Geophysical Research Letters

RESEARCH LETTER
10.1029/2026GL122225

Assessing Zooplankton Communities' Influence on Particulate Organic Carbon Concentration

Thelma Panaïotis[1] **and B. B. Cael**[1,2]

[1]National Oceanography Centre, Southampton, UK, [2]University of Chicago, Chicago, IL, USA

Key Points:

- Three zooplankton community metrics (evenness, mean gray level, abundance) explain ~50% of variance in small particulate organic carbon (POC) concentration gradients
- Taxonomic evenness is the strongest predictor: communities dominated by few groups show stronger small POC attenuation with depth
- Identified relationships motivate testable hypotheses about zooplankton-mediated particle processing and can inform ocean ecosystem models

Correspondence to:
B. B. Cael,
bbcael@uchicago.edu

Citation:
Panaïotis, T., & Cael, B. B. (2026). Assessing zooplankton communities' influence on particulate organic carbon concentration. *Geophysical Research Letters*, *53*, e2026GL122225. https://doi.org/10.1029/2026GL122225

Received 5 FEB 2026
Accepted 18 JUN 2026

Abstract Zooplankton play a key role in setting the particulate organic carbon (POC) distribution in the ocean, but this role remains poorly quantified on large scales due to the complexity of zooplankton ecosystems and the sparsity and variability of observations. We address this by applying boosted regression trees to a global in situ zooplankton image database from the Underwater Vision Profiler 5 and an Argo float and satellite-derived POC product. We compute taxonomic, morphological, and trophic community metrics, then identify three metrics—abundance, mean gray level, taxonomic evenness—that explain almost half of the spatial variability in the vertical concentration gradients of small POC (<100 μm). Partial dependence analysis and the consistency of the identified relationships with known zooplankton-mediated particle-processing pathways suggest mechanistically interpretable ecological linkages. These results advance quantification of zooplankton communities' influence on ocean carbon cycling and indicate carbon-cycle-relevant zooplankton community properties for improving ocean biogeochemical models.

Plain Language Summary Zooplankton influence the distribution of organic carbon in the ocean by consuming particles, fragmenting aggregates, and excreting fecal pellets. Zooplankton communities are diverse, with different organisms playing distinct biogeochemical and ecological roles. Due to the complexity of zooplankton ecosystems and the sparse and variable nature of marine ecological and biogeochemical measurements, how zooplankton affect organic carbon distribution on large scales remains poorly characterized. This study addresses this gap by applying machine learning to a global data set of particulate organic carbon (POC) and zooplankton imaging products. We identify key zooplankton community metrics (community balance, transparency, and abundance) that best explain variability in the vertical concentration gradients of small POC (<100 μm), thereby quantifying the role of zooplankton communities in setting the POC distribution, and with potential application to improving ocean biogeochemical models.

1. Introduction

The ocean's biological carbon pump modulates Earth's climate by transporting biologically fixed carbon from the upper ocean into the ocean interior (Boyd et al., 2019; Siegel et al., 2023; Volk & Hoffert, 1985). Zooplankton produce fecal pellets, fragment and ingest detrital aggregates and fecal pellets, and graze on phytoplankton and other zooplankton, thus playing multiple roles in shaping vertical gradients in particulate organic carbon (POC, [g/m^3]) concentration (Briggs et al., 2020; Iversen, 2023; Steinberg & Landry, 2017; Steinberg et al., 2008; Turner, 2015). Here we investigate the vertical concentration gradient of small POC (<100 μm), which accounts for most of the mesopelagic POC stock (Baker et al., 2017). The vertical distribution of small POC reflects the integrated effects of particle fragmentation, consumption, repackaging, and microbial remineralization. Zooplankton community structure is therefore expected to be especially relevant to the vertical distribution of this particle class. What is not yet understood, however, is which facets of zooplankton community structure are most strongly linked to the vertical distribution of small POC at the global scale.

Zooplankton community and POC concentration or flux measurements are challenging and resultingly sparse and highly variable (Cael et al., 2021; Nocera et al., 2025). This has hampered quantitative assessment of zooplankton communities' role in shaping POC gradients globally (Henson et al., 2022). To mitigate observational limitations, POC dynamics are often estimated via ocean biogeochemical models (Burd, 2024). Many large-scale biogeochemical models necessarily represent zooplankton and particle transformations in highly aggregated ways. This is particularly consequential for small POC, whose concentration profile integrates fragmentation of larger particles, particle consumption, fecal-pellet production and reprocessing, and redistribution among particle size classes. Recent developments have improved zooplankton and particle-size representation in ocean BGC models

(Di Matteo et al., 2025; Luo et al., 2022; Meyjes et al., 2024; Negrete-Garc'ia et al., 2022; Serra-Pompei et al., 2022). A challenge in ocean biogeochemistry is therefore to overcome data limitations to quantify not only the influence of zooplankton communities on attenuation globally, but also how this influence could be parameterized in ocean biogeochemical models.

Fortuitously, in recent years imaging techniques have provided an growing volume of zooplankton observations (Lombard et al., 2019), elucidating global patterns of diversity (Ibarbalz et al., 2019), biomass (Drago et al., 2022), and community structure (Lombard et al., 2024; Panaïotis et al., 2023). Simultaneously, machine learning (ML) techniques have provided means to automatically extract additional information, such as functional or morphological traits (Irisson et al., 2022; Orenstein et al., 2022). These advances are complemented by those in assimilation of autonomous float, remote sensing, and traditional shipboard measurements to generate robust estimates of biogeochemical variables of interest, such as POC (Sauzède et al., 2020). The combination of advances in in situ imaging, ML-based image processing to extract multi-faceted zooplankton community metrics (Acevedo-Trejos et al., 2022), and data assimilation present an opportunity to overcome historical data limitations in assessing the role of zooplankton on POC gradients at the global scale. In this manuscript we utilize these recent advances in ML and observational data products to identify a small set of key zooplankton community metrics influencing small POC gradients, and show that these metrics explain almost half the spatial variation in these gradients.

2. Materials and Methods

2.1. Data Sets

2.1.1. POC Concentration Product

For POC, we used a ML-derived product based on a combination of ARGO floats and satellite data (Sauzède et al., 2020). This product estimates POC concentration from particulate backscattering at 700 nm (bbp700), which primarily captures small particles (<100 μm) (Briggs et al., 2020). In the POC data set we use, POC is predicted using various biogeochemical and physical (e.g., reflectance, photosynthetically active radiation, sea level anomaly) predictors, with values provided on a 0.25° grid for 36 depth levels between the surface and 1,000 m.

For each pixel, we calculated the vertical POC concentration gradient as the equivalent of Martin's b for POC concentration (Lam et al., 2011) between the depth of maximum POC concentration (referred to as the "reference depth" z_0 [m]; the median z_0 is 45 m and the interquartile range is 30–65 m) and 1,000 m: $POC(z) = POC(z_0)(z/z_0)^{-b}$, where $POC(z)$ is the POC concentration at depth z, and b quantifies the steepness of the vertical gradient. Higher b values indicate a steeper decrease of POC with depth, reflecting biological consumption, fragmentation and remineralization in the mesopelagic. Here we focus on b rather than the surface POC concentration because we are interested in how zooplankton consume, produce, and repackage POC and hence affect its vertical distribution. We use a Martin's-b-like exponent as a compact measure of the steepness of the small POC concentration profile over the epipelagic and mesopelagic ocean. Our goal is therefore not to resolve depth-specific mechanisms at individual horizons, but rather to explain large-scale variation in the integrated attenuation pattern of small POC concentration. The spatial pattern of b is shown in Figure 1.

2.1.2. Zooplankton Imaging Data

For zooplankton community metrics, we used zooplankton imaging data from the global Underwater Vision Profiler 5 (UVP5) data set (>3,000 profiles) (Nocera et al., 2025). The UVP5 (Picheral et al., 2010) is a vertically profiling in situ imaging instrument, which captures images of planktonic organisms and marine snow over a nominal 0.6–50 mm size range. Objects have been homogeneously sorted into 28 taxonomic and/or morphological groups (Nocera et al., 2025) (observations of detrital material such as Cnidaria tentacles were discarded, leaving behind only zooplankton). As we focus on POC dynamics between the reference depth and 1,000 m, we only retain organisms within this depth range and vertically integrate each profile accordingly. We only use profiles sampling to at least 1,000 m depth. The locations and times of the observations in Nocera et al. (2025) are shown in Figure 1.

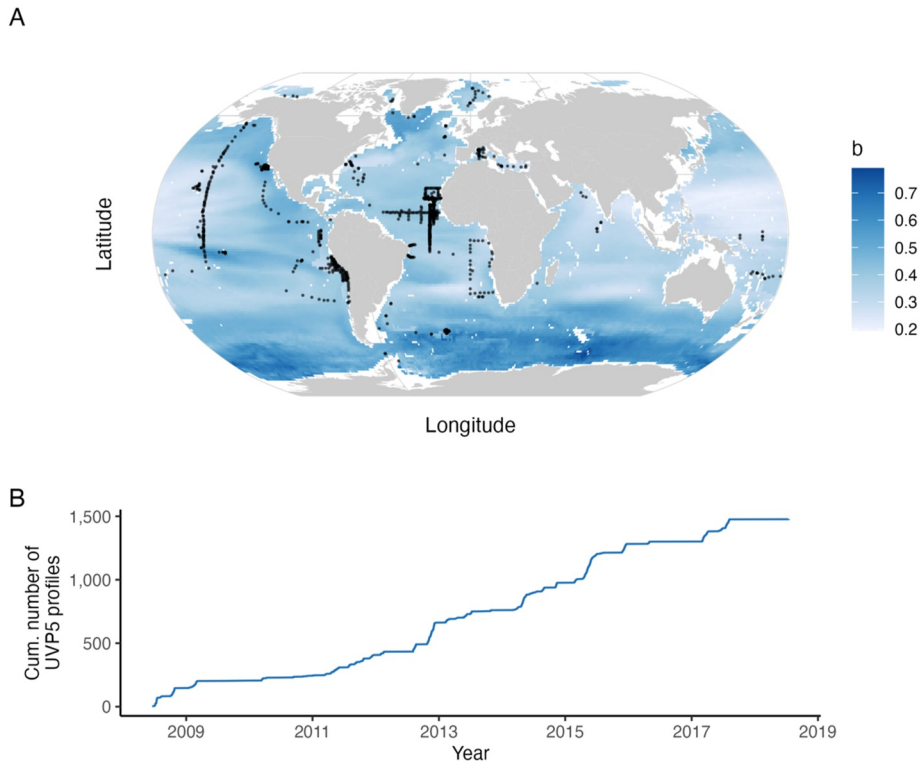

Figure 1. (a) Map of the particulate organic carbon concentration gradient exponent *b* calculated from Sauzède et al. (2020) with locations of zooplankton imaging data from Nocera et al. (2025) overlaid in black, after match-up on a 0.25° grid. Higher *b* values indicate steeper vertical gradients. Axes are longitude and latitude in degrees. (b) Cumulative distribution of the times these profiles were taken.

2.2. Calculating and Selecting Zooplankton Community Metrics

For each UVP5 profile, we computed 16 zooplankton community metrics (summarized in Figure 2), including taxonomic, morphological, and trophic measures. Total abundance and biovolume of planktonic organisms were computed as in Drago et al. (2022). Following Nocera et al. (2025, their Section 2.3), we aggregated these finer-grained taxa into five broader taxonomic groups based on feeding strategies and their ecology: Crustacea, Gelatinous, Mollusca, Rhizaria, and Other. "Other" denotes all zooplankton not classified into the other four groups, which make up a small (<4%) fraction of the total abundance. Taxonomic richness was computed as the number of groups present in the profile. Taxonomic diversity and evenness were computed according to Brillouin indices (Laxton, 1978), because these observations arise from finite sampled profiles rather than from an assumed randomly sampled regional pool.

Morphological community metrics were computed from the images as in Beck et al. (2023). Briefly, a total of 45 image features were extracted from each image, including various gray level statistics such as mean and range, shape statistics such as circularity and fractal dimension, size statistics such as area and Feret diameter, and symmetry statistics such as the contrast symmetry and thickness ratio (see Beck et al., 2023, Table S1 for the full list). These features were used to construct a "morphospace," which was reduced to four principal components that explain 69% of the total variance. These four components correlate strongly with size, mean gray level, elongation, and gray level heterogeneity features respectively (ρ = 0.96/0.95/0.99/0.88), and are labeled as such.

To compute morphological richness, diversity, and evenness, for each profile, we calculated the mean and standard deviation of the object projections on these morphospace principal components. We then applied *k*-means clustering to the objects' coordinates on the morphospace principal components, grouping them into ∼200 sets of morphologically similar organisms, which we found to be a pragmatic choice providing reasonably fine morphological partitioning without over-fragmenting the feature space (our conclusions are not sensitive to this choice). The morphological characteristics of each morph and their distribution across profiles were then used to

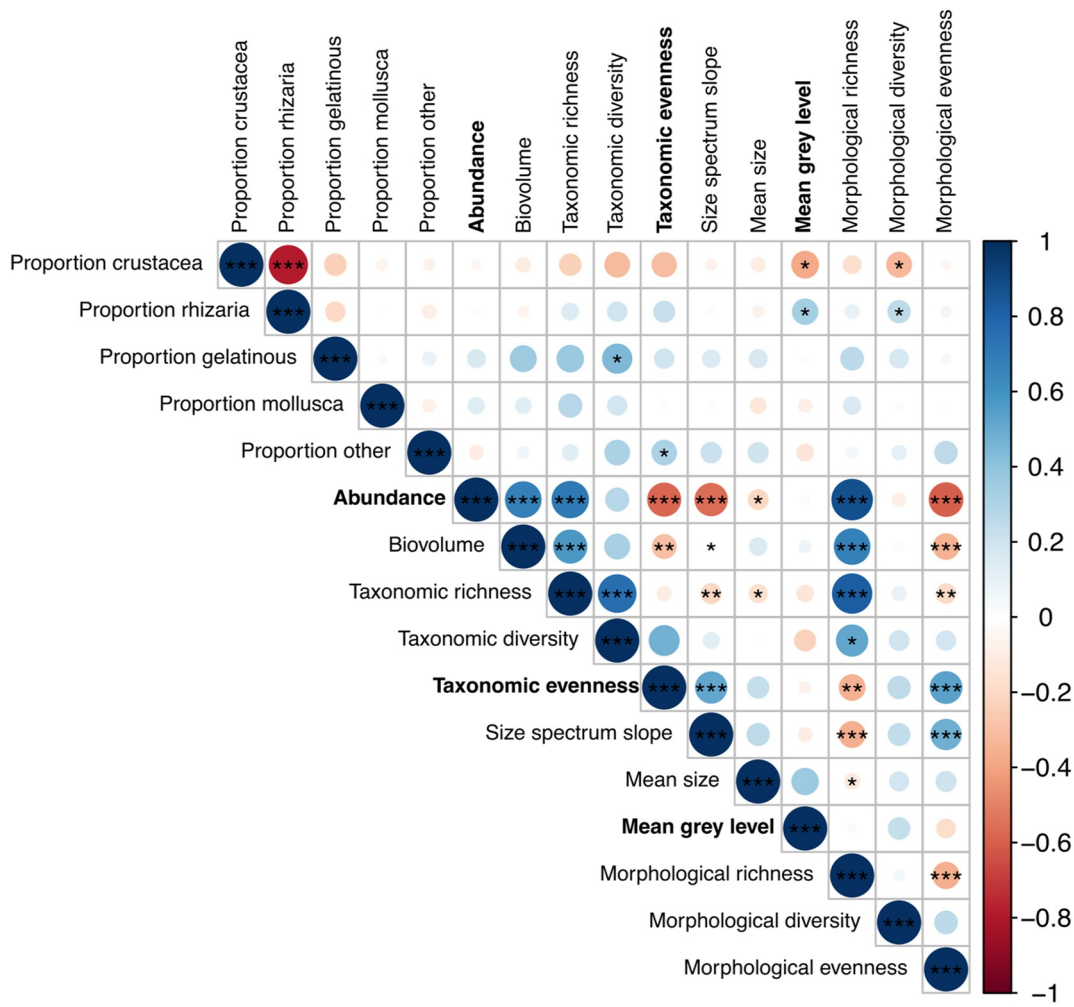

Figure 2. Correlogram of the zooplankton community metrics calculated here. Circle size and color both correspond to correlation strength; color also corresponds to the correlation sign; */**/*** correspond to significance at the 0.1/0.05/0.01 levels of these correlations. The correlogram shows that the 16 candidate community metrics described in Section 2.2 are strongly correlated, and hence that the predictor space can and should be substantially reduced, motivating the choice to use feature permutation to do so in Section 2.3. Bolded predictors are those that dominate predictors' explanatory power in Section 3.

compute morphological equivalents of functional community metrics (Beck et al., 2023; Magneville et al., 2022): morphological richness, diversity, and evenness. Including additional morphological or taxonomic community metrics (e.g., dispersion, divergence, originality, specialization, mean pairwise distance, and mean nearest-neighbor distance) did not affect our conclusions, as these metrics were strongly correlated.

The 28 taxonomic groups in our data set could not be easily assigned to distinct trophic levels. For example, both *Oithona* and *Euchaeta* are classified as copepods, yet one is herbivorous and the other carnivorous. Thus, we calculated the slope of the size spectrum of each profile as a coarse community-structure descriptor related to trophic organization (Lombard et al., 2024), because this slope is determined by the ratio of larger to smaller zooplankton, organisms tend to eat other organisms smaller than them rather than larger than them, and body-size structure may influence particle processing.

Community metrics for each profile were averaged onto a 0.25° grid to match the POC data, yielding 1,076 co-located points (Figure 1).

2.3. Relating POC Gradients to Zooplankton Community Metrics

To relate zooplankton community metrics to POC concentration gradients, we used Boosted Regression Trees (BRT), a ML algorithm that combines regression trees (binary splits that relate a response to predictors) with boosting (iteratively improving by combining many small trees, each working on the residuals of the previous one). Tree-based methods offer several advantages, such as flexibility with predictors (e.g., type, missing values, outliers), the ability to fit complex non-linear relationships, and natural handling of interactions between predictors. BRTs have been successfully used to relate plankton biomass (Drago et al., 2022; Liu et al., 2024) and particles (Denvil-Sommer et al., 2023; Panaïotis et al., 2025) to environmental predictors.

Model training and validation followed a nested cross-validation (CV) (Varma & Simon, 2006) procedure to simultaneously optimize hyperparameters and obtain robust performance estimate. We used 5-fold nested cross-validation (CV), with the outer CV evaluating model performance and the inner CV tuning hyperparameters (number of trees, learning rate, tree depth, and number of randomly selected predictors). Both inner and outer CV were stratified on quintiles of b to ensure a representative distribution across folds. To assess the effect of spatial autocorrelation, we also performed a nested CV using a spatial block CV (Roberts et al., 2017). This approach makes the prediction task more difficult, as spatially separated folds have more dissimilar distribution of the response variable, typically resulting in lower performance estimates. Because of our small sample size compared to what is required by spatial CV, this method did not produce reliable goodness-of-fit estimates, with a coefficient of variation ≥ 1 in the fraction of b variance explained by zooplankton community metrics (as compared to < 0.1 for the stratified CV). Therefore we do not discuss results from this training and validation approach in the next section. However we do note that spatial CV yielded the same ranking order of zooplankton community metrics and partial dependence relationships between b and these metrics, suggesting our central results are robust to spatial autocorrelation.

We used feature permutation (Breiman, 2001) to identify the most important predictors and partial dependence plots (Friedman, 2001) to examine the influence of community metrics on POC gradients. Models were fitted with the xgboost (Chen & Guestrin, 2016) engine, using the tidymodels framework (Kuhn & Wickham, 2020) in R 4.4.1 (R Core Team, 2024).

3. Results and Discussion

We find that zooplankton community metrics collectively explain 44±4% of variance (mean and standard deviation of R^2 across stratified CV folds) in small POC gradient strength (coefficient b), with an root mean square error (RMSE) of 0.0180±0.0003. Variable importance analysis revealed that three metrics dominate this explanatory power: taxonomic evenness, mean gray level, and abundance (Figure 3a). Including other zooplankton community metrics does not appreciably improve how much variation in b can be explained (RMSE decrease of $< 0.005 \pm 0.002$ when using just these predictors rather than the complete predictor set; Figure 3a), indicating these metrics capture the essential dimensions of zooplankton community influence on POC-gradient variation.

Abundance quantifies the total number of zooplankton in the water column, taxonomic evenness measures how balanced the representation of different taxonomic groups is, and mean gray level reflects an image-based opacity/transparency axis of organism appearance. In our data set, higher values tend to correspond to visually more transparent organisms, but this metric is not correlated with the relative abundance of gelatinous taxa; rather, higher mean gray levels correspond to communities with more Rhizaria and fewer Crustacea (Figure 2). Figure 4 illustrates quantitative and visual differences between high- and low-evenness and -mean-gray-level profiles. Taxonomic evenness is negatively related to b, whereas mean gray level and abundance are positively related (Figures 3b–3d). Note that higher b values correspond to steeper gradients.

The BRT analysis thus indicates that almost half of the variation in small POC concentration gradients can be explained by a few key metrics of zooplankton communities. Furthermore, the directions of the partial-dependence relationships are consistent with plausible zooplankton-mediated particle-processing mechanisms. Figure 3 shows steeper POC concentration gradients when zooplankton are more taxonomically uneven, have higher mean gray level, or are more abundant. The negative relationship between b and taxonomic evenness presents a plausible mechanistic hypothesis: taxonomic unevenness indicates dominance by one or a few taxonomic groups; in these circumstances, increased food demand by the dominant group(s) could lead to these

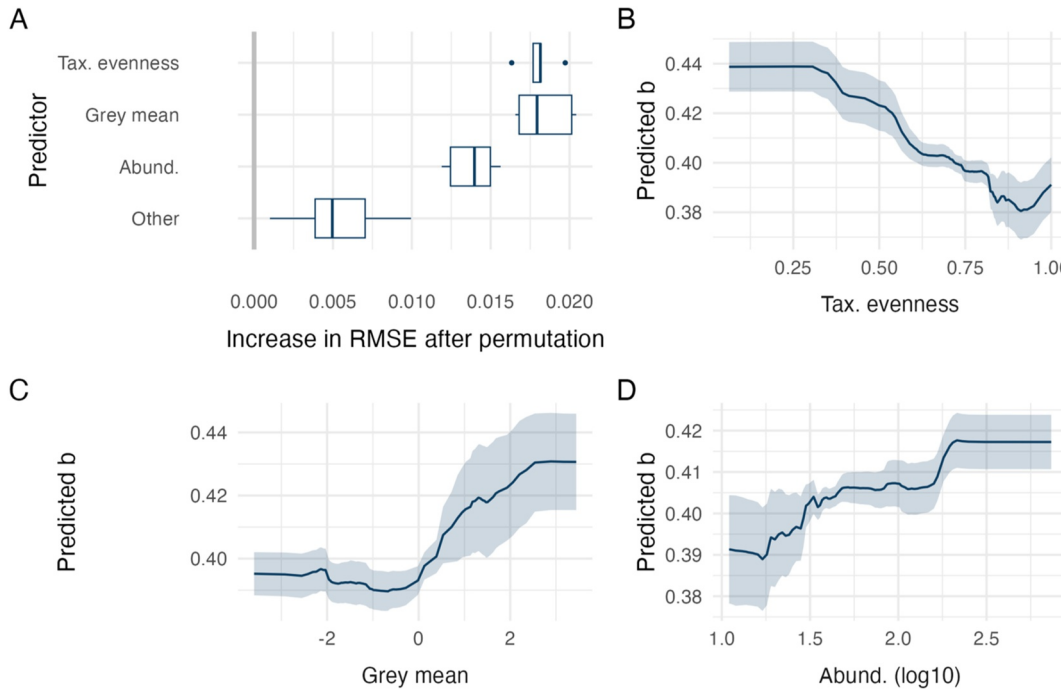

Figure 3. (a) Permutation variable importance, showing the increase in root mean square error when a predictor is removed, compared to the full model (gray vertical line). Bolded predictors are those that dominate predictors' explanatory power in Section 3. (b–d) Partial dependence plots for the top three predictors, showing mean (line) and standard deviation (ribbon) of predicted *b*, calculated across centered Ceteris paribus profiles. Tax. evenness = taxonomic evenness; Abund. = abundance. Note that abundance is $\log_{10}$-transformed in this representation.

groups scavenging for suboptimal food sources, including detrital aggregates and fecal pellets of other zooplankton. Taxonomic composition can indeed determine the fate of fecal pellets (Turner, 2015). More generally, evenness likely influences small POC gradients in multiple ways, including particle consumption and fecal-pellet production and reprocessing. For communities dominated by a few groups, particle consumption may be concentrated in fewer feeding modes, depleting some particle classes. For more even communities, a mixture of feeding, production, and repackaging pathways may redistribute particle processing more heterogeneously, reducing POC depletion rates as POC sinks and leading to weaker vertical concentration gradients.

The positive relationship between *b* and abundance also aligns with established understanding: higher zooplankton abundances result in greater particle encounter rates, and thus higher particle fragmentation and ingestion rates, leading to stronger POC depletion with depth (Cavan et al., 2017; Dilling & Alldredge, 2000).

The positive relationship between *b* and mean gray level is less straightforward to interpret mechanistically. Mean gray level is significantly correlated with the relative abundance of Rhizaria versus Crustacea (Figure 2). This suggests that zooplankton communities with more Rhizaria or less Crustacea are associated with more strongly attenuted POC concentration gradients. Because rhizarians can themselves contribute importantly to sinking particle flux, while crustacean-dominated pathways often involve fecal-pellet production and intensive particle reprocessing, the observed gray-level relationship may reflect a shift between communities more associated with direct particle export and communities more associated with zooplankton-mediated particle transformation (Durkin et al., 2022; Gutierrez-Rodriguez et al., 2019). Alternatively, mean gray level may correlate with unmeasured functional traits; further mechanistic investigation is needed to resolve this relationship.

Because many oceanographic variables covary geographically, broad-scale spatial structure can generate misleading ecological associations. We therefore show the large-scale latitudinal structure of *b* and the leading zooplankton predictors in Figure 5 for spatial context. The weak latitudinal structure of some of these variables are suggestive that the observed relationships are not reducible to a simple shared meridional gradient.

10.1029/2026GL122225

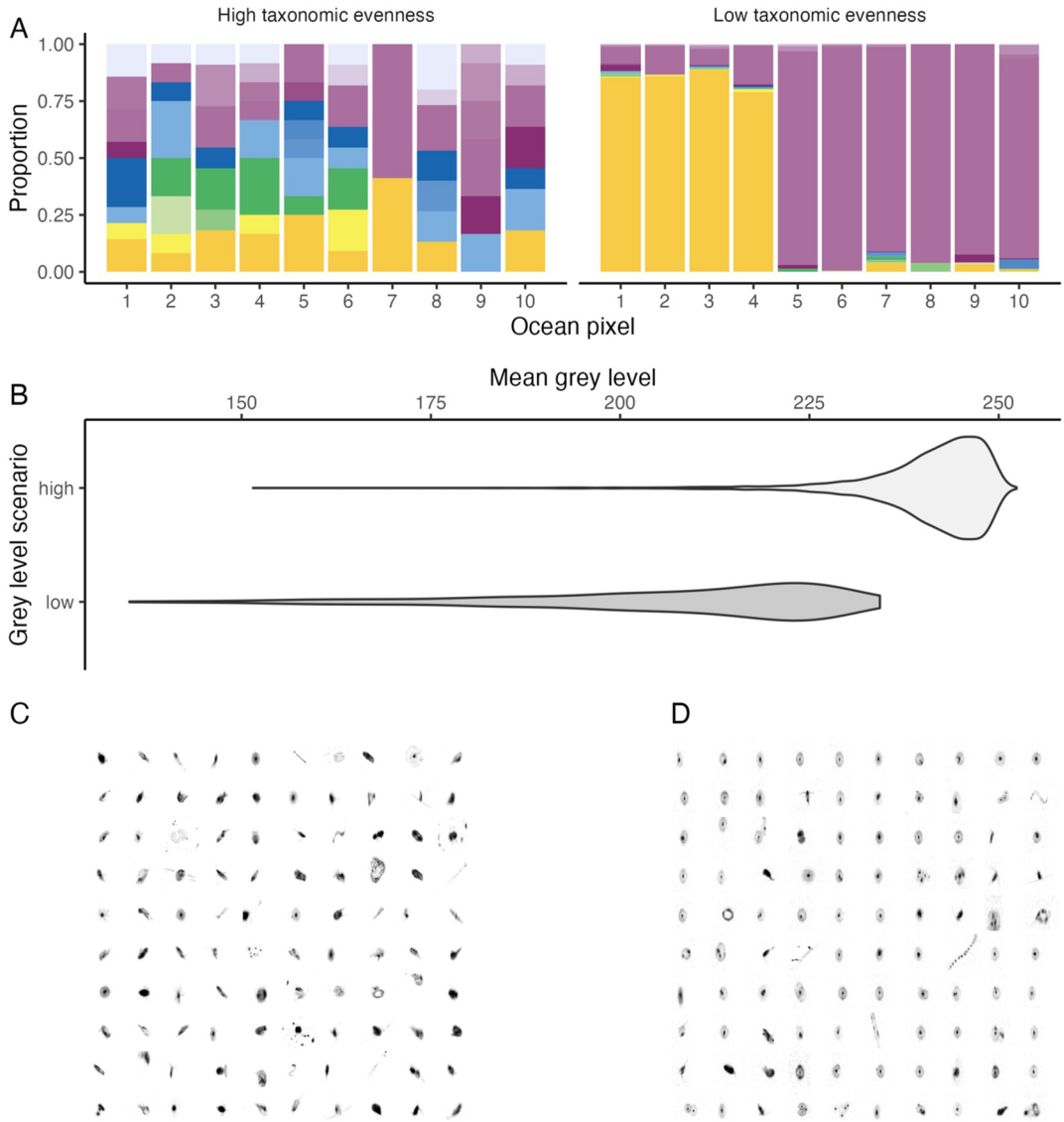

Figure 4. (a) Relative zooplankton abundances in 10 ocean pixels with high taxonomic evenness and 10 ocean pixels with low taxonomic evenness. Color legend for taxonomic groups is intentionally omitted to emphasize that evenness is independent of taxa identity. (b) Probability density distributions of objects' gray levels in 10 ocean pixels with high mean gray level and 10 ocean pixels with low mean gray level. (c, d) Mosaics of zooplankton images sampled from the same ocean pixels as in (b): (c) low mean gray level and high taxonomic evenness; (d) high mean gray level and low taxonomic evenness. Note that this figure is not intended to be representative of the full data set nor to illustrate differences in abundance.

The present study is observational and identifies statistical associations that present testable hypotheses for process-resolving analyses or targeted experiments. We note four primary limitations, each related to the data currently available for analyses such as the one presented here. First, the POC product analyzed here captures primarily small particles (<100 μm) due to the removal of optical spikes during processing (Sauzède et al., 2020). While small POC represents ∼90% of the standing stocks (Baker et al., 2017), it contributes less to instantaneous vertical flux than fast-sinking aggregates. Our results therefore concern the vertical distribution of small POC rather than large-particle export flux. The relationship between small POC concentration gradients and POC flux attenuation depends on what relationship one assumes (or ideally measures) between particle size, sinking speed, and depth (Cael & Bisson, 2018). Second, the UVP5 observations and POC product cannot explicitly resolve diel vertical migration effects. Migrating taxa can alter both particle consumption and pellet production across depth and time of day, potentially modulating small-POC gradients in ways that cannot be resolved without sub-daily zooplankton and POC profile changes. Future work with such observations could test how strongly the

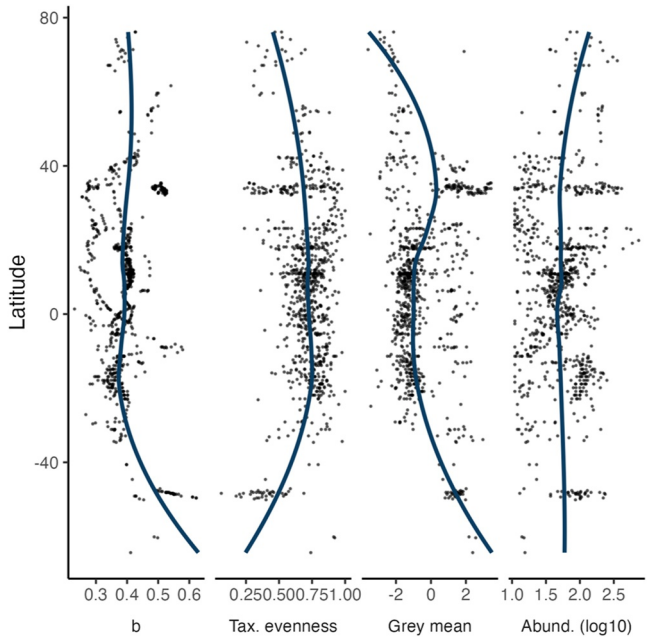

Figure 5. Latitudinal splines of the particulate organic carbon concentration gradient *b* and the three selected zooplankton metrics, shown as descriptive spatial context rather than inferential evidence.

relationships reported here depend on migratory behavior. Third, the POC product yields only a latitudinally and longitudinally resolved map of time-mean POC attenuation, whereas UVP5 data are instantaneous snapshots. This mismatch neglects both seasonal and interannual variability in POC attenuation, meaning our estimate for how much of the variability in POC attenuation are explained by zooplankton community metrics is a lower bound; future work using a temporally resolved POC product could investigate how much of the ~56% of unexplained variability in *b* is due to covariation of *b* with the zooplankton community seasonally or interannually. Fourth, while the UVP5 product here is the largest and most globally representative available data set of imaging-based zooplankton community structure, its taxonomic and trophic resolution is somewhat coarse; other data sets such as ZooSCAN data collected aboard the *Tara* could provide finer taxonomic and trophic resolution, which might sharpen interpretation of which functional groups underlie the relationships identified here, but would likely come with reduced spatiotemporal coverage and thus less global representativeness than the data used in the present study.

Altogether these results indicate that three key zooplankton community metrics—namely taxonomic evenness, mean gray level, and abundance—explain almost half of the variation in small POC concentration gradients in the mesopelagic ocean. These three metrics appear to be key mediators of the influence of zooplankton on the marine carbon cycle, suggesting potential avenues for improving the representation of zooplankton communities in ocean biogeochemical models (Ratnarajah et al., 2023). Specifically, models could benefit from resolving or parameterizing taxonomic evenness, abundance, and image-derived morphology axes related to organism opacity/transparency, along with their effects on small-POC gradients. Our results motivate testable observational hypotheses—for example, whether steeper vertical gradients in small POC are produced by more uneven, more opaque-versus-transparent, or more abundant zooplankton communities due to how such communities consume, produce, and repackage particulate matter. By applying ML to new publicly available data products, this study expands the scope of such studies to the global ocean, thereby presenting novel quantitative evidence that zooplankton community structure—not just biomass—plays an important role in mesopelagic carbon processing.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The code used in this study is available at (Panaïotis, 2026). Zooplankton data are available at https://www.seanoe.org/data/00964/107583/. Particulate organic carbon data are available at https://doi.org/10.48670/moi-00046.

References

Acevedo-Trejos, E., Cadier, M., Chakraborty, S., Chen, B., Cheung, S. Y., Grigoratou, M., et al. (2022). Modelling approaches for capturing plankton diversity (MODIV), their societal applications and data needs. *Frontiers in Marine Science*, *9*, 975414. https://doi.org/10.3389/fmars.2022.975414

Baker, C. A., Henson, S. A., Cavan, E. L., Giering, S. L. C., Yool, A., Gehlen, M., et al. (2017). Slow-sinking particulate organic carbon in the Atlantic Ocean: Magnitude, flux, and potential controls. *Global Biogeochemical Cycles*, *31*(7), 1051–1065. https://doi.org/10.1002/2017GB005638

Beck, M., Cailleton, C., Guidi, L., Desnos, C., Jalabert, L., Elineau, A., et al. (2023). Morphological diversity increases with decreasing resources along a zooplankton time series. *Proceedings of the Royal Society B: Biological Sciences*, *290*(2011), 20232109. https://doi.org/10.1098/rspb.2023.2109

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., & Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, *568*(7752), 327–335. https://doi.org/10.1038/s41586-019-1098-2

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Acknowledgments
TP and BC were supported by Schmidt Sciences Virtual Earth System Research Institute project CALIPSO and Horizon Europe project BIOcean5D (Grant 101059915).

Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of CO_2 by the oceans. *Science*, *367*(6479), 791–793. https://doi.org/10.1126/science.aay1790

Burd, A. B. (2024). Modeling the vertical flux of organic carbon in the global ocean. *Annual Review of Marine Science*, *16*(1), 135–161. https://doi.org/10.1146/annurev-marine-022123-102516

Cael, B., & Bisson, K. (2018). Particle flux parameterizations: Quantitative and mechanistic similarities and differences. *Frontiers in Marine Science*, *5*, 395. https://doi.org/10.3389/fmars.2018.00395

Cael, B., Bisson, K., Conte, M., Duret, M. T., Follett, C. L., Henson, S. A., et al. (2021). Open ocean particle flux variability from surface to seafloor. *Geophysical Research Letters*, *48*(9), e2021GL092895. https://doi.org/10.1029/2021gl092895

Cavan, E. L., Trimmer, M., Shelley, F., & Sanders, R. (2017). The importance of zooplankton-mediated detritus flux for the biological pump. *Biogeosciences*, *14*(1), 177–186. https://doi.org/10.5194/bg-14-177-2017

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system (No. arXiv:1603.02754). *arXiv*. https://doi.org/10.48550/arXiv.1603.02754

Denvil-Sommer, A., Buitenhuis, E. T., Kiko, R., Lombard, F., Guidi, L., & Le Quéré, C. (2023). Testing the reconstruction of modelled particulate organic carbon from surface ecosystem components using PlankTOM12 and machine learning. *Geoscientific Model Development*, *16*(10), 2995–3012. https://doi.org/10.5194/gmd-16-2995-2023

Dilling, L., & Alldredge, A. L. (2000). Fragmentation of marine snow by swimming macrozooplankton: A new process impacting carbon cycling in the sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(7), 1227–1245. https://doi.org/10.1016/s0967-0637(99)00105-3

Di Matteo, L., Benedetti, F., Ayata, S.-D., & Aumont, O. (2025). Including different mesozooplankton feeding strategies in a biogeochemical ocean model impacts global ocean biomass and carbon cycle. *Biogeosciences*, *22*, 7233–7268. https://doi.org/10.5194/bg-22-7233-2025

Drago, L., Panaïotis, T., Irisson, J.-O., Babin, M., Biard, T., Carlotti, F., et al. (2022). Global distribution of Zooplankton biomass estimated by in situ imaging and machine learning. *Frontiers in Marine Science*, *9*, 894372. https://doi.org/10.3389/fmars.2022.894372

Durkin, C. A., Cetinić, I., Estapa, M., Ljubešić, Z., Mucko, M., Neeley, A., & Omand, M. (2022). Tracing the path of carbon export in the ocean though dna sequencing of individual sinking particles. *The ISME Journal*, *16*(8), 1896–1906. https://doi.org/10.1038/s41396-022-01239-2

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gutierrez-Rodriguez, A., Stukel, M. R., Lopes dos Santos, A., Biard, T., Scharek, R., Vaulot, D., et al. (2019). High contribution of Rhizaria (radiolaria) to vertical export in the California current ecosystem revealed by DNA metabarcoding. *The ISME Journal*, *13*(4), 964–976. https://doi.org/10.1038/s41396-018-0322-7

Henson, S. A., Laufkötter, C., Leung, S., Giering, S. L. C., Palevsky, H. I., & Cavan, E. L. (2022). Uncertain response of ocean biological carbon export in a changing world. *Nature Geoscience*, *15*(4), 248–254. https://doi.org/10.1038/s41561-022-00927-0

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., et al. (2019). Global trends in marine plankton diversity across kingdoms of life. *Cell*, *179*(5), 1084–1097. https://doi.org/10.1016/j.cell.2019.10.008

Irisson, J.-O., Ayata, S.-D., Lindsay, D. J., Karp-Boss, L., & Stemmann, L. (2022). Machine learning for the study of Plankton and marine snow from images. *Annual Review of Marine Science*, *14*(1), 277–301. https://doi.org/10.1146/annurev-marine-041921-013023

Iversen, M. H. (2023). Carbon export in the ocean: A biologist's perspective. *Annual Review of Marine Science*, *15*(1), 357–381. https://doi.org/10.1146/annurev-marine-032122-035153

Kuhn, M., & Wickham, H. (2020). Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.

Lam, P. J., Doney, S. C., & Bishop, J. K. B. (2011). The dynamic ocean biological pump: Insights from a global compilation of particulate organic carbon, $CaCO_3$, and opal concentration profiles from the mesopelagic. *Global Biogeochemical Cycles*, *25*(3). https://doi.org/10.1029/2010GB003868

Laxton, R. R. (1978). The measure of diversity. *Journal of Theoretical Biology*, *70*(1), 51–67. https://doi.org/10.1016/0022-5193(78)90302-8

Liu, K., Xu, Z., Liu, X., Huang, B., Liu, H., & Chen, B. (2024). Modelling global mesozooplankton biomass using machine learning techniques and earth system model environmental variables. *Progress in Oceanography*, *229*, 103383. https://doi.org/10.1016/j.pocean.2024.103383

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, *6*, 196. https://doi.org/10.3389/fmars.2019.00196

Lombard, F., Guidi, L., Brandão, M. C., Coelho, M. P., Colin, S., Dolan, J. R., et al. (2024). Ubiquity of inverted 'gelatinous' ecosystem pyramids in the global ocean. https://doi.org/10.1101/2024.02.09.579612

Luo, J., Stock, C. A., Henschke, N., Dunne, J. P., O'Brien, T. D., Bianchi, D., & Siegel, D. A. (2022). Global ecological and biogeochemical impacts of pelagic tunicates. *Progress in Oceanography*, *205*, 102822. https://doi.org/10.1016/j.pocean.2022.102822

Magneville, C., Loiseau, N., Albouy, C., Casajus, N., Claverie, T., Escalas, A., et al. (2022). mFD: An R package to compute and illustrate the multiple facets of functional diversity. *Ecography*, *2022*(1), ecog.05904. https://doi.org/10.1111/ecog.05904

Meyjes, S. A., Petrik, C. M., Rohr, T., Cael, B. B., & Mashayek, A. (2024). Impact of spatial variability in zooplankton grazing rates on carbon export flux. *Global Biogeochemical Cycles*, *38*(6), e2023GB008085. https://doi.org/10.1029/2023GB008085

Negrete-Garc'ia, G., Sailley, S. F., Benincà, E., & Dutkiewicz, S. (2022). Marbl-spectra v1.0: A flexible size-spectrum plankton community model for earth system modeling. *Progress in Oceanography*, *208*, 102899. https://doi.org/10.1016/j.pocean.2022.102899

Nocera, A. C., Stemmann, L., Babin, M., Biard, T., Coustenoble, J., Carlotti, F., et al. (2025). A global consistent database of plankton and detritus from in situ imaging by the Underwater Vision Profiler 5. In *Earth system science data discussions* (pp. 1–37). https://doi.org/10.5194/essd-2025-522

Orenstein, E. C., Ayata, S.-D., Maps, F., Becker, É. C., Benedetti, F., Biard, T., et al. (2022). Machine learning techniques to characterize functional traits of plankton from image data. *Limnology and Oceanography*, *67*(8), 1647–1669. https://doi.org/10.1002/lno.12101

Panaïotis, T. (2026). Code for: Assessing zooplankton communities' influence on particulate organic carbon concentration. *Zenodo*. https://doi.org/10.5281/zenodo.20626647

Panaïotis, T., Babin, M., Biard, T., Carlotti, F., Coppola, L., Guidi, L., et al. (2023). Three major mesoplanktonic communities resolved by in situ imaging in the upper 500 m of the global ocean. *Global Ecology and Biogeography*, *32*(11), 1991–2005. https://doi.org/10.1111/geb.13741

Panaïotis, T., Wilson, J. D., & Cael, B. B. (2025). A machine learning-based dissolved organic carbon climatology. *Geophysical Research Letters*, *52*(7), e2024GL112792. https://doi.org/10.1029/2024GL112792

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(9), 462–473. https://doi.org/10.4319/lom.2010.8.462

Ratnarajah, L., Abu-Alhaija, R., Atkinson, A., Batten, S., Bax, N. J., Bernard, K. S., et al. (2023). Monitoring and modelling marine zooplankton in a changing climate. *Nature Communications*, *14*(1), 564. https://doi.org/10.1038/s41467-023-36241-5

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. *Vienna, Austria*. Retrieved from https://www.R-project.org/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Sauzède, R., Johnson, J. E., Claustre, H., Camps-Valls, G., & Ruescas, A. B. (2020). Estimation of Oceanic particulate organic carbon with machine learning. ISPRS annals of the photogrammetry. *Remote Sensing and Spatial Information Sciences*, *2*, 949–956. https://doi.org/10.5194/isprs-annals-V-2-2020-949-2020

Serra-Pompei, C., Ward, B. A., Pinti, J., Visser, A. W., Kiørboe, T., & Andersen, K. H. (2022). Linking plankton size spectra and community composition to carbon export and its efficiency. *Global Biogeochemical Cycles*, *36*(5), e2021GB007275. https://doi.org/10.1029/2021gb007275

Siegel, D. A., DeVries, T., Cetinić, I., & Bisson, K. M. (2023). Quantifying the ocean's biological pump and its carbon cycle impacts on global scales. *Annual Review of Marine Science*, *15*(1), 329–356. https://doi.org/10.1146/annurev-marine-040722-115226

Steinberg, D. K., Cope, J. S., Wilson, S. E., & Kobari, T. (2008). A comparison of mesopelagic mesozooplankton community structure in the subtropical and subarctic North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(14), 1615–1635. https://doi.org/10.1016/j.dsr2.2008.04.025

Steinberg, D. K., & Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Annual Review of Marine Science*, *9*(1), 413–444. https://doi.org/10.1146/annurev-marine-010814-015924

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205–248. https://doi.org/10.1016/j.pocean.2014.08.005

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*(1), 91. https://doi.org/10.1186/1471-2105-7-91

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In *The carbon cycle and atmospheric CO_2: Natural variations Archean to present* (Vol. 32, pp. 99–110).