Manuscript version following response to reviewer comments.

Published version can be found at:
Scott, J., Pitois, S., Close, H., Almeida, N., Culverhouse, P., Tilbury, J. And Malin, G. 2021. *In situ* automated imaging, using the Plankton Imager, captures temporal variations in mesozooplankton using the Celtic Sea as a case study. Journal of Plankton Research 43(2): 300-313 doi:10.1093/plankt/fbab018.

In situ automated imaging, using the Plankton Imager, captures temporal variations in mesozooplankton using the Celtic Sea as a case study.

James Scott[1,2], Sophie Pitois[2], Hayden Close[2], Nevena Almeida[2], Phil Culverhouse[3], Julian Tilbury[3], Gill Malin[2]

[1] Centre for Environment, Fisheries, and Aquaculture Science, Pakefield Road, Lowestoft, NR33 0HT, United Kingdom

[2] Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ. United Kingdom

[3] Plankton Analytics, Plymouth

james.scott@uea.co.uk

Keywords: Plankton Imager, Image Analysis, Automated Sampling, Mesozooplankton, Celtic Sea

Abstract

The Plankton Imager (PI) is an underway semi-automated, high-speed imaging instrument which takes images of all passing particles and classifies the mesozooplankton present. We used data (temperature, salinity and mesozooplankton abundance) collected in the Celtic Sea in spring and autumn from 2016 to 2019 to assess the ability of the PI to describe temporal changes in the mesozooplankton community and to capture the seasonality of individual taxa. The description obtained using the PI identified both seasonal and interannual changes in the mesozooplankton community. Variation was higher between years than seasons due to the large variation in the community between years in autumn, attributed to the breaking down of summer stratification. The spring community was consistent between years. The seasonality of taxa broadly adhered to those presented in the literature. This demonstrates the PI as a robust method to describe the mesozooplankton community. Finally, the potential future applications and how to make best use of the PI are discussed.

Introduction

The ubiquitous distribution and high abundance of zooplankton makes them fundamental in many ocean processes. They have an essential role in the global carbon cycle and carbon sequestration, regulating the exchange of CO_2 between the atmosphere, surface ocean and ultimately the seabed (Hansell, 2002; Steinberg *et al.*, 2002; Steinberg and Landry, 2017). Zooplankton can be used in global monitoring; providing reliable, sensitive indicators to climate change (Taylor *et al.*, 2002). Furthermore, the adult and juvenile stages of zooplankton are the principal prey for many commercially fished species (Beaugrand *et al.*, 2003; Heath, 2005). Despite this, time-series data for zooplankton are sparse (Mackas and Beaugrand, 2010) and our knowledge of communities is spatially fragmented (Pitois *et al.*, 2016).

At the same time, rising exploitation of our seas is putting increasing pressure on critically assessing and protecting the marine environment (Bean *et al.*, 2017). Resultant policy, such as the EUs Marine Strategy Framework Directive (MSFD – Directive 2008/56/EC), demands increasingly complex metrics for plankton communities (McQuatters-Gollop *et al.*, 2017). However, the capacity to resolve questions posed by policy are hindered by financial ceilings that limit monitoring capacity (Bean et al., 2017; Pitois et al., 2018). To make matters worse, traditional taxonomy by light microscopy, the required analyses on net samples, itself a 'discipline in crisis' (Agnarsson and Kuntner, 2007), is laborious and time-consuming. These

factors place impetus on developing cost-effective methods to obtain sufficient data to accurately describe plankton communities (Danovaro *et al.*, 2016). In response, a range of new devices, often using the latest technology have been developed (Wiebe and Benfield, 2003). For example, acoustic tools can provide high temporal and spatial resolution for assessing total biomass (Wiebe and Benfield, 2003), but cannot answer questions requiring taxonomic information (Stanton *et al.*, 1994; Benoit-Bird and Lawson, 2016). Imaging devices, such as the FlowCam (Sieracki *et al.*, 1998) and ZOOScan (Gorsky *et al.*, 2010) are well established, widely used methods. While these devices can speed up identification and provide data archiving benefits, they are commonly used on captured, preserved samples and therefore suffer the same constraints as the deployment of nets and preservation of specimens.

Semi-automated, in situ, imaging devices take a different approach. Deployable devices such as the Video Plankton Recorder (Davis *et al.*, 2005), Underwater Vision Profiler (Picheral *et al.*, 2010) and The *in situ* Ichthyoplankton Imaging System (Cowen and Guigand, 2008) capture images of passing particles for subsequent classification removing the need for physical sample collection. For a comprehensive review of these devices see: Lombard (Lombard *et al.*, 2019). More recent devices, for example The Scripps Plankton Camera System (Orenstein *et al.*, 2020) and PlanktonScope (Song *et al.*, 2020), are currently employed for routine monitoring. The Plankton Image Analyser (Culverhouse *et al.*, 2015; Pitois *et al.*, 2018) uses a similar image-capture method but is instead connected to the ships clean water inlet. This negates the need for deployment and allows for continuous imaging of particles as they pass through the system as the ship is underway. Continuous sampling allows for high spatial and temporal resolution whilst retaining reasonable taxonomic resolution. The device also has economic advantages: it is easily retrofitted to existing vessels and runs continuously with minimal human interaction after set-up. Furthermore, the use of semi-automated image recognition algorithms has strong potential to significantly reduce analysis time.

The first application of the Plankton Image Analyser (PIA) was a comparison with ring net sampling (Pitois *et al.*, 2018). The study found that the PIA performed well, but noted limitations associated with depth of field issues leading to a higher number of unidentifiable blurred images. Here, the Plankton Imager (PI), an evolution of the PIA, was used with an improved method to reduce the incidence of blurred images. We used the Celtic Sea as a case

study to assess the ability of the PI to describe temporal changes in the mesozooplankton community.

Method

Study area and sampling methods

Data were collected at night during five fisheries surveys from 2016 to 2019 in the Celtic Sea aboard the RV Cefas Endeavour (Fig. 1, Table 1). Autumn data were collected as part of the PELTIC survey (PELagic ecosystems in the western English Channel and eastern celTIC Seas) and spring data aboard the SWECOSS survey (South West ECOSystems Survey). The PI ran continuously and mesozooplankton counts were obtained at 107 stations by extracting from the raw data. Temperature and salinity data were collected in autumn using a SAIV mini Conductivity, Temperature, Depth (CTD) and in spring using the Cefas-built ESM2 data logger at 93 stations. Due to sampling constraints, 14 stations did not have corresponding temperature and salinity data (Fig. 1). PI data are freely available from the Cefas Data Hub: 10.14466/CefasDataHub.101 (Pitois *et al.*, 2020).

Temperature and salinity

Temperature and Salinity data were bin averaged into 1 m depth increments starting at the sea surface. Sea surface temperature (SST) was taken as the shallowest bin available. Although the PI samples at 4 m, the difference in temperature between the surface and 4m was only 0.02 °C on average and thus negligible. Differences in temperature (ΔT) and salinity (ΔS) were calculated between the shallowest and deepest readings.

The Plankton Imager (PI)

The Plankton Imager is an instrument for the continuous, semi-automated, underway sampling of mesozooplankton in surface waters. The PI uses a Basler 2048-70kc line scan camera with a scanning rate of 70,000 lines per second capturing 2,048 10 µm pixels per line (Culverhouse *et al.*, 2015; Pitois *et al.*, 2018, 2021). Water pumped from 4 m depth passes through the flow cell at 22 L min^{-1} equating to an approximate 1.3 m^3 of seawater per hour. The flow cell has an internal depth of 12.8 mm giving a field of view of 10 µm x 20.48 mm. The PI can capture particles sized from 10 µm to 2 cm but was set up with a range of 200 µm – 2 cm. This was to prevent the image capture rate exceeding the hard drive write speed

which would result in lost images. Captured images are classified by a Random Forest machine learning algorithm which sorts images into predefined categories (Breiman, 2001). The algorithm is trained on expert-sorted PI images (Fig. 2). In line with PI software and hardware developments, the classifier training set has been continually improved. Classification accuracy varies between stations but currently all images are checked and, if needed, resorted by an expert taxonomist.

Zooplankton counts were derived from 200 zooplankton images extracted at random from all images obtained during a 1 hour period at each station (Pitois *et al.*, 2018). The actual number of images needed to reach 200 specimens varied based on plankton density and detrital content. All non-target images (e.g. large phytoplankton or particulates) are classified as detritus. This process of subsampling is analogous that used by the Folsom Splitter where data are continually and randomly split until the target number of individuals is reached. Images per hour ranged by an order of magnitude. The PI operates on a semi-automated classification method, similar to that used by ZOOScan (Gorsky *et al.*, 2010). An expert taxonomist validated the output from the machine learning classifier for each station. The PI flow rate, sampling duration, number of images classified (including detritus) and total number of images sampled in the timeframe were used to resolve the scaling factor. The scaling factor was then used to calculate zooplankton abundance (individuals m^{-3}; ind. m^{-3} henceforth).

Zooplankton were classified into 40 taxonomical categories. All images are classified to the maximum discernible taxonomic resolution. In some cases, due to orientation of the specimen or image blur, an image could only be confidently identified to a low taxonomic resolution (e.g. unidentified copepod or decapod larvae). Of the 40 categories, 12 contributed to less than 1 % of the total abundance and were present in less than 5 % of all stations (fish larvae, cladocera, gammaridea, monstrilloida, marine mites, ascidian larvae, siphonophora, *Caligus* spp., caprellidae, ostracoda, physonectae and *Clione* spp.). These were removed prior to data analysis. The abundance data from the remaining 28 taxonomic groups (Fig. 7; Supplementary Table 1) were used to compare the communities across the survey areas and between seasons.

Statistical analyses

All analyses were undertaken in R (version 4.0.2) using the Vegan package (Oksanen *et al.*, 2013; R Development Core Team, 2018).

The non-parametric Spearman's correlation coefficient was used to test for correlation between max depth and ΔT. For those sites with salinity and temperature data, the community was analysed using non-metric dimensional scaling (NMDS) on a Bray-Curtis dissimilarity matrix. Prior to NMDS abundance data were transformed using the Hellinger transformation to reduce data asymmetry (Legendre and Gallagher, 2001). The envfit() function, which fits supplementary variables on the NMDS, was used to determine the correlation and forcing direction of environmental factors. The ordisurf() function, which fits a smooth surface to an ordination using a generalised additive model, was used to visualise the difference in environmental variables between seasons as well as explore their relationship with seasonal groupings.

To test for a significant difference between years and seasons, Permutational Multivariate Analysis Of Variance Using Distance Matrices (PERMANOVA, using the ADNOIS function) were used with 999 random permutations (Anderson, 2001). The betadisper() function, which analyses multivariate homogeneity of group dispersions, was used to determine if a significant result produced by PERMANOVA was the result of the variable being tested (*year* or *season*) or variations within seasons (Anderson *et al.*, 2006). A NMDS was run, exactly as before, for all sites and the envfit() function used to determine the correlation and forcing of each taxa toward a particular survey / season. This was reinforced through use of a SIMPER analyses (Clarke and Ainsworth, 1993).

Results

Physical conditions

There was a clear difference in temperature profiles between seasons (Fig. 3). Across all 4 years autumn temperature (range = 9.52 °C to 16.9 °C, mean ΔT = 4.8 °C) was an average of 3.9 °C warmer and had higher variability in ΔT than spring (range = 9.26 °C to 13.2 °C, mean ΔT = 2.35 °C). Varying degrees of stratification can be seen in autumn where ΔT > 1 °C for a third of stations (Fig. 3). The strength of the stratification was positively correlated with deeper waters across all years ($r_s = 0.4$, $p < 0.001$), such that ΔT was highest at the most westward stations. Near-shore stations, those within the English Channel and in close proximity to the Bristol Channel, had the lowest ΔT values (Fig. 4). Spring profiles show that the water column was well mixed with little to no variation in temperature with depth (i.e. ΔT < 1 °C for 95 % of spring profiles) or between years (Fig. 3).

The NMDS plot (Fig. 5a) was used to explore the relationship between physical variables and plankton distribution. The larger variation in the stratification between years in autumn compared to spring (Fig. 3) was reflected in the mesozooplankton community where variation between years was larger in autumn than spring (indicated by the large spread along the x-axis in Fig. 5a). Autumn 2018 had no overlap in community with either 2016 or 2019 (Fig. 5a) and could reflect the cooler seas in 2018 (Fig. 3). The envit() function suggested that the supplementary physical variables with a more linear relationship to the NMDS scores (indicated by the contour plots: SST, Fig. 5d and ΔS, Fig. 5c) were related with dissimilarity between seasonal groupings (Fig. 5a). ΔT is not shown in Fig. 5a as the model fitted by the envfit() function was not significant. The contour plots for SST (Fig. 5d) and ΔS (Fig. 5c) indicate that cooler SST (below 12 °C), and reduced ΔS (where variation was < 0.08), were found in spring. Higher SST values and a more variable ΔS were found in autumn sites. Figure 5b highlights the non-linear relationship between NMDS site scores and ΔT. Lower stratification (Fig. 3) was associated with spring sites. Most autumn sites had a ΔT > 0.8 °C, although there was high variation between years (Fig. 5b). The spread of points across the ΔT contours may reflect the variation in ΔT between locations (Fig. 4)

Mesozooplankton Community

Mesozooplankton abundance varied greatly between years and seasons (Fig. 6). Autumn 2019 stations had the highest mean abundance (6,780.5 ind. m^{-3}). Conversely, the previous autumn had the lowest mean station abundance (2,323.5 ind. m^{-3}). There appears to be no relationship between mean station abundance and season. Although both surveys in 2019 had almost double the mean station abundance of any other previous survey (Fig. 6).

On average, 9 to 10 taxa contributed to over 95% of the total abundance (Supplementary Table 1). Over all 5 surveys common dominant taxa were ‘Unknown copepods’, copepod nauplii and *Centropages* spp. Unknown copepod tended to be the largest contributor to total abundance of any taxa, but this was inconsistent between years (Fig. 7). The contribution of copepod nauplii to total abundance was fairly consistent with the exception of spring 2019 (mean relative abundance ranged from 1.3 % to 5.85 % excluding spring 2019 where relative abundance was 14.64 %, Fig. 7). *Centropages* spp. also made a consistent contribution to total abundance (mean relative abundance ranged from 1.58 % to 4.17 % with an average value of 3.2 %, Fig. 7). None of these 3 taxa adhered to a particular season. Common to four of five surveys dominant taxa were *Oithona* spp., *Pseudocalanus* spp. and *Acartia* spp.

With the notable exception of radiolaria in autumn 2019, which had an average contribution of 61 % to the total abundance at each station (Fig. 7), copepods dominated the community accounting for > 70 % of the total abundance on average (Fig. 6). Of these, an average 30.9% were classified as “Unknown copepods” (Fig. 7). *Para-Pseudocalanus* spp. were particularly numerous in spring 2019, contributing to a third of the total abundance (mean station abundance 1470.0 ind. m^{-3}, Supplementary Table 1).

Meroplankton constituted a larger portion of the spring community than autumn (Fig. 6). For spring 2017 and 2019, meroplankton made up 30 % and 8 % of the total mesozooplankton abundance respectively, while their contribution was < 1% in all 3 autumn surveys. The higher proportion of meroplankton found in spring comprised different larval forms each year. In spring 2017, the high meroplankton abundance was driven by decapod and barnacle larvae (25 % and 7.23 % of total abundance respectively, Fig. 7). Spring 2019 mainly comprised of echinoderm larvae, followed by Polychaete larvae (6.53 % and 0.71 %, Fig. 7). Some meroplankton, mainly bryozoa and bivalve larvae, were found in high abundance during autumn (Fig. 7).

Statistical analyses were performed on mesozooplankton abundances to determine statistically which taxa were driving variation between communities and add robustness to the prior description. A PERMANOVA suggested that both *season* and *year* were significant factors in causing variations between the communities (*season* $p < 0.05$, $R^2 = 0.08$; *year* $p < 0.05$, $R^2 = 0.42$). It is likely due to heterogeneous dispersion effects (within-survey variation seen in all autumn surveys, Fig. 8) in autumn that *year* yielded a higher R^2 value. Further evidence for this comes from the lack of ellipse overlap for the autumn surveys in the NMDS plot (Fig. 8). Subsequent ANVOVAs on the analysis of multivariate homogeneity of groups (using the betadisp() function) confirmed this. Spring communities had homogeneity among years ($F(2,24) = 1.9145$, $p = 0.17$) whilst autumn were significantly different in groups between years ($F(2,78) = 9.6832$, $p < 0.001$). It is therefore likely that *year* had some influence on our PERMANOVA result and perhaps reduced the seasonal effect.

The taxa loadings on the NMDS plot (Fig. 8) echoes those trends seen in relative abundance (Fig. 7). For example, the high relative abundance of radiolaria in autumn 2019 (Fig. 7) is also seen in the ordination plot where radiolaria are highly correlated with this survey stations. This forcing from an individual, or few key taxa, is characteristic for all autumn surveys (Fig. 8). *Oithona* spp. and *Centropages* spp. are strongly correlated with autumn

2016 where as *Candacia* spp. and *Corycaeus* spp. are strongly correlated with 2018 and neither taxa strongly correlated with other years. Spring stations show the opposite, where there is high overlap in yearly ellipses and discerning taxa that correlate better to one year than another is difficult (Fig. 8). In general, the location of spring stations within the ordination are forced by meroplankton, such as decapod and echinoderm larvae compared to autumn. SIMPER analysis reported that average dissimilarity between spring and autumn was 49.2 %. Six taxa were responsible for driving > 50 % of the difference between spring and autumn communities and twelve for > 75 %. This being said, it is important to consider the heterogeneity of autumn surveys.

Discussion

Environmental drivers of seasonality

The increased stratification offshore in early autumn (Fig. 4) is consistent with the established summer stratification of the Celtic Sea and Western Channel as well as the breaking down of stratification through increased mixing in coastal waters during late summer and autumn (Southward *et al.*, 2004; Harris, 2010). This trend is also consistent between years but is not reflected by consistency in the mesozooplankton community between years. While there is a similar linear spread of each autumn surveys stations across the ΔS contours in the NMDS plot (Fig. 5c), the communities are dissimilar between years (Fig. 8). Conversely, spring had both consistency in the degree of stratification (negligible variation between years and sites, Fig. 4) and in community overlap on the ordination plot (Fig. 8). This suggests that the summer stratification persists into early autumn, before it begins to degrade (Southward *et al.*, 2004; Smyth *et al.*, 2015), and its increased strength with distance offshore is likely to have contributed to the large dissimilarity of autumn stations seen in the ordination analysis (Fig. 8).

The potential processes driving community composition and taxa seasonality at PML's longstanding L4 timeseries station (Harris, 2010), which falls within our study area, are summarised by Atkinson (Atkinson *et al.*, 2018). The authors review a suite of mechanisms that govern the mesozooplankton and suggest that a synergistic combination of mechanism is often responsible. Those most relevant to our study are: the loophole hypothesis, whereby physiochemical changes favour some taxa (Irigoien *et al.*, 2005); changes in net heat flux where stabilisation of the water column promotes the spring bloom (Smyth *et al.*, 2014);

mortality-controlled copepod phenology (Irigoien and Harris, 2003; Maud *et al.*, 2015) and zooplankton feeding traits (Sailley *et al.*, 2015). The high community variation we see in autumn between years may be the result of a complex combination of these processes where each hypothesis is more relevant in a specific year. For example, the persistence of the summer stratification into autumn (Southward *et al.*, 2004), may have been more pronounced in 2019 resulting in the exceptional abundance of radiolaria in 2019, although we do not see this effect in our data (Fig. 3). Radiolarian diversity has been shown to increase with distance offshore and depth of stratification (Biard *et al.*, 2017) and this may explain, in-part, their high abundance. On further investigation, the highest numbers of radiolaria were found at the most stratified, offshore stations. An additional factor may be the coincidence of the autumn bloom with the survey timing. While survey dates tend to be consistent year on year (Table 1), the timing of the environmental phenomena leading to phytoplankton blooms, and thus an increase in mesozooplankton, are not so regular. This potentially resulting in a mismatch between bloom conditions and the survey dates. It has been suggested that survey 'snapshots' might be spatially misleading (Huret *et al.*, 2018) and may be responsible for the large variation between the autumn communities despite similar environmental conditions.

System performance

The community description presented here reveals interannual and seasonal variabilities in both the abundance of individual taxa and the mesozooplankton community structure. Our findings were in line with those found by two time series in the area: The Continuous Plankton Recorder (CPR) (Richardson *et al.*, 2006) and Plymouth Marine laboratories L4 Station (Eloire *et al.*, 2010). Additionally, the seasonality of individual taxa and observation of distinct seasonal communities presented here agrees with previous descriptions of mesozooplankton in the Celtic Sea (Johns, 2006; Eloire *et al.*, 2010; Highfield *et al.*, 2010; Giering *et al.*, 2019). This agreement between devices is found despite that loss of detailed taxonomic information when using the PI compared to traditional methods (i.e. those samples analysed by microscopy). For example, we find a high number of unknown copepods in all surveys (Fig. 7) due to occurrences where a specimen has a non-favourable orientation relative to the camera when imaged (Tang *et al.*, 1998), with copepods being particularly troublesome.

Zooplankton are a morphologically diverse group of organisms in terms of size, shape and behaviour. Therefore, any plankton sampling device will preferentially sample, or be biased

towards, a certain group of organisms (Owens et al., 2013). The PI, like all plankton samplers, suffers gear specific issues such as active and passive avoidance or damage to samples. The radiolaria peak presented here provides an interesting example. Fragile varieties or those that form colonies, such as radiolaria, are difficult to sample with nets or devices that require collection of the individual (Cifelli and Sachs, 1966; Burki and Keeling, 2014). This is less problematic for imaging devices. The high abundance of radiolaria seen in 2019, as well as its consistent appearance within the dominant taxa of PI samples, adds to a growing body of evidence from imaging devices that suggest radiolaria are highly abundant and an important part of marine food webs which are often missed by traditional methods (Dennett *et al.*, 2002; Picheral *et al.*, 2010; Biard *et al.*, 2016). Conversely, the PI was found to underreport certain taxonomic groups when compared to a ring net; Pitois (Pitois *et al.*, 2018) suggested that the fragility of Appendicularia resulted in destruction beyond recognition and the strong swimming ability of chaetognaths resulted in sampling avoidance. Although the seasonality exhibited by these taxa (Fig. 7) agrees with existing literature (Johns, 2006; Eloire *et al.*, 2010). This suggests that while the PI under samples these taxa, it still does so sufficiently to detect seasonal trends. The fixed depth intake from which the PI samples may give rise to variation as well as the choice to only use night stations. Many zooplankton undergo diel vertical migration (Hays, 2003), which may introduce a sampling bias toward those that only occupy the upper water column at night.

Moving forward

As a new device, the PI needs to find its niche amongst existing devices. How it can best complement, build upon, or supplement existing data sets needs to be determined. No single device is able to accurately capture all components of the zooplankton and all systems underestimate parts of the zooplankton community (Owens *et al.*, 2013). Researchers must select a system, or a suite of systems, that is most appropriate to answer the research questions posed (Skjoldal *et al.*, 2013).

The findings presented here suggests that the PI captures the community with sufficient accuracy to describe trends and community structures within the mesozooplankton. The limitations of the PI, mainly the loss of highly detailed taxonomic information and its fixed sampling depth, are balanced by several advantages. From an economic standpoint, the automated nature of the device and ease of integration onto existing surveys make it an attractive option for continuous underway sampling where the level of description of the

community presented here is satisfactory (for example, a potential application may be food web studies). However, the foremost advantages are seen from an ecological point of view, the PI can obtain this information at an unparalleled spatial and temporal resolution due to sampling 24/7 with negligible down time. To date, and to demonstrate the robustness of the PI as a mesozooplankton sampler, the PI stations have been chosen to coincide with ring nets. This does not use the PI to its full potential. For example, in autumn 2018, a representative year in terms of images although with the most stations per survey, the PI captured 8.3 million images (inclusive of detritus) over the whole survey. Of these, only 50,162 (or 0.6 %) were used. On a more recent survey in 2020, not reported here, the PI captured 16.2 million images. With recent improvements, mainly through increasing the processing and storage speeds to keep up with the phenomenal data collection rate, experiments at sea suggest the minimum size (currently 200 μm) can be reduced by half. This would reveal more of the plankton community, although anecdotal evidence suggests the number of images captured would increase by an order of magnitude, in turn bringing its own data processing challenges. To tackle these challenges, new tools must be developed to make best use of the 'big-data' produced by the PI.

Conclusion

We have demonstrated that the PI is able to detect changes in mesozooplankton abundances in line with established devices. While the inherent strength of devices such as the PI (i.e. cost effectiveness and high frequency sampling leading to fine scale spatial data) can be used to address new research questions, they also give rise to new challenges. Mainly, the data collection rate is faster than the processing rate. Progress in the machine learning classifier and the emergence of innovative methods in data analytics will remove the need to subsample images and classify all particles at a modest taxonomic resolution. This will result in truly high mesozooplankton resolution data, able to complement existing large-scale or simple point sampling timeseries for this important group of marine organisms.

Acknowledgements

We would like to thank Jeroen Van Der Kooij and Ian Holmes for the opportunity to participate in the various cruises and the considerable efforts off all scientists, offices and crew aboard the RV Cefas Endeavour.

Funding

This work was supported by the Natural Environment Research Council through the EnvEast Doctoral Training Partnership [grant number NE/L002582/1] and co-funded by Cefas Seedcorn (SP002) as part of the Cefas-UEA Strategic Alliance.

References

Agnarsson, I. and Kuntner, M. (2007) Taxonomy in a changing world: Seeking solutions for a science in crisis. *Syst. Biol.*, **56**, 531–539.

Anderson, M. J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol.*, **26**, 32–46.

Anderson, M. J., Ellingsen, K. E., and McArdle, B. H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecol. Lett.*, **9**, 683–693.

Atkinson, A., Polimene, L., Fileman, E. S., Widdicombe, C. E., McEvoy, A. J., Smyth, T. J., Djeghri, N., Sailley, S. F., *et al.* (2018) Comment. What drives plankton seasonality in a stratifying shelf sea? Some competing and complementary theories. *Limnol. Oceanogr.*, **63**, 2877–2884.

Bean, T. P., Greenwood, N., Beckett, R., Biermann, L., Bignell, J. P., Brant, J. L., Copp, G. H., Devlin, M. J., *et al.* (2017) A Review of the Tools Used for Marine Monitoring in the UK: Combining Historic and Contemporary Methods with Modeling and Socioeconomics to Fulfill Legislative Needs and Scientific Ambitions. *Front. Mar. Sci.*, **4**, 263.

Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S., and Reid, P. C. (2003) Plankton effect on cod recruitment in the North Sea. *Nature*, **426**, 661–664.

Benoit-Bird, K. J. and Lawson, G. L. (2016) Ecological Insights from Pelagic Habitats Acquired Using Active Acoustic Techniques. *Ann. Rev. Mar. Sci.*, **8**, 463–490.

Biard, T., Bigeard, E., Audic, S., Poulain, J., Gutierrez-Rodriguez, A., Pesant, S., Stemmann, L., and Not, F. (2017) Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. *ISME J.*, **11**, 1331–1344.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., *et al.* (2016) In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*, **532**, 504–507.

Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.

Burki, F. and Keeling, P. J. (2014) Rhizaria. *Curr. Biol.*, **24**, R103–R107.

Cifelli, R. and Sachs, K. N. (1966) Abundance relationships of planktonic Foraminifera and Radiolaria. *Deep Sea Res. Oceanogr. Abstr.*, **13**, 751–753.

Clarke, K. R. and Ainsworth, M. (1993) A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Ser.*, **92**, 205.

Cowen, R. K. and Guigand, C. M. (2008) In situ ichthyoplankton imaging system (ISIIS): System design and preliminary results. *Limnol. Oceanogr. Methods*, **6**, 126–132.

Culverhouse, P. F., Gallienne, C., Williams, R., and Tilbury, J. (2015) An Instrument for Rapid Mesozooplankton Monitoring at Ocean Basin Scale. *J. Mar. Biol. Aquac.*, **1**, 1–11.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., *et al.* (2016) Implementing and innovating marine monitoring approaches for assessing marine environmental status. *Front. Mar. Sci.*, **3**, 213.

Davis, C. S., Thwaites, F. T., Gallager, S. M., and Hu, Q. (2005) A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography. *Limnol. Oceanogr. Methods*, **3**, 59–74.

Dennett, M. R., Caron, D. A., Michaels, A. F., Gallager, S. M., and Davis, C. S. (2002) Video plankton recorder reveals high abundances of colonial Radiolaria in surface waters of the central North Pacific. *J. Plankton Res.*, **24**, 797–805.

Eloire, D., Somerfield, P. J., Conway, D. V. P., Halsband-Lenk, C., Harris, R., and Bonnet, D. (2010) Temporal variability and community composition of zooplankton at station L4 in the Western Channel: 20 years of sampling. *J. Plankton Res.*, **32**, 657–679.

Giering, S. L. C., Wells, S. R., Mayers, K. M. J., Schuster, H., Cornwell, L., Fileman, E. S., Atkinson, A., Cook, K. B., *et al.* (2019) Seasonal variation of zooplankton community structure and trophic position in the Celtic Sea: A stable isotope and biovolume spectrum approach. *Prog. Oceanogr.*, **177**, 101943.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.-B. B., Cawood, A., Pesant, S., *et al.* (2010) Digital zooplankton image analysis using the ZooScan integrated system. *J. Plankton Res.*, **32**, 285–303.

Hansell, D. A. (2002) DOC in the Global Ocean Carbon Cycle. In Hansell, D. A. and Carlson, C. A. (eds), *Biogeochemistry of Marine Dissolved Organic Matter*. Academic Press, San Diego, pp. 685–715.

Harris, R. (2010) The L4 time-series: The first 20 years. *J. Plankton Res.*, **32**, 577–583.

Hays, G. C. (2003) A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Migrations and Dispersal of Marine Organisms*. Springer, pp. 163–170.

Heath, M. R. (2005) Regional variability in the trophic requirements of shelf sea fisheries in the Northeast Atlantic, 1973-2000. *ICES J. Mar. Sci.*, **62**, 1233–1244.

Highfield, J. M., Eloire, D., Conway, D. V. P., Lindeque, P. K., Attrill, M. J., and Somerfield, P. J. (2010) Seasonal dynamics of meroplankton assemblages at station L4. *J. Plankton Res.*, **32**, 681–691.

Huret, M., Bourriau, P., Doray, M., Gohin, F., and Petitgas, P. (2018) Survey timing vs. ecosystem scheduling: Degree-days to underpin observed interannual variability in marine ecosystems. *Prog. Oceanogr.*, **166**, 30–40.

Irigoien, X., Flynn, K. J., and Harris, R. P. (2005) Phytoplankton blooms: a 'loophole'in microzooplankton grazing impact? *J. Plankton Res.*, **27**, 313–321.

Irigoien, X. and Harris, R. P. (2003) Interannual variability of Calanus helgolandicus in the English Channel. *Fish. Oceanogr.*, **12**, 317–326.

Johns, D. (2006) The Plankton Ecology of the SEA 8 area. *Strateg. Environ. Assess. Program. UK Dept. Trade Ind.*, 1–44.

Legendre, P. and Gallagher, E. D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J. B., *et al.* (2019) Globally consistent quantitative observations of

planktonic ecosystems. *Front. Mar. Sci.*, **6**.

Mackas, D. L. and Beaugrand, G. (2010) Comparisons of zooplankton time series. *J. Mar. Syst.*, **79**, 286–304.

Maud, J. L., Atkinson, A., Hirst, A. G., Lindeque, P. K., Widdicombe, C. E., Harmer, R. A., McEvoy, A. J., and Cummings, D. G. (2015) How does Calanus helgolandicus maintain its population in a variable environment? Analysis of a 25-year time series from the English Channel. *Prog. Oceanogr.*, **137**, 513–523.

McQuatters-Gollop, A., Johns, D. G., Bresnan, E., Skinner, J., Rombouts, I., Stern, R., Aubert, A., Johansen, M., *et al.* (2017) From microscope to management: The critical value of plankton taxonomy to marine policy and biodiversity conservation. *Mar. Policy*, **83**, 1–10.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., *et al.* (2013) Community ecology package. *R Packag. version*, 2–0.

Orenstein, E. C., Ratelle, D., Briseño-Avena, C., Carter, M. L., Franks, P. J. S., Jaffe, J. S., and Roberts, P. L. D. (2020) The Scripps Plankton Camera system: A framework and platform for in situ microscopy. *Limnol. Oceanogr. Methods*, **18**, 681–695.

Owens, N. J. P., Hosie, G. W., Batten, S. D., Edwards, M., Johns, D. G., and Beaugrand, G. (2013) All plankton sampling systems underestimate abundance: Response to "Continuous plankton recorder underestimates zooplankton abundance" by JW Dippner and M. Krause. *J. Mar. Syst.*, **128**, 240–242.

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., and Gorsky, G. (2010) The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnol. Oceanogr. Methods*, **8**, 462–473.

Pitois, S. G., Bouch, P., Creach, V., and Van Der Kooij, J. (2016) Comparison of zooplankton data collected by a continuous semi-automatic sampler (CALPS) and a traditional vertical ring net. *J. Plankton Res.*, **38**, 931–943.

Pitois, S. G., Graves, C. A., Close, H., Lynam, C., Scott, J., Tilbury, J., Van Der Kooij, J., and Culverhouse, P. (2021) A first approach to build and test the Copepod Mean Size and Total Abundance (CMSTA) ecological indicator using in-situ size measurements from the Plankton Imager (PI). *Ecol. Indic.*

Pitois, S. G., Tilbury, J., Bouch, P., Close, H., Barnett, S., and Culverhouse, P. F. (2018) Comparison sof a Cost-Effective Integrated Plankton Sampling and Imaging Instrument with Traditional Systems for Mesozooplankton Sampling in the Celtic Sea. *Front. Mar. Sci.*, **5**, 1–15.

Pitois, S., Scott, J., Culverhouse, P., Tilbury, J., and Close, H. (2020) Zooplankton abundance data derived from the Plankton Imager system from the Western English Channel and Eastern Irish Sea from 2016 to 2019.

R Development Core Team (2018) A Language and Environment for Statistical Computing. *R Found. Stat. Comput.*, https://www.R-project.org.

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., Stevens, D., and Witt, M. (2006) Using continuous plankton recorder data. *Prog. Oceanogr.*, **68**, 27–74.

Sailley, S. F., Polimene, L., Mitra, A., Atkinson, A., and Allen, J. I. (2015) Impact of zooplankton food selectivity on plankton dynamics and nutrient cycling. *J. Plankton Res.*, **37**, 519–529.

Sieracki, C. K., Sieracki, M. E., and Yentsch, C. S. (1998) An imaging-in-flow system for automated analysis of marine microplankton. *Mar. Ecol. Prog. Ser.*, **168**, 285–296.

Skjoldal, H. R., Wiebe, P. H., Postel, L., Knutsen, T., Kaartvedt, S., and Sameoto, D. D. (2013) Intercomparison of zooplankton (net) sampling systems: Results from the ICES/GLOBEC sea-going workshop. *Prog. Oceanogr.*, **108**, 1–42.

Smyth, T., Atkinson, A., Widdicombe, S., Frost, M., Allen, I., Fishwick, J., Queiros, A., Sims, D., *et al.* (2015) The Western channel observatory. *Prog. Ocean.*, **137**, 335–341.

Smyth, T. J., Allen, I., Atkinson, A., Bruun, J. T., Harmer, R. A., Pingree, R. D., Widdicombe, C. E., and Somerfield, P. J. (2014) Ocean net heat flux influences seasonal to interannual patterns of plankton abundance. *PLoS One*, **9**, e98709.

Song, J., Bi, H., Cai, Z., Cheng, X., He, Y., Benfield, M. C., and Fan, C. (2020) Early warning of Noctiluca scintillans blooms using in-situ plankton imaging system: An example from Dapeng Bay, P.R. China. *Ecol. Indic.*, **112**, 106123.

Southward, A. J., Langmead, O., Hardman-Mountford, N. J., Aiken, J., Boalch, G. T., Dando, P. R., Genner, M. J., Joint, I., *et al.* (2004) Long-term oceanographic and ecological research in the western English Channel. *Adv. Mar. Biol.*, **47**, 1–105.

Stanton, T. K., Wiebe, P. H., Chu, D., Benfield, M. C., Scanlon, L., Martin, L., and Eastwood, R. L. (1994) On acoustic estimates of zooplankton biomass. *ICES J. Mar. Sci.*, **51**, 505–512.

Steinberg, D. K., Goldthwait, S. A., and Hansell, D. A. (2002) Zooplankton vertical migration and the active transport of dissolved organic and inorganic nitrogen in the Sargasso Sea. *Deep. Res. Part I Oceanogr. Res. Pap.*, **49**, 1445–1461.

Steinberg, D. K. and Landry, M. R. (2017) Zooplankton and the Ocean Carbon Cycle. *Ann. Rev. Mar. Sci.*, **9**, 413–444.

Tang, X., Stewart, W. K., Huang, H., Gallager, S. M., Davis, C. S., Vincent, L., and Marra, M. (1998) Automatic plankton image recognition. *Artif. Intell. Rev.*, **12**, 177–199.

Taylor, A. H., Allen, J. I., and Clark, P. A. (2002) Extraction of a weak climatic signal by an ecosystem. *Nature*, **416**, 629–632.

Wiebe, P. H. and Benfield, M. C. (2003) From the Hensen net toward four-dimensional biological oceanography. *Prog. Oceanogr.*, **56**, 7–136.

Table and Figure legends

Figure 1. Location of the 107 zooplankton stations and 93 CTD stations off the south-west coast of the UK. Where Stns. = Stations.

Table 1. Number of stations per survey. For zooplankton stations n = 107. For corresponding CTD stations n = 93: Dash indicates no data. Zoop. = Zooplankton.

Figure 2. Collage of example mesozooplankton images used for training set for the 12 most abundant categories across all surveys.

Figure 3. Temperature profiles for each station per survey. The inverse y-axis shows depth (1 m bins) with the x-axis showing temperature (°C).

Figure 4. ΔT plotted as circle size for all autumn stations across all years.

Figure 5. Analysis of interactions between environmental variables and plankton community site dissimilarity for sites with physical data. All plots show the same non-metric multidimensional scaling plot created using a Bray-Curtis dissimilarity matrix on Hellinger-transformed abundance data. 5a shows the supplementary environmental variables plotted using the envfit() function. Plots 5b (ΔT), 5c (ΔS) and 5d (SST) show contour plots created using the ordisurf() function to explore the relationships between environmental variables and the NMDS site scores.

Figure 6. Mean station abundance for each survey with taxa grouped into four major categories.

Figure 7. Relative abundance (%) (Relative Abun.) for all surveys for taxa that contributed to > 1 % of the total abundance. Axis labels are on bottom left subplot (Hyperiidae) and are the same for all subplots. Categories are arranged in order of decreasing relative abundance from highest in the top left to lowest in the bottom right. Mean abundance and tabulated values are available in Supplementary Table 1.

Figure 8. NMDS plot on Hellinger-transformed abundance data using a Bray-Curtis dissimilarity matrix on Hellinger-transformed abundance data, where stress = 0.18. Supplementary variables taxa were plotted using the envfit() function with a $p < 0.05$.

Tables

Table 1 Number of stations per survey

Year	Spring			Autumn		
	Dates	Zoop. stations	CTD Stations	Dates	Zoop. stations	CTD Stations
2016	-	-	-	3 – 19 Oct	39	36
2017	7 Mar – 5 April	18	13	-	-	-
2018	-	-	-	6 Oct – 10 Nov	24	22
2019	17 – 31 Mar	8	4	1 – 28 Oct	18	18

Supplementary material

Supplementary Table 1. Average abundance (Mean Abd.) and relative abundance (Rel. Abd.) for each survey for those data that contributed to > 1 % of total abundance.

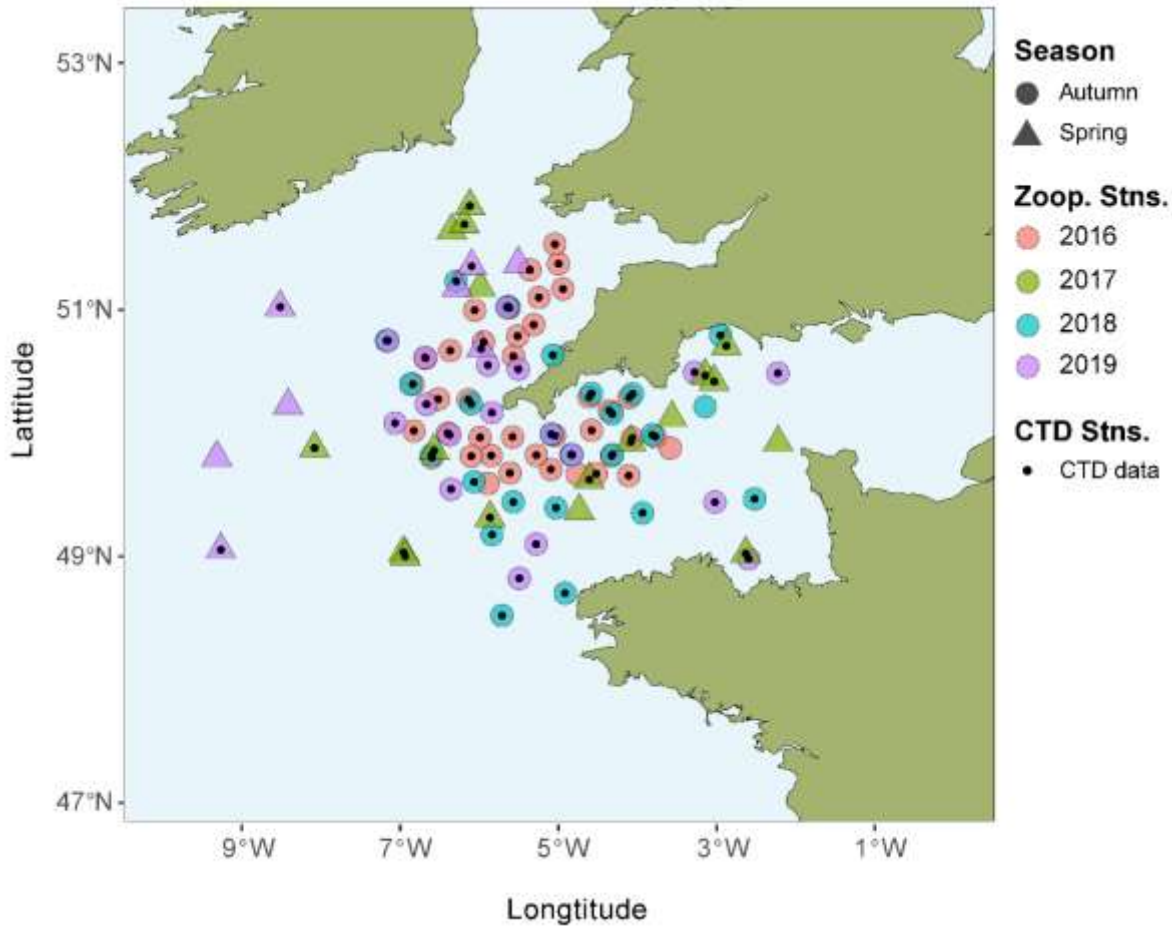
53°N
51°N
49°N
47°N
Lattitude
9°W
7°W
5°W
3°W
1°W
Longtitude
Season
Autumn
Spring
Zoop. Stns.
2016
2017
2018
2019
CTD Stns.
CTD data

Figure 1.

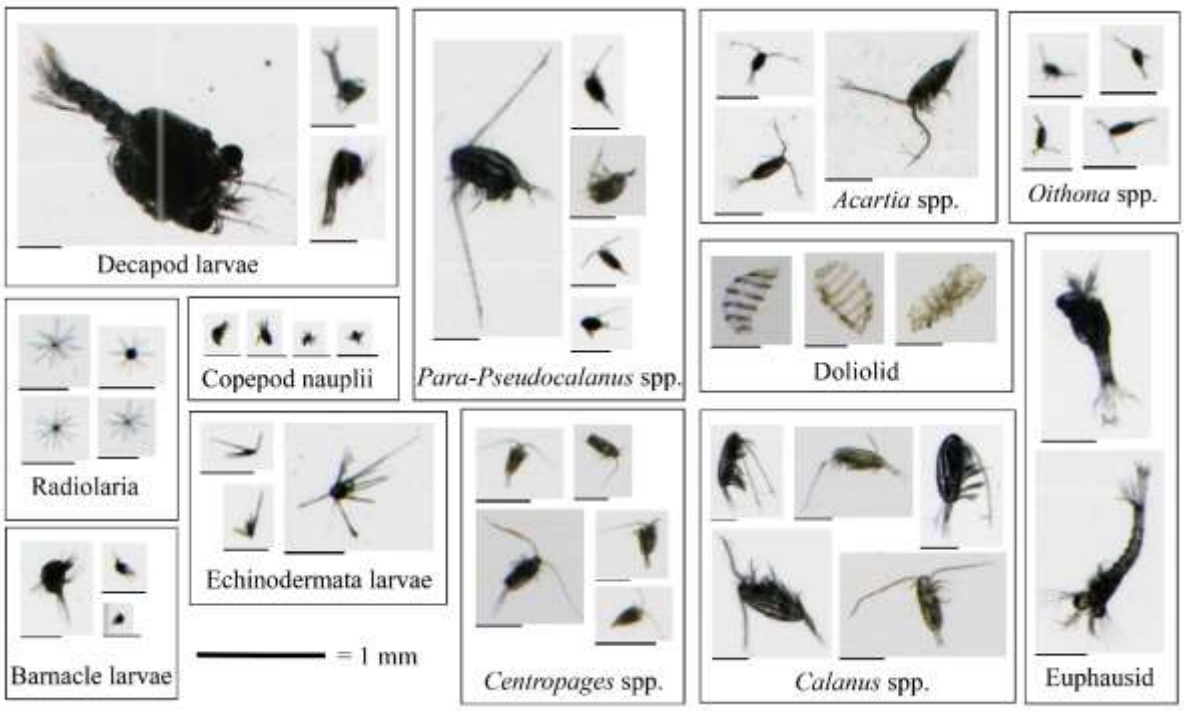
Decapod larvae
Copepod nauplii
Radiolaria
Barnacle larvae
Echinodermata larvae
= 1 mm
Para-Pseudocalanus spp.
Centropages spp.
Acartia spp.
Doliolid
Calanus spp.
Oithona spp.
Euphausid

Figure 2.

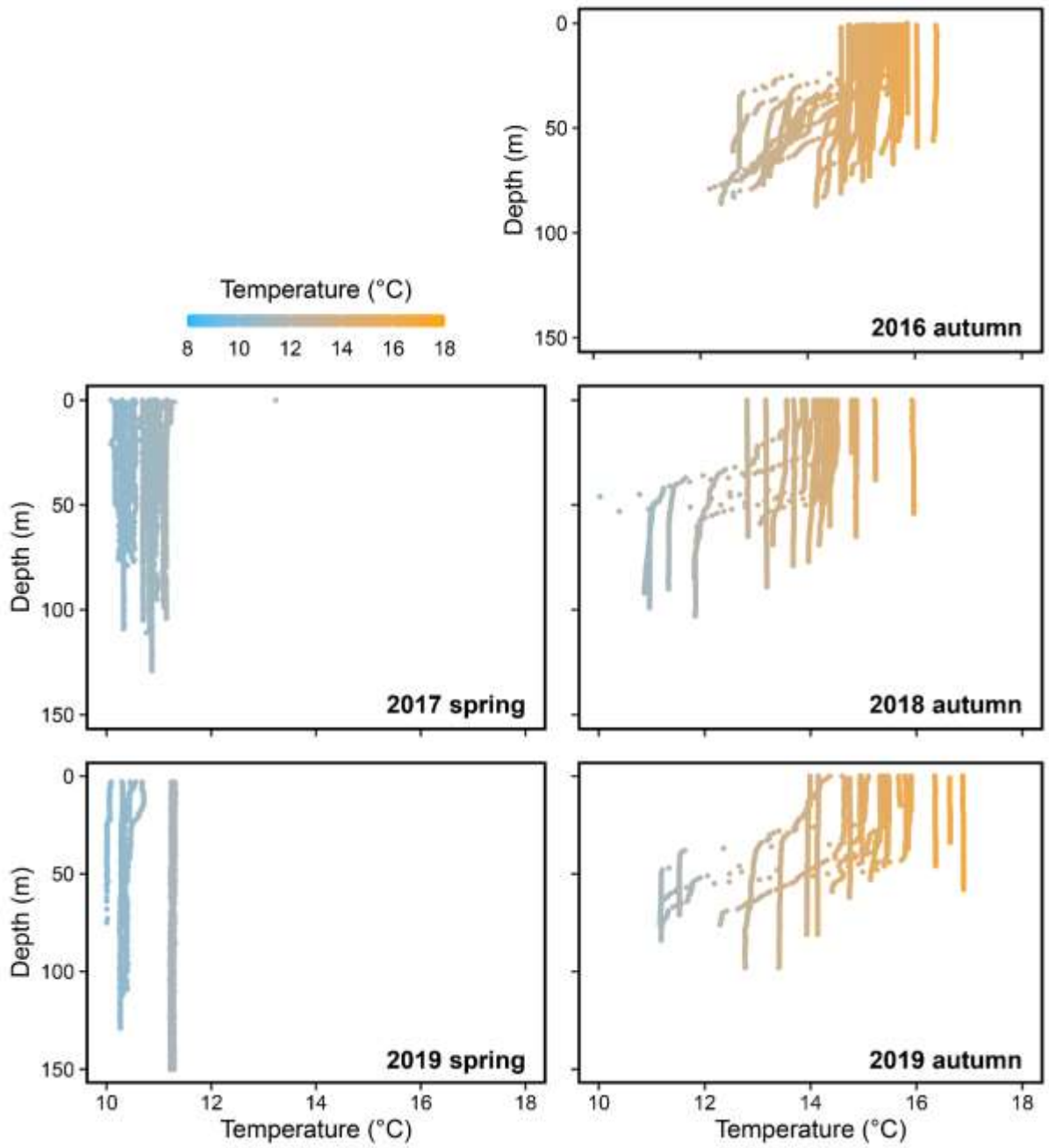
Temperature (°C)
8 10 12 14 16 18
Depth (m)
0
50
100
150
2016 autumn
2017 spring
2018 autumn
2019 spring
2019 autumn
10 12 14 16 18
Temperature (°C)

Figure 3.

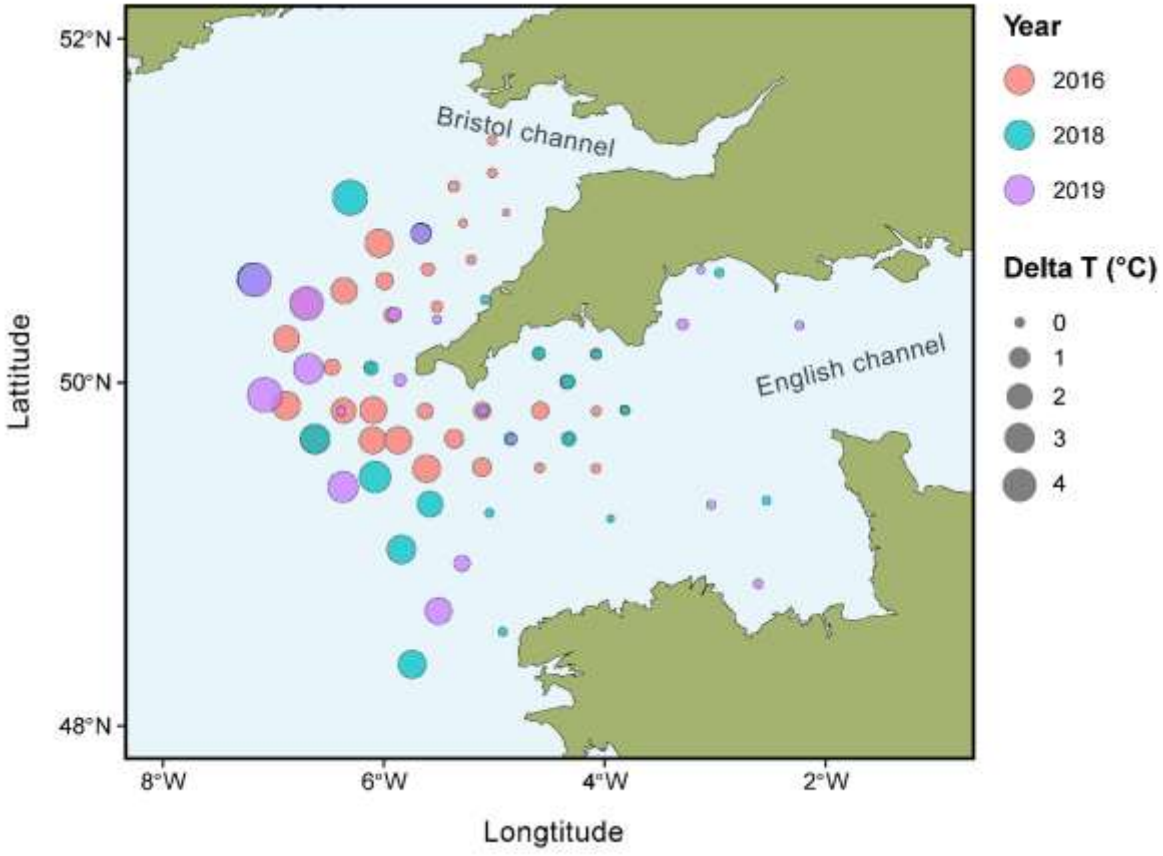
52°N
50°N
48°N
Lattitude
8°W
6°W
4°W
2°W
Longtitude
Bristol channel
English channel
Year
2016
2018
2019
Delta T (°C)
0
1
2
3
4

Figure 4.

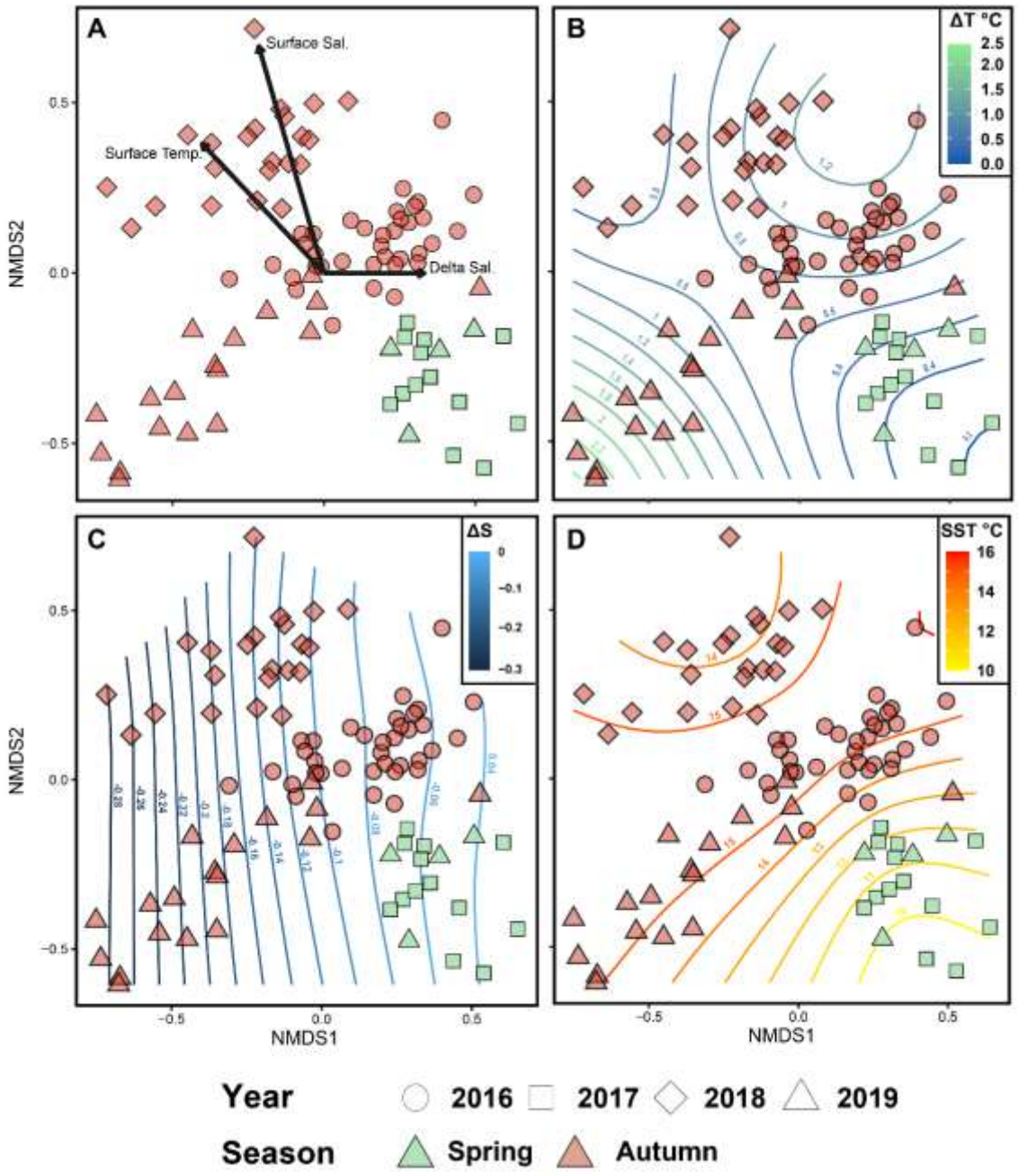
A
Surface Sal.
Surface Temp.
Delta Sal.
B
ΔT °C
2.5
2.0
1.5
1.0
0.5
0.0
C
ΔS
0
−0.1
−0.2
−0.3
D
SST °C
16
14
12
10
NMDS1
NMDS2
Year
2016
2017
2018
2019
Season
Spring
Autumn

Figure 5.

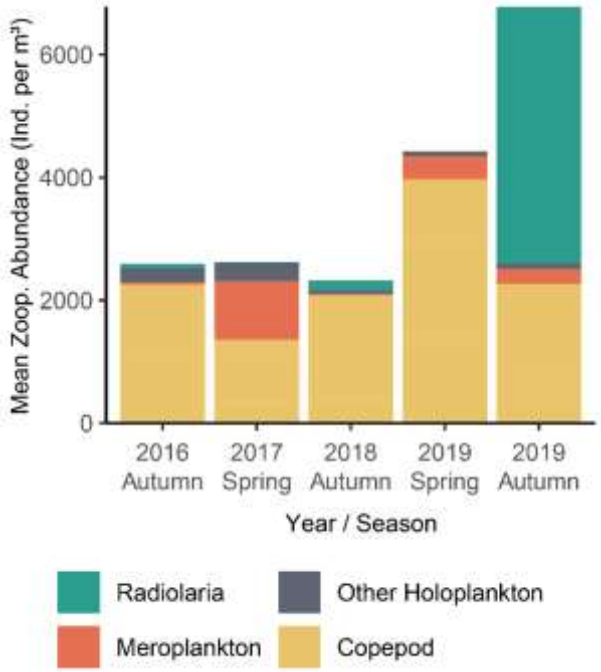
Mean Zoop. Abundance (Ind. per m³)
6000
4000
2000
0
2016 Autumn
2017 Spring
2018 Autumn
2019 Spring
2019 Autumn
Year / Season
Radiolaria
Other Holoplankton
Meroplankton
Copepod

Figure 6.

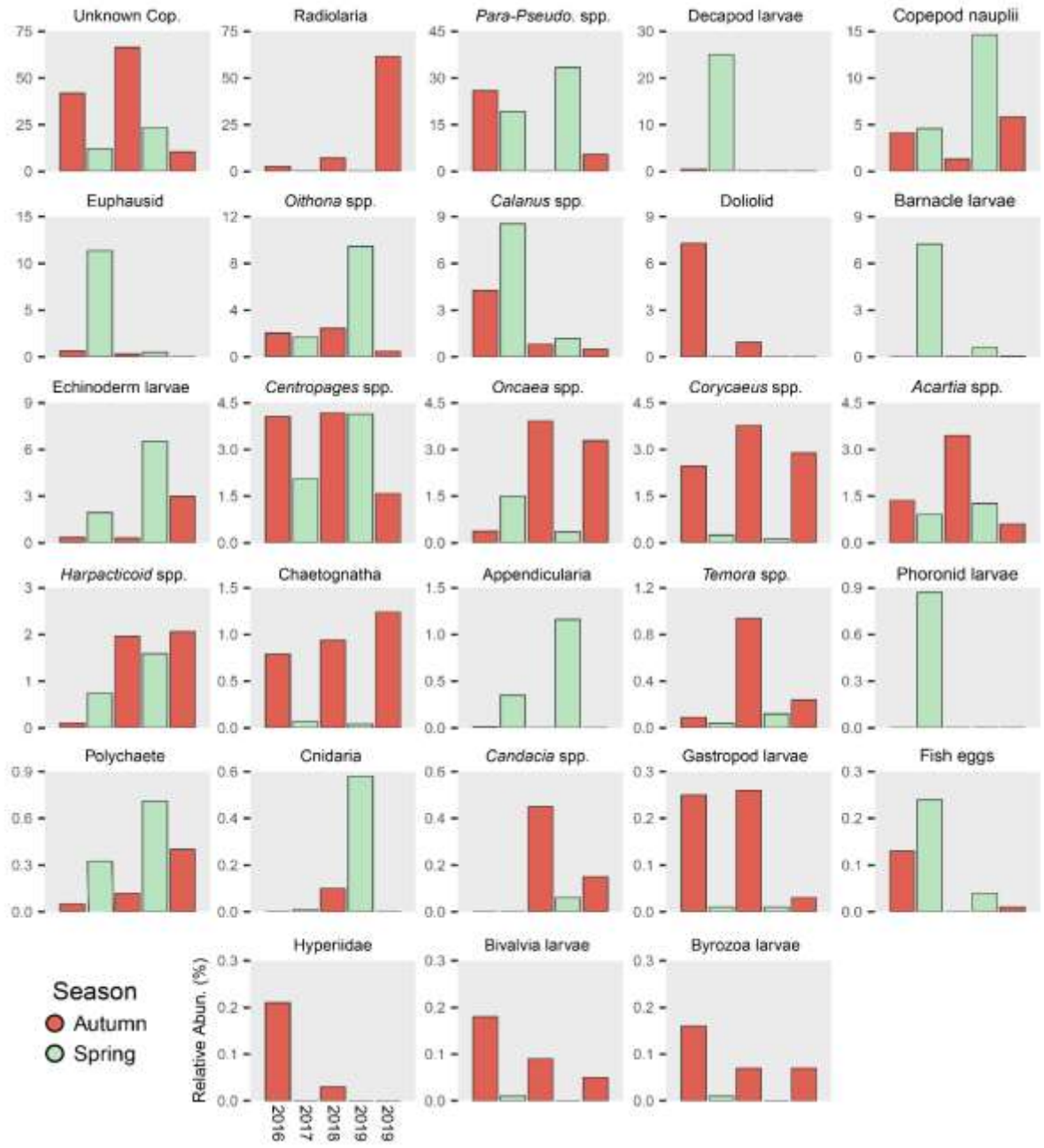
Unknown Cop.
Radiolaria
Para-Pseudo. spp.
Decapod larvae
Copepod nauplii
Euphausid
Oithona spp.
Calanus spp.
Doliolid
Barnacle larvae
Echinoderm larvae
Centropages spp.
Oncaea spp.
Corycaeus spp.
Acartia spp.
Harpacticoid spp.
Chaetognatha
Appendicularia
Temora spp.
Phoronid larvae
Polychaete
Cnidaria
Candacia spp.
Gastropod larvae
Fish eggs
Hyperiidae
Bivalvia larvae
Byrozoa larvae
Season
Autumn
Spring
Relative Abun. (%)
2016
2017
2018
2019
2019

Figure 7.

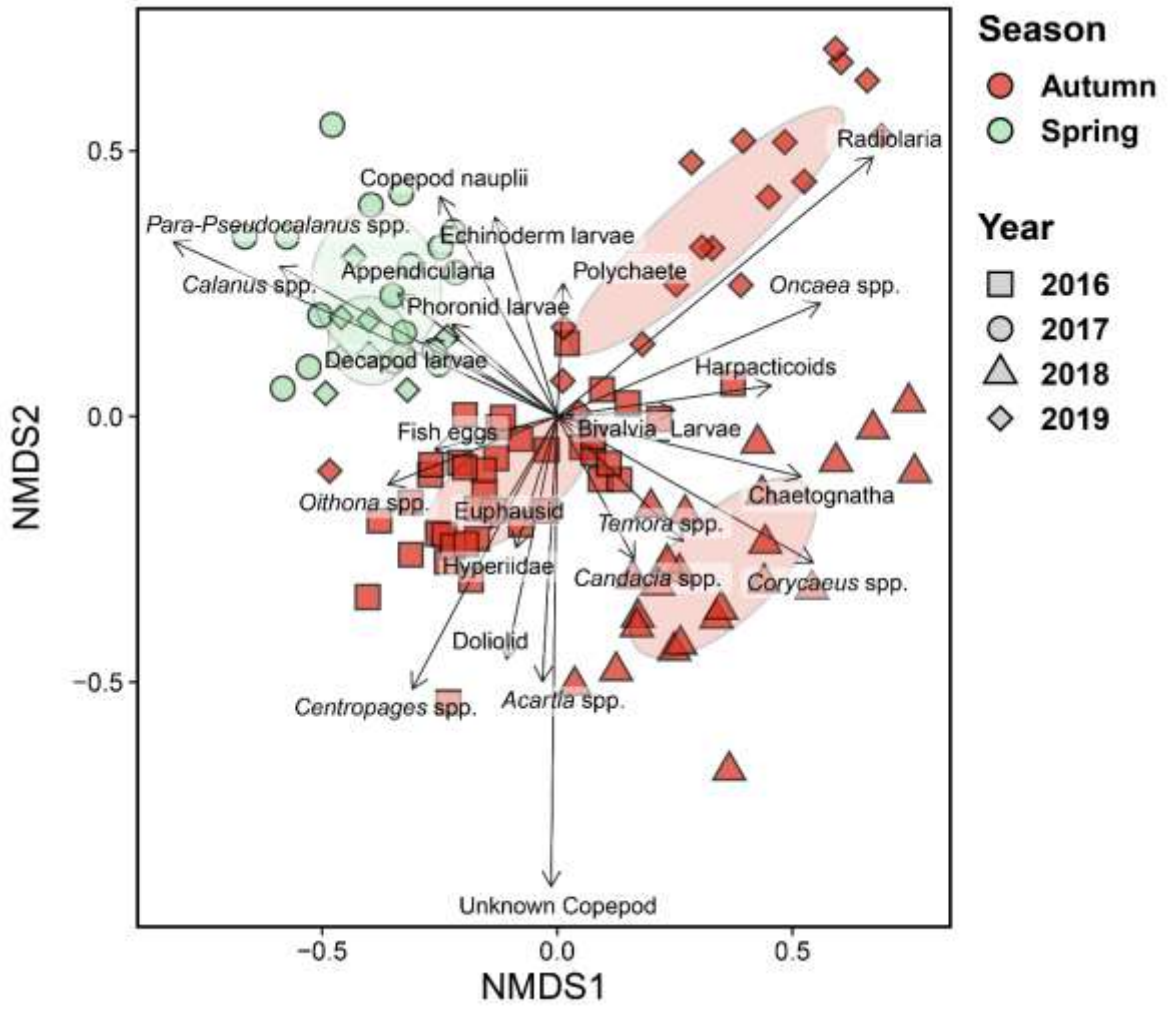
Season
Autumn
Spring
Year
2016
2017
2018
2019
Radiolaria
Copepod nauplii
Para-Pseudocalanus spp.
Echinoderm larvae
Appendicularia
Polychaete
Calanus spp.
Oncaea spp.
Phoronid larvae
Decapod larvae
Harpacticoids
Fish eggs
Bivalvia_Larvae
Oithona spp.
Euphausiid
Chaetognatha
Temora spp.
Hyperiidae
Candacia spp.
Corycaeus spp.
Doliolid
Centropages spp.
Acartia spp.
Unknown Copepod
NMDS1
NMDS2
0.5
0.0
−0.5
−0.5
0.0
0.5

Figure 8.