Work Term Report II

Riley Klaver

B00736874

Data Analyst

Department of Fisheries and Oceans, Ocean and Ecosystems Science Sector

Supervisor: Dr. Kevin Sorochan

Employed from January 2nd, 2019 to April 17th, 2019

**Summary**

Due to the North Atlantic Right Whale (NARW) distribution changing, a cruise was deployed by Fisheries and Oceans Canada in October 2018 to gather data on zooplankton abundance in the southern Gulf of St. Lawrence using a Video Plankton Recorder. Methods were developed to process this data, including image filtering and classification of images, and the data was processed accordingly. The resulting set of classified images was used to qualitatively describe trends in the vertical distribution of the most abundant zooplankton taxa and suspended particulate material, including copepods of the genus *Calanus*, which are key NARW prey. The methods developed, and some information on vertical distribution, will be used to plan for fieldwork and facilitate data processing in the future.

It was found that small, weak-swimming, herbivorous copepods were associated with the mixed layer in the upper water column, strong swimmers such as krill and *Calanus* spp. were found throughout water column, *Calanus* vertical distribution varied considerably based on geographical location, and that chaetognaths were found mostly near the bottom, below the pycnocline. A strange type of detritus, roughly 2 mm long and stick-shaped, was found to be very abundant in surface layers of water, possibly associated with the mixed layer or with closeness to shore.

The work term afforded an extensive opportunity for learning about science in the real world, improving clarity of communication, organisational ability, and learning to work in the R programming language, all in a friendly and collaborative work environment. The work term was of immense value, with the only two negative aspects being an apartment lifestyle and bus rides. This type of employment was extremely interesting and challenging, and would be highly recommended to any other student with a strong background in oceanography and a basic knowledge of coding.

**General Information about the Bedford Institute of Oceanography**

The Bedford Institute of Oceanography was established in 1962 as the communal centre for high-caliber ocean-based research on any and all topics for all of the Maritimes. The building has been expanded significantly since its conception, and the employees have undergone many hierarchical restructurings as well. It currently houses many branches of the government, including the Department of National Defence, the Canadian Coast Guard, the Canadian Hydrographic Service, and the Department of Fisheries and Oceans (DFO). All of these departments have many sectors and divisions, with a very extensive hierarchical structure. I was hired into the Ocean and Ecosystem Science Division of DFO, and the team that hired me is led by Dr. Catherine Johnson. Her lab employs Dr. Catherine Brennan, Dr. Kevin Sorochan, and others with whom I did not work closely. Dr. Brennan supervised Emma Leitao (another co-op student), while Dr. Sorochan supervised me. I was hired to help Dr. Sorochan develop methods to process, plot, and interpret a data set on zooplankton, and to then process that data set accordingly. Marc Ringuette was a notable individual not officially employed by Dr. Johnson’s lab who helped extensively with proper taxonomic classifications.

The dataset that I worked on was collected to determine parameters concerning zooplankton ecology in the Gulf of St. Lawrence; in particular, spatial variation in copepods of the genus *Calanus* (*C. finmarchicus*, *C. hyperboreus*, *C. glacialis*), due to their importance as a food source for the NARW, *Eubalaena glacialis*. *Calanus* spp. are large herbivorous copepods and contain an energy-rich oil droplet, making them nutritious food for higher trophic levels. The presence of NARW in the traditional feeding grounds on the Scotian Shelf and in the Gulf of Maine appears to be related to the abundance of *Calanus* spp., and the distribution of NARWs on feeding grounds is believed to have changed, with potentially more individuals in the Gulf of St. Lawrence (GSL) in recent years (Meyer-Gutbrod et al. 2018). The whales require localized high concentrations of food to feed efficiently, and little is known regarding the processes that may concentrate food in the southern GSL.

**Detailed Project Description**

In response to the believed shift in NARW distribution, data on small scale spatial variation in zooplankton abundance was obtained on a cruise in fall of 2018 in the southern GSL with a Visual Plankton Recorder (VPR) at multiple locations. This data was taken on the scale of hours to days, and metres to kilometres. Zooplankton abundance on this scale can be affected by life history, food distribution, zooplankton behaviour, and flow fields. For example, the life history of *Calanus* spp. includes a diapause stage whereupon they descend to depth in the winter. Small copepods are usually herbivorous, and thus tend to stay near the surface to feed on algae (Longhurst 1985). Behaviour includes activities such as swimming, where weaker swimmers will be more confined by stratification, thus not able to cross the pycnocline as easily or as frequently. Flow fields can interact with swimming behaviours to spread weaker swimmers across the mixed layer, or to concentrate zooplankton at water mass fronts (Franks 1992).

My task on this work term was to develop methods to process the data from the VPR, and qualitatively characterize spatial distributions of the zooplankton (including, but not limited to *Calanus* spp.), in order to glean a basic understanding of the vertical and horizontal structure of abundant taxa in the zooplankton community. My work provided a rare insight into small-scale (1-1000 m) vertical and horizontal variation in zooplankton and suspended particulate matter, pre-emptively expediting processing of further data, and providing critical information with regard to plans and preparations of future field work on zooplankton spatial variations in the southern GSL.

**Detailed Description of Duties and Techniques**

***Methods***

*Data Collection*

A VPR is a camera designed to capture microscopic images of plankton. The camera is attached to a computer and a Conductivity Temperature Depth (CTD) package and fluorometer. The camera faces a strobe light and the camera takes photos only when the strobe flashes. The strobe serves as the primary illumination of the undisturbed water between the camera and the strobe. The resolution used in this case was 1024x1024 pixels equaling 24x24 mm, with a depth of field assumed to be 24 mm. This allows simultaneous capture of zooplankton abundance and water property data on very small scales (< 1 m). A VPR is one of the very few instruments capable of recording zooplankton data on such fine scales, and without sampling destructively, which makes this dataset rare and highly informative.

The VPR was deployed at 7 locations in the southern GSL (Figure 1). At each location the instrument was towed along a transect in a “towyo” pattern for 1-2 hours while the ship moved forward at a speed of ~ 1 m s-1. A single transect was conducted at 4 locations (stations 4.85, 6.6, 7.45, 7.6). Two transects were conducted at a site in Shediac Valley (station ‘SHED’), and a total of 7 transects were conducted at two sites near Cap d’Espoir (stations ‘Cap’) over a period of 48 hours to evaluate variations in zooplankton distribution over time.

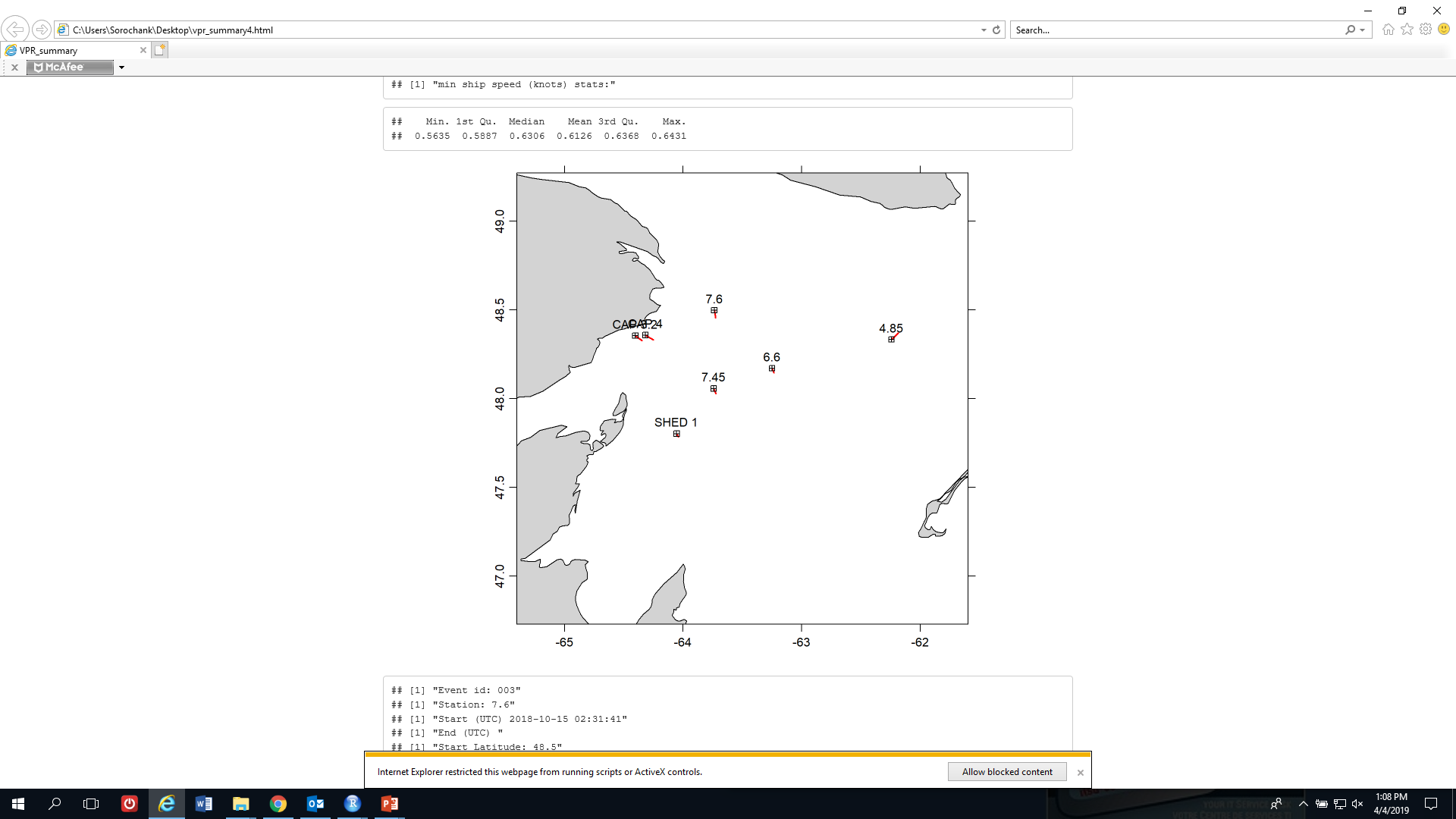


Figure 1 Map of the southern Gulf of St. Lawrence showing VPR transects; Gaspé Peninsula in top left corner

Data processing involved filtering the video footage from the VPR, and extracting salinity, temperature, pressure, and fluorescence data using a software called Autodeck. The filtered images, referred to as “regions of interest” (ROIs), were then sorted into categories based on taxa or morphotype using Visual Plankton (VP), an automated image classification software in Matlab. After automated sorting, ROIs were validated for their correct identification and incorrect classifications were removed to ensure accuracy. The sorted images and corresponding water property data were combined and plotted in the R programming environment to visualize spatial variations in zooplankton and their environment. A large component of my project was on the assessment and development of methods for data processing within the framework described above.

Image filtering

*Basic functioning of Autodeck*

Investigation of basic functioning in Autodeck was crucial in order to ensure high data quality; and to eventually develop a method of data extraction requiring as little human input as possible. In this process, I determined 5 main things: which ROIs were saved if 1) the filtering process was manually stopped; 2) the program “crashed”; 3) duplication of ROIs was attempted; 4) ROIs were saved out of chronological order; and 5) there are limitations with regard to the maximum number of ROIs that could be saved in a folder.

“Crashes” were caused by 1-4 consecutive frames of footage which would cause the Autodeck interface to disappear completely, requiring a full re-boot of the program. The frames causing a crash always stayed the same in a given cast, and are thought to be caused by localized glitches due to an old USB. Autodeck would save everything up to when it crashed, with a few caveats. If stopped, whether via crash or manual stop, Autodeck would randomly skip some frames and not record any ROIs there in. It also may record some ROIs from the frames causing the crash, but this cannot be determined for certain. When frames are processed twice, Autodeck simply overwrites the image, but corresponding data on water properties is duplicated.

The maximum number of ROIs that a folder could contain was found to be 21844; this is a limitation of the PC, not Autodeck, but Autodeck was unable to record any further images in that folder if the limit was reached. There were 2 instances when the maximum storage capacity of a folder was reached. In these instances, the remaining data needed to be considered by Autodeck as a separate transect, and new folders were created to obtain the missing sections of the transects.

The ROI files produced by Autodeck are named according to their timestamp, which is measured in milliseconds since the start of the Julian day, and corresponds directly to the timestamps of CTD observations. These timestamps do not correspond to internal directory structures, local time, or time in UTC, but are still usable to determine durations of transects and to associate CTD data and ROI data. In summary, Autodeck functions very well for its specified task when used correctly, and the only persistent issue was the imposed directory structure caused by full folders.

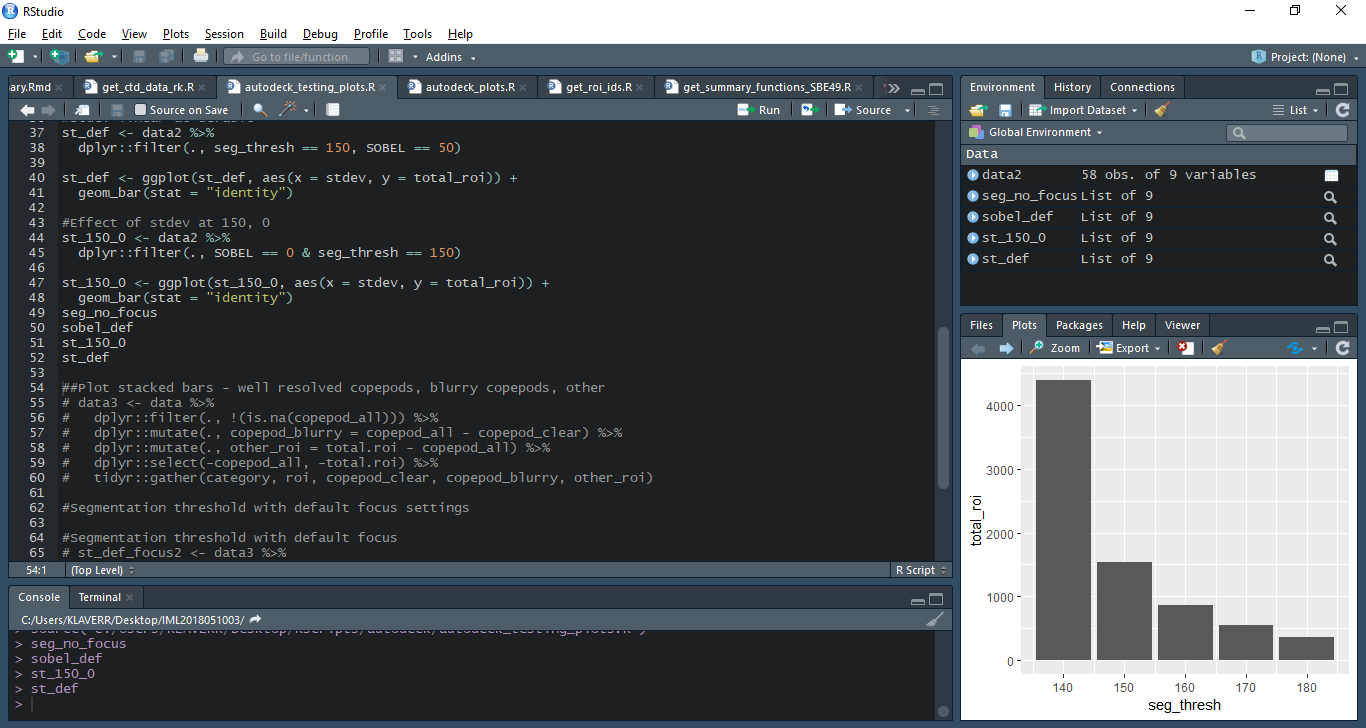
*Sensitivity of number of ROIs to filtering settings*

Image quality can vary drastically depending on water clarity. In order to retain as many useful images from the VPR as possible, while excluding unidentifiable ones, filtering settings needed to be optimised for our specific VPR; due to staying in relatively similar areas of the GSL, only one collection of filtering settings was used to filter all transects taken. Images were filtered in Autodeck (Seascan Inc. 2008) based on the following thresholds: “segmentation threshold” (light intensity), “SOBEL” (“fuzziness” of images, indicating out-of-focus images), and “St.Dev.” (“blown-out” images, indicating loss of highlight detail due to overexposure). Increasing each one of these settings increases the required threshold for an image to be passed on to the next filtering step, and decreases the total number of ROIs found in each video. To determine the sensitivity of the number and quality of ROIs produced during the image filtering process, the number of copepods (clear, no fuzziness), fuzzy copepods, and total number of ROIs was enumerated throughout a range of threshold values for each type of setting. A less rigorous test was conducted on the full range of settings to examine the software’s capabilities, and a rigorous test was performed on values between -20 and +20 (inclusive) of the default settings, in increments of 5. Decreasing either the segmentation threshold or the SOBEL caused an exponential increase in number of ROIs obtained (Figure 2AB). Decreasing the St.Dev. causes a much slower exponential increase in total numbers of ROIs, which is only noticeable at very low SOBEL levels (Figure 2C); at 50 SOBEL, decreasing the St.Dev. appears to only have a linear effect (Figure 2D). This is due to St.Dev being the third of three filters, there being high correlation between fuzzy and blown-out images, and the relative rarity of blown-out images.

The final settings chosen (segmentation threshold, 150; SOBEL, 20; St. Dev., 10, based on number of copepods missing if settings were increased) caused rejection of some dark and fuzzy copepods, but allowed the vast majority of copepods of the genus *Calanus* spp. to be captured. The relatively low SOBEL threshold caused a much higher retention of photos of turbid water, unidentifiable matter, and fuzzy images than desired, but was deemed necessary to avoid a loss of about 128 *Calanus* copepods/transect (specifically, 11 *Calanus* copepods/6000 frames). The St.Dev setting was left at the default value of 10 due to a lower setting causing retention of fuzzy ROIs, but anything higher showing no difference in reducing the total number of ROIs.

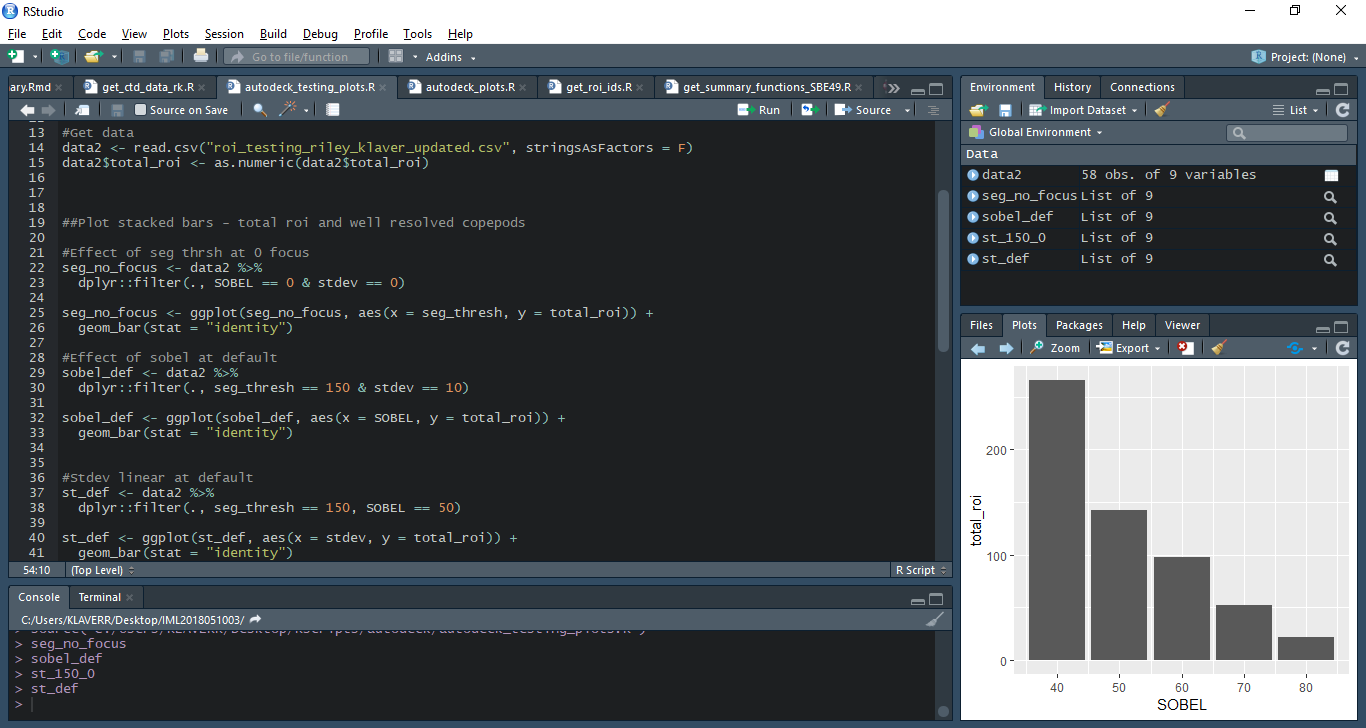
A B

C D



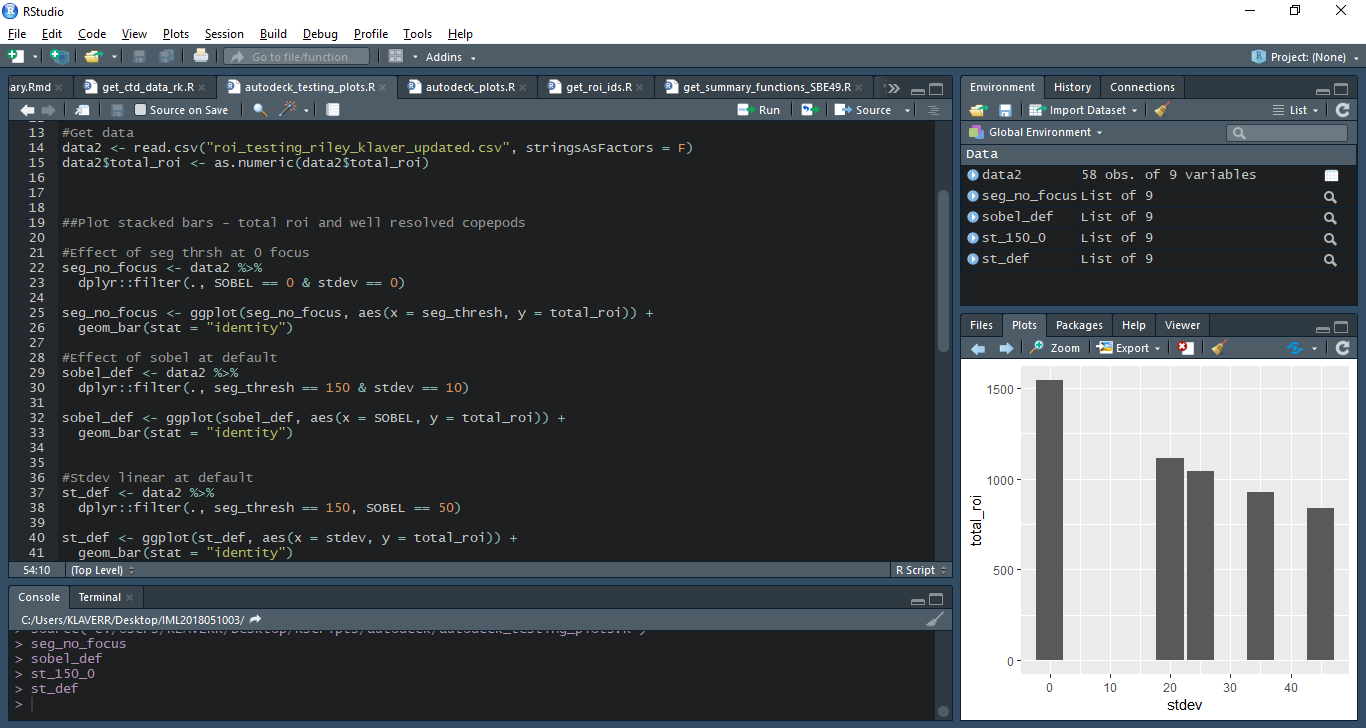
Segmentation threshold

Total ROIs obtained



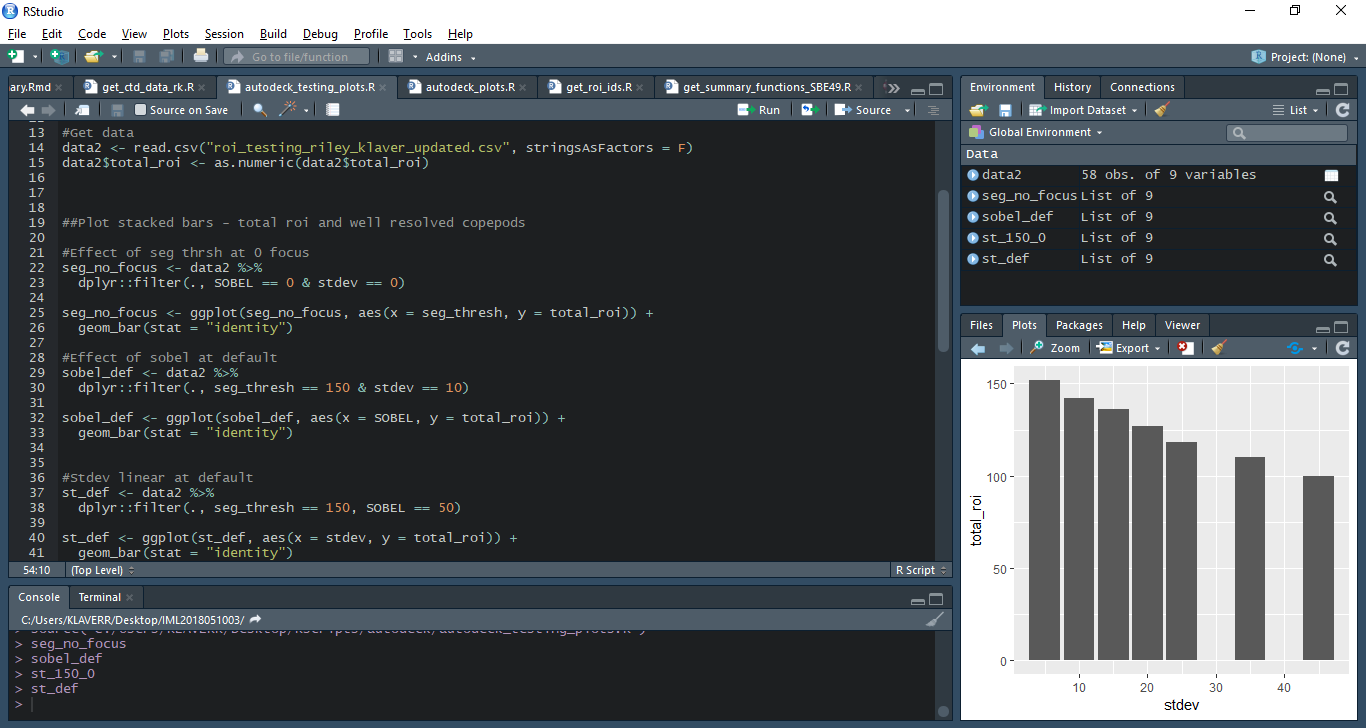
SOBEL

Total ROIs obtained



St. Dev

Total ROIs obtained



St. Dev

Total ROIs obtained

Figure 2 Counts of total ROIs obtained across 2000 frames of video by A) varying segmentation threshold at 0 SOBEL and 0 St.Dev. B) varying SOBEL at segmentation threshold 150 and St.Dev 10 C) varying St.Dev. at segmentation threshold 150 and SOBEL 0, and D) varying St.Dev. at segmentation threshold 150 and SOBEL 50

Imageextractionprocedure

Given the image filtering settings stated above, images were extracted from as much of each VPR transect as possible. A transect was considered to be from the moment the VPR started descending after soaking to the moment before it reached the surface on the final upcast. This definition of a transect avoided capture of extraneous ROIs (*i.e*. contamination) when the VPR was not in the water and soaking in the upper ~5 m of the water column prior to its descent. Image acquisition was manually halted to avoid contamination if the VPR touched bottom, suspended sediment near the bottom, or broke the surface.

Image classification

Automated classification of ROIs was carried out in VP using a combination of Support Vector Machine (SVM) and/or Neural Net (NN) supervised machine-learning techniques. In these algorithms, the computer must first be “trained” using ROIs that have been manually sorted into the categories of interest prior to classification. This training set, with 20 categories containing between 1 and 1700 images in a category was provided by Dr. Sorochan prior to running VP.

It is worth noting that the VP software had no associated documentation or manual. I was responsible for finding many of the connections between the program’s various scripts, while Dr. Sorochan ensured that the program ran properly. This was a large task, which required comprehension of vast amounts of Matlab code.

The accuracy of the following 3 classification methods was tested: NN, SVM, and NN and SVM combined (*i.e.* “Dual”). In the case of the NN and SVM disagreeing under the Dual regime, the image in question would be classified as ‘unknown’. Since *Calanus* spp. were of highest interest for further research, the classification accuracy of this category was used to determine which classifier was used for my project as well. It was found that the Dual method had the highest accuracy: only 18.8% of images categorised as *Calanus* spp. were false positives, 203/5914 photos in ‘unknown’ were false negatives, and 52/5186 remaining images were false negatives; while using the SVM, there was a 53.1% false positive rate, and 255/10937 remaining images were false negatives; and finally for the NN, there was a 56.1% false positive rate, and 199 of remaining 10783 images were *Calanus* spp.

Since the Dual classifier had the lowest rate of false positives, it was used to classify the rest of the images obtained from Autodeck, in order to obtain the best results in the timeframe of my co-op term. The “unknown” folder generated by the Dual classifier not analysed; therefore, concentrations of taxa and detritus that were calculated were underestimated. Following all image classifications, and reviewing the success of all the categories classified, it was determined that only 7 categories were accurate enough and abundant enough to analyse: *Calanus*, *Chaetognatha*, Euphausiacea, larval echinoderms, small copepods (the summation of *Temora* sp., *Triconia* sp., and *Oithona* sp., and other unknown copepod species), marine snow (detritus aggregates), and sticks (a roughly 2 mm long, stick-like type of detritus) (Figure 3). Incorrect images in groupings of *Calanus*, chaetognaths, krill, and larval echinoderm were removed manually, but not the other categories, due to containing too many images to process in a short amount of time, and being of a lower priority level due to containing insufficient high-quality data capable of answering the target question.

A B C

C:\data\cruise_IML2018051\Everything_classified\marine_snow\official1dualaid.d286.h22_ROIS\roi.0539351800.tif

D E F G



Figure 3 VPR ROIs of A) krill B) marine snow C) *Calanus* D) Chaetognath E) Echinoderm larva F) small copepod (with ovisac) and G) a stick

*Data processing and plotting*

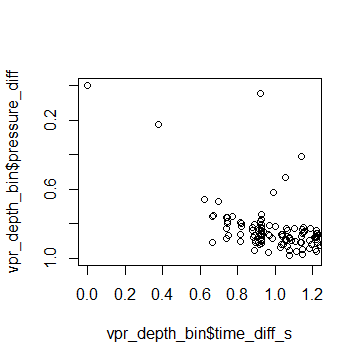
The objectives of data processing and plotting were to calculate concentrations of categorisations, remove any unjustified outliers, and visualize the data for qualitative analysis. The final product of image classification is a text file with a list of ROI identifiers (*i.e.* time stamps in milliseconds), with separate text files for each category of interest. The depth (and associated water properties) of each ROI was determined by matching the ROI identifier to the time in which water properties were recorded by the CTD and fluorometer. Once these data were assembled, we quantified the concentration of ROIs in 1 meter depth bins for each taxon of interest using the following equation:

concentration= N/(V x r x ∆t)

where N is the number of ROIs in a depth bin, V is the image volume, r is the framerate of the camera, and ∆t is the time spent in each depth bin. Depth bins with time differences between 0 and 0.8 seconds had artificially inflated zooplankton concentrations due to small times spent in a bin. These data points were removed, but this only constituted the loss of a few data points (Figure 4). This artificial inflation occurred at the top or bottom of a cast when the VPR dipped into the shallowest or deepest depth bin for only a short period of time or travelled through a bin very quickly.

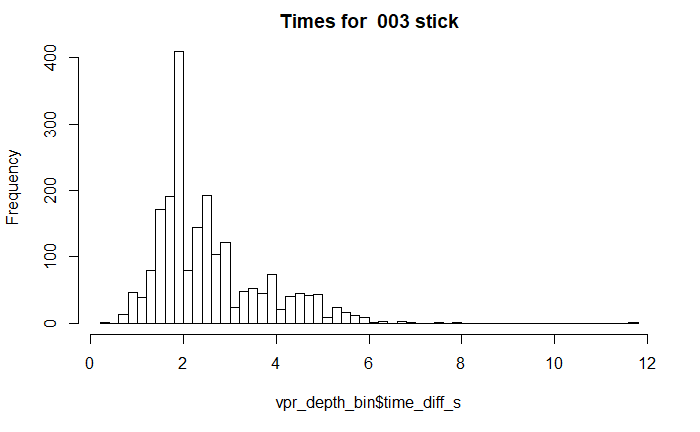
B

A



Time difference across depth bin (s)

Pressure difference across depth bin (dbar)



Time difference across a depth bin (s)

Frequency

Figure 4 A) Time difference across a bin vs. pressure difference across a bin (zoomed in along x-axis) B) histogram of time differences across depth bins

Framerate, *r*, was estimated to be 13 fps, and was calculated as an average rate across all transects. In each transect, the framerate was calculated by dividing the total number of frames from a given cast by the total time of the cast. Framerates ranged from 12.4 fps to 14.4 fps (Figure 5). Using one single framerate for all casts will overestimate concentration for transects with a higher framerate, and underestimate concentrations for transects with a lower framerate. The image volume, *V*= 13.8 x 103 mm3, was calculated assuming a cubic geometry, since the depth of field estimation is still pending from Seascan.

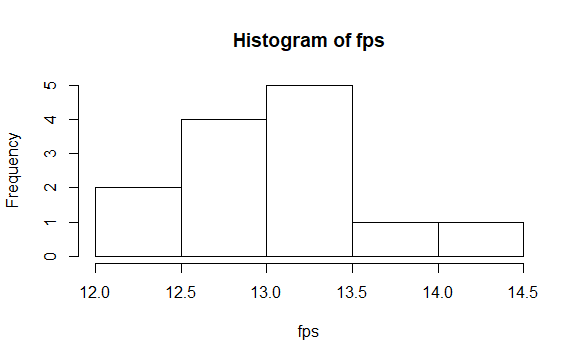


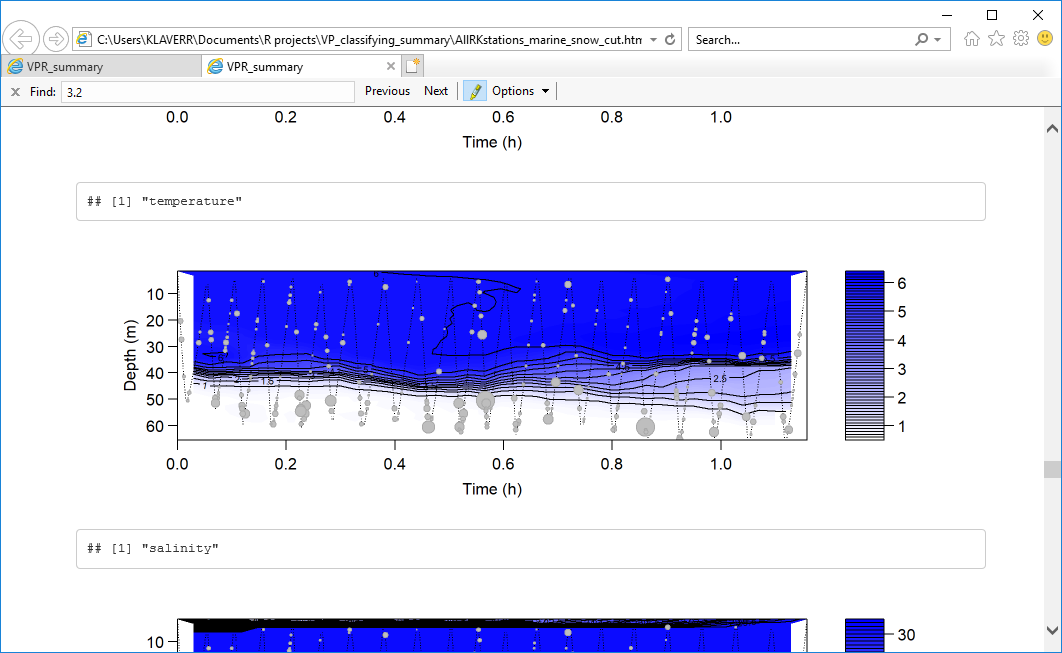
Figure 5 Histogram of calculated framerates of the camera during different transects

This method of calculating concentration assumes that there is no overlap between successive image volumes. To justify this assumption, the minimum forward ship speed required to prevent overlap of image volumes was calculated as:

Ship speed = r x d x q

Where *Ship speed* is the minimum forward velocity in knots, *r* is the framerate, *d is* the length of a frame (24 mm), and *q* is the conversion factor (1 knot/ 514.4 mm s-1) is the conversion ratio. The highest minimum forward speed between all casts was estimated to be 0.67 knots, which was well below the 2 knot towing speed.

Spatial variation in concentration was visualized along the VPR path by plotting concentration as a bubble with its radius proportional to magnitude of concentration, over interpolated fields of water properties (density [sigma-theta], temperature, salinity, fluorescence) (Figure 6). The largest grey bubble is the highest concentration observed of that particular taxon at that particular station. Vertical profiles of 1-m binned concentration and water properties and their averages were also plotted to analyse vertical variation over the entirety of each transect (Figure 7). Bubble plots of horizontal and vertical variation were evaluated for the Cap stations, because these transects crossed substantial gradients in bathymetry and water properties. Vertical profiles were evaluated for the remaining stations.



T (oC)

Figure 6 VPR path along station 6.6, where grey bubble radius is proportional to marine snow concentration, over an interpolated temperature field in degrees Celsius, as shown by the scale bar.



Figure 7 Vertical profiles of water properties (temperature, T, salinity, S; Fluorescence, F; and Density, ρ) and abundance of *Calanus* spp. at station 7.6. Solid lines are 1-m depth bin averages across the whole towyo, and black dots represent horizontal variation in concentration in each depth bin.

***Results***

*Vertical distribution*

Stations 7.6, 7.45, 4.85, 6.6, and SHED 1 all showed limited within-transect horizontal variation in water properties, abundance of zooplankton, and particulate matter. Therefore, patterns only of vertical distribution were described at each station, and comparisons between these transects were based only on vertical distribution. Patterns in the vertical distribution of *Calanus* spp. varied strongly among transects. For example, abundance was highest near the surface (Shed 1), in midwater below the pycnocline (6.6), or near the surface and at depth (4.85) Small copepods were most abundant above the pycnocline in the mixed layer, associated with fluorescence. At stations 4.85 and 7.45, however, small copepod abundance was low and the distribution was even. Krill were usually evenly distributed, but had a bimodal distribution at stations 7.6 and 7.45. Chaetognaths were usually most abundant between the bottom and 20 m above the bottom, and although station 7.6 had chaetognaths in the middle of the water column, they were still found below the pycnocline. Stations 4.85 and SHED 1 had very low abundances of chaetognaths. Larval echinoderms were always most abundant near the surface, with their lower limit corresponding to the pycnocline. However, 4.85 and SHED 1 had very low abundances of echinoderms. Sticks were usually most abundant near the surface, with the limit of their highest abundance being the pycnocline, although some sticks were found at all depth ranges at all stations. Sticks at 6.6 and SHED 1 were much closer to an even distribution. Marine snow increased with depth; and at 7.45 and 7.6, the increase appeared exponential. By contrast, marine snow was evenly distributed at 4.85 and SHED 1 (Figure 8).

In almost all cases, vertical distributions of categories at stations SHED 1 and 4.85 were different than those observed at other stations. Station 4.85 is in the Laurentian channel, and is much deeper than other stations (close to 300 m), which could account for some variability. SHED 1 is quite shallow (about 70 m deep), and is also a location where NARW have been sighted frequently. Day/night cycles could also be a large source of variation: *Calanus* spp. appear to have diurnal variation, but the potential effect of day/night could not be tested given the sample size of only 5 transects.

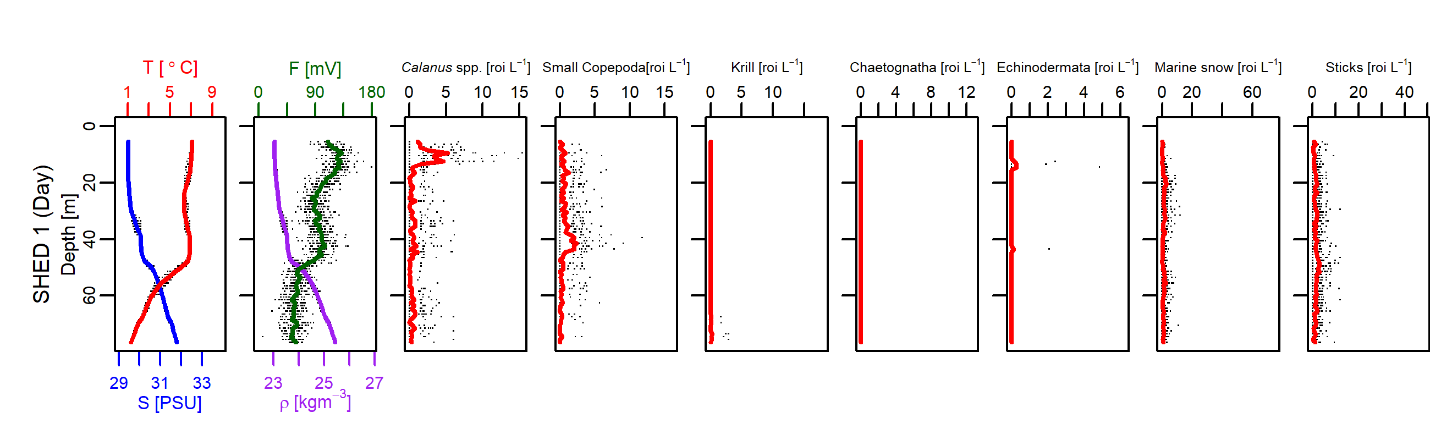
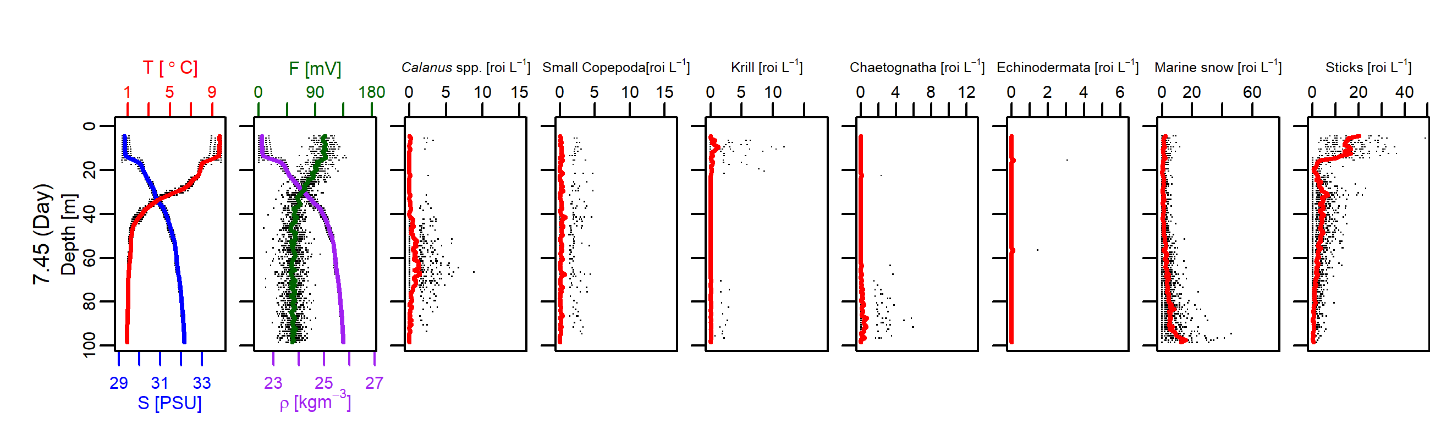
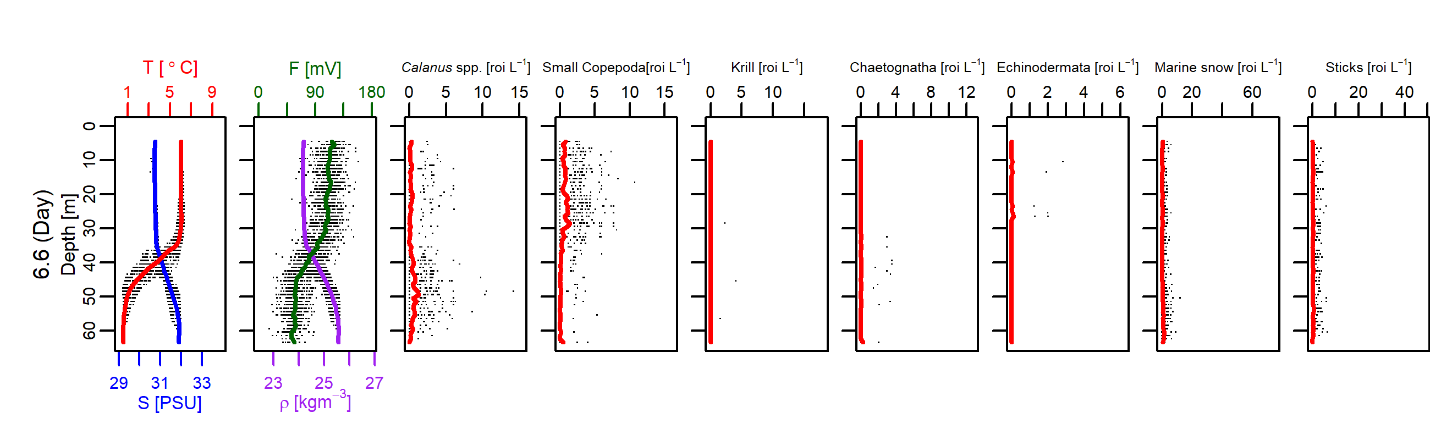
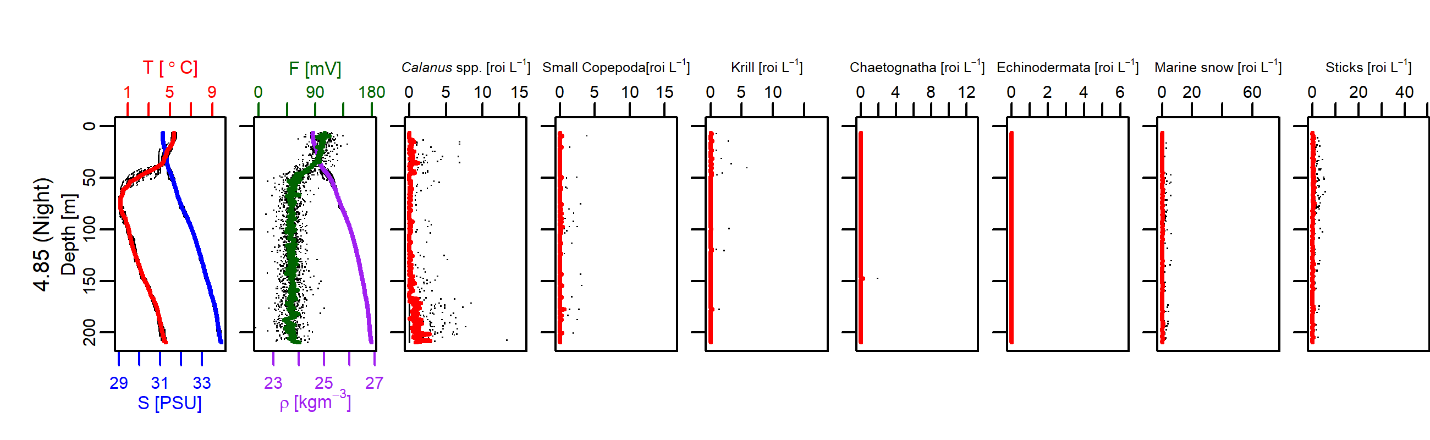
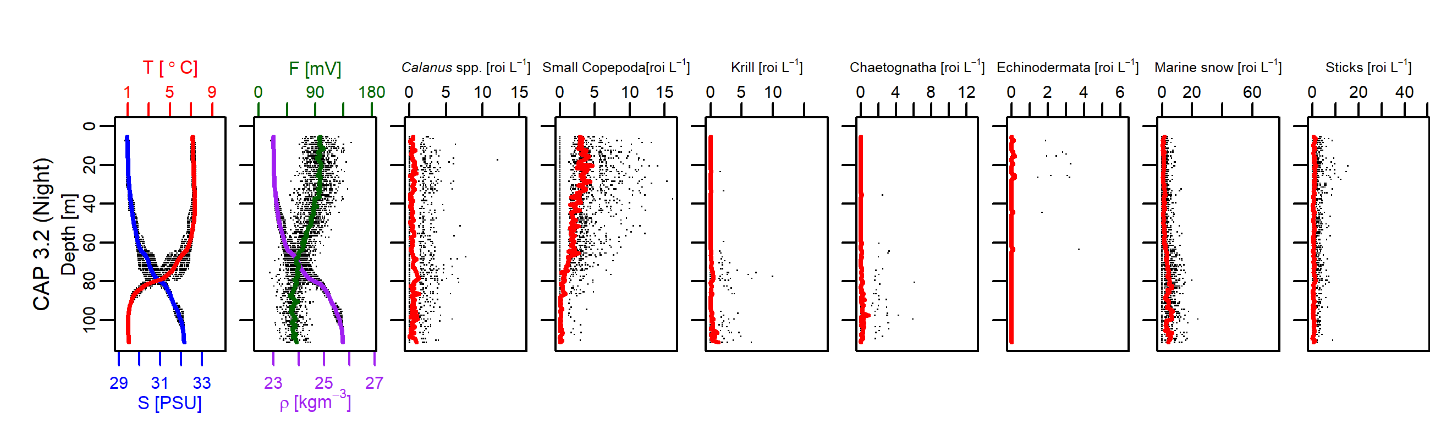
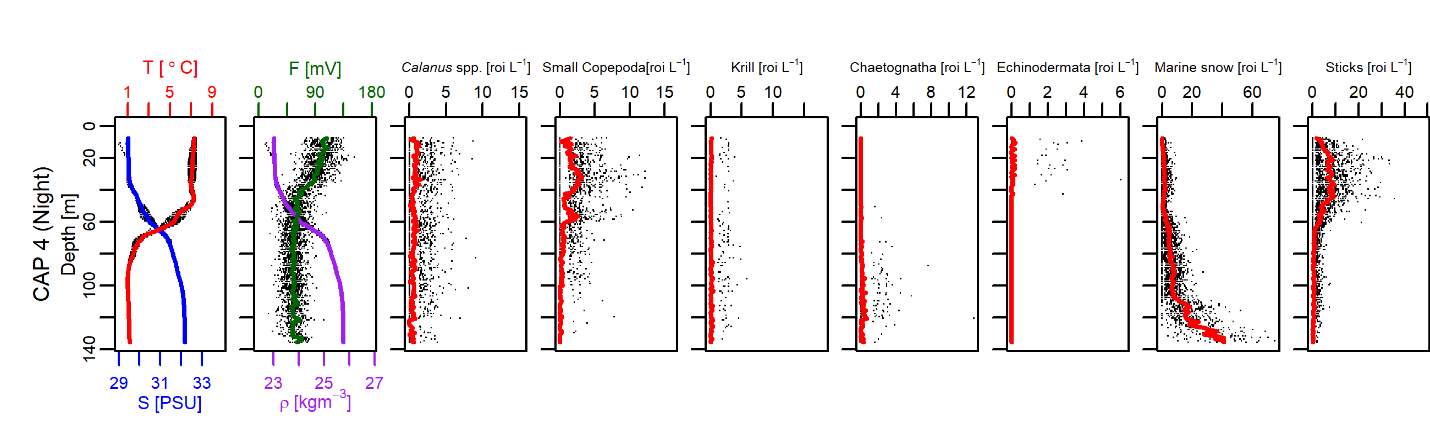
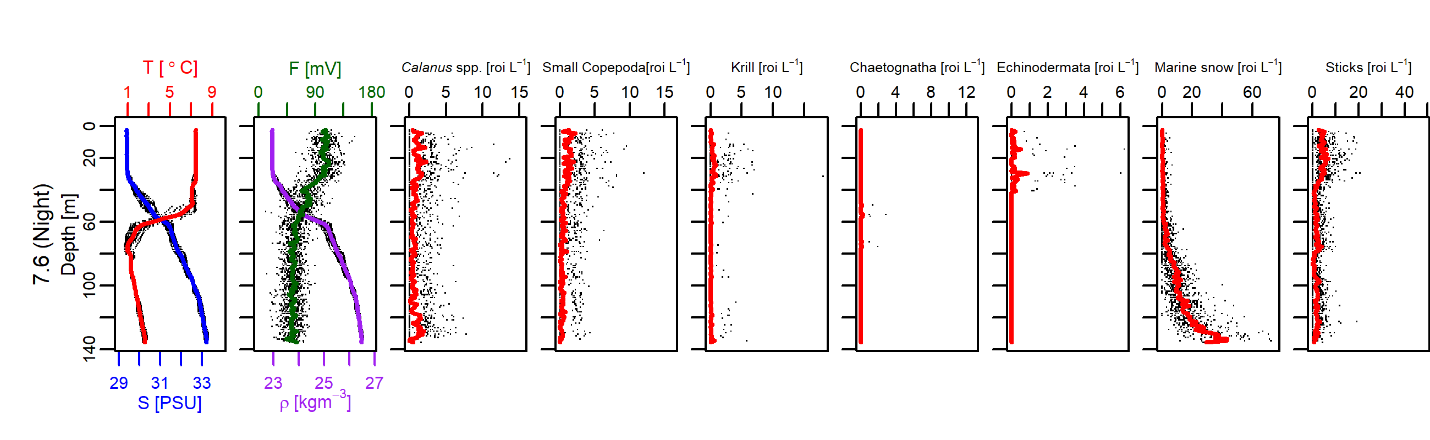
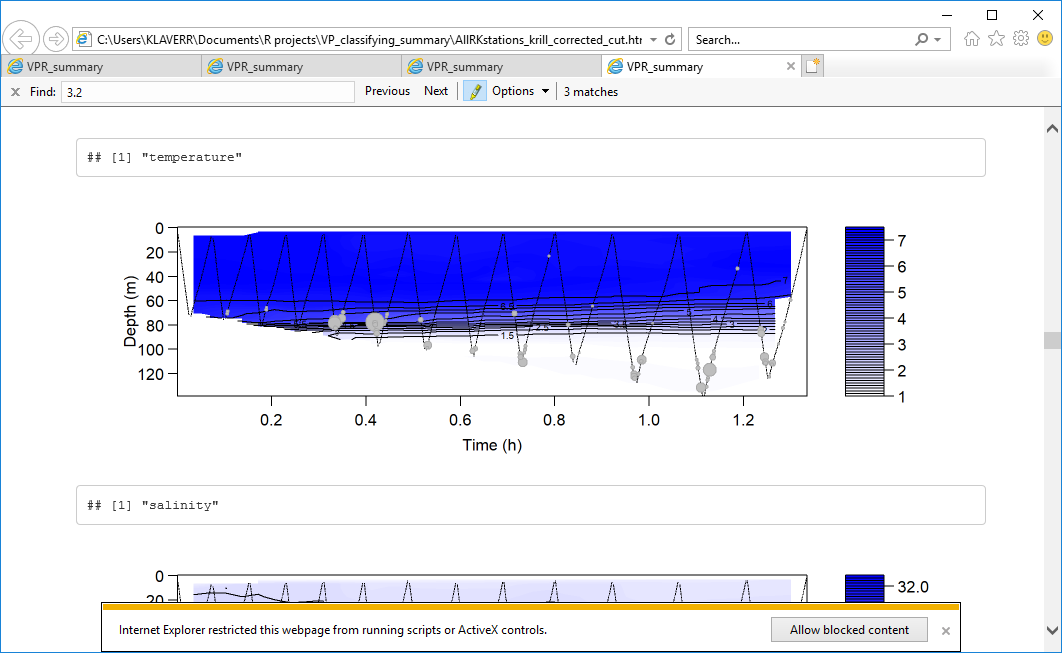
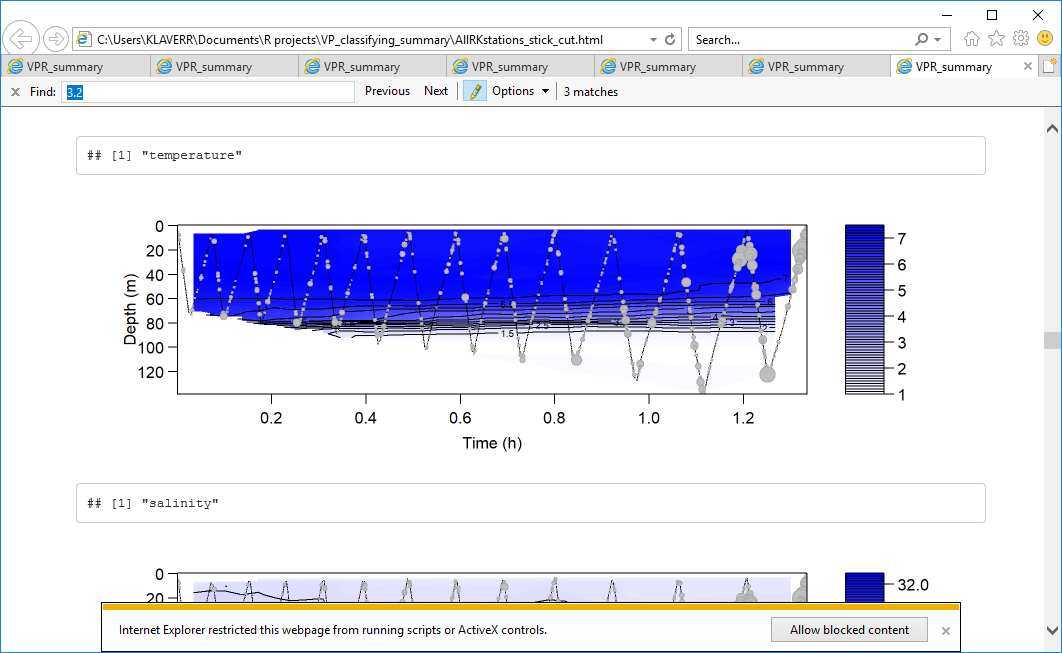
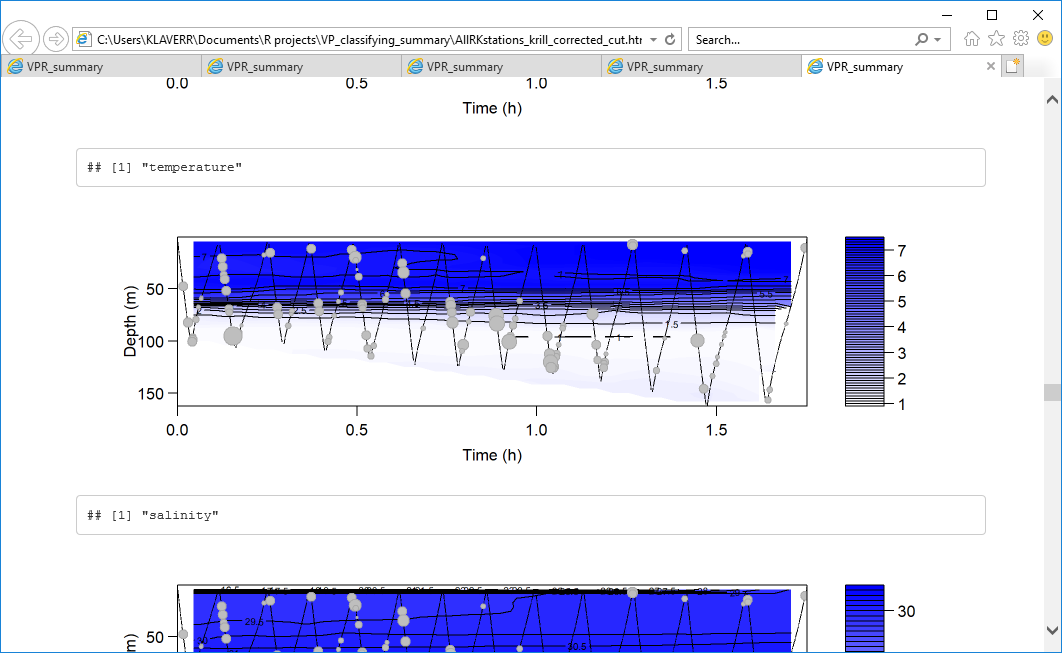
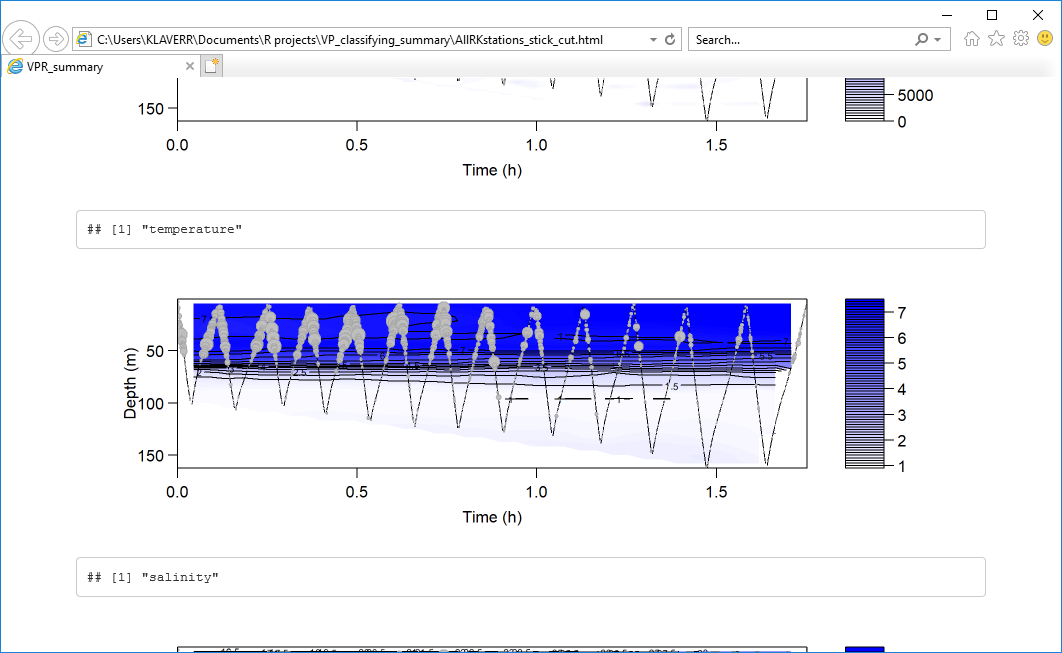
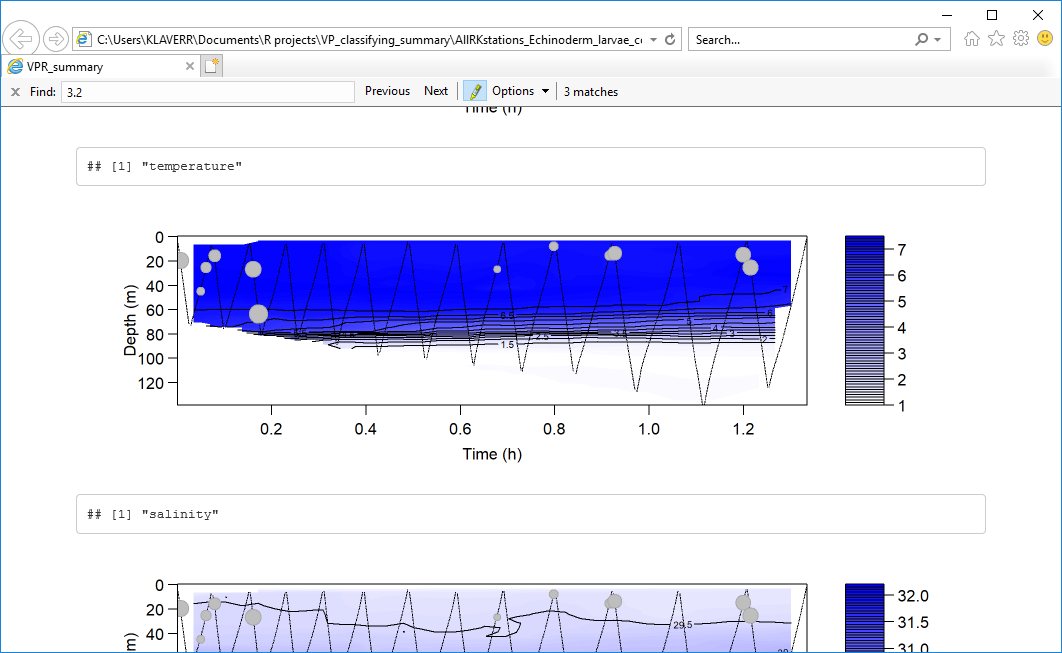


Figure 8 Vertical profiles of water properties (temperature, T, salinity, S; Fluorescence, F; and Density, ρ) and abundance of zooplankton (*Calanus* spp., small copepods, krill, chaetognaths, larval echinoderms) and particulate material (marine snow and “sticks”). Solid lines are 1-m depth bin averages across the whole towyo, and black dots represent horizontal variation in concentration in each depth bin. Each station is a row, transects sampled at night and day are at the top and bottom, respectively. Within day/night groupings, rows are arranged north to south descending with the page.

*Vertical and horizontal distribution (Cap stations only)*

Notable horizontal variation in salinity, temperature, and density was observed at the Cap stations, unlike the other stations. Horizontal variation of depth gradient and taxon distribution was also readily apparent at the Cap stations. Almost all of the vertical patterns described at non-Cap stations remained the same at Cap 3.2 and 4 for each category, although horizontal variation did occur within these patterns. For example, echinoderm larvae were found exclusively above the pycnocline, just as in the flat-bottomed stations, but the larvae were only found at the start and the end of the transect, not in the middle (Figure 9E). Only two exceptions to the vertical patterns occurred: Krill were near the bottom at Cap 3.2, and displayed substantial horizontal and vertical variation at Cap 4 (Figure 9AB). Sticks appeared to be negatively associated with depth, density, salinity, and distance from shore; and positively associated with temperature. This trend is especially clear in Figure 9D.

A

E

D

C

B

T (oC)

T (oC)

T (oC)

T (oC)

T (oC)

Figure 9 Spatial variation in A) concentration of krill at station Cap 3.2 B) Concentration of krill at station Cap 4 C) stick concentration at station Cap 3.2 D) stick concentration at station Cap 4 E) echinoderm larvae concentration at station Cap 3.2. Concentrations are plotted as proportional bubbles over an interpolated temperature field.

**Project conclusions and recommendations**

***Conclusions***

The VPR was very successful for small-scale distributions: lots of fine-scale structure was able to be observed in both zooplankton and CTD data. Small, herbivorous, weak swimmers (Echinoderms, small copepods) were found in the mixed layer, associated with fluorescence as would be expected, while strong swimmers (Krill, *Calanus* spp.) were found throughout water column. *Calanus* spp. distribution has no clear pattern among stations. Chaetognaths were found mostly near the bottom, below the pycnocline. The source of sticks is unclear, and could be fecal pellets or terrigenous detritus. In any case, the abundance of sticks appeared to decrease with distance from shore, depth, density, salinity, and increase with temperature. Finally, stations 4.85 and SHED1 were exceptions to most vertical patterns, possibly due to differences in depth and water column structure.

The goal of this project, to develop methods for VPR data processing and qualitatively characterize spatial distributions of the zooplankton, was met with certainty. It was unclear at the start of term whether there would have been enough time to progress to analysing results in any capacity at all due to the quantity of data to process. All data from all transects has been processed, with software in place to make simple plots of the data as soon it has been processed. There was even time to preliminarily plot and analyse a subset of data.

***Recommendations***

The goal may have been met, but improvements can be made. The folder of ‘unknown’ images from VP should be sorted manually in order to provide more data to plot, and to provide the computer with a second iteration of training images. This will have the effect of reducing both false positives and false negatives. Specifically, a larger library for *Metridia* should be created due to the computer having many difficulties with distinguishing these from *Calanus*. Additionally, there were many rare taxa (1-5 images total) that were not used in the training regime, but may be able to be included with more images. As of yet, the confusion matrix portion of VP is not functional, but the implementation of it will prove very useful in diagnosing categorisation issues.

Intro oceanography classes were critical to even limited success for this work term, and every other higher-level oceanography class was crucial for proper success. Without this background knowledge, many of the papers read during the term would have been unintelligible, and issues during data processing would not have been noticed. A deeper knowledge of coding and machine learning would have proved valuable, although many resources were made available for both these topics. In future, as many more biological oceanography classes (especially zooplankton ecology) as possible would be much needed, as well as courses on coding and computer learning.