

Utilizing Environmental DNA to Investigate the Effects of Hypoxia on Copepod
Abundance

A Thesis
Presented to
The Established Interdisciplinary Committee for Environmental Studies
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Ella Crotty

May 2025

Chapter 1

Methods

1.1 eDNA Index Preparation

Data analysis for this study was conducted using R version 4.3.0 "Already Tomorrow." I downloaded the eDNA sample data with the corresponding metadata and the environmental data from the TH042 mooring during the study period from previous research. Some *P. mimus* samples were misidentified as *P. acuspis*, so I re-identified the copepod sequence samples using NCBI BLAST and corrected the species identifications using those results. Additionally, environmental data from the LiveOcean model at TH042 during the study period was downloaded to fill in a data gap in 2023 during which the TH042 mooring was not functional. eDNA index was calculated in order to obtain a measure of the relative abundance of each species over time. Before calculating the eDNA index, I averaged the number of reads across technical replicates so that each observation represented one sampling time. Technical replicates included multiple samples taken at the same time, as well as multiple polymerase chain reaction (PCR) runs performed on the same sample. Additionally, each species has multiple characteristic variants of each gene that was used in the metabarcoding process, and each of those variants is called an ESV (Exact Sequence Variant). Because each individual observation in the data was one ESV, there were many observations of each species per sample. I combined these by adding up the total number of reads across all ESVs for each species in each sample.

I calculated eDNA index by grouping by sample, then calculating the proportion of reads of each species in each sample. I then normalized each species' proportions on a scale of 0 to 1, such that the sample with the most reads of that species had a value of 1, and any sample where that species was not detected had a value of 0. It is important to note that eDNA index cannot be used to make comparisons between species, but can be used to make comparisons between different time points in the same species. After calculating eDNA index, the data was filtered to only include copepods by selecting any observations in the genera *Calocalanus*, *Clausocalanus*, or *Paracalanus*, as well as any observations in the class Hexanauplia, which contains Thecostraca and Copepoda (Oakley et al. 2013). While the class Hexanauplia is no longer recognized (*WoRMS - World Register of Marine Species - Copepoda* 2024;

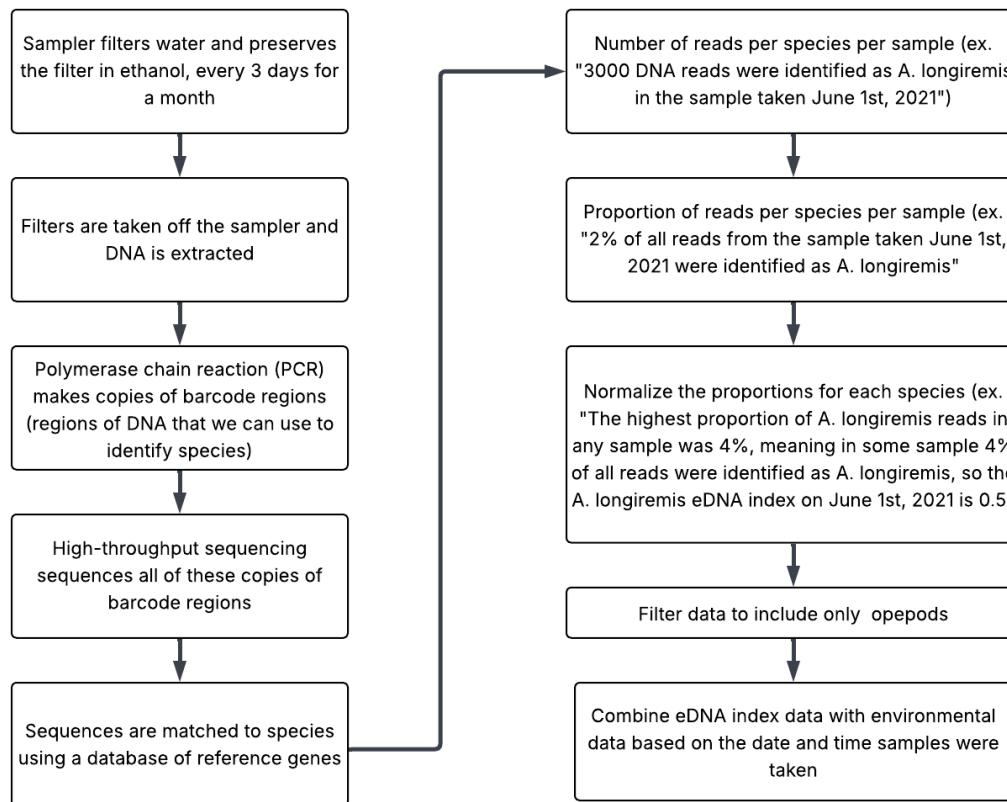


Figure 1.1: Flowchart detailing the steps to calculating eDNA index as laid out in sections 0.2 and 1.1.

Lozano-Fernandez et al. 2019), and Thecostraca and Copepoda are now recognized as classes, it was used when the eDNA for this study was processed, so I used it for data processing as well. This study will not discuss species in Thecostraca.

I combined the eDNA index data with its corresponding oceanographic data, using code from the previous study to attach each sample ID to its corresponding date and time, then joining it with the closest mooring observation. The TH042 mooring broke in 2023, but eDNA sampling continued. Therefore, I acquired data from the location of the TH042 mooring from the LiveOcean model by the UW Coastal Modeling Group, and used the R package ncdf4 to make the model compatible with the existing data (*LiveOcean Homepage* 2025; D. Pierce 2025). I matched the LiveOcean model's hourly dissolved oxygen and temperature data to the eDNA sample data by date and time, and also compared the results of the LiveOcean model's 2021-22 data to the TH042 mooring's 2021-22 data. The model data matched the mooring data closely enough to be useable (See Figure 1.2). When mooring and model data from the same hour was compared via linear model, the dissolved oxygen data had an R^2 value of 0.39 and a root mean square error (RMSE) of 0.98, and the temperature data had an R^2 of 0.82 and a RMSE of 0.41. The 2021-22 mooring-only data is still used for many of

the statistical analyses in this thesis, as the mooring data captures short and extreme hypoxic events more accurately (See Figure 1.2, orange spikes downwards are hypoxic events captured by the mooring but not the model). When modeling data is used in this thesis, it is used for all years 2021-23, in an effort to ensure that the different years are comparable and increase the number of useable eDNA sample dates.

1.2 Data Visualization

In order to determine the relationships between copepod abundance and environmental factors, eDNA index was compared to temperature and dissolved oxygen. Due to the mooring issue in 2023, I performed all analyses on the 2021-22 data, then on a combination of the 2021-22 mooring data and the 2023 model data, then on the 2021-23 modeling data. Preliminary analysis was done using time series plots, where dissolved oxygen was plotted as a time series and each sample was plotted as a dot on top of the associated dissolved oxygen measurement, with color and size corresponding to eDNA index. This allowed for visual identification of hypoxic events in July 2021, October 2021, September 2022, and October 2023, as well as a heatwave in September 2021. Analysis of abundance relative to both dissolved oxygen and temperature was first conducted using scatter plots of dissolved oxygen and temperature at each sampling time, with color again corresponding to eDNA index. To compensate for the different sampling frequencies in different environmental conditions, I then made density plots of the detection frequency within bins of environmental conditions, each spanning 0.5 mg/L of dissolved oxygen and 0.5 degrees Celsius. Each

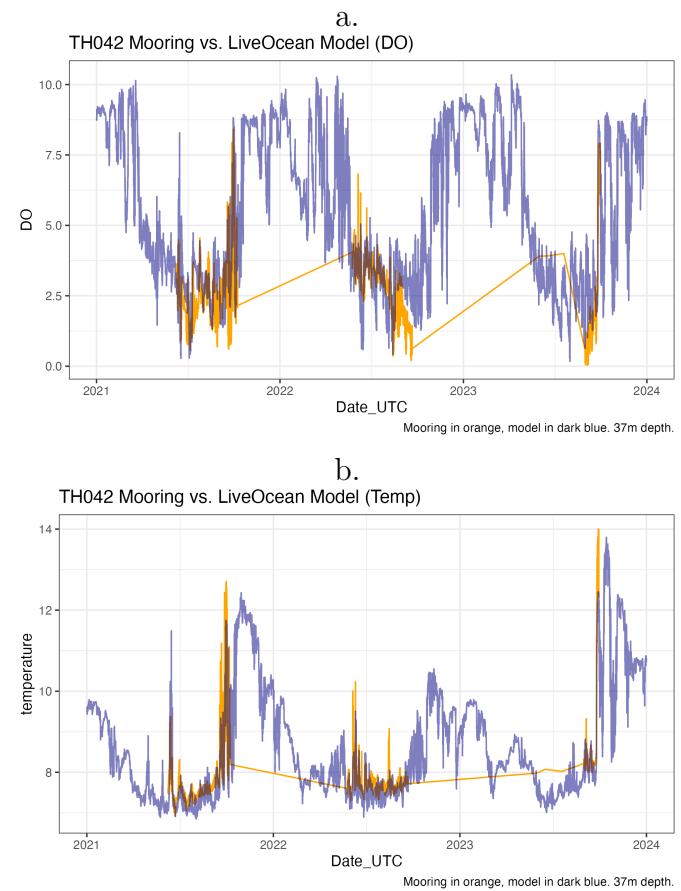


Figure 1.2: Time series of (a) dissolved oxygen (mg/L) and (b) temperature (Celsius) over the study period. TH042 mooring data is displayed in orange, while LiveOcean model data is displayed in blue.

I then made density plots of the detection frequency within bins of environmental conditions, each spanning 0.5 mg/L of dissolved oxygen and 0.5 degrees Celsius. Each

square of the plot represents a varying number of eDNA samples, and the darker squares have a higher proportion of species detections relative to non-detections.

I divided copepods into size bins based on the size ranges listed in the Marine Planktonic Copepods database (*WoRMS - World Register of Marine Species - Copepoda* 2024). Copepods smaller than 1.5 mm were classified as small, 1.5-2.5 mm copepods were classified as medium, 2.5-4 mm copepods were classified as large, and copepods exceeding 4 mm were classified as huge.

1.3 Statistical Analysis

I conducted statistical analysis of the relationship between species abundance and dissolved oxygen by using binomial regression models and zero-inflated beta regression models. I generated binomial regression models relating dissolved oxygen to species presence for each species of copepod in the dataset, excluding one data point that occurred during a short heatwave in September 2021. Because only one sample was taken during the heatwave and no other samples were from temperatures that high, it was difficult to draw conclusions about the effects of the heatwave on species presence, and the datapoint was excluded from the binomial regressions. Using the LiveOcean model data, two of the data points fall during the heatwave, as the model estimated that the heatwave started days earlier than the mooring indicated, and so those two were excluded from statistical analyses. I then used the p-values for these binomial regressions to analyze individual species responses to hypoxia.

I then attempted to generate zero-inflated beta regression models using the zoib R package, but due to difficulties in visualizing the model output I switched to the brms package, a more general modeling package that uses Stan (Bürkner et al. 2024). I chose a zero-inflated beta regression because eDNA index ranges from zero to one, necessitating a beta regression, and the data is zero-inflated because each species was absent from many samples. Using modelr, tidybayse, broom, and broom.mixed, I used the output of these models to generate and plot mean predicted data. I then compared these models to previously calculated binomial regression models, which related the presence and absence of eDNA to dissolved oxygen. Based on the Akaike information criterion (AIC), the binomial regression models performed better (lower AIC) for *A. longiremis*, *C. pacificus*, *O. similis*, *Paracalanus sp. C AC-2013*, *P. mimus*, and *P. newmani*. Based on AIC again, the zero-inflated beta regression models performed better for *C. parapergens*, *Lucicutia flavidornis*, *Metridia lucens*, *M. pacifica*, and *P. acuspis*. The zero-inflated beta regression for *C. pacificus* performed better with the mooring data, but the binomial regression model for *C. pacificus* performed better with the modeled data. In species with 5 or fewer detections, the binomial regression models performed better, but in species with more than 5 detections, the zero-inflated beta regression models performed better.

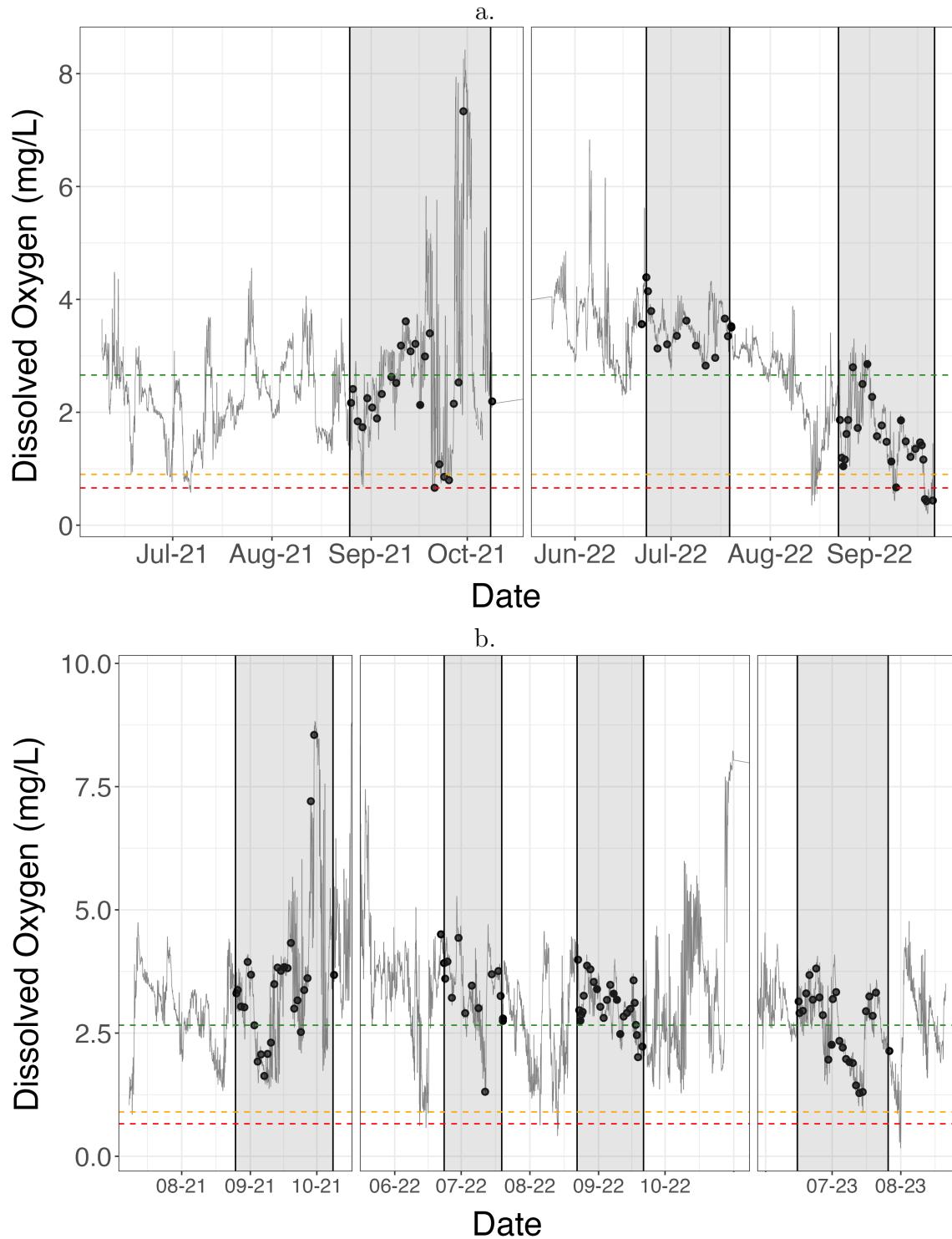


Figure 1.3: Time series of dissolved oxygen (mg/L) over the study period. Plot (a) displays TH042 mooring data, plot (b) displays LiveOcean model data. Colored dots represent eDNA index of *A. longiremis* to serve as an example. Gray boxes represent automated eDNA sampling periods in 2021 and 2022. Red line represents 0.6 mg/L DO, a level known to kill most copepods in lab experiments, orange line represents 0.9 mg/L DO, a level known to kill approximately 50% of copepods in lab experiments, and green line represents 2.66 mg/L DO, a level above which copepods have not shown negative effects in lab studies.

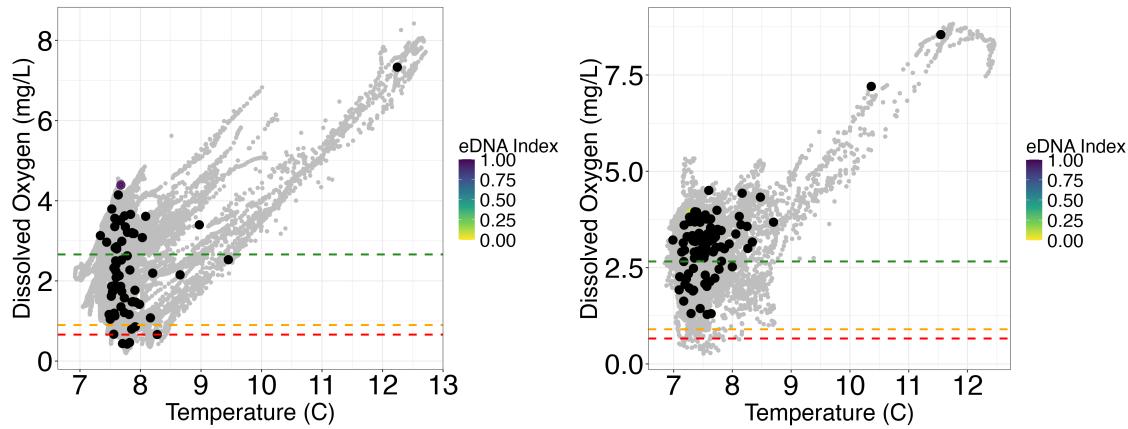


Figure 1.4: Distribution of samples versus the distribution of temperature and dissolved oxygen during the entire sampling period, of May-October 2021-23. LiveOcean model output from winter months is not shown, since the eDNA sampler and TH042 mooring were not deployed during winter months. Gray dots indicate any environmental data sampling time, from the TH042 mooring (left) or from the LiveOcean model (right). Black dots indicate environmental conditions at which eDNA samples were taken. Note the high outliers, which are during a short heatwave in September 2021 and will not be shown in future scatterplots or considered in statistical analyses due to their small sample size. These plots only utilize TH042 mooring data from 2021 to 2022, because the LiveOcean model data does not contain any dissolved oxygen values below 1.0 mg/L.

Chapter 2

Results

After bioinformatics and quality filtering, the total number of reads with the CO1 primer was 2,454,571, with 611,924 from copepods. From the CO1 primer reads, 569 unique species were identified, of which 16 were copepods with useable DNA reads in the dataset. The highest number of reads in one sample was 62,940, and the highest number of copepod reads in a sample was 19,335. The copepod eDNA detections dataset contained 26 samples from 2021, 35 samples from 2022, and 26 samples from 2023. In total, there were 556,748 DNA reads, 364,980 of which were from 2021 and 2022.

Of the copepod species detected in the data, 5 species had more than 10 detections: *A. longiremis*, *P. mimus*, *Paracalanus* sp. C AC-2013, *P. newmani*, and *O. similis*. *C. abdominalis* had 7 detections overall, with one in 2023. *Calanus pacificus* had 5 detections, none of which were in 2023, and *P. acuspis* had 5 detections, 2 of which were in 2023. All other copepod species had fewer than 5 detections overall.

2.1 Dissolved Oxygen and Species Abundance

Most of the northern copepods had higher eDNA index at higher dissolved oxygen levels, with a positive correlation between eDNA index and DO according to zero-inflated beta regressions (See Appendix B). All northern copepods except *P. newmani* were more abundant at higher dissolved oxygen levels according to zero-inflated beta regressions based on the TH042 mooring data from 2021-22.

cite zoib
ask sam

2.2 Dissolved Oxygen and Species Prevalence

The data suggests that northern copepods have a comparatively lower tolerance to hypoxia. Southern copepods and *O. similis*, which is present year-round, were detected proportionally more at dissolved oxygen levels below 0.9 mg/L (see Figure 2.2a). The northern copepod *P. newmani* had fewer detections at high dissolved oxygen levels according to a binomial regression comparing LiveOcean modeled DO to species presence, but the effect was not significant ($p = 0.3$, see Figure 2.1). ($p < 0.001$, $df =$). *P. newmani* also had a relatively consistent eDNA index across DO

add df

Species	Group	Detections		Regression Results	
Species	Group	2021-22	2021-23	2021-22	2021-23
<i>Acartia longiremis</i>	Cold-water	61	86	Positive	Neutral
<i>Pseudocalanus mimus</i>	Cold-water	28	46	Positive	Negative
<i>Pseudocalanus newmani</i>	Cold-water	20	25	Neutral	Neutral
<i>Centropages abdominalis</i>	Cold-water	6	7	Positive	Neutral
<i>Pseudocalanus acuspis</i>	Cold-water	3	5	Positive	Neutral
<i>Oithona similis</i>	Year-round	11	11	Negative	Neutral
<i>Paracalanus</i> sp. C. AC-2013	Warm-water	33	33	Neutral	Positive
<i>Calanus pacificus</i>	Warm-water	5	5	Negative	Negative

Table 2.1: Trends of zero-inflated beta regressions relating dissolved oxygen concentration to eDNA index for different species of copepods. Regressions were computed using the TH042 mooring data from 2021-22, and again using LiveOcean model data from 2021-23, so both are shown. See Appendix B for plotted regression results. Cold-water copepods are northern copepods, warm-water copepods are southern.

levels according to a zero-inflated beta regression.

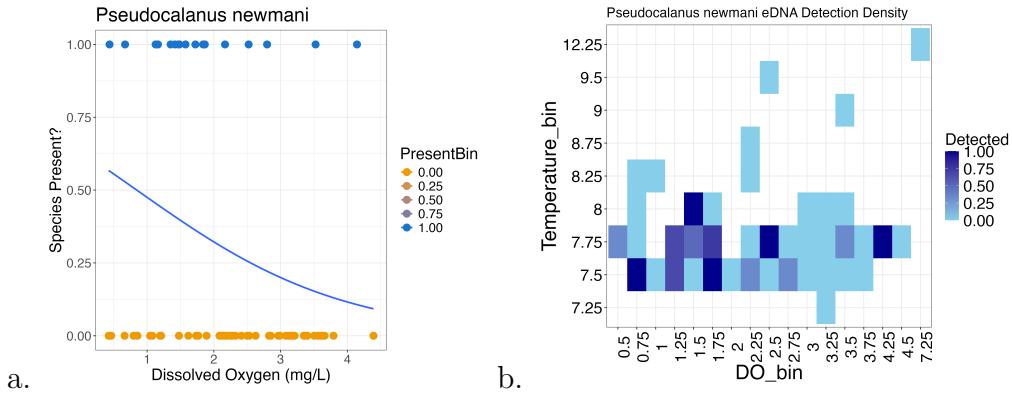


Figure 2.1: (a) *P. newmani* binomial regression comparing species presence to dissolved oxygen. $P = 0.025$, significant at alpha = 0.05. *P. newmani* was detected significantly less frequently at higher dissolved oxygen levels. (b) Density plot of *P. newmani* detections. Each pixel represents a range of dissolved oxygen and temperature values, and the shade of each pixel represents the proportion of eDNA samples in that environmental range in which *P. newmani* was detected. Dissolved oxygen data comes from the TH042 mooring in 2021-22.

Small copepod species showed a higher hypoxia tolerance than large copepod species in this study. Large and huge copepods (*C. pacificus* and *M. pacifica*) were rarely detected below 0.9 mg/L DO, although *C. pacificus*, the only huge (larger than 4 mm) copepod in the dataset, was detected once below 0.6 mg/L DO with a high eDNA index. Medium copepods were rarely detected below 0.9 mg/L DO compared to small copepods (See Figure 2.2b). Small copepods were frequently detected below

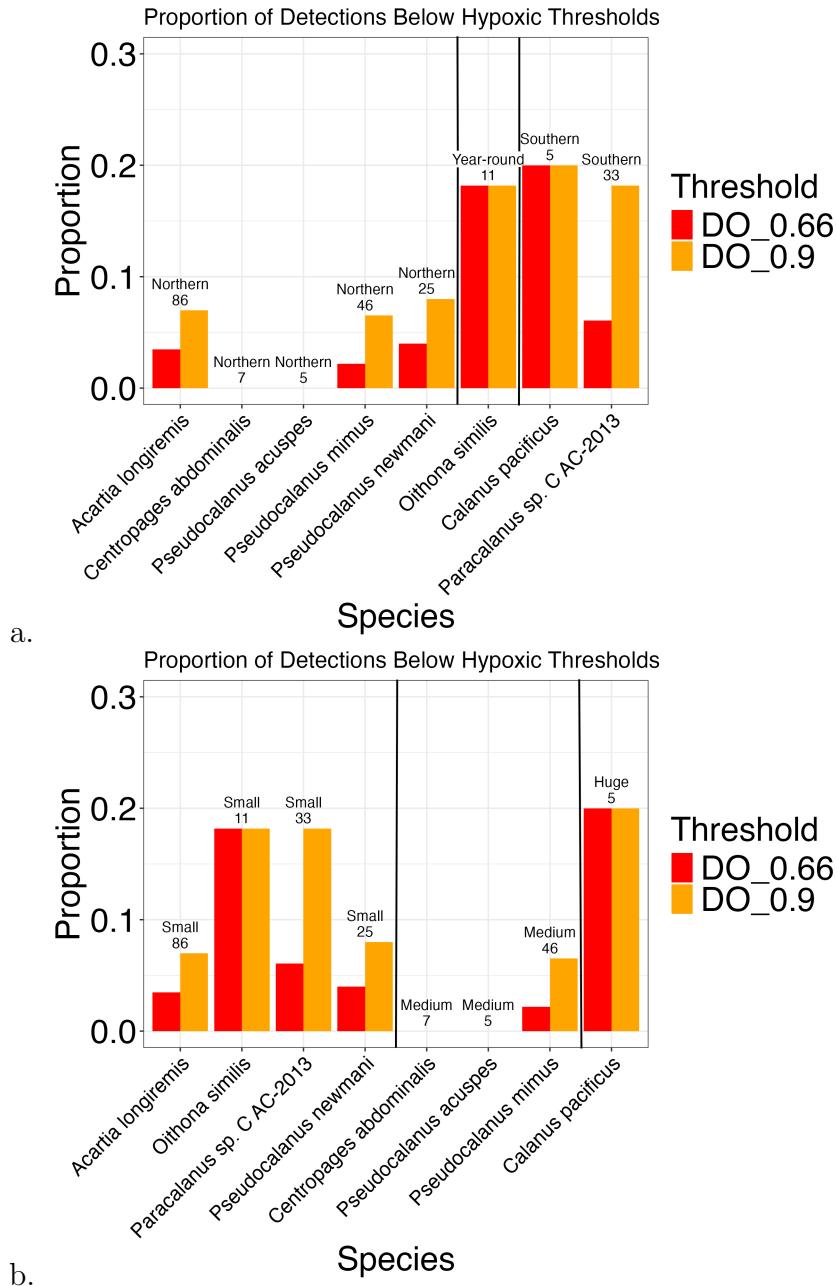


Figure 2.2: A comparison of different copepod groups and their rates of occurrence below certain hypoxic thresholds. Red bars indicate the proportion of detections of that species that were in conditions below 0.66 mg/L of oxygen, and orange bars indicate the proportion of detections of that species that were in conditions below 0.9 mg/L of oxygen. Numbers above bars indicate total number of detections. In plot (a) copepods are also labeled by seasonal groupings. In plot (b), copepods are also labeled by approximate size. Species with a maximum size of less than 1.5 mm are Small, maximum size 1.5-2.5 mm is Medium, maximum size 2.5-4 mm is Large, and maximum size >4 mm is Huge (WoRMS - World Register of Marine Species - Copepoda 2024)

0.6 mg/L DO, especially *A. longiremis* and *Paracalanus* sp. (see Figure 2.2b). *O. similis*, a small copepod, was present below 0.6 mg/L DO, but generally had a higher

eDNA index at higher values of DO.

2.3 Temperature

During this study, northern copepods were generally detected at lower temperatures than southern copepods. Most of the northern copepods in this study were not detected at temperatures above 8 °C, except for *P. mimus*, which was detected above 8 °C but was not detected at the highest temperature in the dataset, 12.2 °C. The southern copepods were primarily detected above 7.5 °C, and the southern copepod *Paracalanus* sp. C AC-2013 had higher eDNA index values measured above 8 °C, despite being detected many times below 8 °C.

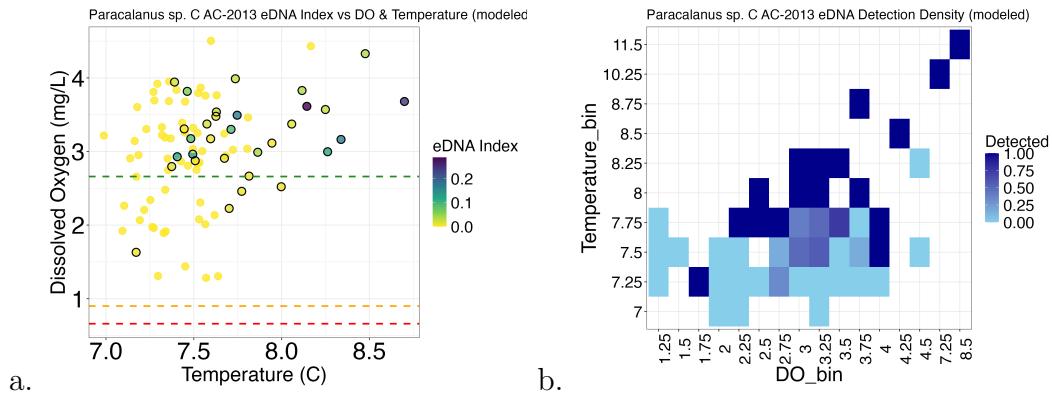


Figure 2.3: (a) Scatter plot of dissolved oxygen (mg/L) and temperature (Celsius) for each environmental DNA sampling time, with *Paracalanus* sp. C AC-2013 scatterplot detections circled in black. Color of dots represents eDNA index, a normalized measure of abundance that can compare abundance within a species. Yellow, uncircled dots represent sampling dates when *Paracalanus* sp. C AC-2013 scatterplot was not detected. (b) Density plot of *Paracalanus* sp. C AC-2013 detections. Each pixel represents a range of dissolved oxygen and temperature values, and the shade of each pixel represents the proportion of eDNA samples in that environmental range in which *Paracalanus* sp. C AC-2013 was detected. These plots use LiveOcean model data from 2021-23.