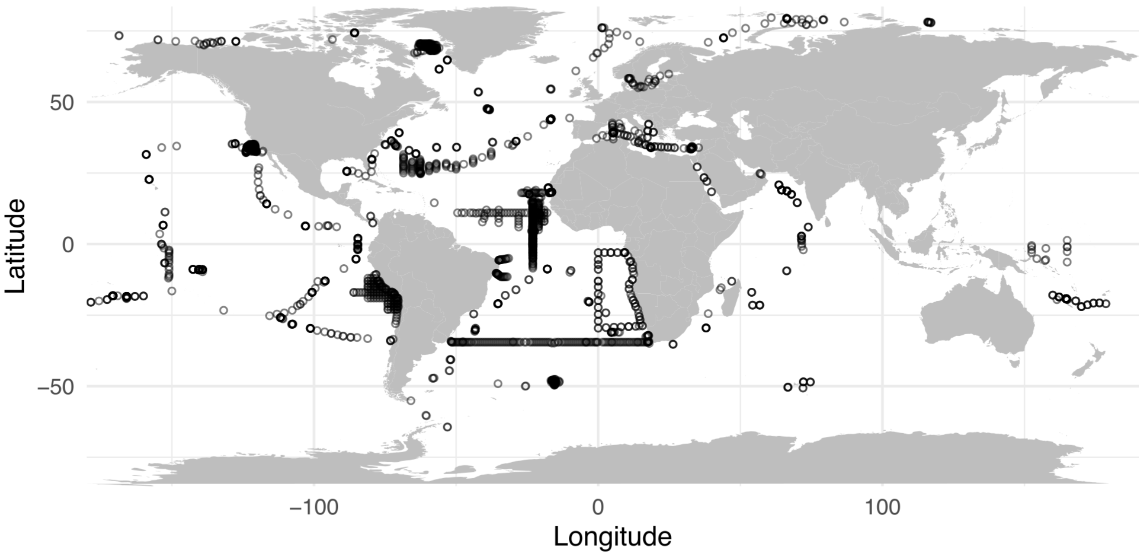

Figure S1: World map of included stations (whether in the epipelagic or mesopelagic layer).

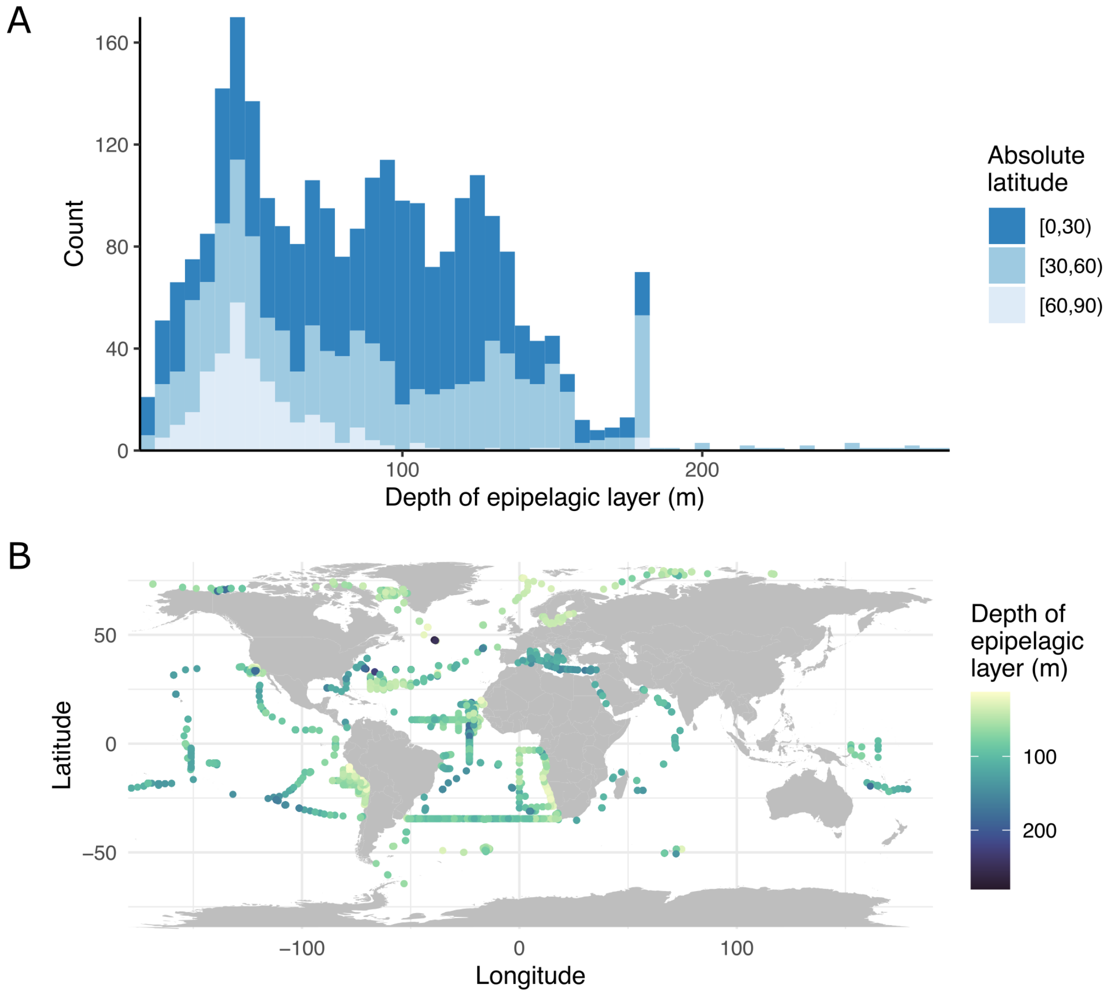

Figure S2: Depth of the dynamic epi-mesopelagic boundary, computed as the deepest value among the mixed layer depth and the euphotic depth. (A) Histogram of the epipelagic layer depth per 30° of absolute latitude bands. The peak at 180 m highlights cases of euphotic depth at 180 m and shallower mixed layer depth. (B) World map of the epipelagic layer depth.

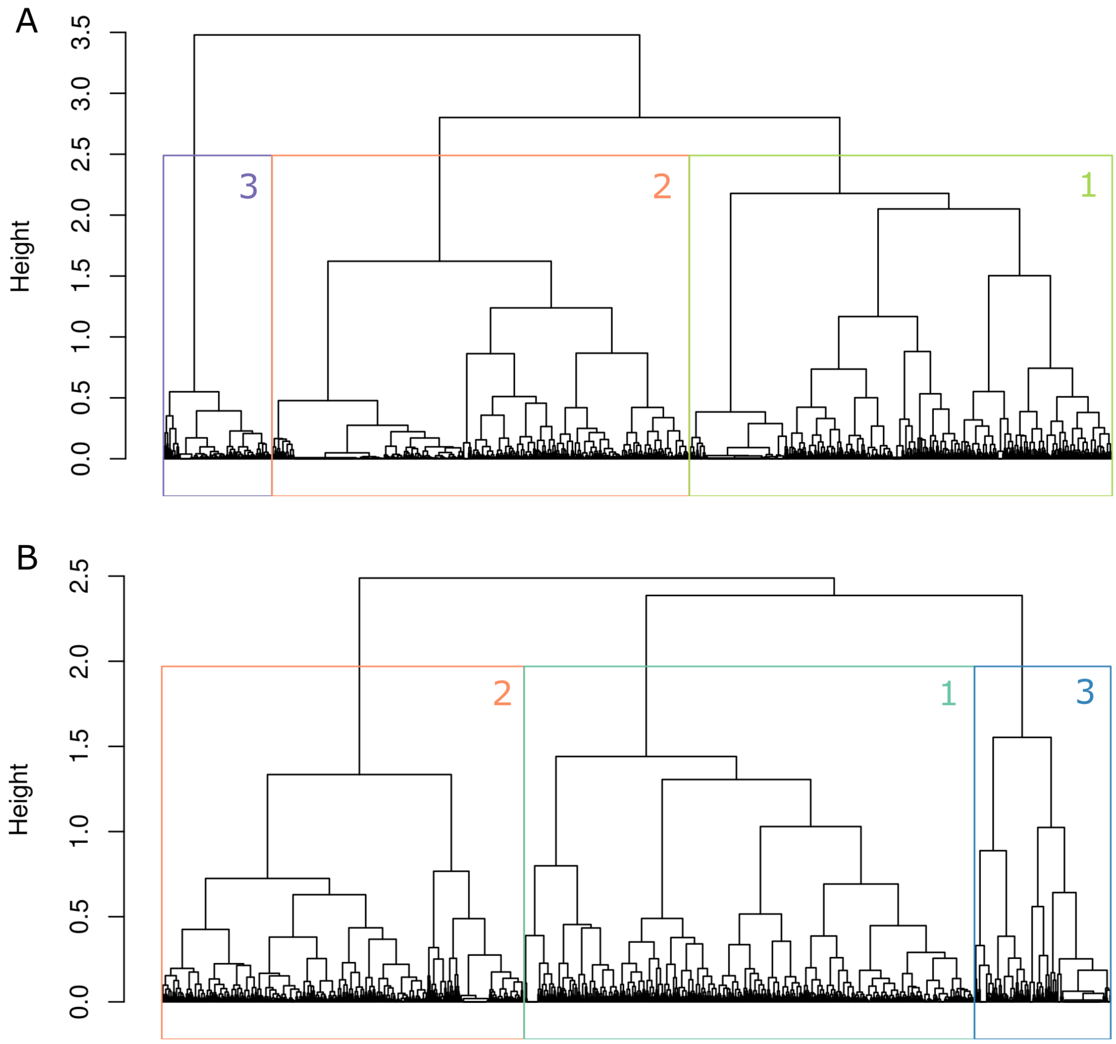

Figure S3: HAC dendrograms based on the first five principal components of profiles projection in the Hellinger-transformed plankton PCA data, for (A) epipelagic and (B) mesopelagic layers. Generated clusters are shown in the same colours and numbers as they appear on figures 3, 4 and 5.

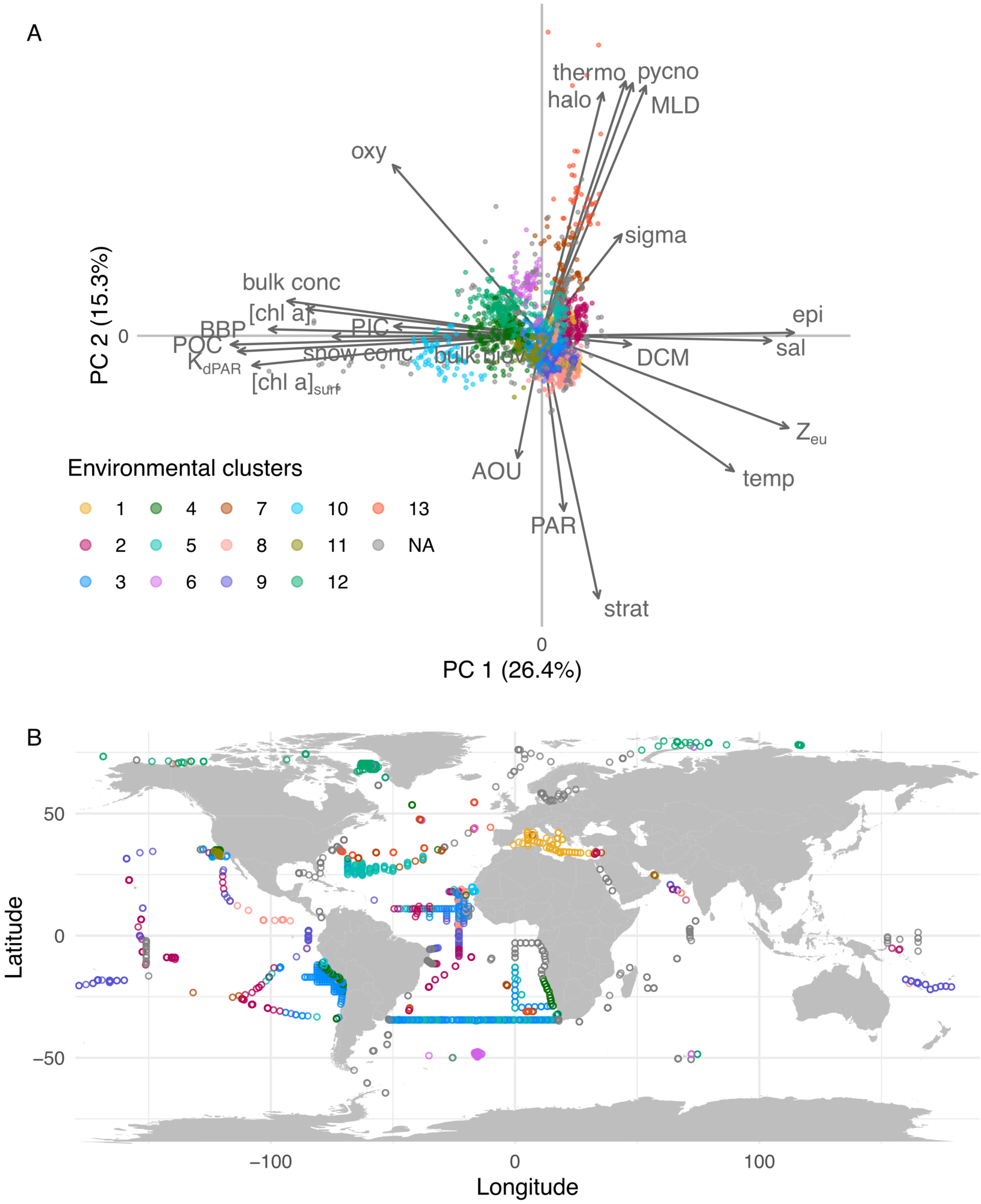

Figure S4: Local environment clustering for the epipelagic layer. (A) PCA performed on environmental variables, illustrated by a biplot in scaling 2. Points represent profiles and are coloured according to the cluster defined by the *k*-means algorithm. NA represents profiles that could not be associated wuth a cluster with more than 25 profiles. (B) Map of epipelagic profiles, coloured as in A.

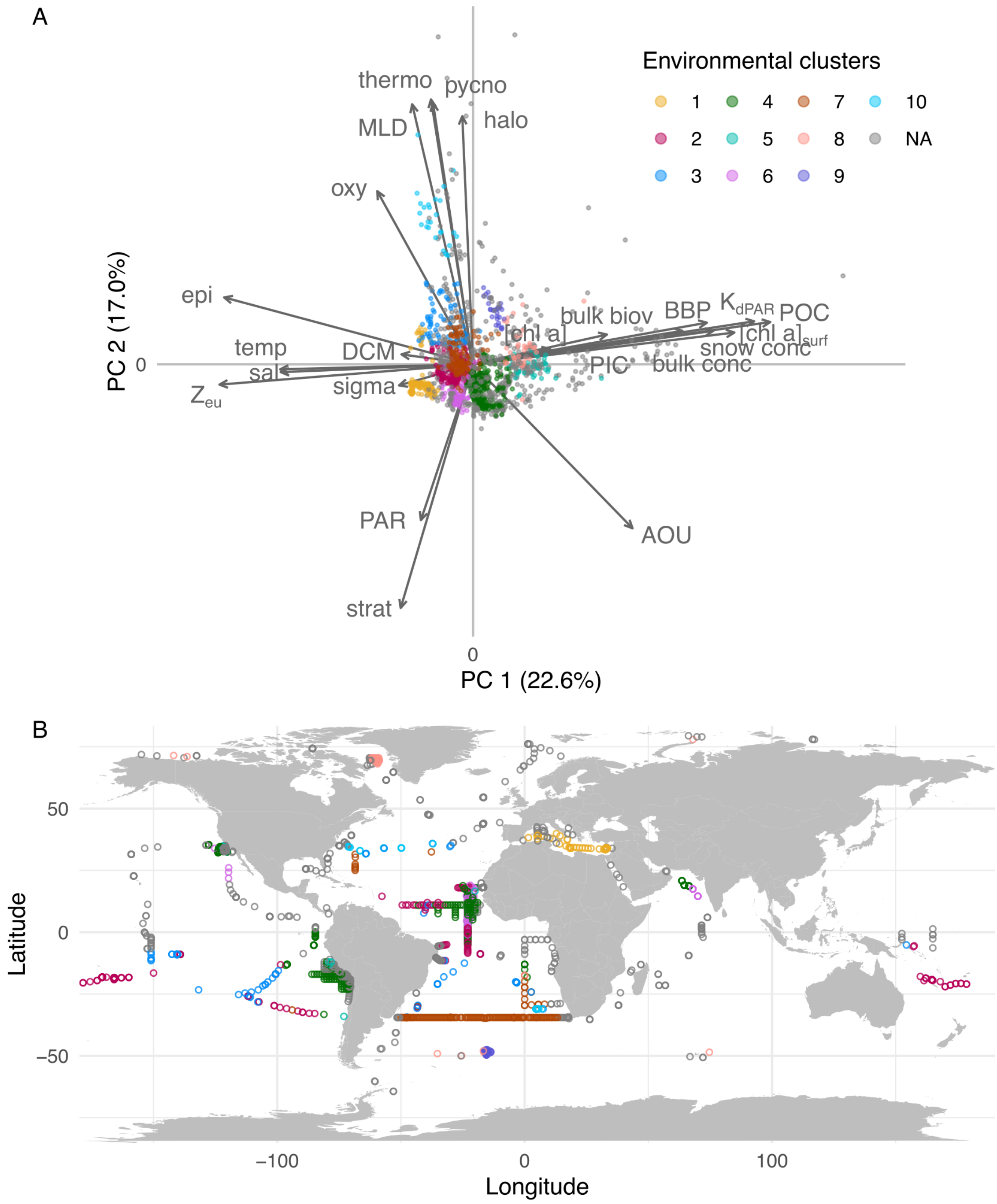

Figure S5: Local environment clustering for the mesopelagic layer. (A) PCA performed on environmental variables, illustrated by a biplot in scaling 2. Points represent profiles and are coloured according to the cluster defined by the *k*-means algorithm. NA represents profiles that could not be associated with a cluster with more than 25 profiles. (B) Map of mesopelagic profiles, coloured as in A.

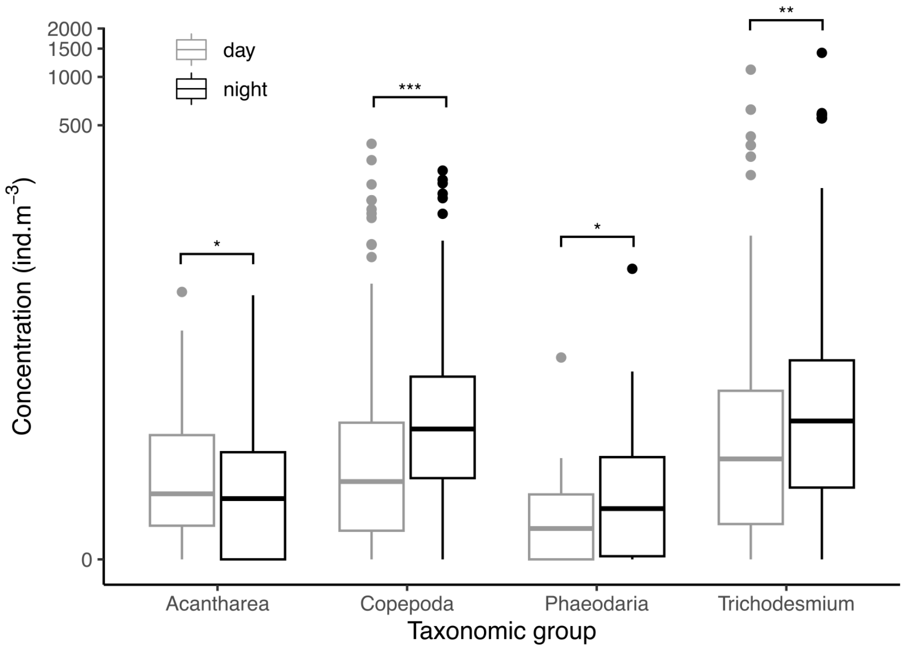

Figure S6: Average epipelagic concentration of the four most abundant taxa in 172 pairs of day/night stations after removing double zeros (i.e. pairs of stations where concentrations are null at day and night for a given taxon). Stations were paired according to both geographical distance (< 2 km) and time (< 24h). Note that the Y axis is log-transformed. Differences were tested with a paired Wilcoxon-Mann-Whitney test. * = 0.05, ** = 0.01, *** = 0.001.

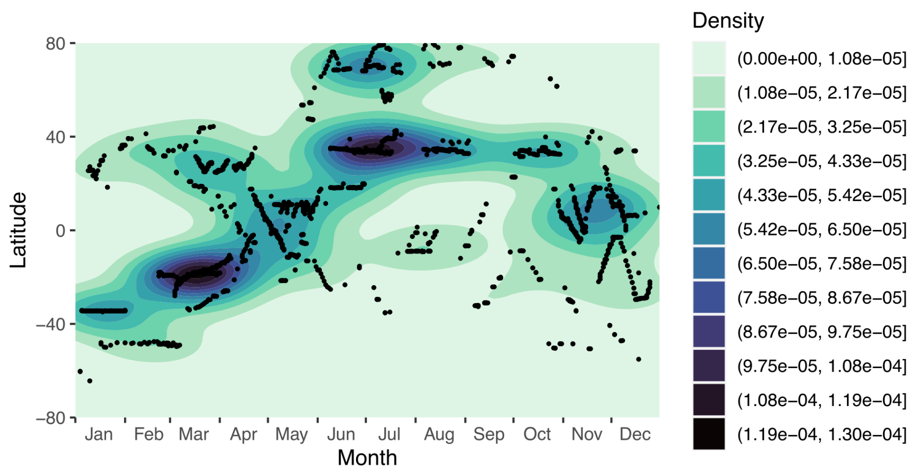

Figure S7: Time versus latitude Hovmöller diagram of sampled stations.

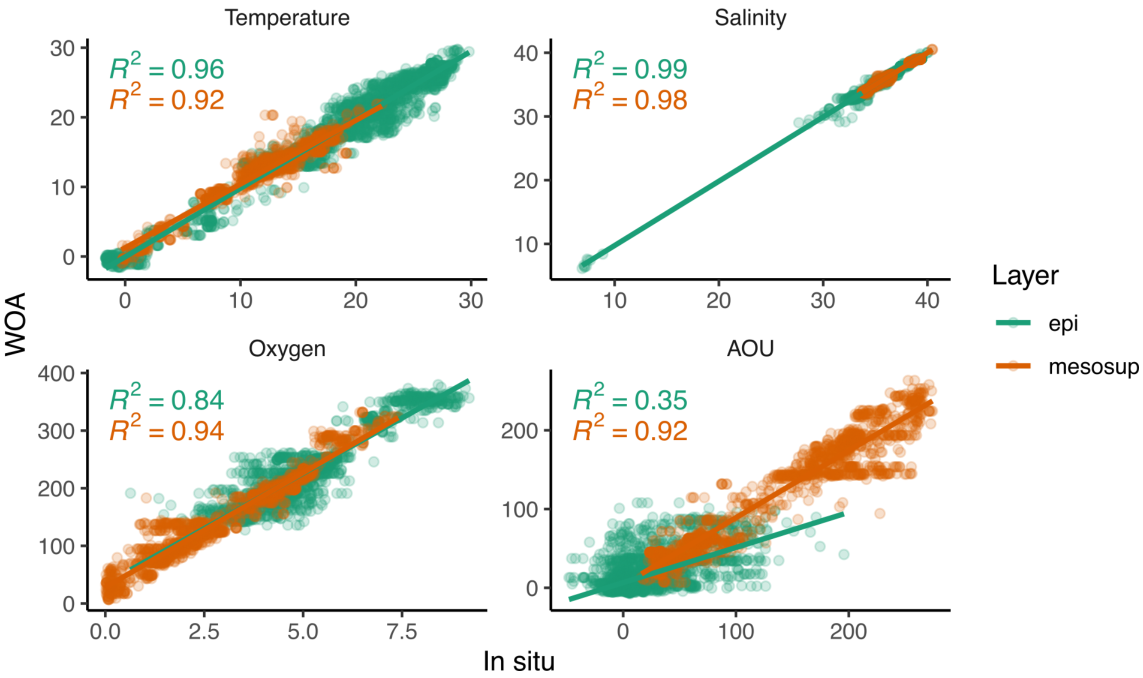

Figure S8: Correlation between *in situ* and annual WOA data at UVP5 profiles locations in the epipelagic and mesopelagic layers.

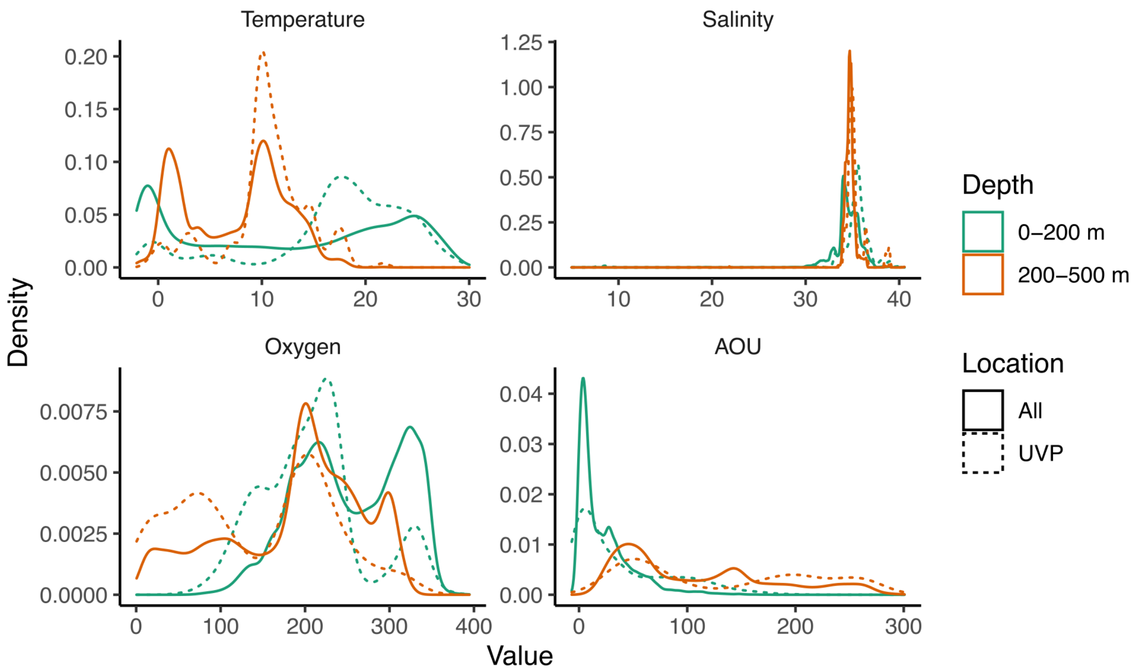

Figure S9: Distribution of annual WOA data all over the globe and at UVP5 profiles locations.

Tables

Table S1: List of oceanographic campaigns included in the study.

Campaign	Year	Nb profiles	UVP5
BOUM	2008	177	sd
CASSIOPEE	2015	13	sd
CCELTER 2008	2008	73	sd
CCELTER 2011	2011	56	zd
CCELTER 2012	2012	59	sd
CCELTER 2014	2014	60	sd
CCELTER 2017	2017	68	hd
DEWEX	2013	1	sd
MSM22	2012	101	sd
MSM23	2012	64	sd
M105	2014	8	sd
M106	2014	114	sd
M107	2014	71	sd
PS88b	2014	36	sd
M116	2015	74	sd
M121	2015	84	sd
M135	2017	138	sd
GreenEdge 2016	2016	121	hd
MSM060	2017	126	hd
IPS Amundsen 2018	2018	6	sd
JERICO 2017	2017	24	sd
KEOPS	2011	13	zd
LOHAFEX	2009	55	sd
MALINA	2009	16	sd
MooseGE[1]	2015	3	sd
NAAMES02	2016	21	hd
OUTPACE	2015	193	sd
P16N	2015	14	sd
Sargasso	2014	84	sd
SOMBA	2014	6	sd
Tara Oceans	2009-2013	643	sd

[1]https://doi.org/10.18142/235

Table S2: Definition of productive (1) and non-productive (0) seasons based on latitude and month.

Latitude band	Month											
	J	F	M	A	M	J	J	A	S	O	N	D
90°N - 66.5°N	0	0	0	0	0	1	1	1	0	0	0	0
66.5°N - 23.5°N	0	0	1	1	1	0	0	0	1	1	0	0
23.5°N - 23.5°S	0	0	0	0	0	0	0	0	0	0	0	0
23.5°S - 66.5°S	0	0	1	1	0	0	0	0	1	1	1	0
66.5°S - 90°S	1	1	0	0	0	0	0	0	0	0	0	1

This model is based on light intensity and nutrients availability. In polar regions, light availability is often limited (namely in winter) but becomes sufficient after the summer ice breakup, allowing productivity. In mid-latitudes, both light and nutrients become available in spring and autumn, generating phytoplankton blooms. In tropical regions, productivity is limited all year by nutrients and remains low.

Table S3: Variance in plankton community composition explained by diel and seasonal cycles computed from RDA. n = number of profiles included in each modality (day / night profiles for diel cycle; non-productive / productive season for seasonal cycle.) All p-value < 0.001.

Regionalisation	Epipelagic		Mesopelagic	
	n	R^2	n	R^2
Diel	1595 / 922	1.1%	1088 / 659	0.9%
Seasonal	1925 / 592	1.3%	-	-

References

Behrenfeld, M. J., & Boss, E. S. (2014). Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Annual Review of Marine Science*, *6*(1), 167–194.

Boyer, T. P., Garcia, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., Weathers, K. A., Baranova, O. K., Paver, C. R., Seidov, D., & Smolyar, I. V. (2018). World Ocean Atlas 2018. NOAA National Centers for Environmental Information. Dataset.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Functional Ecology*, *3*(1), 21–27.

Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.

McGillicuddy, D. J., Anderson, L. A., Bates, N. R., Bibby, T., Buesseler, K. O., Carlson, C. A., Davis, C. S., Ewart, C., Falkowski, P. G., Goldthwait, S. A., Hansell, D. A., Jenkins, W. J., Johnson, R., Kosnyrev, V. K., Ledwell, J. R., Li, Q. P., Siegel, D. A., & Steinberg, D. K. (2007). Eddy/Wind Interactions Stimulate Extraordinary Mid-Ocean Plankton Blooms. *Science*, *316*(5827), 1021–1026.

Soviadan, Y. D., Benedetti, F., Brandão, M. C., Ayata, S.-D., Irisson, J.-O., Jamet, J. L., Kiko, R., Lombard, F., Gnandi, K., & Stemmann, L. (2022). Patterns of mesozooplankton community composition and vertical fluxes in the global ocean. *Progress in Oceanography*, *200*, 102717.

Westberry, T. K., & Siegel, D. A. (2006). Spatial and temporal distribution of Trichodesmium blooms in the world's oceans. *Global Biogeochemical Cycles*, *20*(4).