Applicability of in situ imaging to investigate copepod DVM patterns and morphology

Alex Barth   Joshua Stone

02 Feb 2023

1 Introduction

- Diel Vertical Migration Importance and Background

- Challenges of

- Application of in-situ imaging

- Hypotheses:

In this study, we investigate the applicability of in-situ imaging data to study DVM behavior. We aim to test two hypotheses. First, (H1) Overall copepod morphology, including size and transparency, will impact DVM behavior. It is predicted that copepods which are more visually detectable, e.g. larger and darker, will have to migrate larger distances. Second, (H2) DVM behavior will vary based on environmental conditions related to the risks and rewards of occupying surface waters. Specifically, increased prey availability will increase a copepod's likelihood to travel to surface waters, while increased light availability will force copepods to deeper sections of the water column. To test these hypotheses we build on novel statistical

tools to describe morphology of objects sampled by in situ imaging Vilgrain et al. (2021). Additionally we evaluate multiple approaches to describing DVM using in situ imaging data and how to estimate the impact of environmental factors on DVM behavior.

2 Methods

2.1 CTD profiles and UVP imaging of copepods

Data were collected aboard the R/V Atlantic Explorer in collaboration with the Bermuda Atlantic Time-series Study (BATS) (Steinberg et al. 2001). In situ images of plankton were acquired using an Underwater Vision Profiler (UVP5-DEEP, sn: 209, Hydroptic, (Picheral et al. 2010)). The original sampling methodology and instrument specification followed details described in (Barth and Stone 2022). A brief summary of new and relevant sampling methodology are described below. The UVP was attached to the CTD rosette aboard the R/V Atlantic Explorer. The UVP was equipped for monthly cruises into the Sargasso Sea from June 2019 - Aug 2019 and October 2020 - December 2021. A typical monthly cruise included 13 profiles with descents typically to 1200m. Full cast details are available in Supporting Information. Only profiles within the BATS region were included (Figure 1). Generally, as described in analyses below, it was assumed that profiles within the BATS region during a cruise sampled a similar water mass. While mesoscale features are known to drive variability in this region, there were no major variations within a cruise (Supplemental Figure 1). The CTD collected continuous environmental parameters on down-casts, including temperature (Dual SBE-03f,

Sea Bird Scientific), conductivity (Dual SBE04, Sea Bird Scientific), dissolved oxygen (SBE43, Sea Bird Scientific) and fluorescence (Chelsea Instruments). Because the purpose of this study was to investigate day/night differences in copepod vertical distributions, it was important that there were day and night casts during each cruise. It would be ideal to investigate each profile, without pooling the entire region by cruise, however, the small sampling volume of the UVP and low abundance of plankton in this region create the need to pool or average profiles (Barth and Stone 2022). Profiles were assigned to be day or night based on locally calculated nautical dawn and nautical dusk times using the R package `suncalc 0.5.1`.
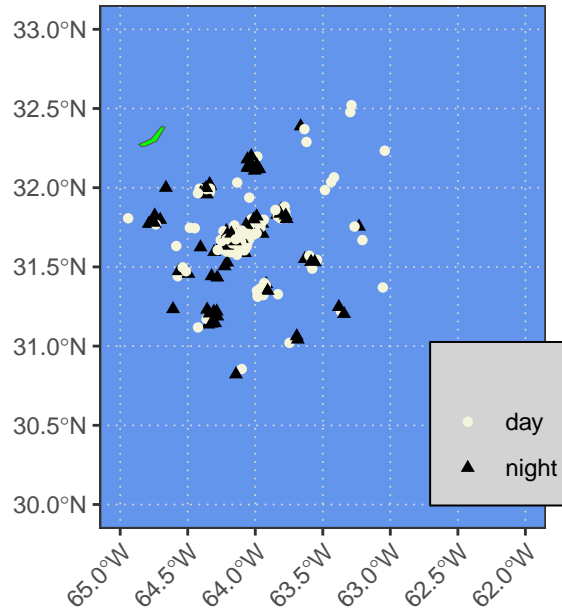

Figure 1: Figure 1. Map of CTD and UVP profiles used in the study. Day and night points from all cruises are shown. Bermuda is shown in green.

The UVP sampled during down-casts of all profiles at a rate of ~15Hz. The UVP records counts of small particles ($>184\mu$m Equivalent Spherical Diameter, ESD) and images of large

particles ($>600\mu$m ESD). However, living particles are not reliably identifiable below 0.9 mm (Barth and Stone 2022). All recorded images were processed using Zooprocess (Gorsky et al. 2010), which provides several metrics related to size, grey value, and shape complexity. These features were then used to automatically sort images using the Ecotaxa application (Picheral et al.). While automatic classification increases the overall speed of identification (Irisson et al. 2022), all images were manually verified by the same trained taxonomist. In total, 294913 images were recorded. Of these, 85.2% were images of debris, marine-snow, or artefacts, while 14.8% were identified as living. In the present study, only copepods were investigated. The smallest observed and identified copepod was 0.940 mm ESD and the largest was 5.904 mm ESD. Across all casts, copepods were the most common observed organism at 58.7% of all identified, living particles. In total, there were 4151 individual copepods investigated in this study.

2.2 Identification of morphogroups classification

To test the hypothesis of morphological influences on DVM magnitude, copepod images were classified into distinct morphological-groups (referred to as morphogroups). The UVP automatically measures and collects several morphologically relevant parameters. To create relevant groups of copepods, a dimension reduction and clustering approach was used. Similar methods have been successfully utilized to provide novel insights to marine snow (Trudnowska et al. 2021) and copepod dynamics in the Arctic (Vilgrain et al. 2021). First, 18 parameters were selected to be included in a Principle Components Analysis (PCA) which represent

major copepod morphological features following (Vilgrain et al. 2021). Parameters can be described as relating to size (e.g. major axis, feret diameter, ESD), grey intensity (e.g. mean grey value), shape (e.g. elongation, symmetry), and shape complexity (e.g. fractal dimension, perimeter/feret). All parameters and model structure are available in Supporting Information. The PCA was weighted by the volume sampled in a 1-m depth bin for each observation included in the model. This approach provides a correction for the UVP's variable descent speed which can cause duplicate imaging of individuals. While this phenomena has a minor impact on overall results(Barth and Stone 2022), we used the weighted approach to assure that no individual features were over represented. An unweighted PCA did not produce substantially different results, which supports the notion that duplicate images are a minor phenomena. All morphological descriptors were scaled and centered prior to inclusion in the analysis. The model was constructed using the R package `FactoMineR 2.7`. Principle components were deemed to be significant if their eigenvalues were greater than 1. This approach yielded 4 clusters which describe 87.3% of the total variation in morphological parameters, with 34.5% and 26.5% in the first two components respectively.

Then to classify distinct morphogroups, a k-means clustering algorithm was used on the principle components (Trudnowska et al. 2021). While, there were 4 significant PCs, only the first two were used in the k-means clustering. Similar to Vilgrain et al. (2021), the third axis could be described by the orientation of the copepod to the camera and the fourth axis could be used to describe appendage visibility. Presumably, the both these factors would be random with respect to the depth at which a copepod was found. Possibly, appendage visibility could

have a relationship with feeding activity, yet when included in the cluster analysis, no group clearly separated along this axis. Thus for simplicity, only the first two axes were included in the k-means clustering algorithm. To select the optimal number of clusters, clusters were added until the addition of a new cluster would not produce an increase the overall explanatory power ($\frac{WithinSS}{TotalSS}$) by more than 10%. This process identified four distinct morphogroups. The k-means clustering algorithm was ran with 1000 random starts. All code for the PCA and k-means models can be found in the supporting information. Additional model diagnostics are available in Supporting Information 2.

2.3 Investigating copepod vertical structure

Copepods in this system are well documented to undergo DVM (Steinberg et al. 2000; Schnetzer and Steinberg 2002; Maas et al. 2018) However, direct measurements of DVM using in situ imaging are not common. Thus, there are not established methods to describe DVM differences between groups or relation to environmental parameters. Here, we use and evaluate several approaches to specifically test DVM variation between morphogroup and in response to environmental factors. These include both classic and novel methods.

2.3.1 Vertical distribution of copepods

To visualize copepod vertical structure across morphogroups, the concentration of each morphogroup was calculated in 20m depth-bins for each UVP profile.

These binned-profiles were then averaged together based on time of day. This average-cast approach is useful as it provides a general pattern with information as to the variability between casts. Then to better estimate the day-night differences within each morphogroup, a proportional difference was calculated in each depth bin. A morphogroup's proportional difference was defined for each i depth bin as $Prop\ Diff_i = \frac{Night_i - Day_i}{\sum_i^N Night_i + Day_i}$, where $Night_i$ and Day_i are the average concentrations in that bin. By scaling each bin's diel difference by the total morphogroup concentration, it is possible to compare the profiles of proportional difference without the influence of difference in total concentration between morphogroups or depth bins. The proportional difference profile then was used to select a maximum depth (600m) to include for the weighted mean depth calculations.

2.3.2 Weighted mean depth variability

Weighted mean depth (WMD) is a common metric to describe vertical structure and DVM in zooplankton **CITE**. However, with in-situ imagery, this approach presents a few challenges. WMD cannot be calculated individually for each profile then averaged because each profile had a different descent depth. Similarly, profiles had different sampling efforts throughout the water column. So to calculate the WMD for each morphogroup and each time of day, all 20-m binned concentrations from (0,20] and (580, 600], were pooled together. Then a bootstrap method was used to describe the variability in profiles and vertical structure. A traditional bootstrap approach would not work however, because there were more instances of shallow depth bins (as not all casts went as deep) so it would bias estimates. To correct for this, we

define a bin-constrained approach to estimate a bootstrapped WMD^*.

$$WMD^*_{mg,tod} = \sum_i^{N=60} \frac{d_i(conc^*_{i,mg,tod})}{\sum_i^{N=60} conc^*_{i,mg,tod}}$$

For each morphogroup (mg) and each time of day (tod), the WMD^* was calculated by randomly sampling a concentration from the pooled distribution of concentrations for each depth bin i. This was done for all 60 depth bins above 600m. The concentration was then multiplied by the midpoint of the depth bin and divided by the sum of all randomly sampled concentrations. This process was iterated 999 times to create a distribution of WMD^* for each morphogroup and time of day. For each iteration, the concentrations for a depth bin were randomly drawn with replacement for the next iteration. This approach effectively simulates multiple even profiles over the entire study region using the data from existing profiles. Then a bootstrapped mean and 95% confidence interval were calculated from that distribution of WMD^*. The the bootstrapped confidence interval can then be interpreted as wider for both variability between UVP profiles and the overall spread of the zooplankton in the water column. Consequentially, it can be said a morphogroup exhibits a clear/significant signal of DVM if the confidence intervals are non-overlapping between day and night.

2.3.3 Occupancy modelling of vertical structure

While the bootstrapped WMD depth is useful for investigating broad-scale DVM patterns across the whole data set, it is a less effective when separating the data set into specific months

and seasons. Again issues arise with the compounding effect of the UVP's low sampling volume and the rarity of organisms in this oligotrophic system. As described in detail by Bisson et al., when sampling volumes are low estimating the theoretical probability of correctly estimating particle concentration greatly decreases. In the present system, when specific morphogroups are quite rare for a given season or month, the estimate of concentration greatly depends on the presence/absence of that taxa in a depth bin and is not reliable as a result. To investigate the environmental impact on DVM, the vertical structure needs to be compared across cruises, or at least seasons. Therefore, using the WMD approach is not applicable because the split data sets would be too small.

Here, we investigate the novel application of site-occupancy models to this data set. Site-occupancy models are a hierarchical-model popularized in biogeography and conservation biology Kéry and Schmidt (2008). Often used with presence/absence data, the fundamental concept behind site-occupancy models is that the process of observing an organism at a given location is a function of the probability the organisms is there (occupancy/occurrence probability, ψ) and the probability of detecting the organism given the sampling methodology (detection probability, p). These are an attractive class of models to apply to zooplankton in-situ imaging data because the detection probability is likely influenced by the sampling volume of the instrument while the occupancy is likely influenced by environmental/biological factors. To construct a site-occupancy model which estimates detection probability, replicate "surveys" must be done to each "site". For our data set, we define a "site" as a depth-bin on a particular cruise and a "survey" as a single UVP profile on that cruise. Conceptually

this assumption is similar to pooling-casts together over a study area. To first validate the application of this model, we constructed a simple model using the presence/absence of all copepods (regardless of morphogroup).

$$Y_{ij,db,tod} \sim Bernoilli(p_{ij,tod} * z_{i,db,tod})$$

$$z_{i,db,tod} \sim Bernoulli(\psi_{db,tod})$$

$$logit(p_{ij,tod}) \sim a_0 + a_{vs,tod} * vs_{ij}$$

In this model, the event of observing a copepod at site i during survey j , for each depth bin db and time of day tod, is a function of the observation process which is modeled by a Bernoulli process. The observation process occurs with the probability of the product of the detection probability, p_ij, tod, and the occupancy, $z_{i,db,tod}$. This model was ran for each depth-bin (20m intervals from 0-1200m) and each time of day (day or night). The occupancy (ψ) , was assumed to be a unique, independent probability for each depth bin while the detection probability was modelling using a logistic regression to include the effect of volume sampled ($\alpha_{vs,tod}$). If the site occupancy model is effective in this system, it would be expected that there would be a positive effect of volume sampled on the detection probability. The models were ran three separate times, to estimate detection probability differences between the epipelagic (0-200m), upper mesopelagic (200-600m), and lower mesopelagic (600-1200m). Detection probability was likely to differ between these regions due to differences in UVP descent speed and thus sampling volume through the water column as described in (Barth and Stone 2022).

Following the general all copepod model, we repeated the model yet separating out the different morphogroups. This allowed the test of the method to describe DVM. If effective, it would be expected there to be a difference in occupancy between daytime and nighttime in regions where migration is occurring (near-surface to upper mesopelagic). This additionally provides another method to test the hypothesis (H1) that morphology will increase DVM behavior, as it would predict the difference in day/night occupancy would be larger for more visually conspicuous morphogroups.

Finally, to evaluate the hypothesis that environmental conditions would influence DVM behavior, a site-occupancy model was constructed with included environmental parameters into the biological process for occupancy. Several metrics were selected to be included in the model. Two metrics corresponding to light availability were included, photosynthetically active radiation (PAR) and the diffuse attenuation coefficient ($\frac{1}{k_{490}}$). Ohman and Romagnan (2016) described a clear relationship between increasing $\frac{1}{k_{490}}$, or deeper light penetration into the water column, and increased DVM amplitude. As such, in the present system, it was predicted that increases in PAR and $\frac{1}{k_{490}}$ would have a negative effect on daytime occupancy in depth bins near the surface. PAR and $\frac{1}{k_{490}}$ were both obtained from the Aqua MODIS sensor (NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group 2014) in 4km resolution in a grid encompassing the entire study area (Figure 1). Satellite data were collected with daily resolution yet averaged together over a cruise period because they were being modeled as "site" covariates. Metrics related to food availability were included as well. The depth of the deep chlorophyll-a maximum (DCM_d) was included with the pre-

diction that a deeper DCM_d would have a positive effect on deeper depth bin occupancy and a negative effect on shallower depth bin occupancy. This logic suggests that copepods are not likely to travel to the surface if there is not food available there. The DCM_d was calculated from CTD profiles concurrent with the UVP then averaged together over the course of a cruise as a site covariate. CTD data were accessed from (http://batsftp.bios.edu/BATS/ctd/) using the R package `batsFtpReadR 0.1.0` (https://github.com/TheAlexBarth/batsFtpReadr). Finally the particle concentration $<450\mu$m was included as it can be a useful representative of the zooplankton prey field (Whitmore and Ohman 2021; Gastauer et al. 2022). Here, we use the integrated epipelagic (<250m) prey field with the prediction that a decrease in epipelagic food would have a negative impact on occupancy in that region of the water column. All particle data were recorded with the UVP and the integrated epipelagic prey field was averaged across all casts on a cruise as a site covariate.

The full environmental model can largely follows the general site-occupancy model described above. However, in this model, the biological process that a site i would follow the occupancy, or occurrence probability as modeled by a logistic regression. This was conducted with distinct intercepts (separate models) for each time of day, depth bin, and morphogroup.

$$z_{i,db,tod,mg} \sim Binomial(\psi_{i,db,tod,mg})$$

$$logit(\psi_{i,db,tod,mg}) \sim \beta_{0_{db,tod,mg}} + \beta_{dcm\_d_{db,tod,mg}} * dcm\_d_i + \beta_{prey_{db,tod,mg}} * prey_i$$
$$+ \beta_{par_{db,tod,mg}} * par_i + \beta_{\frac{1}{k_{490}}_{db,tod,mg}} * \frac{1}{k_{490}}_{db,tod,mg}$$

Thus there was a separate effect for each parameter β on a depth bin (db), time of day (tod), and morphogroup (mg) . This allows for the investigation of how the posterior credible interval for the slope of each effect changes with depth.

All site-occupancy models follow a Bayesian hierarchical structure. These models were simulated using JAGS ver 4.3.1 using the R package `jagsUI 1.5.2`. All parameters were given vague priors. For fixed probabilities (α_0, ψ in the general model, β_0), a Beta(1,1) prior was used. For all slope parameters (α_{vs}, all β's), a weak normal ($\mu =0$, $\sigma = 1/10$) was used. The MCMC algorithm was ran with three chains for 10,000 iterations. Chains were thinned at a rate of 2. All data used to inform slope parameters were standardized prior to inclusion to reduce autocorrelation.

2.4 Data availability

All data and code are made available via (https://github.com/TheAlexBarth/DVM_Migration-Morphology). All supplemental figures, tables, and analyses are hosted on a public static site (https://thealexbarth.github.io/DVM_Migration_Morphology). This content is also available in the Supporting Information.

3 Results

3.1 Morphogroup classification

The PCA revealed four major axis of variability (Figure 2). The first axis (PC1, 34.23% of variability) was largely explained by along a gradient of increasing values related to size, such as perimeter (loading score = 0.927) and feret diameter (loading score = 0.910).

The second axis (PC2, 27.24% of variability) can be interpreted primarily by a spectrum of transparent to dark/shaded individuals. PC2 was largely anticorrelated with mean grey value (higher values indicate a more transparent individual) (loading score = -0.920). As noted in the methods, PC3 and PC4 were both related to the orientation of the copepod and the appendage visibility respectively (Supplemental 3). As a result, these axis were not included in k-means clustering analysis.

The k-means clustering analysis yielded 4 morphogroups which represent over 64% of the total variation in the two-component morphospace (Figure 2). Morphogroup 1 primarily includes small-to-medium, dark individuals with moderate PC1 scores and high PC2 scores (Figure 2, Figure 3B,D). Morphogroup 2 is best described by the smallest copepods, with low values on PC1 (Figure 2, Figure 3A,C). Morphogroup 3 is associated with moderate-to-big values of PC1 and negative values of PC2. Thus, morphogroup 3 largely contains more transparent, medium-to-large copepods (Figure 2, 3A,D). Finally, morphogroup 4 constituents are the largest copepods, spanning a wide range of shade values (Figure 2, Figure 3A,B,D).

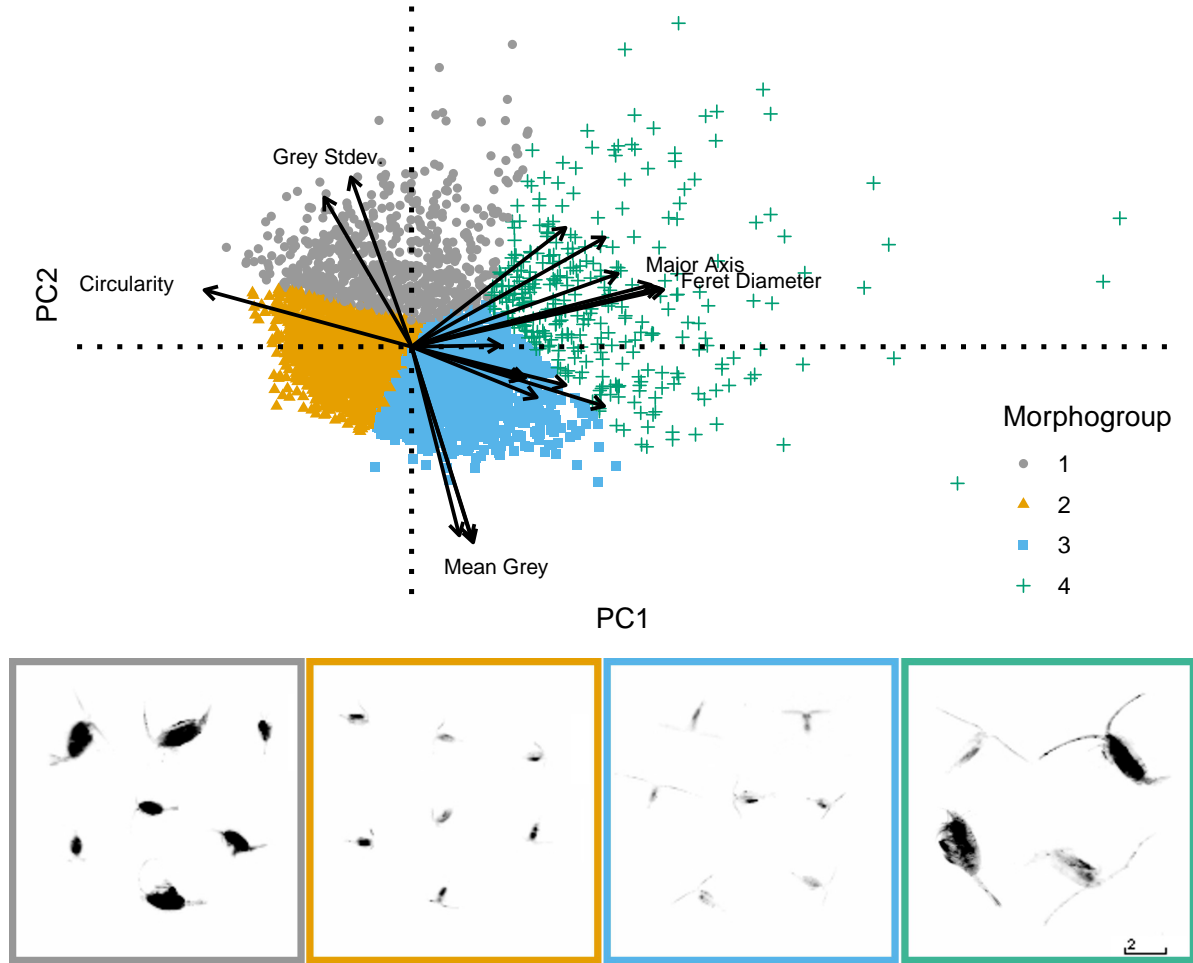To verify that the morphogroups were not arbitrary, a Dunn's pairwise comparison test was

Figure 2: Morphogroups shown in the 2-component morphospace. PC1 represents 34.23% of total variability, PC2 represents 27.24%. Representative images from morphogroups 1-4 as identified by k-means clustering are shown in numerical order at bottom of the plot. Colored borders correspond to morphospace. 2mm scale bar is present in the bottom right.

used with a Bonferrioni correction. It was found that all groups were significantly different from one another (p < 0.001) along both principle components (Figure 3). To illustrate the difference along morphological parameters, morphogroups were plotted against mean feret diameter and mean grey value (Figure 3C,D). From these figures, it is evident that the morphogroups were different along morphologically specific parameters.
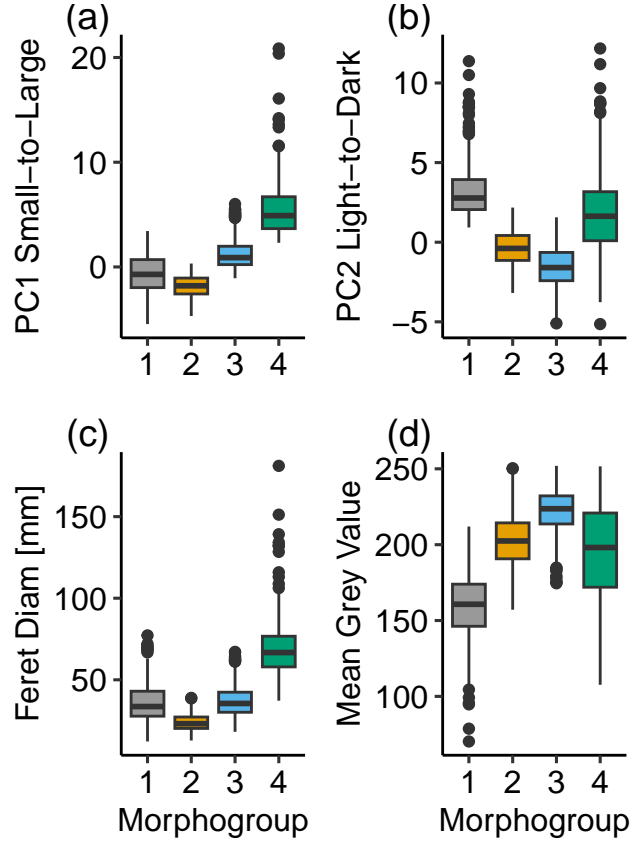

Figure 3: Distribution of morphogroups across (A) PC1, (B) PC2, (C) Feret Diameter, and (D) Mean Grey Value. Morphogroups were significantly different from another along PC1 and PC2 (Dunn test for multiple comparison, p < 0.001).

3.2 Morphogroup average concentration and vertical distribution

Across all cruises, the overall integrated (0-1200m) abundance of morphogroups followed a similar pattern. Abundance was higher during the spring-summer months and lowest during the fall and winter (Supplemental 4). Generally, the largest copepods (morphogroup 4) were less abundant in all cruises. The darker copepods (morphogroup 1) were also in lower abundance at some times of the year (Supplemental 4). The smaller copepods were more abundant (morphogroup 2), which is consistent with the notion this system is dominated by smaller individuals. Interestingly, morphogroup 3 tended to be more abundant than morphogroup 1 despite the presumption that morphogroup 3 consists of larger copepods. In some cruises, morphogroup 3 was the most abundant category (Supplemental 4).

By morphogroup, the overall patterns of average abundance are reflected in each average vertical profile (Figure 4A-D). Average abundance throughout the water column was higher in morphogroup 2 and morphogroup 3 then lowest in morphogroup 4. All morphogroups displayed a nighttime peak in average abundance near the lower epipelagic (100 - 200m) (Figure 4A-D). While all morphogroups had some concentration of copepods near the surface during the daytime, the proportional difference clearly showed nighttime abundances were higher, on average (Figure 4E-F). Additionally, copepod average abundances were low around around 200m-400m. Then from about 400m-600m, the average abundance of all morphogroups increased for daytime bins, but not nighttime. Below 600m, day/night differences were more stochastic, which is consistent with a reasonable exception for DVM to not extend this deep into the water column. It should be noted that the standard deviation in each depth bin's

abundance is quite large (Figure 4A-D). The large variability both within and between cruises,

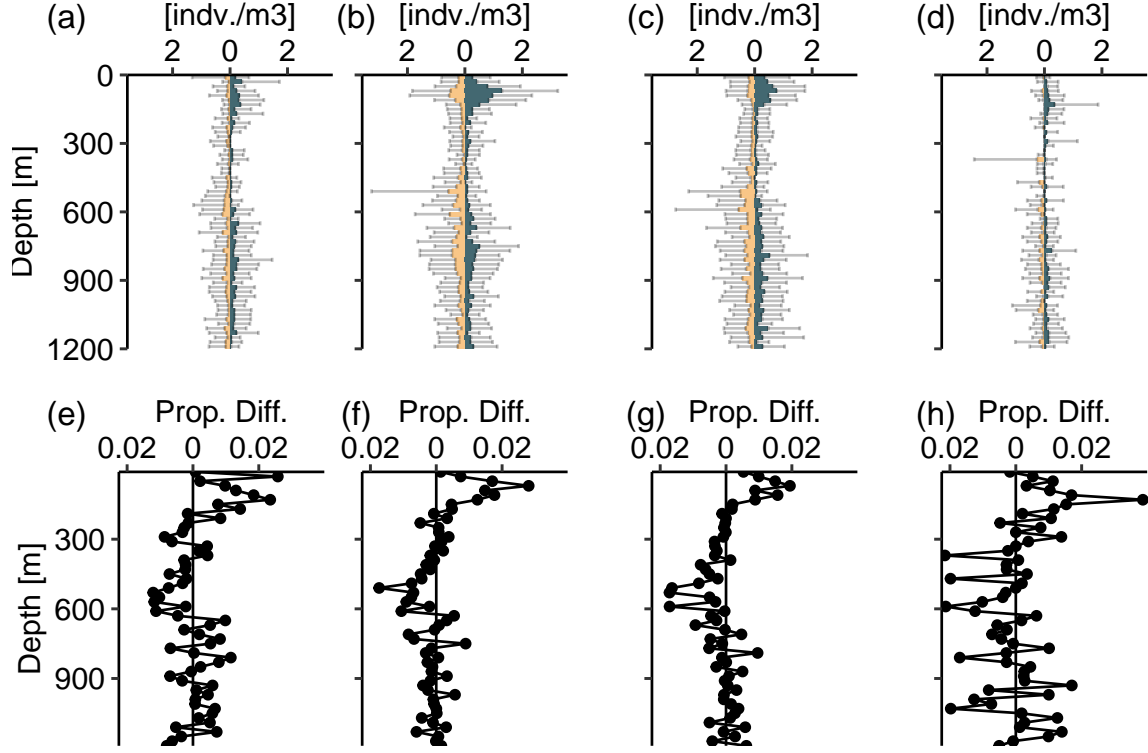likely drives the large variation in bin average abundance.

Figure 4: (A-D) Vertical profiles of average copepod concentration in 20m depth bins. Daytime concentrations are light-colored and shown left of the vertical axis, nighttime values are dark-colored and shown right of the vertical axis. (E-H) Proportional nighttime/daytime difference in 20m bins. A nighttime-dominated bin is right of the vertical axis, daytime-dominated left of the vertical axis. Vertically paired, panels correspond to an individual morphogroup; (A,E) Morphogroup 1, (B,F) Morphogroup 2, (C,G) Morphogroup 3, (D,H) Morphogroup 4