THE PLANKTON IMAGER. A NOVEL TOOL FOR THE AUTOMATED AND CONTINUOUS SAMPLING OF ZOOPLANKTON.

A thesis submitted to the School of Environmental Sciences
of the University of East Anglia in partial fulfilment
of the requirements for the degree of Doctor of Philosophy

JAMES ALISTAIR SCOTT

MARCH 2023

ABSTRACT

Marine zooplankton have global ubiquitous distribution and are fundamental in the ocean carbon cycle, as prey for planktivores and use as indicators for ecosystem health. Recent impetus has been on developing cost effective methods to better sample the plankton. As a result, imaging devices are becoming synonymous with plankton sampling. This study contributes to the development, and demonstrates the ecological application, of a novel plankton imaging instrument: the Plankton Imager (PI). The PI is a continuous, automated instrument that uses water pumped onboard a ship and images all particles present. The images can be resolved to a moderate (family-level) taxonomic resolution by experts. This method revealed strong temporal changes in the zooplankton community of the Celtic Sea where interannual variation was greater than seasonal. In order to better harness the continuous nature of the PI, temporal subsampling (classifying 1 in 10 images) allowed for greater spatial coverage at finer resolution. This approach revealed that the choice of sampling resolution must be appropriate to the scale of the ecological process as decreasing spatial resolution had a considerable effect on the strength and significance of the relationship between zooplankton biomass and their phytoplankton prey. Concurrently with development of the instrument, machine learning classifiers, capable of classifying the millions of images the PI collects per day, have been developed. Application of a machine learning classifier to PI images resulted in zooplankton dataset with very fine spatiotemporal scales where data could be resolved to minutes or meters. These data were aligned with other continuous datasets to re-evaluate relationships with predatory commercial pelagic fish using finer scale data. This thesis demonstrates the PI, and similar instruments, are a cost effective method that can provide a similar description to existing methods as well as provide new insight into plankton ecology by yielding fine spatiotemporal data.

ACKNOWLEDGEMENTS

Foremost I would like to thank Sophie Pitois. Since meeting over four years ago for a preinteview-interview, Sophie has been an invaluable mentor in not just how to write and 'do' science but how to become a professional, develop and explain my ideas and understand the workings of the academic world and my place in it. I am deeply grateful for everything Sophie has done for me during the PhD and continues to do post-PhD. Thank you.

I have been incredibly lucky with rest of my supervisory panel and I am grateful to each of them. Phil has helped my refine and challenged my ideas and the resulting science has benefited hugely from it. Working together on developing tools are some of the fondest memories of the PhD. Gill consistently provided big-picture thinking, encouraged me to step back from work and provided invaluable support on how to get through the PhD. The coffees over the last four years were always needed, and will always be appreciated. Julian has kept the PhD, the research and instrument tangible. Collaborating and tinkering on the instrument in the 'dungeon' are times I shall never forget and am extremely grateful for.

My parents have provided all the support one could ever need over the last four years. I am hugely grateful to all they that have done, probably without knowing it, to make the PhD journey that much easier. They have been and will continue to be a great source of inspiration and strength. My brother, and best friend, Ben, has provided all the distractions I could need to enjoy life alongside the PhD. The numerous pub trips listening to PhD students moan about their PhDs is no doubt an unenviable task. They are appreciated, as is the proof reading, but above all I am truly grateful for our relationship. Finally to my late Granddad, who told me 'education can never be taken away', a statement and a person that helped motivate me everyday over the last four years.

To my friends who have provided so many enjoyable memories alongside the PhD. To recent friends who have shared the PhD journey and to old friends, both of whom have provided distraction, catharsis, encouragement and motivation over the last four years.

PUBLISHED WORKS

Chapter 3 : Community Ecology

This chapter was published in the Journal of Plankton Research in 2021.

> 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.

MY CONTRIBUTION

I was responsible for harmonising the data across different years (e.g. merging taxonomic classes) and collected some of the data aboard the RV Cefas Endeavour. I did all the data analysis, created all figures and wrote the manuscript.

OTHERS CONTRIBUTION

Sophie Pitois has led the work on the application of the PI system on-board the RV Cefas Endeavour and edited the manuscript. Hayden Close and Nevena Almedia were responsible for the PI aboard the RV Cefas Endeavour for several years and chiefly responsible for the manually classifying the images as well as quality controlling my taxonomy. Julian Tilbury is the lead software developer for the PI system and contributed to data management. Phil Culverhouse is the lead scientist responsible for the development of the PI technology and contributed to the writing of the manuscript. Gill Malin contributed to editing the manuscript.

Chapter 4 : Fine spatial zooplankton data

This chapter is under review in Progress in Oceanography at time of submission.

> Scott, J., Pitois, S., Creach, V., Malin, G., Culverhouse, P., Tilbury, J. (*in-review*). Resolution changes relationships: optimizing sampling design using small scale zooplankton data. *Progress in Oceanography*.

MY CONTRIBUTION

I collected the data aboard the RV Cefas Endeavour. I led the writing of the manuscript and did all data analysis and created all figures

OTHERS CONTRIBUTION

Sophie Pitois has led the work on the application of the PI system on-board RV Cefas Endeavour. She has managed the project, organized the field work and contributed to editing the manuscript. Veronique Creach was responsible for the chlorophyll methodology, the conversion of fluorescence to chlorophyll and editing the manuscript. Julian Tilbury is the lead software developer for the PI system and contributed to data management. Phil Culverhouse is the lead scientist responsible for the development of the PI technology and contributed to the editing of the manuscript. Gill Malin contributed to editing the manuscript.

CO-AUTHORED ARTICLES

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). Ecological Indicators, 123:107307.

MY CONTRIBUTION

I collated the zooplankton data for the project. I wrote the code to extract the size data for each copepod and created figures for the manuscript.

RELEVANCE TO THESIS

Part of this code is used for the data analysis in Chapters 4 and 5. The statistical analysis in the publication is recreated in Chapter 5 to test the benefit of using continuous zooplankton data compared to the point, sub-sampled data presented in this article.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

Zooplankton (Figure 1.1) have a ubiquitous global distribution in the world oceans and seas. The range in their morphological and behavioural adaptations finds them in the worlds coldest and deepest oceans to the warmest and shallowest seas (Swadling et al., 1997; Wiebe et al., 1988). These organisms represent the most abundant multicellular animals on Earth, outnumbering insects by more than ten orders of magnitude and amassing to a collective biomass 50 times larger than the human population (Schminke, 2007). The word zooplankton, and subsequently the community it describes, is the combination of the Greek words for animal (*Zoon*) and wanderer (*Planktos*). A more modern definition of Zooplankton are those organisms that have the ability to invoke partial movement against a weak current but are unable to move contradictory to a stronger current. The Zoon, or Zoo- prefix signifies the organisms as heterotrophs; their energy is derived from consumption of autotropic phytoplankton, smaller zooplankton or organic particulate matter. Zooplankton further describes both permanent members of the plankton (holoplankton), and temporary visitors (meroplankton), such as fish eggs and larvae. They are crucially positioned between autotrophic phytoplankton and marine species at higher tropic levels, such as fish, marine mammals and sea birds.

This range in morphology and behaviour also poses the most significant challenge to zooplankton sampling. The smallest zooplankton can measure less than 2 µm (e.g. protozoa) long compared to the largest jellyfish which can measure over 2 m with tentacles over 60 m long (Steinberg and Landry, 2017). This range in morphologies makes designing instruments and effective sampling strategies difficult: it is impossible to sample the variety of zooplankton with a single device. This challenge is made more difficult by 'plankton patchiness'. This phenomenon is the seemingly random distribution of plankton in space and time over small spatial scales (Mackas et al., 1985; Abraham, 1998). The devices traditionally used for sampling zooplankton - ring nets, are the gold standard and responsible for the majority of our current understanding of plankton ecology. Although, they are both costly and timely, often involve use of hazardous chemicals and require a division of the sample (subsampling) to make analyse feasible. These limitations are explored in this chapter. Unfortunately, a decline in the taxonomic expertise

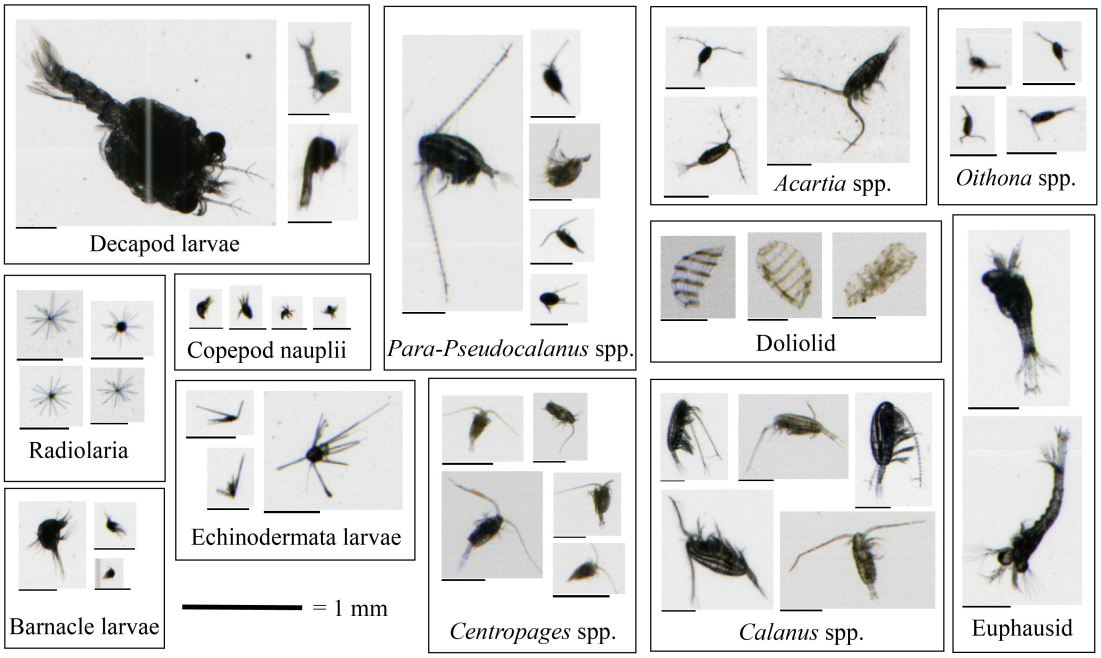

Figure 1.1: Collage of example mesozooplankton images captured by the Plankton Imager (this figure is taken from the published Chapter 3 Scott et al., 2021).

required to identify the zooplankton species (Agnarsson and Kuntner, 2007) means analysing the net-caught samples is increasingly difficult. While these methods have resulted in good understanding of the taxonomy (Mitra et al., 2008; Castellani and Edwards, 2017), life history (Allan, 1976; Varpe et al., 2007) and response to stressors (Garzke et al., 2016; Zervoudaki et al., 2017; Lewis et al., 2013). Although spatial distributions remain uncertain although recent studies showed that the combination of nets, imaging, machine learning and complex factorial analysis methods enabled the characterisation of regional scale communities distributions (Scott et al., 2023; Grandremy et al., 2023). Their importance, in combination with these challenges (both ecological and financial), have placed impetus on developing cost and time effective methods.

The Plankton Imager (PI) is a novel tool for the continuous sampling of zooplankton and has been developed in response to these challenges. The PI has the potential to, in-part, address some of these challenges and provide new insight into the spatial dynamics of zooplankton. This thesis aims to use the PI to overcome some of the challenges associated with zooplankton sampling and in doing so further

understand zooplankton ecology. These challenges are first explored in more detail to provide a context for the PI among other sampling devices in this chapter. The potential benefits of the instrument to gaining new insight into zooplankton ecology are also discussed. A detailed methodology for the instrument, including a standard operating procedure for deployment aboard the RV Cefas Endeavour is detailed in Chapter 2. Following this there are three data analysis chapters that explore the PIs ability to describe the zooplankton. The first chapter explores seasonal changes in the zooplankton community of the Celtic Sea (Chapter 3). The following chapter begins to take advantage of the PIs continuous nature by using smaller scale data to explore the relationship between zooplankton, physical variables and chlorophyll (Chapter 4). Finally, an application of a machine learning classifier is used to obtain zooplankton data at an unprecedented spatial scale and these data are used to examine the relationship with fisheries (Chapter 5). These data chapters are largely self contained and provide short, specific introductions to maintain readability. To conclude, the key findings from the thesis, a discussion of the thesis limitations and potential future applications are synthesised (Chapter 6).

1.1 ZOOPLANKTON'S GLOBAL IMPORTANCE

Zooplankton are found globally and encompass a great diversity of phyla with a plethora of ecological adaptations to a range of environments. Their body size spans over 15 orders of magnitude (Hirst, 2017) and include species that live from less than a 1 day, such as flagellates (WETZEL, 2001), to *immortal* jellyfish (Matsumoto et al., 2019). Zooplankton include both holoplankton and meroplankton. Holoplankton describe those organisms that reside within the plankton for their whole life and include copepods, decapods and gelatinous organisms such as ctenophores. Meroplankton describe those organisms that only exist within the plankton for a short time. These are mostly larval forms of a multitude of animals, for example: fish, crabs and corals.

Zooplankton are crucial prey for a range of planktivorous species. These include a large number commercial important fishes (Confer and Blades, 1975; Beaugrand

et al., 2003), such as cod larvae (Pinnegar et al., 2003) and mackerel (Reid et al., 2001). They are also prey to the worlds largest animals, the blue whale, and other marine megafauna (Sims and Quayle, 1998; Sims, 1999), as well as sea birds (Pakhomov and McQuaid, 1996; Lauria et al., 2012). Zooplankton are also a key component of the global carbon cycle or the 'Biological pump' Hansell (2002); Steinberg et al. (2002); Steinberg and Landry (2017); Boyd et al. (2019). They sequester carbon through excretion of fast-sinking fecal pellets (Ducklow et al., 2001) and contribute to the passive sinking of organic detritus on death (Buesseler et al., 2007). Their vertical migration to the deep ocean (>1000 m) has been shown to help deliver carbon rich lipids below the permanent thermocline (Huld et al., 2015). Zooplankton also have use as an indicator species due to their short life cycles and are proving a reliable indicator for climate change (Taylor et al., 2002; Field and Barros, 2014; Chiba et al., 2018).

Due to their ecological position, zooplankton form an integral component of both ocean biogeochemical and fisheries models. These models can advise the sustainable use of fisheries (Mitra et al., 2014) or inform global climate change predictions (Steinberg and Landry, 2017). The complexity and diversity of zooplankton behaviours, size and spatial distribution, in addition to the variation in life history between taxa, makes accurate parameterisation of the zooplankton component challenging, regardless of model size (Travers et al., 2007; Mitra et al., 2014). Zooplankton monitoring is also mandated by regulatory policy written into domestic and international law. For example in the EU, the Marine Strategy Framework Directive (MSFD, 2009) requires member states to achieve 'Good Environment Status' by 2020. Two of the 11 descriptors of Good Environment Status directly relate to zooplankton, these are: (1) community-level monitoring of zooplankton to inform maintained biodiversity and (2) normal occurrence of food web characteristics. Many policy orientated studies debate how to best achieve these policy goals and what type of monitoring would best inform our progress toward them (Bedford et al., 2019; Romagnan et al., 2016; McQuatters-Gollop et al., 2019).

This brief overview only touches on zooplankton's multitude of roles within the global oceans but demonstrates the socioeconomic, climatic and ecological

motivations for monitoring zooplankton. The introduction contained in each data chapter explains in more detail the rationale for the specific study.

1.1.1 MESOZOOPLANKTON AND THE COPEPODS

Zooplankton are typically divided into three operationally defined classes based on size (Steinberg and Landry, 2017). Those smaller than 200 μm are the microzooplankton, comprised functionally of protistan and juvenile mesozooplankton (Paffenhöfer, 1998; Quevedo and Anadón, 2000). The mesozooplankton are those species that reside within 0.02 – 2 cm, comprised of larger protists, true multicellular animals and Radiolaria (Stoecker and Gustafson, 1996). Those organisms larger than 2 cm, mainly gelatinous varieties and krill, form the macrozooplankton. The PhD research was focused on mesozooplankton. This results from the size range the PI can capture (discussed in Chapter 2.3). Thus, all zooplankton data presented in the thesis is mesozooplankton.

There is exclusive focus on copepods in Chapters 4 & 5. The motivation for this is two fold. Firstly, copepods tend to have relatively homogeneous morphology across different species and a oval distinct body shape. This makes them easier for machine learning classifiers (and manual taxonomists) to correctly classify (see Chapter 2.6.2). The second reason is copepod's ecological significance. Copepods are small crustaceans found both in fresh and marine waters and are the most abundant and diverse multicelluar organisms on earth (Mauchline et al., 1998). They can be hetrotrophic, cannibalistic and parasitic (between themselves and higher fauna) (Huys and Boxshall, 1991). They commonly constitute 55-95% of mesozooplankton biomass (Longhurst, 1985). They perform all functions performed by the general zooplankton, discussed above (e.g. movement of carbon from primary producing phytoplankon, explored in Chapter 4) but have an important role in fisheries which is examined in more detail in Chapter 5. Commercial pelagic fish (e.g. mackerel or herring) primarily, preferentially predate copepods (Garrido et al., 2008) with the size of the copepods being important in driving prey selection, controlling food quality and availability (Pitois et al., 2016; Barton et al., 2013; Van Deurs et al., 2015).

1.2 ZOOPLANKTON SAMPLING

1.2.1 NETS AND MICROSCOPY

The first methodology for the capture and analysis of plankton is usually credited to Thompson, who in 1828 invented a thin-mesh net for qualitatively sampling crustacea larvae (Fraser, 1968). The capture of plankton through netting, and subsequent analysis by microscopy, remains relatively unchanged to the present day and is the gold standard for zooplankton sampling. In 1895, Hensen proposed the first comprehensive methodology in an attempt to steer sampling towards quantitative assessment to begin resolving questions, such as plankton patchiness (Mackas et al., 1985), that still puzzle ecologists today (Wiebe and Benfield, 2003). The Hensen net and the diverse net-type descendants that have evolved over the history of modern oceanography (Skjoldal et al., 2013), have become the cornerstone of zooplankton sampling and are principally responsible for our current understanding of zooplankton ecology. Improvements, adaptations and sampling strategies are discussed in detail in an exhaustive review paper detailing the history of the plankton sampling devices by Wiebe et al. (2017). Ring nets variants are still used routinely in surveys and continue to provide new insight into plankton ecology (Capuzzo et al., 2022; Kajiwara et al., 2022; Head et al., 2022). Despite the limitations discussed below and the subsequent arguments for new devices, ring nets and analysis of a physical specimen by light microscopy will remain a staple tool for sampling zooplankton.

Aside from the complexities that arise from designing effective sampling of the seemingly random, patchy distribution of zooplankton using point sampling (Mackas et al., 1985; Abraham, 1998), issues are inherent with both the netting of plankton and the requisite analysis by microscopy. By design, net samples are subject to a range of uncontrollable physical limitations, predominately escapement. To achieve a minimal level of escapement a compromise is needed to balance extrusion and avoidance (Vannucci, 1968). Extrusion is the escape of organisms which are smaller than, or equal too, the mesh opening (Kofoid, 1897). When the net begins to fill with biota, extrusion pressures within the net can lead to increased escapement as

organisms are pushed through the mesh or lead to the destruction of delicate species in the net periphery (Pitois et al., 2016). Avoidance can occur both passively and actively. Passive avoidance occurs due to hydrodynamic effects that arise from the physical net design; particles in the water are pushed away from the opening as water flows past the device. Active avoidance is performed by those organisms which can detect the approach of the device, usually by detecting the 'bow-wave' of the instrument (Smith et al., 1968). Active avoidance can be described as a factor of swimming speed and net size, the smaller the net size, the lower swimming speed needed and lesser distance for an organism to swim to escape capture (Smith et al., 1968). The speed at which a net can be towed is a function of the mesh and aperture size. Nets have a maximum filtering speed, which, if exceeded results in limited or no new water entering the net, and thus reducing the overall effectiveness of the sampling operation. Vessel or tow speed and mesh size must be considered in parallel for optimal sampling of the desired group (Skjoldal et al., 2013). In effect, the towing vessel must adjust its speed, usually by reducing, to accommodate the sampling.

Issues with the analysis of these samples arise due to the complex, difficult nature of zooplankton taxonomy. In almost all cases, to compromise between time and samples processed, analysis of zooplankton samples consists of sub sampling an aliquot containing 200 individuals and analysis by an expert using light microscopy. In order to resolve the distribution of the full sample, aliquots are simply scaled up by the split factor (Wiebe et al., 2017). Although plankton have been demonstrated to follow a Poisson distribution within a single sample (Postel et al., 2000), allowing for confidence when scaling up, sub sampling typically results in the under or over reporting of rare species (Mack et al., 2012). In addition, analysis by light microscopy, is considered an arduous task (McQuatters-Gollop et al., 2017), where continued concentration is paramount and a high-level of personal diligence is essential. Unquestionably, expert taxonomists, defined by Culverhouse et al. (2014) as individuals that can expect 80 percent consistent repetition rate between samples, have a difficult job. The reliability of experts has been questioned. Culverhouse et al. (2014) empirically assessed both self-consistency and consistency between labs. The study demonstrated significant differences between laboratories working

on an identical sample, describing this as 'cause for concern'. However, the count error is usually eclipsed by the high sampling error that arises between replicate tows, with variability ranging up from $1/6^{th}$ to 6 times per tow (Wiebe and Wiebe, 1968). Inescapable human physiological factors present a further hurdle. Vreugdenhil (1989), summarised by Culverhouse et al. (2003), highlighted several inherent factors that can impair identification: short term memory limit of 5 to 9 items; bias toward recently used labels; positivity bias, fatigue and boredom. Regardless of where error arises, and in absence of an improved method, the continued practice of manual taxonomy is crucial to maintaining time series and furthering understanding. Furthermore, taxonomy is currently facing additional issues, described by multiple authors as a discipline in crisis (Agnarsson and Kuntner, 2007; Pearson et al., 2011; McQuatters-Gollop et al., 2017). These papers highlight reduced funding, the unattractive nature of taxonomy and difficulty it poses for recruitment, as well as the economic and time cost of training.

1.2.2 IMAGING, ACOUSTICS AND DNA

In response to limitations, mainly the financial and time incurred cost associated with traditional ring nets, a multitude of acoustic and optical devices have been developed over the last 30 years. Different mesh sizes in ring nets will sample different portions of the plankton. This is analogous to more 'modern' approaches where each device (acoustic, optical or DNA) will only sample a part of the zooplankton (Lombard et al., 2019). They may be limited be size, depth, or face similar avoidance issues as nets. The variations, as well as what size spectra each device can sample are comprehensively reviewed by Wiebe et al. (2017). The key distinguishing difference between the majority of these 'modern' methods and ring nets is the need to collect a physical sample (excluding DNA, although more recent eDNA approaches negate this, as discussed below). Some zooplankton, particularly copepods, can only be speciated by manipulating the organism to discern key morphological traits (McQuatters-Gollop et al., 2017). Although as DNA identifications are matched with reference species, an ever more detailed taxonomic resolution can likely be gained

without needing a physical sample. Thus, newer methods are not designed to replace ring nets but rather offer a more cost effective solution as well as potentially providing new high-frequency, fine spatial and temporal data. Below is a brief review of each the major 'modern' approaches with an emphasis on imaging devices to better frame the niche for the PI amongst other devices.

DNA

DNA is revolutionising many areas of ecology and its application to zooplankton is growing. DNA / RNA sequencing of zooplankton has resulted in an explosion of new species (Metzker, 2010). Although to successfully describe and categories these species, reference species are needed which do not currently exist for all DNA-found species (MacLeod et al., 2010; Djurhuus et al., 2018). Traditional DNA approaches requires physical capture of the organism which is frequently achieved with ring nets. Thus, these approaches are bound by the same constraints. Furthermore, successful DNA analysis demands unique preservation protocols as standard preservation by formalin, used for light microscopy, can damage DNA (Bucklin, 2000; Bucklin and Allen, 2004; Wiebe et al., 2017). More recently environmental DNA (eDNA) is being demonstrated as a powerful, less complex (e.g. does not require capturing of the individual and only requires a water sample) alternative to sampling zooplankton (Yang and Zhang, 2020; Djurhuus et al., 2018; Sun et al., 2018). Environmental DNA is DNA that is has been shed by an organisms rather than being DNA collected directly from the organism (Shokralla et al., 2012). For example, a water sample might be taken from the ocean and its contents analysed for the DNA present. Like all methods, eDNA has it's own associated limitations. A major hurdle with eDNA, not specifically to zooplankton, is not being able to obtain with confidence where the individual who 'shed' the DNA was spatially or temporally as eDNA is influenced by water movements such as tides or currents (Barnes et al., 2014).

The application of DNA and eDNA to zooplankton is growing (MacHida et al., 2021) and will likely prove to be a staple in future zooplankton sampling. The growing consensus among both traditional taxonomists and genetic analysts is for traditional descriptive taxonomy to work in unison with the added resolution genomics brings,

working towards an integrated approach of the methods (Saunders and McDevit, 2012; McQuatters-Gollop et al., 2017; Will et al., 2005).

ACOUSTICS

High frequency acoustics provide another approach for describing the zooplankton, following the sample principles as their use with fish (e.g. fishing vessel's fish finding devices). The echo (or 'backscatter') from an acoustic signal can be used to discern biological data. Different acoustic frequencies can be used to target specific organisms or size ranges (Foote and Stanton, 2000). Within the plankton these devices can target the mesozooplankton, the macrozooplankton (> 20 mm) and micronekton (mainly juvinelle pelagic fish) (Blanluet et al., 2019; Brophy and Danilowicz, 2002; Sieburth et al., 1978; Mair et al., 2005). Through recent developments in acoustic methods, broadband acoustics are allow its application to improve specific species or group identification (Stanton, 2012; Jech et al., 2017). For a recent example of the application of acoustics to zooplankton see (Blanluet et al., 2019) (2019). This study was able to identify several taxonomic groups (Siphonophores, Copepods, Pteropods and Euphausiids).

One of the distinct advantages of acoustic devices over nets, (e)DNA and optical devices is their ease of deployment facilitated by low power demand. This means acoustics can be used to describe the zooplankton by deployment on remote vehicles, e.g. gliders (Berge et al., 2020; Ohman et al., 2019) or underwater vehicles (Guihen et al., 2014). These vehicles can often sample places that are inaccessible or unsafe to sample, e.g. under sea ice (Berge et al., 2020).

In its current state the technology can make rapid estimates of biomass, size and numbers of zooplankton but identification of individual species cannot be made (Wiebe et al., 2017; Benoit-Bird and Lawson, 2016). These techniques and instruments are quickly developing and it is becoming increasingly common to find these devices used in tandem with optical devices (Ohman et al., 2019) and nets (Blanluet et al., 2019). Acoustics can provide fast, cheap data on zooplankton, including size distributions (Holliday et al., 1989; Tanaka et al., 2021) and its continued use will be essential to further understanding zooplankton ecology.

IMAGING

Zooplankton imaging is becoming an increasingly common approach to sampling zooplankton. There are several approaches to imaging the zooplankton with the main difference being where or when the zooplankton are imaged, these are bench top, in situ (deployed) and in situ (abord ship). For a in depth review of modern imaging devices see: Romagnan et al. (2016); Wiebe et al. (2017) and Lombard et al. (2019).

The closest imaging systems to traditional microscopy are those devices that image the zooplankton post sample collection, these are refereed to as 'bench top'. The sample is collected in a traditional way - with a ring net. Stored and then analysed back ashore in a laboratory. The two most prominent examples of these are the ZooScan (Gorsky et al., 2010; Grosjean et al., 2004; Naito et al., 2019; Schultes et al., 2013) and FlowCam (Alvarez et al., 2011, 2014; Kerr et al., 2020; Buskey and Hyatt, 2006). These devices have been the principle method in 100s of zooplankton studies. The ZooScan is a flat bed scanner that images a sampled poured onto the scan bed. In the FlowCam a poured sample flows past a camera. Due to their reliance on physical sample collection they are subject to many of the constraints associated with physical sample collection (discussed above). Furthermore, they are also associated with many of the limitations of imaging, such as orientation of a specimen (discussed below). These devices are globally established and imaging sharing platforms (mainly Ecotaxa: https://ecotaxa.obs-vlfr.fr/) (Picheral et al., 2017) provide an easy and collaborative approach to image classification.

The vast majority of image devices are deployed and towed. The VPR is one of the original, widely used devices to image the plankton (Davis et al., 1992; Ollevier et al., 2022). The assembly of the VPR is representative of the many optical devices that have been developed since, each with varying adaptations, often to a specific target taxon (e.g. the In Situ Ichthyoplankton Imaging System) (Cowen and Guigand, 2008). Due to the challenges associated with photographing plankton, devices are often tailored (or limited by focus) to a specific size or taxa. In order to mitigate the uncertainty of a passing particles distance from the camera and thus ensure a focused image, devices primarily consisted of two designs.

Closed systems are those where water flows through an aperture and particles in the water are imaged. Open-range system is simply a camera pointing out into the open water. The VPR, ZOOVIS (Benfield et al., 2001) and Under Water Vision Profiler (Picheral et al., 2010) are popular examples of open-range systems where the position of the organism relative to the camera is not certain and this brings issues with image quality. Blurry images are the result of a particle at an uncapturable distance, either too near or too far and thus outside the focus range (Marini et al., 2015; Colas et al., 2018). While closed systems remove this chance for error by assuring particles are imaged at a consistent, focused distance, they are restricted by the maximum size of the inlet hole (Colas et al., 2018). Image resolution can also be limiting factor. Smaller resolution images do not allow for zooming on the image. The higher the image resolution the more likely small, distinguishing morphological traits will be captured which are needed to speciate individuals. Another consideration, or compromise, is the increase storage cost (in bytes) incurred by increased resolution, discussed below. Although, newer camera technologies and increased storage on smaller physical disks should result in less of a compromise in image resolution. The device inlet size is the main control on determining the maximum size of imaged particles.

Optical tools have a few key advantages over traditional methods, arising from the capture of images as opposed to the physical capture of the individual. Foote and Stanton (2000) highlights the increased vertical and spatial resolution of optical sampling systems, suggesting optical systems have the potential to provide abundance data at shorter temporal intervals when compared to net sampling. Optical devices also have the ability to sample delicate taxa that would otherwise be destroyed by external pressures during capture by netting (Davis et al., 2004; Biard et al., 2016). The capture of images makes the use of hazardous preservative chemicals, commonly formalin, redundant and means the collected digital data can be easily transferred, copied, archived and validated.

Established examples of towed imaging devices include the Optical Plankton Recorder (Herman, 1988). Towed devices have similar requirements to nets; they have a maximum speed and maximum sea state, where if either are exceeded the instrument ceases to function or can become damaged or unsafe to operate (Pitois

et al., 2018). They often require adjustments to the vessel behaviour or hardware, e.g. reduced speed or suitable gear for towing and deployments (Culverhouse, 2015; Pitois et al., 2016). Ideally these devices are towed alongside the ship and require robust side gantry equipment infrequently found on smaller vessels making integration onto these vessels difficult (Pitois et al., 2016). Instead the device is often towed in the wake, which wholly disrupts small-scale plankton distributions and adds to passive avoidance effects (Davis et al., 2005).

Finally there are those devices that reside aboard ship (Colas et al., 2018). These are much less common than their towed of bench top counterparts. The Plankton Imager, the device central to the this thesis is introduced below and the various ecological applications are explored in the following chapters.

TOO MANY IMAGES?

Optical devices also share in unavoidable issues, such as the discussed active and passive avoidance and hydrodynamic effects of the instrument. Alongside, a more specific hurdle has arisen from zooplankton imaging: the sheer number of images. For example, a system imaging at 30 Hz may collect over 2.5 million images per day (Wiebe et al., 2017). Sorting all images manually would be an impossible task similar to sampling all organisms caught in a ring net. Recent developments in applying machine learning to zooplankton images means there is the potential to analyse and classify all images. This would yield zooplankton data at unprecedented spatial and temporal scales and provide new insight into plankton ecology. This potential and the absolute size of the data imaging devices yield requires new techniques to sort, analyse and classify the data.

1.2.3 MACHINE LEARNING CLASSIFICATION

Machine learning (Jordan and Mitchell, 2015) is being increasingly used in nearly all research disciplines, from economics to physics to... zooplankton (Irisson et al., 2022; Orenstein et al., 2022). Machine learning classifiers are statistical algorithms that can be fine tuned to try to match unknown data to known data categories. A training set is required to 'train' these algorithms for use on real-world data. The algorithm

fine tunes various statistical parameters to best match the training data. This fine tuning can be considered the learning element of the classifier - it's learning which statistical tweaks are needed to ensure a best fit of the known data (Irisson et al., 2022). These training sets comprise data manually classified by experts. Typically a machine learning algorithm will be trained on two thirds of the training set with the final third used for testing the accuracy of the classifier (Faillettaz et al., 2016). Several limitations arise with classifiers. Firstly, training sets are limited by the number of classified data and the accuracy of those data manually classified. The larger the training set, in terms of balanced example images per category, the higher the success of the classifier (Faillettaz et al., 2016). Furthermore, the success of a classifier is therefor largely subject to the quality of the training set, where quality is defined by size of the training set and the absence of errors (images in the wrong categories).

APPLICATION TO ZOOPLANKTON

Machine learning (ML) is revolutionising many areas of ecology (Prasad et al., 2006; Cutler et al., 2007) and may present a solution to analyse the entirety of the newly amassed, and continually expanding, vast image archive. It is commonly paired with imaging devices as it presents a viable solution to sorting and classifying millions of images with a high degree of accuracy in a relatively short space of time. In the four years of PhD research the application of ML to zooplankton has significantly matured, for a review of the current application and detailed history of ML in zooplankton science see: Irisson et al. (2022). Although in the current state, many optical records still require or undergo manual classification (Culverhouse et al., 2016; Ellen et al., 2019; Irisson et al., 2022) and exclusive use of imagery and machine learning are unlikely replace traditional methods (MacLeod et al., 2010; Giering et al., 2022).

The are several hurdles that pose specific challenges to the application of ML to zooplankton classification, mainly associated with the diverse morphologies and total abundance ('they all look the same'). The limits of machine classifiers can be described as both internal and external. Internal limitations are dependent on the chosen algorithm. Classifiers such as convolutional neural networks (CNN) and support-vector machine (SVM) are subject to over fitting. This is a phenomenon

whereby a classifier is over trained, and bias are introduced within the algorithms statistical mechanism resulting in incorrect classification of new data. Although CNNs are the recently the favourable approach to zooplankton classification which can reports accuracys of up to 70-80% with modest taxonomic resolution (e.g. family level) (Irisson et al., 2022; González et al., 2019; Orenstein et al., 2020).

External issues are those that concern both the data being passed to the classifier (in our case zooplankton images), and the quality of the training set (Fernandes et al., 2009). Zooplankton classification itself presents a set of unique challenges to machine learning, presenting large sources of external errors (or user-dervied errors). Irisson et al. (2007) highlight the predominant challenges associated with the taxa themselves, independent of image capture, mainly: (1) the morphological diversity of the zooplankton; (2) the presence of non-living particulate matter within the water column; (3) intraspecies size variation; (4) intraspecies ontogenetic morphological variation. These issues are exacerbated by the problematic task of imaging zooplankton.

The application of ML to zooplankton remains somewhat silo-ed, where each institute of group applies its own algorithms and big-data pipelines. Breaking down these barriers and sharing information across databases would greatly imporove the progress in sucesfful application of ML. For a review of several different approaches to plankton science and their comparative success, see Rani et al. (2021) and The Turing Centre (2021).

1.3 THE PLANKTON IMAGER

The Plankton Imager (PI) is a relatively new tool for sampling zooplankton in the open ocean. The motivation for its creation was to create a cost effective, easy-to-integrate, continuous, automated plankton sampler which harnesses the discussed benefits of imaging while negating the discussed issues associated with nets or towed imaging devices and manual taxonomy. The ultimate ambition is to use machine learning to automatically classify the images in near-real time. 'Easy-to-integrate' means the ability to deploy the instrument with relative ease onto multi disciplinary surveys.

It requires no modification to the ship's speed, direction or course and uses the ship's existing infrastructure (a continuous flow pump found on all research ships). 'Automated' means the devices require little to no human interaction, the device is turned on at the start of the survey and left running till the end. The methodology is explained in detail in Chapter 2. In brief, the PI is a line scan camera that images all passing particles using an in-flow system from the ship's continuous water supply.

The PhD research has helped the project realise these aims, although the thesis mainly demonstrates the ecological applications of the Plankton Imager and explains the methodology in detail. What is not included, is the continual back and fourth between the ecology team and the instrument designers to make small and larger improvements to realise these aims and bring the instrument to a commercially viable version. The development and this collaboration started before the PhD research. A very brief summary of the existing literature and know limitations of the PI are detailed below.

1.3.1 THE PI PRE-PHD

The following is a brief review of the ecological studies published using predecessors of the PI. The known limitations of the PI are briefly reviewed. These studies helped inform the direction of the PhD research.

The initial prototype that would go on to become the PI and finally the commercial version: PI-10, see Chapter 2.8, was called the Line-Scanning Zooplankton Analyser (LiZA) (Culverhouse, 2015; Culverhouse et al., 2016). The LiZA was developed in response to the discussed challenges associated with zooplankton sampling. The LiZA was used once on the Atlantic Meridian Transect program (Aiken et al., 2000) in 2011 in conjunction with the Optical Plankton Recorder (Herman, 1992) for truthing purposes. The name: Plankton Image Analyser (PIA), was coined for the software that contained the machine learning classifier that attempted to identity the captured images. Following a new collaboration with Cefas the name 'PIA' was adopted to describe all parts of the imaging device (e.g. the camera and the machine learning software). Hardware and software upgrades are described in Chapter 2.2. Prior to

the PhD, the PIA was used for a similar ground truthing exercise aboard the RV Cefas Endeavour.

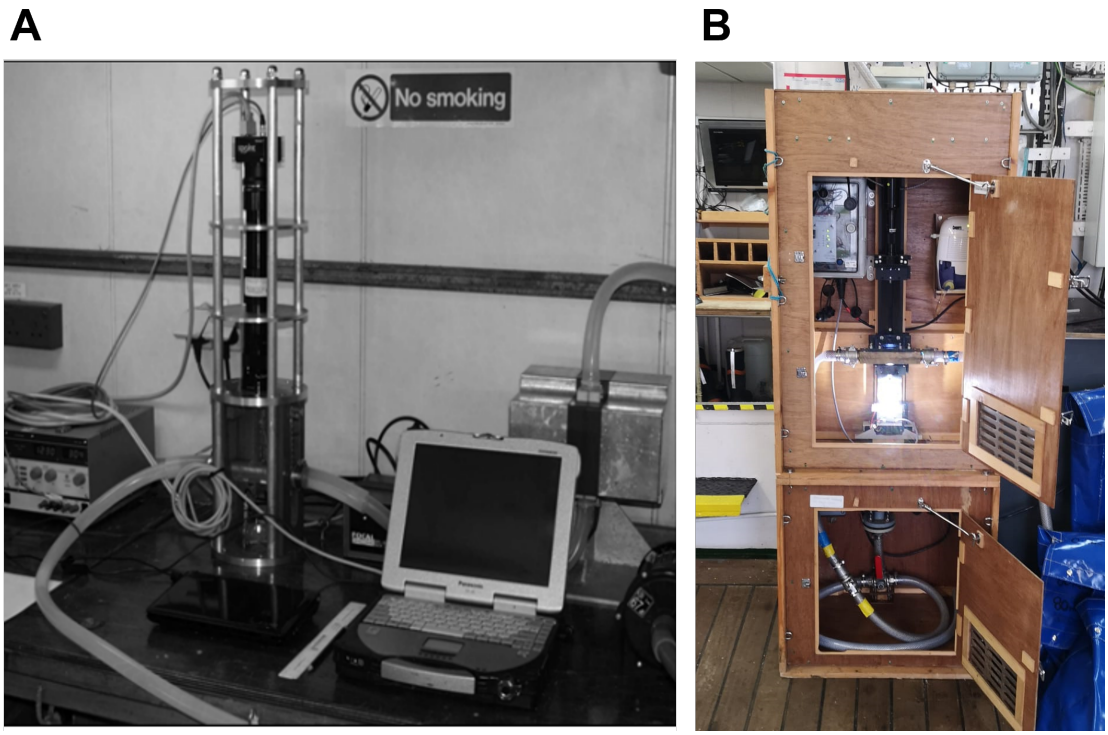

Figure 1.2: A: LiZa set up aboard the AMT21 survey, line scan camera (central) and pumped water supply hoses (right). **B** Plankton Imager Analyser aboard the RV Cefas Endeavour showing the flow cell (gold, middle of the unit) and line scan camera (black, atop the flow cell). **A** was reproduced from Culverhouse et al. (2016).

LINE-SCANNING ZOOPLANKTON ANALYSER, LIZA

The LiZa was used both as a continuous sampler, much the same as the PI, and a discreet sampler to image the contents of ring nets which were pumped through the instrument from a holding tank (Culverhouse, 2015) (Figure 1.2**A**). Although this type of discreet sampling is a possibility with the PI, it is only used in this fashion for instrument testing. The continuous sampling, from the ship's continuous flow, was at a reduced rate, compared to the PI, of 12.5 litres per minute. This method forms the basis of the modern PI and the detailed description in Culverhouse (2015) still applies for much of the image processing used in the PI. The then-termed 'Plankton Imager Analyser' software was used to attempt to classify the 600,000 imaged particles. The LiZA showed marginally higher counts in medium size classes and reduced counts

in smaller and larger size classes when compared to the Optical Plankton Recorder (Herman, 1988). The study reported a 67 % classifier success rate across 7 pre-defined taxonomic classes, rising to 90 % with an expert validation step, with 4 % of captured images being blurred. These studies also served to reveal areas for improvement with the device, mainly: (1) under-reporting of rarer classes due to subsampling, (2) confusion between artefacts, detritus, blurred images and plankton, (3) issues with specific morphological types, mainly Chaetognatha, a long, thin looking taxa which is often cropped by the image capture process, and (4) human derived training set errors impinging classifier performance.

From an ecological perspective, the LiZa articles (Culverhouse, 2015; Culverhouse et al., 2016) begin to show the benefit of sampling zooplankton continuously and the advantages of imaging over capture *then* store *then* subsample. Culverhouse et al. (2016) highlights the potential importance of capturing species such as filamentous cyanobacteria *Trichodesmium* spp., a species that is usally destroyed by ring nets and preservation by formalin. The articles also emphasises the speed between sampling and results and comparisons are drawn the Continuous Plankton Recorder which typically takes three months from sample collection to finished data. The sentiments of the paper, in terms of the need for a device such as the PI and its benefits for zooplankton sampling, are echoed and built on in this chapter.

PLANKTON IMAGE ANALYSER

In 2016, following a new collaboration between Cefas and the LiZa instrument designers, the Plankton Imager Analyser (Figure 1.2**B**), now used to describe both the camera system and the machine learning classifier, was installed on the RV Cefas Endeavour as part of the PELagic ecosystem in the western English Channel and eastern CelTIC Sea (Working Group of International Pelagic Surveys, 2015) survey. This version of the instrument, excluding a few smaller upgrades (see Chapter 2.2), is the version used for all data collected for the PhD. These data, collected for the publication by Pitois et al. (2018), were also used as part of the analysis in Chapter 3. This study compared the PI with the Automatic Litter and Plankton Sampler (CALPS) and traditional ring net vertical hauls. The deployment was also to test

and demonstrate the ease of the PIAs integration as a low-cost, easily integrated sampling device. The study found good agreement across all devices but suggested the image capture and analysis step was partly responsible for discrepancies between the datasets.

These publications, prior to the PhD research, proved the device. Knowing the PI provided an accurate description of the mesozooplankton framed the work for the thesis and subsequent publications. Work is now focused on harnessing the PIs continuous nature and the development of machine learning classifiers.

The limitations of the PI are detailed in Chapter 2.5. The PhD thesis does address some of these issues which are reviewed in the final synthesis chapter 6.

1.4 CONCLUSION

This chapter briefly summarises zooplankton's global importance, their key role within the global ecosystem and carbon cycle and their mandated monitoring by policy. The overview of the challenges associated with zooplankton are described followed by a brief review of the existing, major methods. The limitations of these methods coupled with zooplankton's global importance are used to demonstrate the need and potential benefits of the PI. On thesis submission, the PI is an unique instrument in terms of its speed and it's continuous, automated nature which requires little attention to vessel operation or installation.

Each chapter builds on this review in its respective introduction, giving more detail as to why the PI provides a good solution to the chapter's specific question. The global importance of zooplankton is only briefly mentioned in each introduction and there are infrequent examples or discussion of other specific devices or methods. This chapter can therefor be used for reference as to the ecological or economical (in terms of sampling ease) importance of the study. Finally, the aims and objectives for the thesis are listed below.

1.5 AIMS

The overarching aim of the thesis is to contribute to the continued development of the PI and demonstrate its novel applications to zooplankton ecology, mainly through the continuous and automated sampling. The thesis also aims to explore the data from an ecological perspective, identify potential future applications as well as address the instrument's limitations.

1.6 OBJECTIVES

The thesis aims can be broken down into the following objectives which corresponds with the chapter structure:

1. Through literature review provide a context for the PI amongst other zooplankton samplers and demonstrate its potential value to zooplankton research.

2. Contribute to ongoing development of the system and its methodology for full operational and optimal deployment.

3. Evaluate the PI's ability for community and biodiversity applications.

4. Develop an analytic method to best use the PI's ability to collect continuous, high frequency zooplankton data.

5. To help develop and test a machine learning algorithm on image data obtained from the PI to yield unprecedented fine scale data.

6. To summarise the major findings from the thesis to provide context and suggestions for future studies.

1.7 THESIS STRUCTURE

The thesis is split into several sections. This chapter has provided the context for the study, reviewed the PI literature to date and explored ecological applications. This chapter addresses objective 1.

Chapter 2 provides a detailed method for the PI. The standard operating procedure for deployment aboard the RV Cefas Endeavour is explored. The stages involved in image capture and processing are explained. The machine learning classifiers are also detailed in this chapter. There is some overlap with the short methodologies in the data chapters. This chapter addresses objective 2.

The data chapters are written as stand alone chapters. This is more true for Chapter 3 and 4 as both Chapters were written for publication (see *Published Works*). Therefore there is minor overlap in the introductions and methodologies and limited internal reference within the thesis. Chapter 5 was not written for publication but will form part of a paper following the PhD research reviewing the applications of machine learning to plankton. For this reason there is reduced overlap with the method with a higher incident of internal referencing. Chapter 5 uses the PI in a similar way to a traditional ring net through sub-sampling and more detailed taxonomic resolution. It asses the PI's capacity to detect changes in the zooplankton community and addressees objective 4. Chapter 4 begins to take advantage of the PI's continuous nature. Here data are temporally subsampled and compared with other continuous physical and chlorophyll data. The final data chapter, Chapter 5, takes advantage of a machine learning classifier developed during the course of the PhD research. Here continuous, non subsampled data are used to compare with continuous fisheries data.

The final part of the thesis, Chapter 6 discusses the limitations of the instrument and the data. Finally the results of the thesis are synthesised and suggestions made for future applications of the PI.

1.8 BIBLIOGRAPHY

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, 391(6667):577–580.

Agnarsson, I. and Kuntner, M. (2007). Taxonomy in a changing world: Seeking solutions for a science in crisis. *Systematic Biology*, 56(3):531–539.

Aiken, J., Rees, N., Hooker, S., Holligan, P., Bale, A., Robins, D., Moore, G., Harris, R., and Pilgrim, D. (2000). The Atlantic Meridional Transect: Overview and synthesis of data. *Progress in Oceanography*, 45(3-4):257–312.

Allan, J. D. (1976). Life history patterns in zooplankton. *The American Naturalist*, 110(971):165–180.

Alvarez, E., Lopez-Urrutia, A., Nogueira, E., and Fraga, S. (2011). How to effectively sample the plankton size spectrum? A case study using FlowCAM. *Journal of plankton research*, 33(7):1119–1133.

Alvarez, E., Moyano, M., Lopez-Urrutia, A., Nogueira, E., and Scharek, R. (2014). Routine determination of plankton community composition and size structure: a comparison between FlowCAM and light microscopy. *Journal of plankton research*, 36(1):170–184.

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., and Lodge, D. M. (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology*, 48(3):1819–1827.

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., Kiørboe, T., and Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, 16(4):522–534.

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(6967):661–664.

Bedford, J., Johns, D., and McQuatters-Gollop, A. (2019). A century of change in North Sea plankton communities explored through integrating historical datasets. *ICES Journal of Marine Science*, 76(1):104–112.

Benfield, M., Schwehm, C., and Keenan, S. (2001). ZOOVIS: a high resolution digital camera system for quantifying zooplankton abundance and environmental data. *American Society of Limnology and Oceanography*, pages 12–17.

Benoit-Bird, K. J. and Lawson, G. L. (2016). Ecological Insights from Pelagic Habitats Acquired Using Active Acoustic Techniques. *Annual Review of Marine Science*, 8(1):463–490.

Berge, J., Geoffroy, M., Daase, M., Cottier, F., Priou, P., Cohen, J. H., Johnsen, G., McKee, D., Kostakis, I., Renaud, P. E., Vogedes, D., Anderson, P., Last, K. S., and Gauthier, S. (2020). Artificial light during the polar night disrupts Arctic fish and zooplankton behaviour down to 200 m depth. *Communications Biology*, 3(1):1–8.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., and Not, F. (2016). In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*, 532(7600):504–507.

Blanluet, A., Doray, M., Berger, L., Romagnan, J. B., Le Bouffant, N., Lehuta, S., and Petitgas, P. (2019). Characterization of sound scattering layers in the Bay of Biscay using broadband acoustics, nets and video. *PLoS ONE*, 14(10):1–19.

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., and Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, 568(7752):327–335.

Brophy, D. and Danilowicz, B. S. (2002). Tracing populations of Atlantic herring (Clupea harengus L.) in the Irish and Celtic Seas using otolith microstructure. *ICES Journal of Marine Science*, 59(6):1305–1313.

Bucklin, A. (2000). Methods for population genetic analysis of zooplankton. In *ICES Zooplankton Methodology Manual*, pages 533–570. Elsevier.

Bucklin, A. and Allen, L. D. (2004). MtDNA sequencing from zooplankton after long-term preservation in buffered formalin. *Molecular Phylogenetics and Evolution*, 30(3):879–882.

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., Bishop, J. K. B., Casciotti, K. L., Dehairs, F., and Elskens, M. (2007). Revisiting carbon flux through the ocean's twilight zone. *Science*, 316(5824):567–570.

Buskey, E. J. and Hyatt, C. J. (2006). Use of the FlowCAM for semi-automated recognition and enumeration of red tide cells (Karenia brevis) in natural plankton samples. *Harmful Algae*, 5(6):685–692.

Capuzzo, E., Wright, S., Bouch, P., Collingridge, K., Creach, V., Pitois, S., Stephens, D., and Kooij, J. v. d. (2022). Variability in structure and carbon content of plankton communities in autumn in the waters south-west of the UK. *Progress in Oceanography*, 204(October 2021):102805.

Castellani, C. and Edwards, M. (2017). *Marine Plankton: A Practical Guide to Ecology, Methodology, and Taxonomy*. Number February.

Chiba, S., Batten, S., Martin, C. S., Ivory, S., Miloslavich, P., and Weatherdon, L. V. (2018). Zooplankton monitoring to contribute towards addressing global biodiversity conservation challenges. *Journal of Plankton Research*, 40(5):509–518.

Colas, F., Tardivel, M., Perchoc, J., Lunven, M., Forest, B., Guyader, G., Danielou, M. M., Le Mestre, S., Bourriau, P., Antajan, E., Sourisseau, M., Huret, M., Petitgas, P., and Romagnan, J. B. (2018). The ZooCAM, a new in-flow imaging system for fast onboard counting, sizing and classification of fish eggs and metazooplankton. *Progress in Oceanography*, 166:54–65.

Confer, J. L. and Blades, P. I. (1975). Omnivorous zooplankton and planktivorous fish. *Limnology and Oceanography*, 20(4):571–579.

Cowen, R. K. and Guigand, C. M. (2008). In situ ichthyoplankton imaging system (ISIIS): System design and preliminary results. *Limnology and Oceanography: Methods*, 6(2):126–132.

Culverhouse, P. F. (2015). An Instrument for Rapid Mesozooplankton Monitoring at Ocean Basin Scale. *Journal of Marine Biology and Aquaculture*, 1(1):1–11.

Culverhouse, P. F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M., and Picheral, M. (2014). An empirical assessment of the consistency of taxonomic identifications. *Marine Biology Research*, 10(1):73–84.

Culverhouse, P. F., Williams, R., Gallienne, C., Tilbury, J., and Wall-Palmer, D. (2016). Ocean-Scale Monitoring of Mesozooplankton on Atlantic Meridional Transect 21. *Journal of Marine Biology and Aquaculture*, 2(1):1–13.

Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., and González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, 247:17–25.

Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11):2783–2792.

Davis, C. S., GALLAGER, S. M., Berman, M. S., Haury, L. R., and Strickler, J. R. (1992). The Video Plankton Recorder (VPR): Design and initial results. *Arch. Hydrobiol. Beih.*, 36:67–81.

Davis, C. S., Hu, Q., Gallager, S. M., Tang, X., and Ashjian, C. J. (2004). Real-time observation of taxa-specific plankton distributions: An optical sampling method. *Marine Ecology Progress Series*, 284(May 2016):77–96.

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. *Limnology and Oceanography: Methods*, 3(2):59–74.

Djurhuus, A., Pitz, K., Sawaya, N. A., Rojas-Márquez, J., Michaud, B., Montes, E., Muller-Karger, F., and Breitbart, M. (2018). Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. *Limnology and Oceanography: Methods*, 16(4):209–221.

Ducklow, H. W., Steinberg, D. K., and Buesseler, K. O. (2001). Upper ocean carbon export and the biological pump. *Oceanography*, 14(SPL.ISS. 4):50–58.

Ellen, J. S., Graff, C. A., and Ohman, M. D. (2019). Improving plankton image classification using context metadata. *Limnology and Oceanography: Methods*, 17(8):439–461.

Faillettaz, R., Picheral, M., Luo, J. Y., Guigand, C., Cowen, R. K., and Irisson, J. O. (2016). Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60–77.

Fernandes, J. A., Irigoien, X., Boyra, G., Lozano, J. A., and Inza, I. (2009). Optimizing the number of classes in automated zooplankton classification. *Journal of Plankton Research*, 31(1):19–29.

Field, C. B. and Barros, V. R. (2014). *Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects*. Cambridge University Press.

Foote, K. G. and Stanton, T. K. (2000). Acoustical methods. In *ICES Zooplankton methodology manual*, pages 223–258. Elsevier.

Fraser, J. H. (1968). The overflow of oceanic plankton to the shelf waters of the north-east Atlantic. *Sarsia*, 34(1):313–330.

Garrido, S., Ben-Hamadou, R., Oliveira, P. B., Cunha, M. E., Chícharo, M. A., and van der Lingen, C. D. (2008). Diet and feeding intensity of sardine Sardina pilchardus: correlation with satellite-derived chlorophyll data. *Marine Ecology Progress Series*, 354:245–256.

Garzke, J., Hansen, T., Ismar, S. M. H., and Sommer, U. (2016). Combined effects of ocean warming and acidification on copepod abundance, body size and fatty acid content. *PLoS ONE*, 11(5):1–22.

Giering, S. L. C., Culverhouse, P. F., Johns, D. G., Mcquatters-gollop, A., and Pitois, S. G. (2022). Are plankton nets a thing of the past ? An assessment of in situ imaging of zooplankton for large-scale ecosystem assessment and policy decision-making. (November):1–16.

González, P., Castaño, A., Peacock, E. E., Díez, J., Del Coz, J. J., and Sosik, H. M. (2019). Automatic plankton quantification using deep features. *Journal of Plankton Research*, 41(4):449–463.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.-B. B., Cawood, A., Pesant, S., García-Comas, C., and Prejger, F. (2010). Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, 32(3):285–303.

Grandremy, N., Romagnan, J.-B., Dupuy, C., Doray, M., Huret, M., and Petitgas, P. (2023). Hydrology and small pelagic fish drive the spatio–temporal dynamics of springtime zooplankton assemblages over the Bay of Biscay continental shelf. *Progress in Oceanography*, 210:102949.

Grosjean, P., Picheral, M., Warembourg, C., and Gorsky, G. (2004). Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science*, 61(4):518–525.

Guihen, D., Fielding, S., Murphy, E. J., Heywood, K. J., and Griffiths, G. (2014). An assessment of the use of ocean gliders to undertake acoustic measurements of zooplankton: The distribution and density of Antarctic krill (Euphausia superba) in the Weddell Sea. *Limnology and Oceanography: Methods*.

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

Head, E. J., Johnson, C. L., and Pepin, P. (2022). Plankton monitoring in the Northwest Atlantic: a comparison of zooplankton abundance estimates from vertical net tows and Continuous Plankton Recorder sampling on the Scotian and Newfoundland shelves, 1999-2015. *ICES Journal of Marine Science*, 79(3):901–916.

Herman, A. W. (1988). Simultaneous measurement of zooplankton and light attenuance with a new optical plankton counter. *Continental Shelf Research*, 8(2):205–221.

Herman, A. W. (1992). Design and calibration of a new optical plankton counter capable of sizing small zooplankton. *Deep Sea Research Part A, Oceanographic Research Papers*, 39(3-4):395–415.

Hirst, A. G. (2017). Zooplankton Productivity. *Marine Plankton - A Practical Guide to Ecology, Methodology and Taxonomy*, page 678.

Holliday, D. V., Pieper, R. E., and Kleppel, G. S. (1989). Determination of zooplankton size and distribution with multifrequency acoustic technology. *ICES Journal of Marine Science*, 46(1):52–61.

Huld, J. S., W., V. A., Katherine, R., and R., H. M. (2015). Seasonal copepod lipid pump promotes carbon sequestration in the deep North Atlantic. *Proceedings of the National Academy of Sciences*, 112(39):12122–12126.

Huys, R. and Boxshall, G. A. (1991). *Copepod evolution*.

Irisson, J. O., Ayata, S. D., Lindsay, D. J., Karp-Boss, L., and Stemmann, L. (2022). Machine Learning for the Study of Plankton and Marine Snow from Images. *Annual Review of Marine Science*, 14(January):277–301.

Jech, J. M., Lawson, G. L., and Lavery, A. C. (2017). Wideband (15–260 kHz) acoustic volume backscattering spectra of Northern krill (Meganyctiphanes norvegica) and butterfish (Peprilus triacanthus). *ICES Journal of Marine Science*, 74(8):2249–2261.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Kajiwara, K., Nakaya, M., Suzuki, K., Kano, Y., and Takatsu, T. (2022). Effect of egg size on the growth rate and survival of wild walleye pollock Gadus chalcogrammus larvae. *Fisheries Oceanography*, 31(3):238–254.

Kerr, T., Clark, J. R., Fileman, E. S., Widdicombe, C. E., and Pugeault, N. (2020). Collaborative deep learning models to handle class imbalance in flowcam plankton imagery. *IEEE Access*, 8:170013–170032.

Kofoid, C. A. (1897). On some Important Sources of Error in the Plankton Method Author (s): C . A . Kofoid Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/1623279. 6(153):829–832.

Lauria, V., Attrill, M. J., Pinnegar, J. K., Brown, A., Edwards, M., and Votier, S. C. (2012). Influence of Climate Change and Trophic Coupling across Four Trophic Levels in the Celtic Sea. *PLoS ONE*, 7(10).

Lewis, C. N., Brown, K. A., Edwards, L. A., Cooper, G., and Findlay, H. S. (2013). Sensitivity to ocean acidification parallels natural pCO2 gradients experienced by Arctic copepods under winter sea ice. *Proceedings of the National Academy of Sciences*, 110(51):E4960–E4967.

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J. B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O., Miloslavich, P., Lara-Lopez, A., Kudela, R. M., Lopes, R. M., Karp-Boss, L., Kiko, R., Jaffe, J. S., Iversen, M. H., Irisson, J. O., Hauss, H., Guidi, L., Gorsky, G., Giering, S. L. C., Gaube, P., Gallager, S., Dubelaar, G., Cowen, R. K., Carlotti, F., Briseño-Avena, C., Berline, L., Benoit-Bird, K. J., Bax, N. J., Batten, S. D., Ayata, S. D., and Appeltans, W. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6(MAR).

Longhurst, A. R. (1985). The structure and evolution of plankton communities. *Progress in Oceanography*, 15(1):1–35.

MacHida, R. J., Kurihara, H., Nakajima, R., Sakamaki, T., Lin, Y. Y., and Furusawa, K. (2021). Comparative analysis of zooplankton diversities and compositions estimated from complement DNA and genomic DNA amplicons, metatranscriptomics, and morphological identifications. *ICES Journal of Marine Science*, 78(9):3428–3443.

Mack, H. R., Conroy, J. D., Blocksom, K. A., Stein, R. A., and Ludsin, S. A. (2012). A comparative analysis of zooplankton field collection and sample enumeration methods. *Limnology and Oceanography: Methods*, 10(1):41–53.

Mackas, D. L., Denman, K. L., and Abbott, M. R. (1985). Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science*, 37(2):653–674.

MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312):154–155.

Mair, A. M., Fernandes, P. G., Lebourges-Dhaussy, A., and Brierley, A. S. (2005). An investigation into the zooplankton composition of a prominent 38-kHz scattering layer in the North Sea. *Journal of Plankton Research*, 27(7):623–633.

Marini, S., Corgnati, L., Mazzei, L., Ottaviano, E., Isoppo, B., Aliani, S., Conversi, A., and Griffa, A. (2015). GUARD1: An autonomous system for gelatinous zooplankton image-based recognition. *MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World*, (October).

Matsumoto, Y., Piraino, S., and Miglietta, M. P. (2019). Transcriptome characterization of reverse development in Turritopsis dohrnii (Hydrozoa, Cnidaria). *G3: Genes, Genomes, Genetics*, 9(12):4127–4138.

Mauchline, J., Blaxter, J. H., Southward, A. J., and Tyler, P. A. (1998). *The Biology of Calanoid Copepods*. Number 33.

McQuatters-Gollop, A., Atkinson, A., Aubert, A., Bedford, J., Best, M., Bresnan, E., Cook, K., Devlin, M., Gowen, R., Johns, D. G., Machairopoulou, M., McKinney, A., Mellor, A., Ostle, C., Scherer, C., and Tett, P. (2019). Plankton lifeforms as a biodiversity indicator for regional-scale assessment of pelagic habitats for policy. *Ecological Indicators*, 101(February):913–925.

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

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1):31–46.

Mitra, A., Castellani, C., Gentleman, W. C., Jónasdóttir, S. H., Flynn, K. J., Bode, A., Halsband, C., Kuhn, P., Licandro, P., Agersted, M. D., Calbet, A., Lindeque, P. K., Koppelmann, R., Møller, E. F., Gislason, A., Nielsen, T. G., and St. John, M. (2014). Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. *Progress in Oceanography*, 129(PB):176–199.

Mitra, A., K. Banarjee, and Gangopadhyay, A. (2008). *Introduction to Marine Plankton*. Daya Publishing House Delhi.

MSFD (2009). Marine Strategy Framework Directive Newsletter.

Naito, A., Abe, Y., Matsuno, K., Nishizawa, B., Kanna, N., Sugiyama, S., and Yamaguchi, A. (2019). Surface zooplankton size and taxonomic composition in Bowdoin Fjord, north-western Greenland: a comparison of ZooScan, OPC and microscopic analyses. *Polar Science*, 19:120–129.

Ohman, M. D., Davis, R. E., Sherman, J. T., Grindley, K. R., Whitmore, B. M., Nickels, C. F., and Ellen, J. S. (2019). Zooglider: An autonomous vehicle for optical and acoustic sensing of zooplankton. *Limnology and Oceanography: Methods*, 17(1):69–86.

Ollevier, A., Mortelmans, J., Vandegehuchte, M. B., Develter, R., De Troch, M., and Deneudt, K. (2022). A Video Plankton Recorder user guide: Lessons learned from in situ plankton imaging in shallow and turbid coastal waters in the Belgian part of the North Sea. *Journal of Sea Research*, 188:102257.

Orenstein, E. C., Ayata, S., Maps, F., Becker, E. C., Benedetti, F., Biard, T., de Garidel-Thoron, T., Ellen, J. S., Ferrario, F., and Giering, S. L. C. (2022). Machine learning techniques to characterize functional traits of plankton from image data. *Limnology and oceanography*, 67(8):1647–1669.

Orenstein, E. C., Ratelle, D., Briseño-Avena, C., Carter, M. L., Franks, P. J., Jaffe, J. S., and Roberts, P. L. (2020). The Scripps Plankton Camera system: A framework and platform for in situ microscopy. *Limnology and Oceanography: Methods*, 18(11):681–695.

Paffenhöfer, G.-A. (1998). On the relation of structure, perception and activity in marine planktonic copepods. *Journal of Marine Systems*, 15(1-4):457–473.

Pakhomov, E. A. and McQuaid, C. D. (1996). Distribution of surface zooplankton and seabirds across the Southern Ocean. *Polar Biology*, 16(4):271–286.

Pearson, D. L., Hamilton, A. L., and Erwin, T. L. (2011). Recovery Plan for the Endangered Taxonomy Profession. *BioScience*, 61(1):58–63.

Picheral, M., Colin, S., and Irisson, J. O. (2017). EcoTaxa, a tool for the taxonomic classification of images.

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. *Limnology and Oceanography: Methods*, 8(SEPT):462–473.

Pinnegar, J. K., Trenkel, V. M., Tidd, A. N., Dawson, W. A., and Du Buit, M. H. (2003). Does diet in Celtic Sea fishes reflect prey availability? *Journal of Fish Biology*, 63(SUPPL. A):197–212.

Pitois, S. G., Bouch, P., Creach, V., Van Der Kooij, J., Kikkawa, T., Kita, J., Ishimatsu, A., Pitois, S. G., Tilbury, J., Bouch, P., Close, H., Barnett, S., and Culverhouse, P. F. (2016). Comparison of zooplankton data collected by a continuous semi-automatic sampler (CALPS) and a traditional vertical ring net. *Journal of Plankton Research*, 38(4):931–943.

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

Postel, L., Fock, H., and Hagen, W. (2000). Biomass and abundance. In *ICES Zooplankton Methodology Manual*, pages 83–192. Elsevier.

Prasad, A. M., Iverson, L. R., and Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, 9(2):181–199.

Quevedo, M. and Anadón, R. (2000). *Spring microzooplankton composition, biomass and potential grazing in the central Cantabrian coast (Southern Bay of Biscay)*, volume 23.

Rani, P., Kotwal, S., Manhas, J., Sharma, V., and Sharma, S. (2022). Machine Learning and

Deep Learning Based Computational Approaches in Automatic Microorganisms Image Recognition: Methodologies, Challenges, and Developments. *Archives of Computational Methods in Engineering*, 29(3):1801–1837.

Reid, P. C., De Fatima Borges, M., and Svendsen, E. (2001). A regime shift in the north sea circa 1988 linked to changes in the north sea horse mackerel fishery. *Fisheries Research*, 50(1-2):163–171.

Romagnan, J. B., Aldamman, L., Gasparini, S., Nival, P., Aubert, A., Jamet, J. L., and Stemmann, L. (2016). High frequency mesozooplankton monitoring: Can imaging systems and automated sample analysis help us describe and interpret changes in zooplankton community composition and size structure — An example from a coastal site. *Journal of Marine Systems*, 162(2016):18–28.

Saunders, G. W. and McDevit, D. C. (2012). Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. *Methods in Molecular Biology*.

Schminke, H. K. (2007). Entomology for the copepodologist. *Journal of Plankton Research*, 29(SUPPL. 1):i149–i162.

Schultes, S., Sourisseau, M., Le Masson, E., Lunven, M., and Marié, L. (2013). Influence of physical forcing on mesozooplankton communities at the Ushant tidal front. *Journal of Marine Systems*, 109-110(SUPPL.):S191–S202.

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.

Scott, J., Pitois, S., Creach, V., Malin, G., Culverhouse, P., and Tilbury, J. (2023). Resolution changes relationships: Optimizing sampling design using small scale zooplankton data. *Progress in Oceanography*, 210(December 2022):102946.

Shokralla, S., Spall, J. L., Gibson, J. F., and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular ecology*, 21(8):1794–1805.

Sieburth, J. M., Smetacek, V., and Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. *Limnology and oceanography*, 23(6):1256–1263.

Sims, D. W. (1999). Threshold foraging behaviour of basking sharks on zooplankton: Life on an energetic knife-edge? *Proceedings of the Royal Society B: Biological Sciences*, 266(1427):1437–1443.

Sims, D. W. and Quayle, V. A. (1998). Selective foraging behaviour of basking sharks on zooplankton in a small-scale front. *Nature*, 393(6684):460–464.

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. *Progress in Oceanography*, 108:1–42.

Smith, P. E., Counts, R. C., and Clutter, R. I. (1968). Changes in filtering efficiency of plankton nets due to clogging under tow. *ICES Journal of Marine Science*, 32(2):232–248.

Stanton, T. K. (2012). 30 years of advances in active bioacoustics: A personal perspective. *Methods in Oceanography*, 1-2:49–77.

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-Sea Research Part I: Oceanographic Research Papers*, 49(8):1445–1461.

Steinberg, D. K. and Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science*, 9(1):413–444.

Stoecker, D. and Gustafson, D. (1996). Micro- and mesoprotozooplankton at 140*W in the equatorial Pacific: heterotrophs and mixotrophs. *Aquatic Microbial Ecology*, 10(3):273–282.

Sun, J., Yang, J., and Zhang, X. (2018). Identification and biomass monitoring of zooplankton cladocera species with eDNA metabarcoding technology. *Asian Journal of Ecotoxicology*, (5):76–86.

Swadling, K. M., Gibson, J. A., Ritz, D. A., Nichols, P. D., and Hughes, D. E. (1997). Grazing of phytoplankton by copepods in eastern Antarctic coastal waters. *Marine Biology*, 128(1):39–48.

Tanaka, M., Genin, A., Endo, Y., Ivey, G. N., and Yamazaki, H. (2021). The potential role of turbulence in modulating the migration of demersal zooplankton. *Limnology and Oceanography*, 66(3):855–864.

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

The Turing Centre (2021). DSG Report 2021. Technical report.

Travers, M., Shin, Y. J., Jennings, S., and Cury, P. (2007). Towards end-to-end models for investigating the effects of climate and fishing in marine ecosystems. *Progress in Oceanography*, 75(4):751–770.

Van Deurs, M., Jorgensen, C., and Fiksen, O. (2015). Effects of copepod size on fish growth: A model based on data for North Sea sandeel. *Marine Ecology Progress Series*, 520:235–243.

Vannucci, M. (1968). Loss of organisms through the meshes. In TRANTER, D. and FRASER, J., editors, *Zooplankton sampling, UNESCO monographs on oceanographic methodology.*, pages 77–86. Paris: UNESCO.

Varpe, O., Jorgensen, C., Tarling, G. A., and Fiksen, O. (2007). Early is better: seasonal egg fitness and timing of reproduction in a zooplankton life-history model. *Oikos*, 116(8):1331–1342.

Vreugdenhil, H. (1991). *Bias in human reasoning. Causes and consequences*, volume 77.

WETZEL, R. G. (2001). Planktonic Communities: Zooplankton and Their Interactions With Fish. In WETZEL, R. G. B. T. L. T. E., editor, *Limnology*, pages 395–488. Academic Press, San Diego.

Wiebe, P. H. and Benfield, M. C. (2003). From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography*, 56(1):7–136.

Wiebe, P. H., Bucklin, A., and Benfield, M. (2017). Chapter 10. Sampling, preservation and counting of samples II: Zooplankton. *Marine Plankton: A practical guide to ecology, methodology, and taxonomy*, pages 1–3.

Wiebe, P. H., Copley, N., Van Dover, C., Tamse, A., and Manrique, F. (1988). Deep-water zooplankton of the Guaymas basin hydrothermal vent field. *Deep Sea Research Part A, Oceanographic Research Papers*, 35(6):985–1013.

Wiebe, P. H. and Wiebe, P. H. (1968). Plankton Patchiness: Effects on Repeated Net Tows. *Limnology and Oceanography*, 13(2):315–321.

Will, K. W., Mishler, B. D., and Wheeler, Q. D. (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, 54(5):844–851.

Working Group of International Pelagic Surveys (2015). Manual for International Pelagic Surveys (IPS). Technical report, ICES, Copenhagen.

Yang, J. and Zhang, X. (2020). eDNA metabarcoding in zooplankton improves the ecological status assessment of aquatic ecosystems. *Environment International*, 134:105230.

Zervoudaki, S., Krasakopoulou, E., Moutsopoulos, T., Protopapa, M., Marro, S., and Gazeau, F. (2017). Copepod response to ocean acidification in a low nutrient-low chlorophyll environment in the NW Mediterranean Sea. *Estuarine, Coastal and Shelf Science*, 186:152–162.

2

METHODOLOGY FOR THE PLANKTON IMAGER

2.1 ABSTRACT

The following Chapter details the methodology for the Plankton Imager. The various generations of the instrument are first explained. The Standard Operating Procedure (SOP) for deployment aboard the RV Cefas Endeavour is detailed. These procedures were responsible for all the data collected for the thesis which are used in the subsequent data analysis chapters. The SOP includes all stages from image capture to image analysis and explanation of the bespoke software used to generate ecological data. The calculation of inferred variables, size and biomass, are also detailed. The current limitations of the Plankton Imager are described, mainly the challenges associated with collecting millions of images per day. The various classification algorithms used to sort the images are described as well as their application within the thesis and their construction. Finally, there is a brief report from the sea test of the new commercially available PI-10.

2.2 THE PIA, THE PI AND THE PI-10

The Plankton Imager (PI) is introduced, the publications reviewed and the PIs method in the context of other devices is discussed in Chapter 1.

The PI has been installed on the RV Cefas Endeavour intermittently since 2016. The Plankton Imager Analyser (PIA, 2016-2020) was the original name for the device. From 2020 onward, the device was renamed to Plankton Imager (PI). There were no hardware changes from the PIA to the PI with the exception of upgrading the glass in the flow cell to sapphire to reduce scratching from passing particles. All data for the PhD and publications to date have been collected using the PIA and the PI from 2016-2020. **From here onward the instrument is solely referred to as the PI**. The exception is with software where program names are prefixed by 'PIA_'. This is a remnant from the PIA era. The PI is a prototype instrument used for research and development. Over the 2016-2020 period there have been several software changes resulting from the collaborative relationship between the instrument engineers and designers and the ecological end users. A survey report was written by the attending

scientist on conclusion of the survey detailing desired changes to the software. During the survey, small, ad hoc, revisions were made through a GitHub repository. The finished product, the soon to be commercially available, PI-10 was sea tested in June 2022. The report of the sea test, inclusive of hardware and software improvements are detailed at the end of the Chapter (Section 2.8). The following method is for the most recent version of the PI software.

2.2.1 DEVICE OWNERSHIP

The PI, PIA and PI-10 are owned by Plankton Analytics in Plymouth, United Kingdom. For research and development the device is loaned free of charge to Cefas.

2.2.2 DATA OWNERSHIP

All data collected using the PI are owned by Cefas. These data are freely shared with Plankton Analytics for instrument development and machine learning research. Some output data is available on the Cefas Data Hub: 10.14466/CefasDataHub.101, Table 2.1 provides an overview of the data collected by the PI across various surveys.

2.3 STANDARD OPERATING PROCEDURE

The following sections comprise the Standard Operating Procedure (SOP) for the PI. This SOP outlines the installation on the vessel and describes each stage of the PI method from image capture to classification.

2.3.1 CONNECTION TO THE RV

All versions of the PI followed the same installation procedure aboard the RV Cefas Endeavour. The PI is installed in the CTD Garage (an area used for deploying and housing wet-gear such as gliders, ring nets or rosettes) of the RV where it is connected to one of two continuous flow pumps. These pump water continuously from 4 m below sea level throughout the survey. The inlet is located off the ships starboard side

towards the bow of the vessel. The water passes enters the PI through the right side of Box 1 and flows through a flow cell (Figure 2.1**D**, 2.2**D**). The water is piped down into Box 2 where it passes through a flowmeter (Figure 2.1). Finally the water is piped vertically back through Box 1 where it is connected to an overboard drain (Figure 2.1). This generates >1 m hydrostatic head which is required by the flowmeter for proper functioning. The PI samples at 22 L/ min which equates to a through flow cell speed of 1.4 $m s^{-1}$.

The PIs external unit (Box 1 and 2, Figure 2.1 & 2.2**A**) requires two mains AC 240 volts power supplies for the controlling unit in Box 1 and the flowmeter in Box 2 (Figure 2.1**A** & **B**). These are connected to the ships AC circuit. The PIs camera, housed in Box 1, is connected to the controlling Linux computer via two CameraLink cables (Figure 2.1**C**) which pass through a cable access hole in the bulkhead to the right of the PI. The flowmeter house in Box 2 is connected to the controlling unit (Box 1) via a USB data cable (Figure 2.1**E**)

The controlling Linux computer (Figure 2.2**G**) is housed in the 'CTD Annex' adjacent to the CTD Garage. The computer requires several mains AC 240 volts power supplies. The CameraLink cables are connected to the controlling Linux computer. A serial line from the ships network is connected to the controlling PC which provides GPS and date-time.

2.3.2 CAMERA AND IMAGE ACQUISITION

The primary task is image capture. Water passes through a purpose built flow cell (Figure 2.2**D** & Figure 2.3). The flow cell has two windows (Figure 2.3); the bottom window allows for illumination from a high-power light emitted diode (LED, Figure 2.2**E**); whilst the top window allows for image capture. The flowcell has a width of 20.48 mm and internal depth of 12.8 mm.

The flow cell is positioned perpendicular to a line scan camera (Figure 2.2). The camera lens (Figure 2.2**C**) is used to focus the camera on the middle depth of the flowcell. The PI uses a colour 12-bit Basler 2048-70kc line scan camera with a scanning rate of 70,000 lines per second. The lines consist of 2048 × 1 rows of Bayer

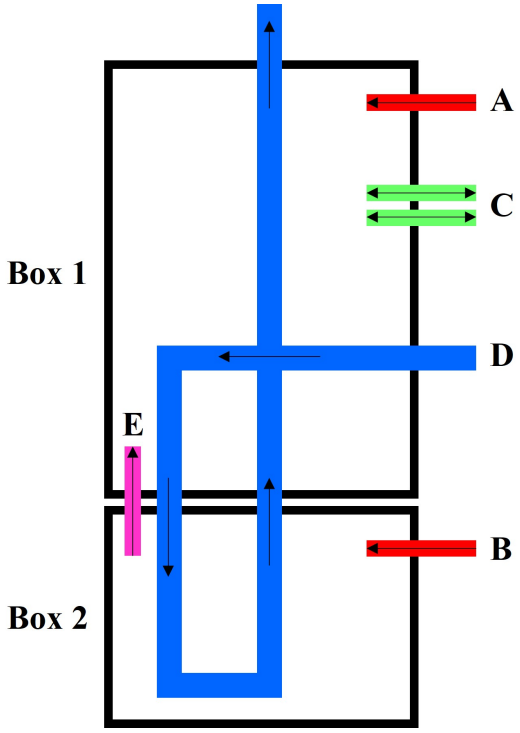

Figure 2.1: PI Schematic 1 - Cable and plumbing required for installation on a research vessel. **Box 1** houses the camera, flow cell and control unit (for more detail see Figure 2.2). **Box 2** houses flowmeter and drain valve. **A** and **B**: Mains 240 V AC power supplies. **C**: PCI Express cables connecting to linux PC. **D**: Water piping. **E**: Data cable from flowmeter to control unit

pixels. The camera was focused through the lens (Figure 2.2**C**) such that 1 Bayer pixel captures 20 μm^2 of seawater. In order to correctly capture square pixels, the flow rate is matched the to line scan speed. This is monitored by a Bell electro-magnetic flow meter (Box 2, Figure 2.1) and can be calculated using equation 2.1.

$$Flowrate = CW * CD * L * Px * 60 \tag{2.1}$$

where

FR = Flow rate (L / m^2)

CW = Flow cell width (m)

CD = Flow cell depth (m)

L = Line scanning rate

Px = Pixel size (m)

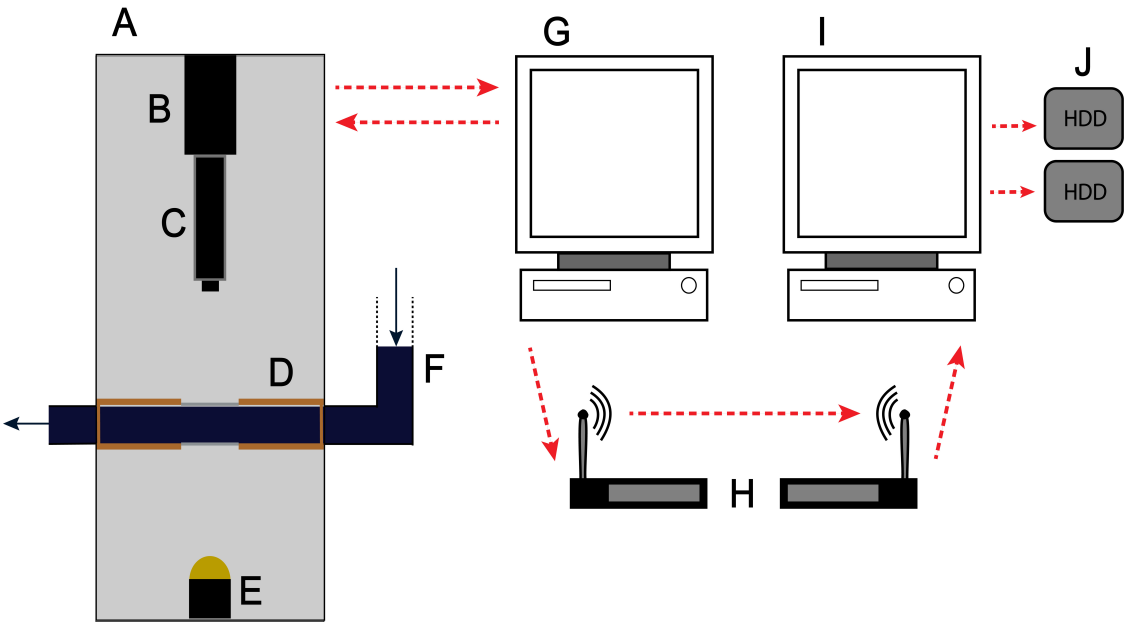

Figure 2.2: PI Schematic 2: Detailed method. Image capture stage (**A**), consisting of a flow cell (**D**), camera lens (**C**), camera (**D**), LED (**E**) with water flowing from pumped supply (**F**). Image archiving and control of the image capture stage is undertaken on desktop (**G**). Images are broadcast to a listening laptop or computer (**I**) using a local area network (**H**). Images are stored on (multiple) external hard drives (**J**). Red arrows indicate data flow. Blue arrows indicate water flow. Note, this figure is published in Pitois et al. (2021).

2.3.3 RAW IMAGE PROCESSING

Bayer lines are received on the controlling PC (Figure 2.2**G**) from the camera unit and using bespoke software, the lines are stacked against the previous line to produce a continuous image of potentially unlimited length (Figure 2.4**A**). This removes the potential for overlapping frames, or gaps between frames. The joined scan lines are separated into blocks of 128 lines. The contents of the blocks are analysed for the presence of shadows which represent particles in the water (Figure 2.4B). The presence of these shadows is determined by contrasting pixel illumination against a background image. The background image is taken at the start of every ten minutes. If a shadow is detected, the region of interest (ROI) is extracted from multiple blocks using a stitching algorithm and saved as 16-bit tagged image file (TIFF / .tif, Figure 2.4**C**).

The .tif is time, date and GPS stamped. These data are stored in the .tif metadata and received at 1 second intervals from a serial line connected to the vessel's GPS

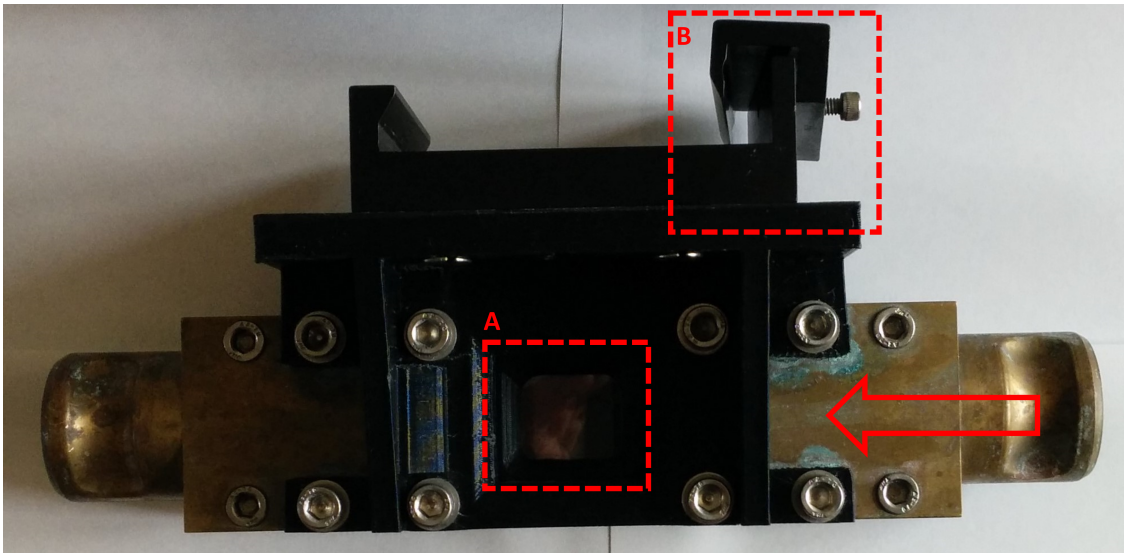

Figure 2.3: The brass flow cell used in the PI. The water enters from the right side (red arrow). **A**: The glass windows (located centrally) allows for illumination and image capture. **B**: A quick release clamp allows for the flowcell to be removed for routine cleaning.

system. The .tif filename includes system name, the time and the image sequence number.

2.3.4 IMAGE DISTRIBUTION

A local area network (LAN) was used to connect 'listening' computers (Figure 2.2I) to the controlling Linux computer (Figure 2.2**G**). This was established using a domestic router (Figure 2.2**H**). The User Datagram Protocol (UDP) is used to broadcast images over the LAN via Ethernet from the Linux computer. In theory, these may be directed to a designated Ethernet port anywhere on the ship's network so the device can be monitored anywhere on the vessel. Although this option was not used. Any machine (running any operating system), and any number of machines can listen and save the broadcast .tif files when connected to the same LAN. The benefit of using this system is that more than one 'listening' device can collect PI images and secondly, PI functionality can be checked anywhere aboard.

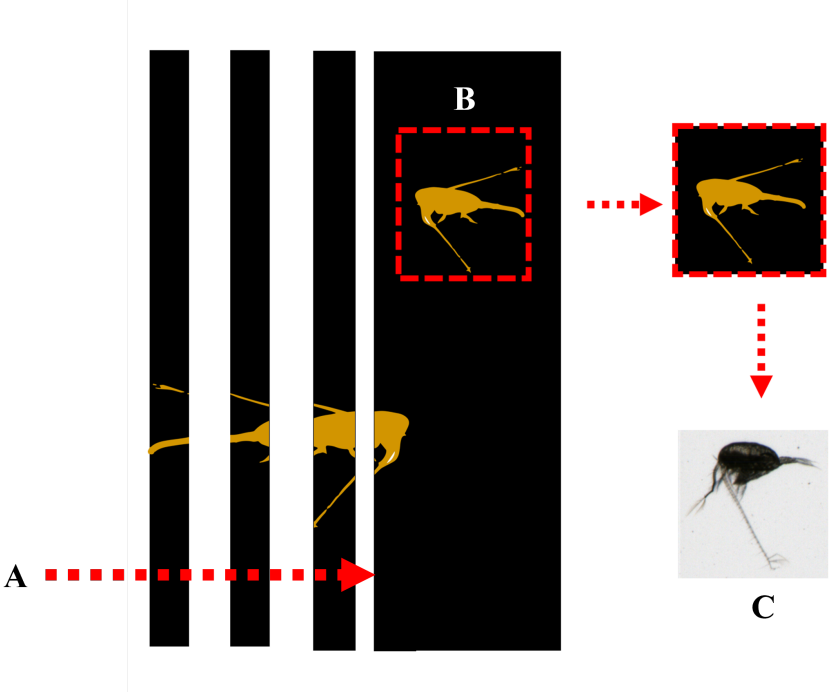

Figure 2.4: Bayer lines (**A**) measuring 2048 ∗ 1 rows of Bayer pixels are received from the camera unit. These are stacked to form scan line blocks of 128 lines. Regions of interest (**B**) are identified. Finally images are converted to 8-bit RGB images (**C**). Images are for example purposes only.

2.3.5 IMAGE ARCHIVING

Images are received on listening computers (Figure 2.2I) from the LAN using bespoke software (PI_Backup.exe). These images are saved to removable storage devices (Figure 2.2J). Images are binned first into date directories and then into 10 minute directories for ease of data management. These are automatically created at the start of each day and 10 minute interval. Post survey the images are backed up on large external hard drives. PI Archive 1 is kept and maintained by James Scott. PI Archive 2 is kept and maintained by Plankton Analytics.

Prior to 2018, a consistent archiving protocol had not been implemented. This resulted in subsequent compromises in both current analysis and training set improvements. Therefore, the principal task on joining the project was to ensure a stringent archiving system was implemented. The predominant hurdle was the size of the PIA raw data; survey data can total 2 Tb of data. Cost of a shared drive space (e.g. at Cefas or at a University) or cloud storage with sufficient upload/download

speeds was deemed to be uneconomical and untimely for raw data backup. Instead, only analysed data (commonly small .csv or .R files) and device metadata are stored on a communal cloud drive for remote access.

2.3.6 TYPICAL CRUISE SAMPLING

Motivated by creation of a new time series in the Celtic Sea, sampling was planned to be spring and autumn every year starting autumn 2016. (Table 2.1). So data could be aligned with other data sources in the area, mainly the Continuous Plankton Recorder Survey (Johns, 2006; Atkinson et al., 2018) and Plymouth Marine Laboratory L4 Station (Eloire et al., 2010), data were sampled in a manner similar to ring net samples. An hour was selected to align with other sampling (CTDs, fishing transects, etc.), was randomly subsampled and images classified to the highest discernible taxonomic resolution by experts. Various complications arising from instrument development resulted on three occasions where the instrument was not sea faring. The Covid-19 pandemic also resulted in cancellation of two surveys. Finally, in 2022 a shorter survey was selected to test the PI-10 (Section 2.8)

Table 2.1: Overview of PI point sampled stations from first survey to date. Data hub refers to the publicly viewable data on the Cefas website.

	Spring		**Autumn**	
Year	Stations	Data hub	Stations	Data hub
2016	pre-pi		40	uploaded
2017	53	uploaded	15	pending upload
2018	Hardware fault		45	uploaded
2019	20	uploaded	29	uploaded
2020	Covid-19		35	pending upload
2021	Covid-19		Instrument not sea faring	
2022	Instrument not sea faring		*summer, testing PI-10	

2.4 IMAGE ANALYSIS

Image analysis follows collection of raw images and includes the extraction, subsampling (if required) and classification of the raw images. Image analysis for all chapters used bespoke software developed by Plankton Analytics for the PI. Chapter 5 also used a machine learning classifier produced in collaboration with the Turing Centre, London UK (see Chapter 2.4.2). In the following section, the bespoke PI programs are detailed followed by a description of how the Turing Classifier was used on PI data. The construction of the classifier, and other in-house PI classifiers is detailed later in section 2.6.

2.4.1 PI BESPOKE PROGRAMS

There are three main programs and these must be run in series, these are, in order: Raw2RGB.exe, PIA_Analyse and PIA_LabelSort. Each program may be used independently and can be run as an executable (.exe) from the shell command line. These commands can easily be batched (using .bat files) to run the programs over several images or target directories. Raw2RGB.exe and PIA_Analyse was used for Chapters 4 and 5. More detail on the three main stages is given below.

STAGE 1: RAW2RGB

The raw saved images are 12-bit Bayer pattern images. These appear as black squares on most common image viewers (e.g. Windows explorer) and are consequently unusable by humans. This program converts these to 8-bit RGB colour images. During conversion each 20 μm^2 Bayer pixel is interpolated to four 10 μm^2 RGB pixels. This program has arguments that allow it to be run over a whole survey, day or a single ten minute directory and also allows the user to skip n files (e.g., for temporal subsampling in Chapter 4).

STAGE 2. PIA_ANALYSE.

This is the computer vision part of the program. Image features are extracted and saved as a .csv which are used in Stage 3. Important size parameters, such as MajAx

and MinAx are extracted at this stage (Figure 2.6). Similarly, to stage 1, the program can be run on both ten minute bins or day samples as well as skip *n* images.

STAGE 3. PIA_LABELSORT

A random forest machine learning classifier uses the features .csv to attempt to classify the images into predefined taxonomic classes. The classifier is discussed in detail in Section 2.6. The classified images must be checked by an expert.

PIA_SUBSAMPLE

The programs can also be run through software with a graphical user interface (GUI) called PIA_Subsample (Figure. 2.5). This automatically runs the programs based on user-defined inputs on the GUI. PIA_Subsample was used for Chapter 3.

A broad user guide to PIA_Subsample is described below to provide the context for how the software was typically used which is shown in italics.

- **System Directory** - This points the software to the parent cruise directory.

  These are typically kept on removable storage or on one of the PI Archives with the directory names matching Cefas cruise codes.

- **Begin and End time** - Allows the user to specify the time range for the target sample.

  These are commonly matched to other ship events, such as a physical plankton sample (for example, see Chapter 3) or fishing trawl. Usually an hour is sampled to approximately match the amount of water sampled by a vertical ring net trawl with a 70 cm opening.

- **Training set** - This points the software to the relevant extensible markup language (.xml) file. The .xml contains algorithm parameters needed for classification. The parent directory of the .xml must also contain file label.txt. This file contains a list of the predefined taxonomic categories into which the classifier will sort the images.

  There have been several iterations of the training sets, described in 2.6. These have

been kept in communal cloud drives and are a result of collaboration between various ecologists and the instrument and software engineers.

- **No. of specimens** - The number of images to be extracted by PIA_Subsample.

  In order to align with ring net samples, the target number of zooplankton to be classified was 200 (Wiebe et al., 2017). Samples would frequently be done in sections. Initially 1000 images were sorted to provide an idea of the number of plankton per 1000 images. The number of specimens for the next batch could then be accordingly adjusted to reach target zooplankton.

- **Ignore dirs. with more images than *n*** - Directories over a certain number of images may be skipped.

  During a survey the instrument may be infrequently swamped by high particular matter, bubbles of phytoplankton, see Section 2.5. This can be used to exclude those directories with an unrealistic number of images to save processing time. This number is arrived at anecdotally through trial and error and varies between surveys.

- **Station name** and **Prime ID** - These dictate the name of the output directory where all images and results are output to.

  These have been typically matched to the plankton stations on the survey and follow the Cefas station name codes for ease of comparison with other data.

- **Flow rate** The flow rate into the instrument.

  Always set at 22 L/min

Following completion of the listed fields above, there are several functions that may be run and each function can be run multiple times:

- **Preview Sample** - This estimates the total number of images within the time range as well as showing the length of time to be sampled.

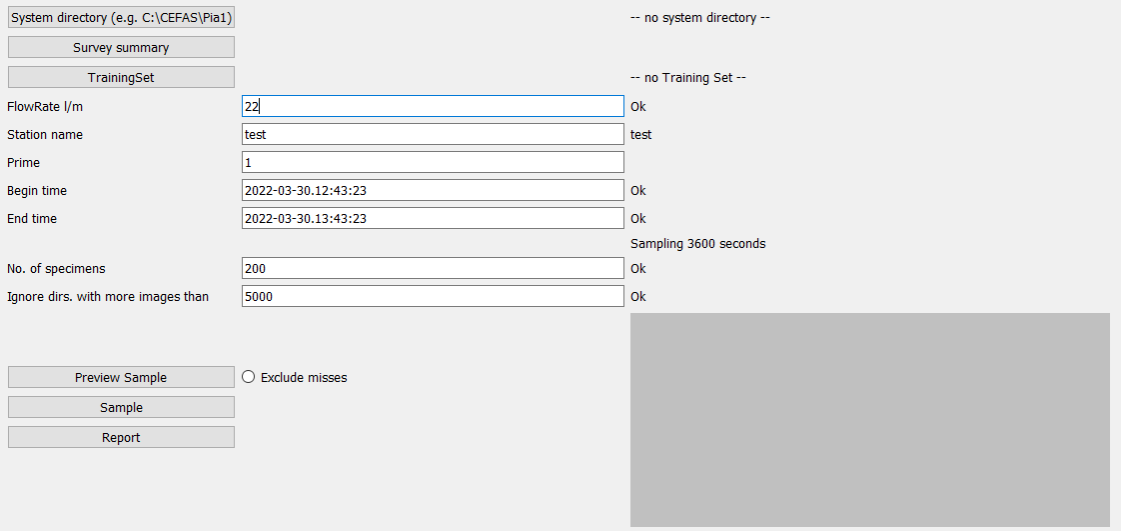

Figure 2.5: Screenshot of PIA_Subsample.

The number of images directly correlates to amount of time taken to run the Sample function, thus it can be used as an estimate to run time.

- **Sample** - This runs through all stages detailed below. The specified number of specimens (/images) are first extracted randomly. These are converted to RGB images and then a classifier attempts to classify each image to taxonomic categories.

  This is often run multiple times in batches to ensure the minimum number of images are sampled to reach the target 200 specimens.

- **Report** - This is the last stage of image analysis using the PI_Subsample software. Following manual validation of the images by an expert after the **Sample** function has been performed, **Report** generates a file that summarises the station. This contains the following data: PrimeID (as defined above), Station Name (as defined above), Pumped Volume (calculated from flow rate and time), Filename (each row contains the image filename in the sub sample), Lat, Long, Date-time, Class (the taxonomic category decided by the classifier, p-value (the statistical likelihood of the Class being correct, note these are 1 if the image had to be manually corrected by an expert), file_index (reference), Major Axis (in pixels, longest distance between two pixels), Minor Axis (in pixels,

smallest distance between two pixels) (10 μum), Area (in pixels, total number of pixels (10 μum) in the particle).

This is the resultant usable ecological data. These are often read and merged by the script which produces the final data.

2.4.2 TURING CLASSIFIER

The Turing Classifier, developed as part of a collaborative data study group between Cefas and the Turing Institute was used for Chapter 5. The construction of the training set for this classifier is detailed in Section 2.6.2. The Turing classifier only works on the converted RGB images. Raw2RGB was batch run over all target directories to convert the raw images to images that can be handled by the classifier. On completion, a python script which classifies each image was looped over the converted images. For each directory this reports a spreadsheet which contains the determined class and the probability that the classifier has correctly sorted the image, this can also been seen as the 'confidence' of the classifier. The performance of the classifier is detailed in Section 2.6.2.

2.4.3 ABUNDANCE, SIZE AND BIOMASS

Abundance follows standard zooplankton subsampling procedure. Three size measurements are available from the PI images described above: MajAx, MinAx and Area. These are are extracted by the PIA_Analyse software. Biomass is derived from size. The following size and biomass methods are a more detailed version than those described in Pitois et al. (2021) and Scott et al. (2021), see *Published Works*.

ABUNDANCE (INDV. M^{-3})

Due to the vast number of plankton collected, physical or imaged, they are nearly always subsampled (as discussed in the introduction, Chapter 1.2.1). For Chapter 3 data were randomly subsampled and for Chapter 4 data were temporally subsampled. For Chapter 5 there was no subsampling and all images were used. Random

subsampling is similar to the methods used on a physical sample where a folsom (or another device) is used to randomly split the sample (Postel et al., 2000) as many times are necessary to get the target number of individuals. Although this assumes the individual abundances of each taxa present follow a Poisson distribution (Postel et al., 2000). As a discussed in 1.2.1 rare species are underrepresented. The abundance or density (number of plankton per meter cubed) is resolved using the following equation 2.2.

$$D = \frac{N * S * 1000}{F * T} \tag{2.2}$$

D = Density of organisms per cubic meter (Indv. m^{-3})

N = Number of organisms (n)

S = Split factor (also known as raising factor)

F = Flow rate (L min^{-1})

T = Time (minutes)

In Chapter 3 the split factor varied and was dependent upon the ratio of plankton to detritus. This was calculated by using the total number of images within the time range divided by the total number of images classified, thus the split factor varied. In Chapter 4 we temporally subsampled 1 in every 10 images resulting in a consistent split factor of ten. In Chapter 5 no subsampling was used. Therefore, the split factor was simply 1 in the equation.

SIZE (μm)

Size parameters from the image, extracted by PIA_Analyse can be used to infer organism sizes. This was only performed on copepod images. PIA_Analyse returns three size parameters: MajAx (Figure. 2.6), MinAx and Area. Only MajAx was used as it approximates the full length of the copepod. All size parametes are given in pixels. Each pixel equates to 10 μm. To obtain the length of the copepod in μm the MajAx was multiplied by 10 (Figure. 2.6).

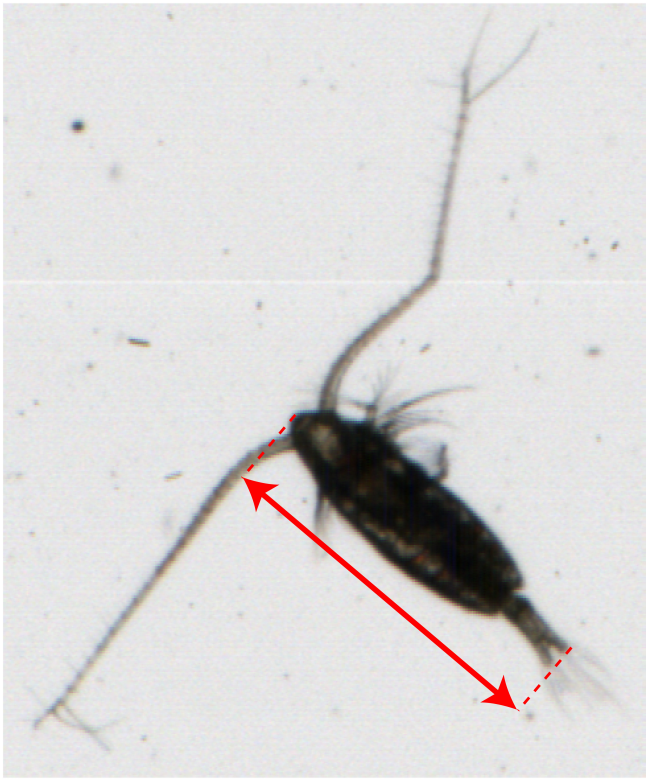

Figure 2.6:
Copepod length measurement from PI, extracted as one measurement between the two end points of the red arrow. This figure is published in Pitois et al. (2021), see *Published Works*.

BIOMASS

To date, biomass has only been calculated for copepods due to their symmetrical and ellipsoidal shape. The following methods are used in Chapters 4, 5 and Pitois et al. (2021).

To calculate biomass the longest size parameter from PI, the MajAx was converted to copepod length (mm). Individual taxa biomass was calculated using a regression to determine wet-weight (WW, µg) as a function of length. Equation 2.3 was derived from regression and applied to each copepod (Pitois et al., 2021).

$$WW = 0.299 * TL^{2}.8948 \tag{2.3}$$

WW = Wet-weight (µg)

TL = Prosome length (mm)

For the length-mass regression, we used the method described in Pitois and Fox (2006) where information on the copepod promsome and urosome sizes were gathered from the literature to convert length to wet weights (Pitois et al., 2016; Chojnacki, 1983) using equation 2.4.

$$Vcopepod = \frac{\pi w_p^2 l_p}{6} + \frac{\pi w_u^2 l_u}{4} \tag{2.4}$$

V = Body volume

w_p = Prosome with

w_u = Urosome width

l_p = Prosome length

l_u = Urosome length

To derive biomass values (mg WW m^{-3}) using individual image sizes, the total observed copepod wet weight for each station was summed, and then scaled by the number of images analysed and the water volume sampled at each station. Using the average size of each copepod group, the mean wet weight of the group was multiplied by the abundance of that group in ind. m^{-3}.

2.5 LIMITATIONS

2.5.1 DISK WRITE SPEEDS

Currently the PI can collect images at a faster rate than disk write speeds. Although storage media with faster write speeds are commercially available, consumer units are used for accessibility, compatibility and ease of purchase. There is negligible gain in speed when using new solid state drives although more reliable. It is difficult to put a maximum value on the number of images the PI can record per hour as the number of images written per second varies on image size, number of images and background processes between the operating system and the disk. If the minimum

possible image size (10 µm) the PI can capture was saved the hard drives would flood instantly (also these images would purely consist of noise and small particulate detritus). Therefore, the current balance is between desired organism size range and disk write speeds. Across all Chapters and Publications the minimum size was 180 µm and the maximum size was 20 mm. We found this to be a good compromise to ensure the target images within the size range are mostly captured. Although, images are still lost when in areas of high particulate matter, e.g. bubbles or silt. For the duration of the PhD these data and the number of images missed are lost as the protocol for recording these has not yet been implemented.

EXAMPLE SURVEY

During the Autumn 2020 survey we collected 71 million images across 38 days. Figure 2.7 provides an overview of the number of images per ten minute bin. The maximum number of images per bin was 111,326 averaging to 185 images per second. The mean number of images per bin was 25,010 with the third quartile at 37,727. The vast majority of the time the image range for this survey (180 µm - 20 mm) being captured (Figure 2.7**A**. There are several bins > 80,000 images. It is likely for these bins images were lost as a result of the disk flooding.

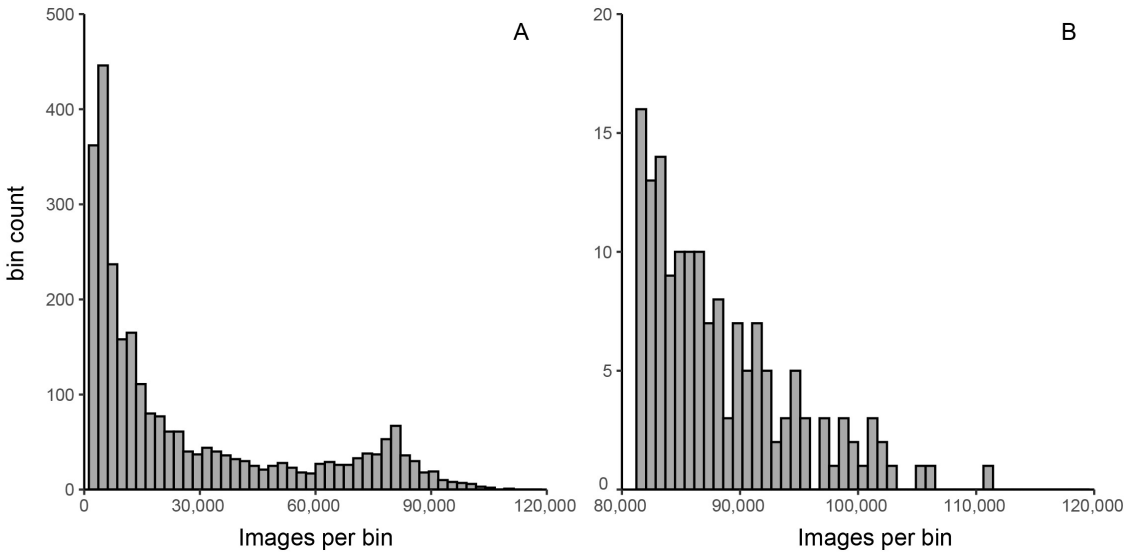

Figure 2.7: The number of images per bin for the Peltic autumn 2020 survey captured over 38 days in The Celtic Sea. All bins across full range (A) and bins larger than 80,000 (B). Number of bins for both panels = 50

HIGH PARTICULATE AREAS

The flooding results from areas where the PI sees a phenomenal number of images within a short time period, these tend to be lots of smaller images. There are several main causes: Bubbles, Phytoplankton or high detrital matter. Aside from system flooding these can also making subsampling a very laborious processes as the number of images required to obtain the target number of individual (see Section 2.3.6) is significantly increased.

Bubbles at a capturable size typically occur due to rough seas where air is mixed into the surface layers of the ocean with ease through wave breaking. This has been a particular problem in the April survey following rough winter seas. Anecdotally, one can find themselves finding only 1 target plankton in 1000 bubbles. While machine learning solutions may eliminate this task and easily sort bubbles from plankton, the disk flooding is a trickier issue to resolve.

This principle is similar to both Phytoplankton and areas of high detritus. Areas of high density phytoplankton are difficult to predict or mitigate against although these are usually acute incidents as the vessel steams through the area relatively quickly. Areas of high particulate matter tends to be a little more predictable and associated with fluvial sources due to turbulence from saline and freshwater mixing. In 2020 the instrument was stopped when entering the Bristol Channel to stop the removal storage filling with images of sand or mud.

2.5.2 VARIABLE FLOW RATE

The flow rate can vary due to unforeseen circumstance. On the Autumn 2019 survey, bio fouling impinged the pump resulting in a reduced flow rate. The flow rate quickly fell from the target 22 l/min to around 15 l/min. The flow rate of 22 l/min is essential for capturing square pixels by matching the line scanning speed (Section 2.3.2). Although, this can be adjusted to mitigate for reduced flow rates by reducing the line scanning rate. This reduced flow rate must be manually recorded as it is required in resolving plankton samples when splitting (Section 2.4.3)

2.6 MACHINE LEARNING TRAINING SETS

To use machine learning for classification purposes, it must be trained on known, example images. Machine learning classifiers and their required training sets have been developed alongside the instrument since 2016. Their application to zooplankton and the associated advantages are introduced and discussed in Chapter 1.2.3. Prior to Spring 2022 the only classifier used on PI data was in the in-house Plankton Analytics random forest classifier. The training sets were created by myself with help from taxonomists at Cefas. These were audited by Plankton Analytics before being used to train the classifiers. All images classified by this classifier underwent manual validation by an expert. In 2022 a collaborative data study group with the Turing set produced a copepod only classifier, this was used for Chapter 5. Images classified by the Turing classifier have not been manually validated as it was used to sample 10 million images.

2.6.1 PLANKTON ANALYTICS TRAINING SETS

The images that comprise the Plankton Analytics training sets result from manual classification of survey images. Thus, there is a very small pool of images to choose from, especially for rarer species. This is the primary limitation for all classifiers used (see Chapter 1.2.3). There have been several iterations of the training set with both minor and major changes, only the major changes are detailed here. Minor changes might include small changes to the algorithm weighting performed by Plankton Analytics. For example they audited the images to ensure best performance (Bad Images column, Table 2.2). None of the training sets performed reliably on real data with an accuracy usually lower than 30 %. The accuracy seemed to vary class by class but all images required manual sorting. The Plankton Analytics machine learning classifier was only used for Chapter 3 as part of the PIA_Subsample software. The major revisions are detailed below. Development of the classifier slowed in 2021 as attention was focused on developing the PI-10. Therefore the most recent classifier is dated 26/01/2021. The categories and respective counts are detailed in Table 2.2.

POST SURVEY BOLSTERING

Following completion of a survey, and validation of images by experts, the classified images are added into the training set. As a result, the training set grows and increases the classifier accuracy. Although the gains are marginal. This method ensures there are as many images of rarer species as possible.

ADDITION OF DETRITUS AND BUBBLES

Images of plankton make up a very small proportion of the total images captured by the PI. This ratio depends on both the local abundance of plankton and non target particles (Table 2.2). The not target particulates are highly diverse in their appearance. Examples include dead flora and fauna, particulate aggregates or phyotoplankton. These easily confused the classifier as they may look similar to target zooplankton. Some not target particles are more consistent. For example bubbles are nearly always almost-perfectly spherical and very opaque meaning they are easily identified and disregarded. In the hope of gaining a improved performance detritus and other classes were added to the training sets and populated.

2.6.2 TURING CLASSIFIER

A machine learning classifier was produced as the result of a collaborative Data Study Group (DSG) between Cefas and Turing Institute, London. The resultant classifier was a CNN within the ResNet50 architecture. The classifier was used for the Chapter 5. The DSG platform invites participates to solve various applied data science problems. Classifying PI images to increasing taxonomic resolution was a challenged proposed by Cefas. 58,791 images were provided by Cefas, compiled by myself and Plankton Analytics and audited by Cefas data science staff (duplicates removed etc.).

The DSG format requires objectives of increasing difficulty. The difficulty associated with classifying plankton images to an increasing taxonomic resolution provided several challenges. The data set was split into three subsets. The crudest subset, in terms of taxonomic resolution consisted only of 17,069 zooplankton images and 40,000 images of detritus (label 1, Table 2.3). The second set had a divsion of the

Table 2.2: Overview of Categories and images per category to the most recent Plankton Analytics training set (26/01/2021). Target taxa are those taxa of ecological interest. 'Bad Images' are those than have been removed after audit by Plankton Analytics to try and improve performance. Example may include blurry or unfavourable orientations. 'Used to Train' shows the actual images used to train the classifier.

Target Taxa?	Category	Image count	Bad Images?	Used to train?
yes	Amphipoda	2	yes	no
yes	Amphipoda	18	no	yes
yes	Amphipoda_Hyperiidae	6	no	yes
yes	Annelida_Polychaeta	7	yes	no
yes	Annelida_Polychaeta	141	no	yes
yes	Appendicularia	5	yes	no
yes	Appendicularia	80	no	yes
yes	Bivalvia-larvae	6	yes	no
yes	Bivalvia-Larvae	43	no	yes
yes	Byrozoa-Larvae	15	yes	no
yes	Byrozoa-Larvae	97	no	yes
yes	Chaetognatha	14	yes	no
yes	Chaetognatha	232	no	yes
yes	Cirripedia_Barnacle-Cyprid	11	no	yes
yes	Cirripedia_Barnacle-Nauplii	5	yes	no
yes	Cirripedia_Barnacle-Nauplii	77	no	yes
yes	Cladocera	37	no	yes
yes	Cladocera_Evadne-spp	32	no	yes
yes	Cnidaria	17	yes	no
yes	Cnidaria	63	no	yes
yes	Cnidaria_Siphonophorae	2	no	yes
yes	Cnidaria_Siphonophorae_Physonectae	22	no	yes
yes	Copepod	875	no	no
yes	Copepod	2076	no	yes

yes	Copepod_Calanoida	542	yes	no
yes	Copepod_Calanoida	2053	no	yes
yes	Copepod_Calanoida_Acartia-spp	68	yes	no
yes	Copepod_Calanoida_Acartia-spp	483	no	yes
yes	Copepod_Calanoida_Calanus-spp	356	no	yes
yes	Copepod_Calanoida_Candacia-spp	41	no	yes
yes	Copepod_Calanoida_Centropages-spp	858	no	yes
yes	Copepod_Calanoida_Centropages-spp-	58	yes	no
yes	Copepod_Calanoida_Para-	2075	no	yes
	Pseudocalanus-spp			
yes	Copepod_Calanoida_Para-	1133	no	no
	Pseudocalanus-spp-Notused			
yes	Copepod_Calanoida_Temora-spp	12	yes	no
yes	Copepod_Calanoida_Temora-spp	178	no	yes
yes	Copepod_Cyclopoida	9	yes	no
yes	Copepod_Cyclopoida	100	no	yes
yes	Copepod_Cyclopoida_Corycaeus spp	1217	no	yes
yes	Copepod_Cyclopoida_Oithona-spp	139	yes	no
yes	Copepod_Cyclopoida_Oithona-spp	548	no	yes
yes	Copepod_Cyclopoida_Oncaea-spp	783	no	yes
yes	Copepod_Harpacticoida	70	yes	no
yes	Copepod_Harpacticoida	697	no	yes
yes	Copepod_Nauplii	302	no	no
yes	Copepod_Nauplii	1663	no	yes
yes	Decapoda-larvae	10	no	yes
yes	Decapoda-larvae_Brachyura	40	no	yes
yes	Echinodermata	10	no	yes
yes	Echniodermata-larvae	222	yes	no
yes	Echniodermata-larvae	910	no	yes
yes	Euphausiid	41	no	yes
yes	Euphausiid_Nauplii	141	no	yes

yes	Fish-eggs	7	yes	no
yes	Fish-eggs	52	no	yes
yes	Fish-larvae	10	no	yes
yes	Gastropoda-larva	1	yes	no
yes	Gastropoda-larva	40	no	yes
yes	Marine Mite	2	no	yes
yes	Mysidacea	3	yes	no
yes	Mysideacea	34	no	yes
yes	Ostracoda	47	no	yes
yes	Para-Pseudocalanus	126	yes	no
yes	Para-Pseudocalanus	4176	no	no
yes	Phoronida-larva	10	no	yes
yes	Radiolaria	398	yes	no
yes	Radiolaria	2854	no	yes
yes	Radiolaria-Notused	1018	no	no
yes	Tintinnida	32	no	yes
yes	Tunicata_Doliolida	45	yes	no
yes	Tunicata_Doliolida	332	no	yes
no	Detritus	134	yes	no
no	Detritus-needs-resorting	37	no	no
no	Bubbles	43	no	yes
no	Detritus	1407	no	yes
no	Diatoms	6	no	yes
no	Phyto_Ceratium-spp	750	no	yes
no	Phyto_Rhizosolenia-spp	185	no	yes

zooplankton into 10,346 copepods and 6,723 noncopepod as well as detritus (label 2, Table 2.3). The final set consisted of zooplankton classes with detritus (label 3, Table 2.3). Each label posed increasing difficulty. Participants were challenged to classify the images to increasing taxonomic resolutions. These are similar to the training set created for the Plankton Analytics training sets 2.2.

Only the copepod vs noncopepod vs detritus (label 2) classifier was used on real data for Chapter 5. The final classifier has a reported accuracy of 97% when compared with expert labelled images. The classifier was run from a python script and can only be run on images post conversion using Raw2RGB. A script was written by Plankton Analytics and modified by myself to loop through all target images.

Table 2.3: Overview of Categories and images per category for the Turing Centre Classifier. These images were curated by myself with help from Plankton Analytics as the training images. This table is reproduced from The Turing Centre (2021).

Label 1		
	detritus	40000
	zooplankton	17069
Label 2		
	detritus	40000
	copepod	10346
	noncopepod	6723
Label 3		
	detritus	40000
	copepod para-pseudocalanus-spp	1988
	copepod unknown	1853
	radiolaria	1810
	copepod calanoida	1665
	copepod nauplii	1380
	copepod cyclopoida corycaeus-spp	1117
	echniodermata-larvae	799
	copepod calanoida centropages-spp	773
	copepod cyclopoida oncaea-spp	710
	copepod harpacticoida	643
	copepod cyclopoida oithona-spp	492
	nt-phyto ceratium-spp	459
	copepod calanoida acartia-spp	451
	nt-bubbles	354
	copepod calanoida calanus-spp	345
	nt phyto chains	298
	tunicata doliolida	291
	nt-phyto rhizosolenia-spp	184
	copepod calanoida temora-spp	168

chaetognatha	158
annelida polychaeta	141
euphausiid nauplii	139
copepod cyclopoida	100
byrozoa-larvae	96
appendicularia	79
cirripedia barnacle-nauplii	77
cnidaria	63
fish-eggs	50
ostracoda	47
bivalvia-larvae	43
euphausiid	41
copepod calanoida candacia-spp	41
gastropoda-larva	40
decapoda-larvae brachyura	40
cladocera	37
mysideacea	33
tintinnida	32
cladocera evadne-spp	32

PERFORMANCE

As part of the DSG the accuracy of the classifier was assessed. These are available in the full report which is pending publication on the Turing website: Turing Data Study Group Reports. Only the level 2 classifier with detritus (Table 2.3) was used and only these results are show in in Table 2.4.

BIG DATA PIPELINES

Most of the published literature resolves plankton images to finer taxonomic resolution (Dai et al., 2016; Orenstein et al., 2020; Kerr et al., 2020; Schröder et al., 2019). ML is only used in Chapter 5 where images are only classified to copepod / non copepod level. This mainly due to not having the resources to construct a more detailed training set (see Chapter 6). Although copepod / non-copepod still provides usual taxonomic information (Scott et al., 2023) but it is difficult to find comparable studies in the literature where a lower taxonomic resolution is used. We report a high accuracy of 95% (Table 2.4) but this is most likely due to the lack of taxonomic resolution. CNNs, specifically the ResNet Architecture are being used to get classification acurracies of > 70% to moderate taxonomic resolution (Irisson et al., 2022).

Handling the image data often demands a bespoke data pipeline. Using the PI as an example (Section 2.3): Images are captured on an external drive; the drive it put in another computer; images are archived and classified later in time. Other data flow pipeline specifics used in various plankton ML studies are not often well reported (including the material published on the PI) (Pitois et al., 2020; Scott et al., 2021), especially when it concerns instruments that collects many images. Although there are few instruments instruments that have the same potential capacity as the PI. The device has the capacity to collect over 1 billion objects per day, limited by a size range and a gigabit Ethernet cable. Thus, solutions in the literature on how to handle these sorts of data sets are not always applicable to data rate of the PI. Although the issues developing big data infrastructure is the next challenge for zooplankton imaging. Key recommendations from a review on ML's application to Zooplankton by Irisson et al. (2022) suggests: *"Effort should be directed toward the development of significant infrastructure to store and curate images, with the help of machine learning, at a scale that is beyond what plankton ecology laboratories are used to."* and perhaps more importantly *"To leverage such an infrastructure and the data therein, oceanographers should be trained in data science, not only to leverage machine learning methods but also simply to deal with the massive data sets involved."*

Table 2.4: The performance of the machine learning classifier used in Chapter 5 (ResNet CNN) against a RandomForest classifier. Accuracy measures the ratio of correctly predicted observations to the total observations. Precision indicates the fraction of relevant positive instances among all retrieved positive instances. Recall, also known as sensitivity, is the fraction of relevant positive instances that were retrieved by the classifier. Reproduced from The Turing Centre (2021).

Label 2: Copepod vs Non-Copepod vs Detritus

Metric	RandomForest	ResNet CNN
Accuary	84%	97%
Precision (average)	71%	95%
Recall (average)	70%	95%
F1 (average)	70%	95%

2.7 CONTRIBUTIONS TO DEVELOPMENT

The PhD has assisted in the development of the instrument through user testing and feedback with the instrument designers. These contributions are listed in Appendix C. Additionally various tools have be written to explore or manage the data. the PI Metadata explorer was an early tool to try to get a handle on the large data the PI collects, this is detailed in Appendix A. Another notable tool is a program for spatiotemporally aligning the data (Appendix B) which was used heavily in Chapters 4 & 5). These are just two examples in a range of tools written in R, Python and PowerShell to wrangle analyses and process the PI data.

2.8 PI-10

The PI-10 was sea tested in June 2022 as part of the Nephrops survey (Dobby et al., 2021). The survey was only to test the commercial version of the instrument and none of the images (and thus ecological) will be used. The instrument is essentially the same as the PI with a few key upgrades. Primarily the camera unit (Figure 2.8**D**) is now a Dalsa 4096 pixel line scan camera operating at 100,000 lines per sec in full rgb colour (an upgrade from 70,000 in the PI). This increased line scan rate means

an increased water flow of 34 L per min can be achieved. The camera also has twice the resolution resulting in clearer images. The water inlet (Figure 2.8**A**) and drain (Figure 2.8**C**) remains similar although a new, larger flow cell (Figure 2.8**B**) to suit the camera is used. These are all now housed in an air and water tight acrylic waterproof reduce dust on the flow cell (dust can result in artefact images as the dust is imaged). This unit (the blue box in Figure 2.8) is removable and when deployed attaches to a aluminium frame that is secured the vessel bulkhead.

The instrument was run with moderate success during the survey. The issue with the data rate (Chapter 2.5.1) is made worse by the higher resolution (and thus larger) images. Otherwise the instrument performed as expected and for the majority of the survey, was fully automated.

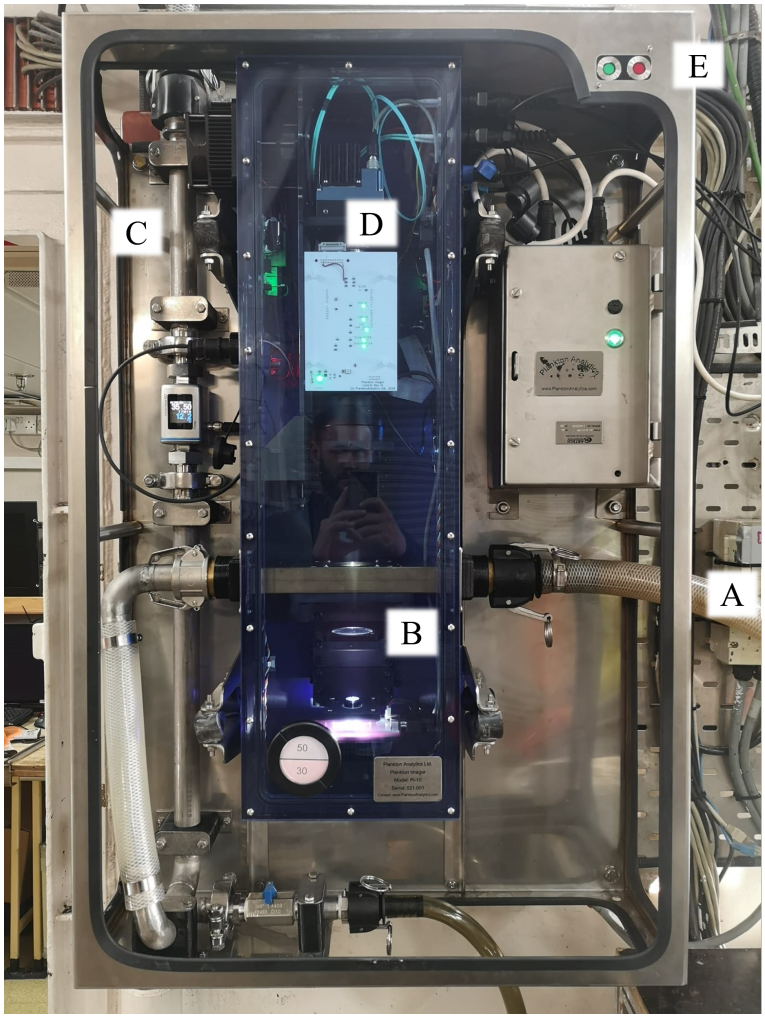

Figure 2.8: The PI-10 aboard the RV Cefas Endeavour. The PI-10 is now wall mounted. The external aluminium frame (silver metal surrounding instrument) is attached to the bulkhead. The major components are the water inlet (**A**), the flow cell (**B**), the drain (**C**), the camera unit and control board (**D**) and a new LED error reporting (**E**). These are now housed in a waterproof compartment (blue box, centre of figure).

2.9 BIBLIOGRAPHY

Atkinson, A., Polimene, L., Fileman, E. S., Widdicombe, C. E., McEvoy, A. J., Smyth, T. J., Djeghri, N., Sailley, S. F., and Cornwell, L. E. (2018). Comment. What drives plankton seasonality in a stratifying shelf sea? Some competing and complementary theories. *Limnology and Oceanography*, 63(6):2877–2884.

Chojnacki, J. (1983). Standardgewichte der Copepoden in der Pommerschen Bucht. Standard Weights of the Pomeranian Bay Copepods. *Internationale Revue der*

gesamten Hydrobiologie und Hydrographie, 68(3):435–441.

Dai, J., Wang, R., Zheng, H., Ji, G., and Qiao, X. (2016). ZooplanktoNet: Deep convolutional network for zooplankton classification. *OCEANS 2016 - Shanghai*.

Dobby, H., Doyle, J., Jónasson, J., Jonsson, P., Leocádio, A., Lordan, C., Weetman, A., and Wieland, K. (2021). ICES Survey Protocols–Manual for Nephrops underwater TV surveys, coordinated under ICES Working Group on Nephrops Surveys (WGNEPS).

Eloire, D., Somerfield, P. J., Conway, D. V., 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. *Journal of Plankton Research*, 32(5):657–679.

Irisson, J. O., Ayata, S. D., Lindsay, D. J., Karp-Boss, L., and Stemmann, L. (2022). Machine Learning for the Study of Plankton and Marine Snow from Images. *Annual Review of Marine Science*, 14(January):277–301.

Johns, D. (2006). The Plankton Ecology of the SEA 8 area. *Strategic Environmental Assessment Programme, UK Dept. of Trade and Industry*, pages 1–44.

Kerr, T., Clark, J. R., Fileman, E. S., Widdicombe, C. E., and Pugeault, N. (2020). Collaborative deep learning models to handle class imbalance in flowcam plankton imagery. *IEEE Access*, 8:170013–170032.

Orenstein, E. C., Kenitz, K. M., Roberts, P. L. D., Franks, P. J. S., Jaffe, J. S., and Barton, A. D. (2020). Semi-and fully supervised quantification techniques to improve population estimates from machine classifiers. *Limnology and Oceanography: Methods*, 18(12):739–753.

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. Technical report, Cefas, UK.

Pitois, S. G., Bouch, P., Creach, V., Van Der Kooij, J., Kikkawa, T., Kita, J., Ishimatsu, A., Pitois, S. G., Tilbury, J., Bouch, P., Close, H., Barnett, S., and Culverhouse, P. F. (2016). Comparison of zooplankton data collected by a continuous semi-automatic sampler (CALPS) and a traditional vertical ring net. *Journal of Plankton Research*, 38(4):931–943.

Pitois, S. G. and Fox, C. J. (2006). Long-term changes in zooplankton biomass concentration and mean size over the Northwest European shelf inferred from Continuous Plankton Recorder data. *ICES Journal of Marine Science*, 63(5):785–798.

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). *Ecological Indicators*, 123:107307.

Postel, L., Fock, H., and Hagen, W. (2000). Biomass and abundance. In *ICES Zooplankton Methodology Manual*, pages 83–192. Elsevier.

Schröder, S.-M., Kiko, R., Irisson, J.-O., and Koch, R. (2019). Low-Shot Learning of Plankton Categories: 40th German Conference, GCPR 2018, Stuttgart, Germany, October 9-12, 2018, Proceedings. (January):391–404.

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.

Scott, J., Pitois, S., Creach, V., Malin, G., Culverhouse, P., and Tilbury, J. (2023). Resolution changes relationships: Optimizing sampling design using small scale zooplankton data. *Progress in Oceanography*, 210(December 2022):102946.

The Turing Centre (2021). DSG Report 2021. Technical report.

Wiebe, P. H., Bucklin, A., and Benfield, M. (2017). Chapter 10. Sampling, preservation and counting of samples II: Zooplankton. *Marine Plankton: A practical guide to ecology, methodology, and taxonomy*, pages 1–3.

3

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

This chapter was published in the Journal of Plankton Research in 2021 (Scott et al., 2021), see Published Works. The following is a direct copy of the manuscript.

3.1 ABSTRACT

The Plankton Imager is an underway, semi-automated, high-speed imaging instrument which takes images of all passing particles and attempts to classify the zooplankton 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.

3.2 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 analysis 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 et al. (2019). More recent devices, for example The Scripps Plankton Camera System (Orenstein et al., 2020) and PlanktonScope (Pollina et al., 2020), are currently employed for routine monitoring. The Plankton Image Analyser (Culverhouse, 2015; Pitois et al., 2018) uses a similar image-capture method but is instead connected to the ship's 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 high 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.

3.3 METHOD

3.3.1 STUDY AREA SAMPLING METHODS

Data were collected at night during five fisheries surveys from 2016 to 2019 in the Celtic Sea aboard the RV Cefas Endeavour (Figure 3.1, Table 3.1). Zooplankton data were sampled at night to reduce the effect of vertical migration (Lampert, 1989; Pitois et al., 2018). 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 (Figure 3.1). PI data are freely available from the Cefas Data Hub: 10.14466/CefasDataHub.101.

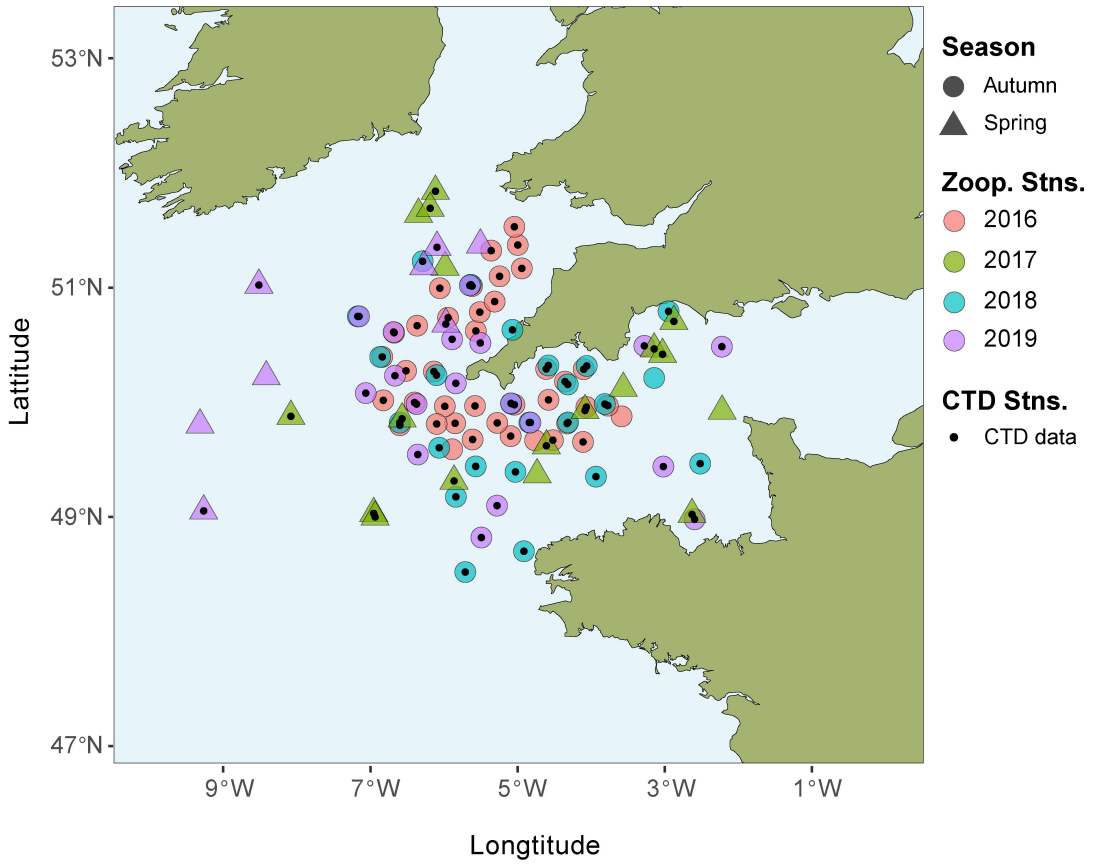

Figure 3.1: Location of the 107 zooplankton stations and 93 CTD stations off the southwest coast of the UK. Zoop. = Zooplankton. Stns. = stations.

Table 3.1: Number of available zooplankton stations per year

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

For zooplankton stations, $n = 107$. For corresponding CTD stations, $n = 93$. Dash indicates no data. Zoop., zooplankton.

3.3.2 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.

3.3.3 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, 2015; Pitois et al., 2018). 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 to 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 (Figure 3.2). 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

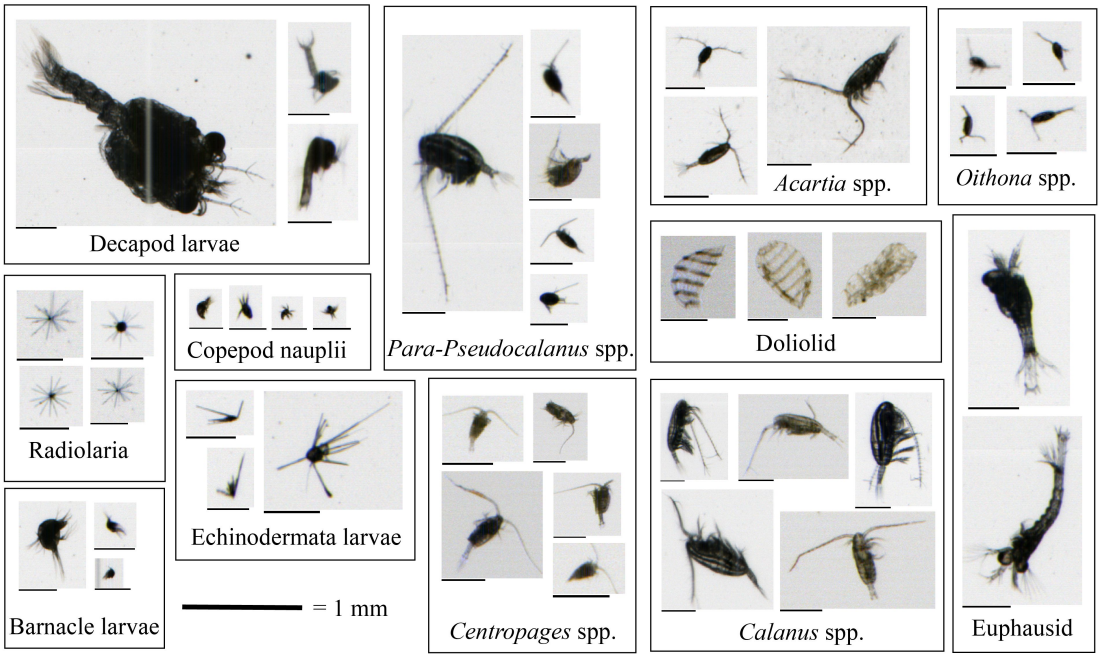

Figure 3.2: Collage of example mesozooplankton images used for training set for the 12 most abundant categories across all surveys.

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 (Figure 3.7) were used to compare the communities across the survey areas and between seasons.

3.3.4 STATISTICAL ANALYSIS

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, 2001). 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).

3.4 RESULTS

3.4.1 PHYSICAL CONDITIONS

There was a clear difference in temperature profiles between seasons (Figure 3.3). 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 higher 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 (Figure 3.3). 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 (Figure 3.4). 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 (Figure 3.3).

The NMDS plot (Figure 3.5**A**) was used to explore the relationship between physical variables and mesozooplankton distribution. The larger variation in the stratification between years in autumn compared to spring (Figure 3.3) 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 Figure 3.5**A**). Autumn 2018 had no overlap in community with either 2016 or 2019 (Figure 3.5**A**) and could reflect the cooler seas in 2018 (Figure 3.3). The envit() function suggested that the supplementary physical variables with a more linear relationship to the NMDS scores (indicated by the contour plots: SST, Figure 3.5**D** and ΔS, Figure 3.5**C**) were related with dissimilarity between seasonal groupings (Figure 3.5**A**). ΔT is not shown in Figure 3.5**A** as the model fitted by the envfit() function was not significant. The contour plots for SST (Figure 3.5**D**) and ΔS (Figure 3.5**C**) 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 (3.3) was associated with spring sites. Most autumn sites had a ΔT >

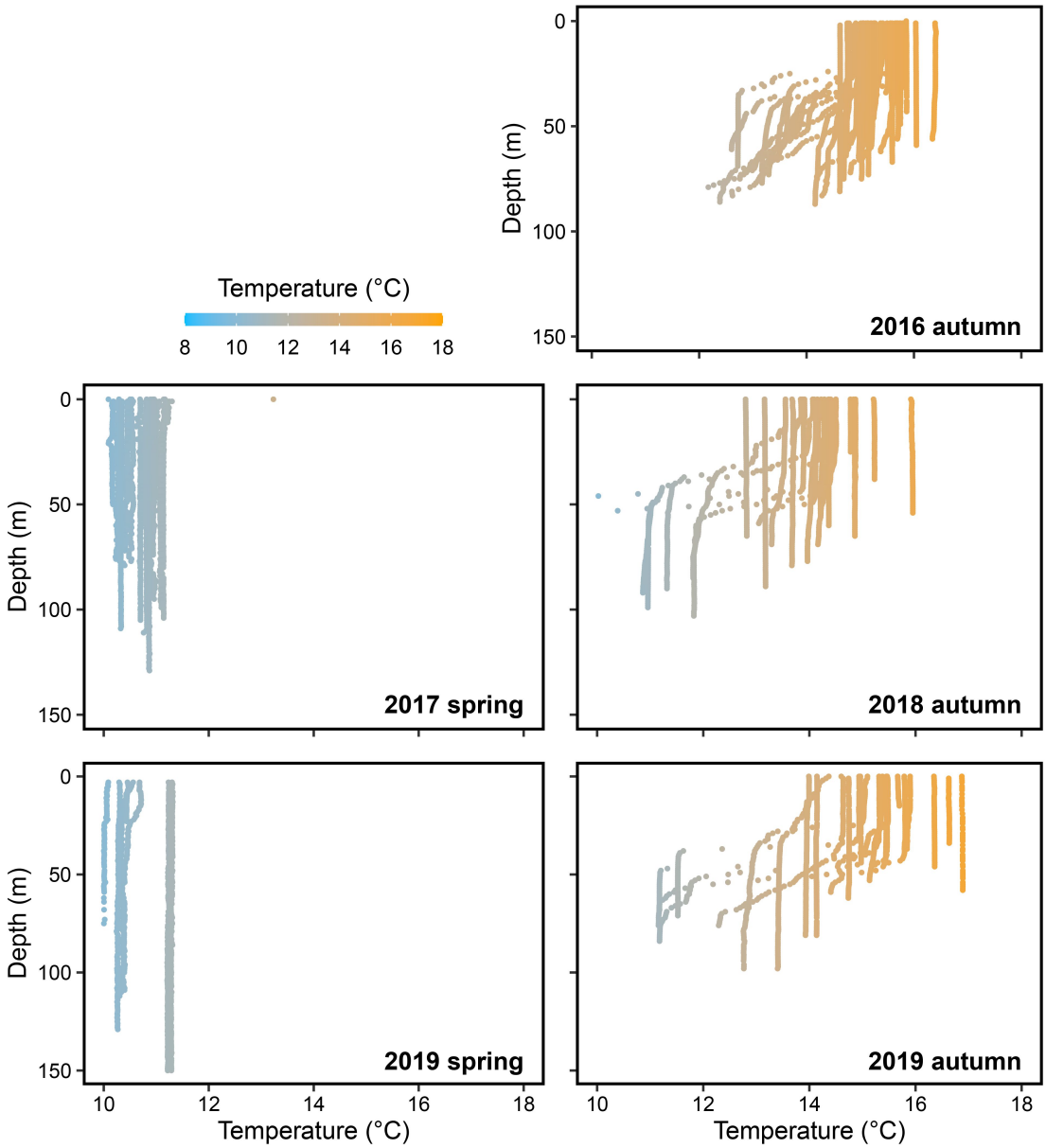

Figure 3.3: Temperature profiles for each station per survey. The inverse y-axis shows depth (1 m bins) with the x-axis showing temperature (∘C).

0.8 °C, although there was high variation between years (Figure 3.5**B**). The spread of points across the ΔT contours may reflect the variation in ΔT between locations (Figure 3.4)

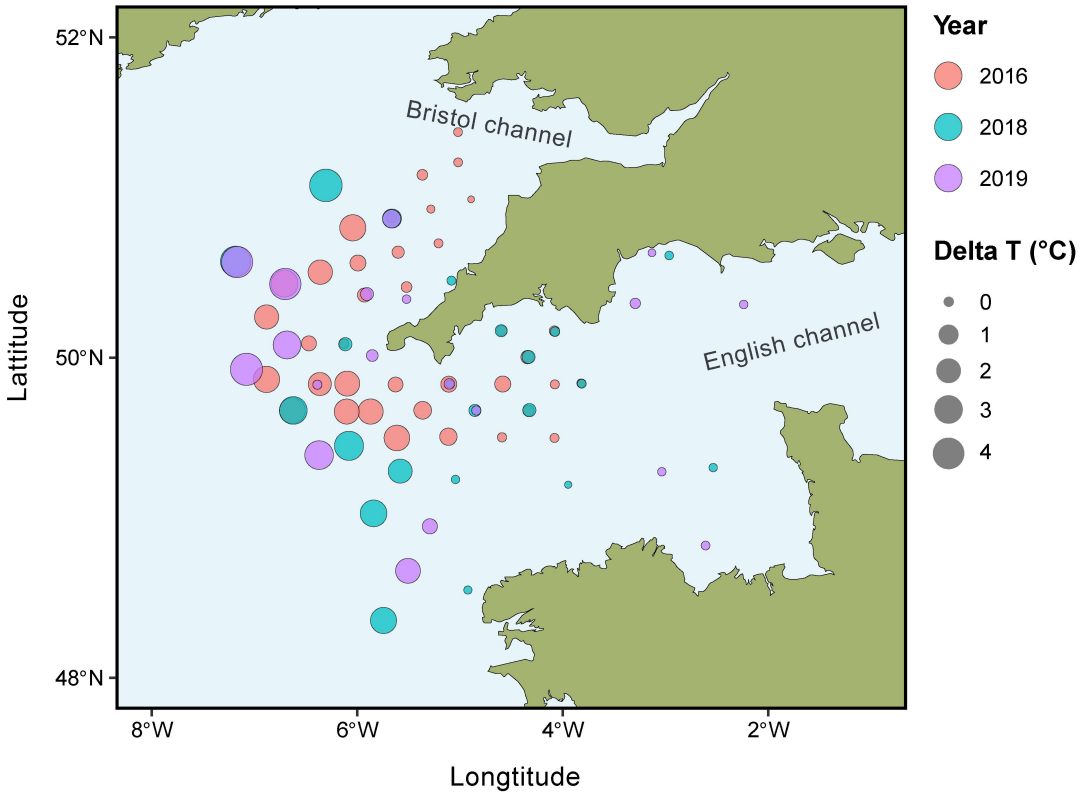

Figure 3.4: ΔT plotted as circle size for all autumn stations across all years.

3.4.2 MESOZOOPLANKTON COMMUNITY

Mesozooplankton abundance varied greatly between years and seasons (Figure 3.6). 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 (Figure 3.6).

On average, 9 to 10 taxa contributed to over 95 % of the total abundance. 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 (Figure 3.7). The contribution of copepod nauplii to total abundance was fairly consistent with the exception of spring 2019 (mean relative abundance ranged from 1.32 % to 5.85 % excluding spring 2019 where relative abundance was 14.64 %, Figure 3.7).

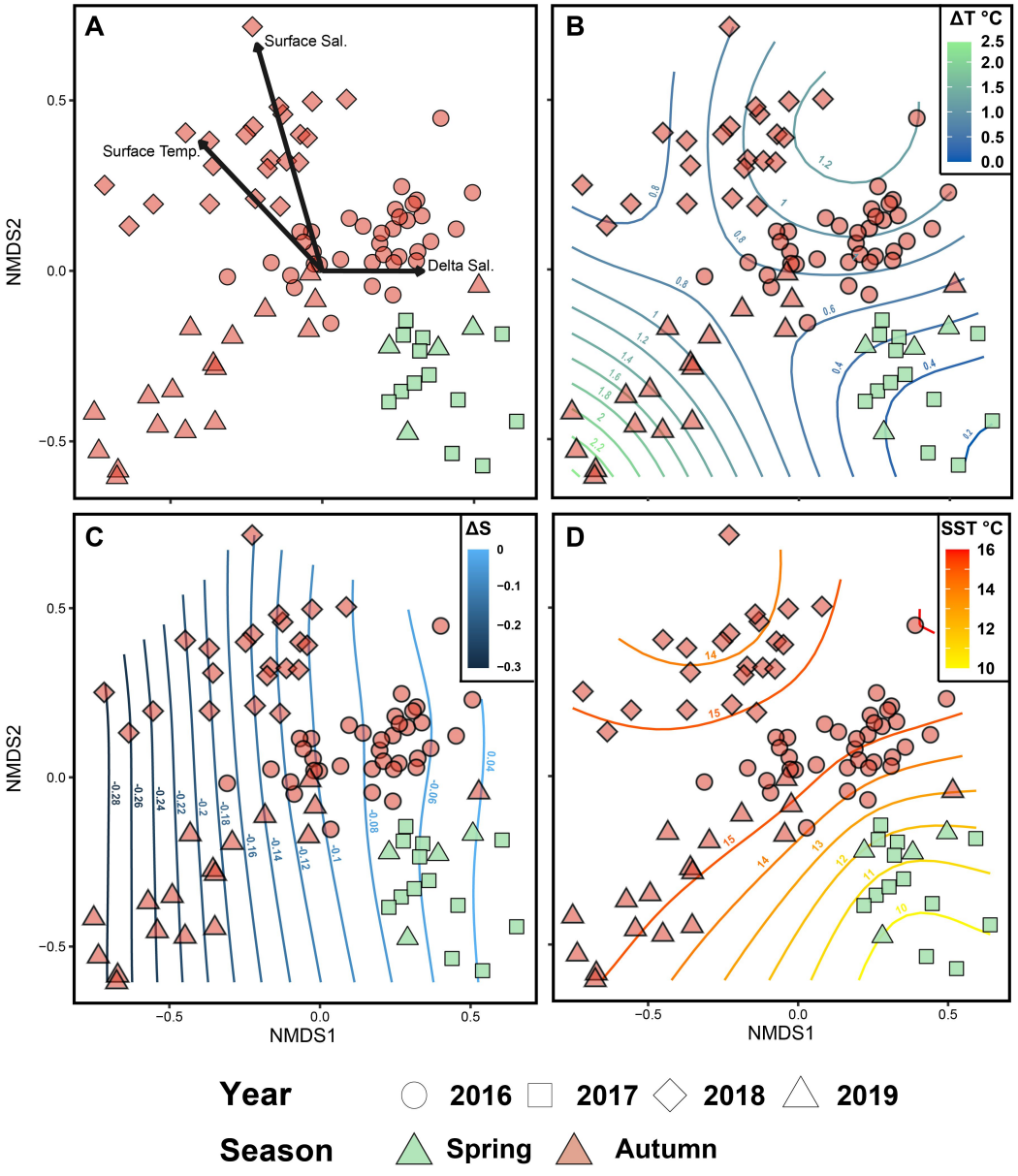

Figure 3.5: Analysis of interactions between environmental variables and plankton community site dissimilarity for sites with physical data. All plots show the same non-metric multidimensional scaling plot created using a Bray–Curtis dissimilarity matrix on Hellinger-transformed abundance data. (**A**) Shows the supplementary environmental variables plotted using the envfit() function. Plots (**B**) (ΔT), (**C**) (ΔS) and (**D**) (SST) show contour plots created using the ordisurf() function to explore the relationships between environmental variables and the NMDS site scores.

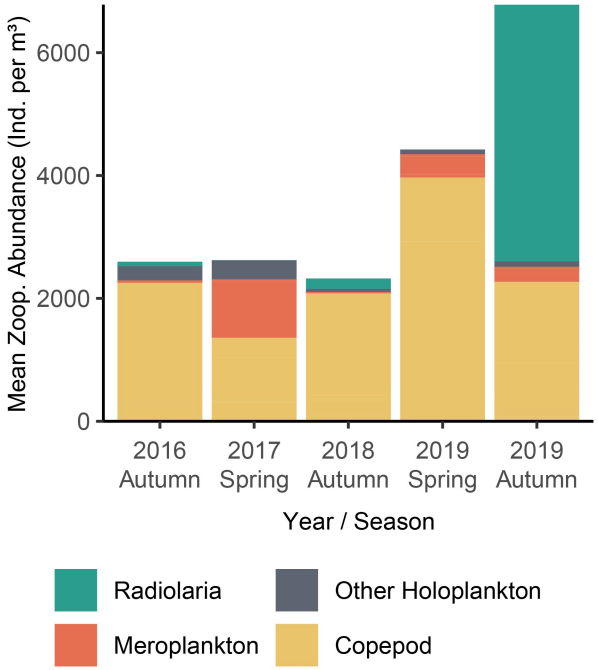

Figure 3.6: Mean station abundance for each survey with taxa grouped into four major categories.

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 3.21 %, Figure 3.7). None of these 3 taxa adhered to a particular season. Common to four of five surveys dominant taxa were *Oithona* spp., *Para-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 (Figure 3.7), copepods dominated the community accounting for > 70 % of the total abundance on average (Figure 3.6). Of these, an average 30.9% were classified as "Unknown copepods" (Figure 3.7). *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}).

Meroplankton constituted a larger portion of the spring community than autumn (Figure 3.6). 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

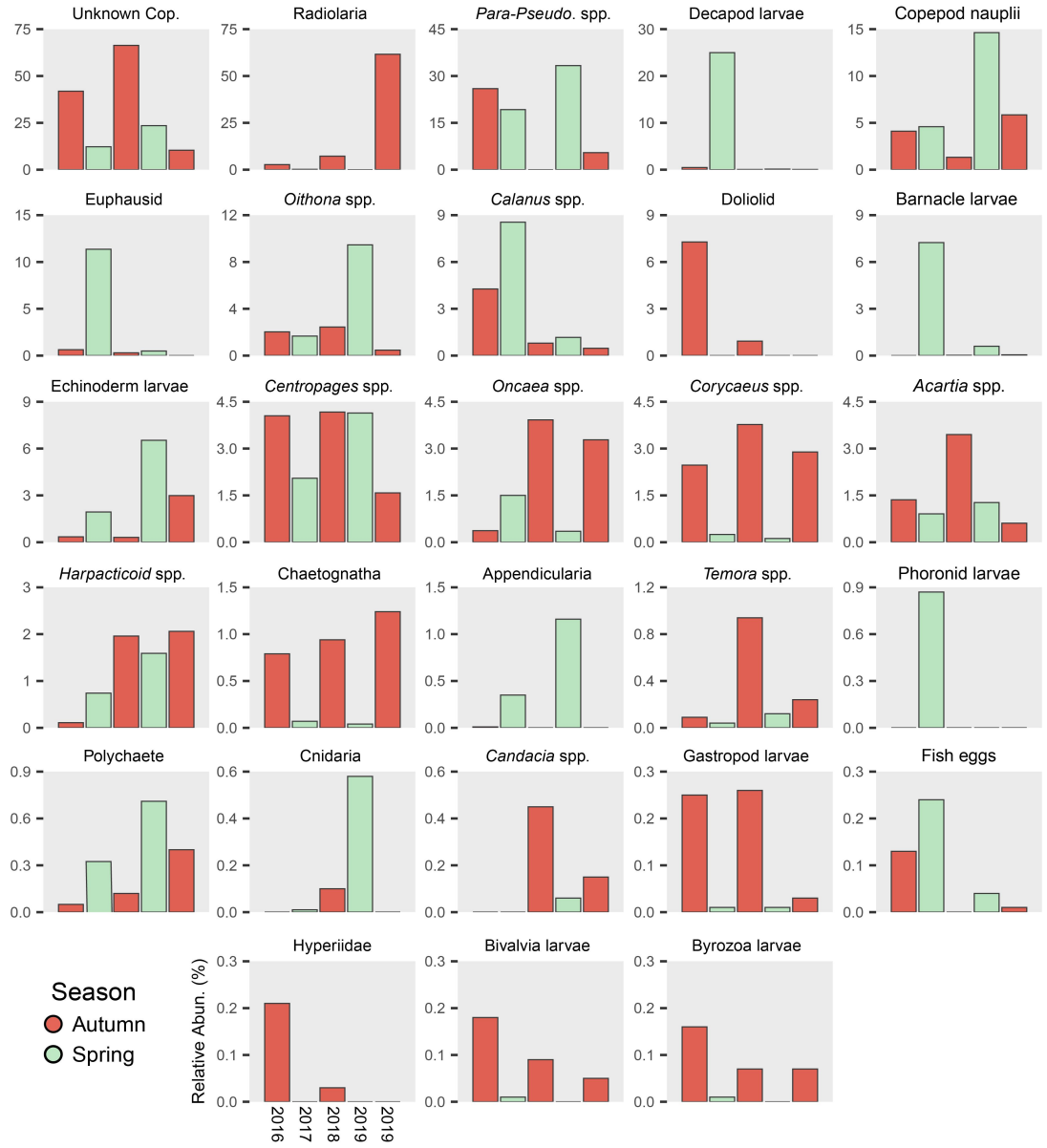

Figure 3.7: Relative abundance (%) (Relative Abun.) for all surveys for taxa that contributed to >1% of the total abundance. Axis labels are on bottom left subplot (Hyperiidae) and are the same for all subplots. Categories are arranged in order of decreasing relative abundance from highest in the top left to lowest in the bottom right.

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, Figure 3.7). Spring 2019 mainly comprised of echinoderm larvae, followed by Polychaete larvae (6.53 % and 0.71 %, Figure 3.7). Some meroplankton, mainly bryozoa and bivalve larvae, were found in high abundance during autumn (Figure 3.7).

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, Figure 3.8) 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 (Figure 3.8). 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 (Figure 3.8) echoes those trends seen in relative abundance (Figure 3.7). For example, the high relative abundance of radiolaria in autumn 2019 (Figure 3.7) is also seen in the ordination plot where radiolaria are highly correlated with autumn 2019 stations. This forcing from an individual, or a few key taxa, is characteristic for all autumn surveys (Figure 3.8). *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 (Figure 3.8). 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.

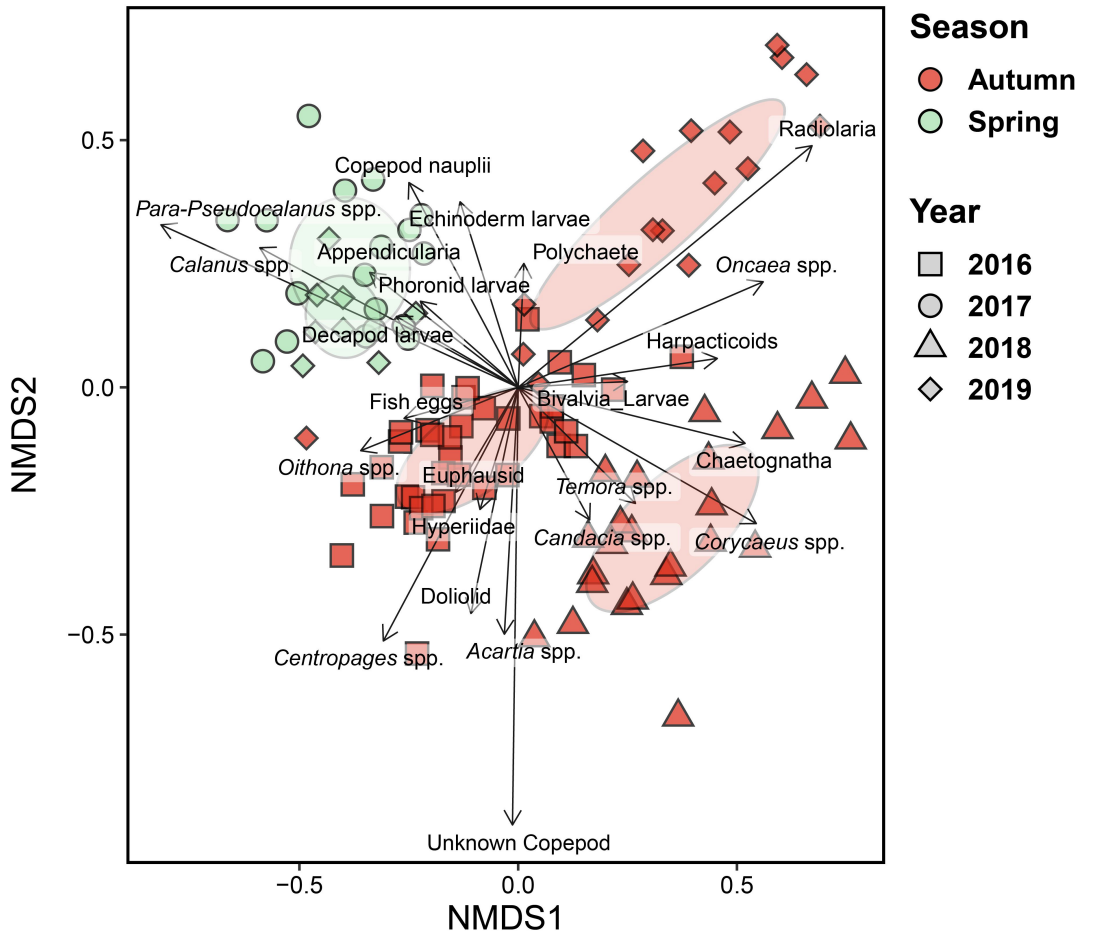

Figure 3.8: NMDS plot on Hellinger-transformed abundance data using a Bray–Curtis dissimilarity matrix on Hellinger-transformed abundance data, where stress =0.18. Supplementary variables taxa were plotted using the envfit() function with a P <0.05.

3.5 DISCUSSION

3.5.1 ENVIRONMENTAL DRIVERS OF SEASONALITY

The increased stratification offshore in early autumn (Figure 3.4) 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 survey's stations across the ΔS contours in the NMDS plot (Figure

3.5**C**), the communities are dissimilar between years (Figure 3.8). Conversely, spring had both consistency in the degree of stratification (negligible variation between years and sites, Figure 3.4) and in community overlap on the ordination plot (Figure 3.8). 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 (Figure 3.8).

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 et al. (2018). The authors review a suite of mechanisms that govern the mesozooplankton and suggest that a synergistic combination of mechanisms is often responsible. Those most relevant to our study are: the loophole hypothesis, whereby physiochemical changes favour some taxa (Irigoien et al., 2009); 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 (Figure 3.3). 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 3.1), 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.

3.5.2 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 (Figure 3.7) 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., 2017). 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 (Figure 3.7) 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.

3.5.3 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.

3.6 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.

3.7 BIBLIOGRAPHY

Agnarsson, I. and Kuntner, M. (2007). Taxonomy in a changing world: Seeking solutions for a science in crisis. *Systematic Biology*, 56(3):531–539.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1):32–46.

Atkinson, A., Polimene, L., Fileman, E. S., Widdicombe, C. E., McEvoy, A. J., Smyth, T. J., Djeghri, N., Sailley, S. F., and Cornwell, L. E. (2018). Comment. What drives plankton seasonality in a stratifying shelf sea? Some competing and complementary theories. *Limnology and Oceanography*, 63(6):2877–2884.

Bean, T. P., Greenwood, N., Beckett, R., Biermann, L., Bignell, J. P., Brant, J. L., Copp, G. H., Devlin, M. J., Dye, S., Feist, S. W., Fernand, L., Foden, D., Hyder, K., Jenkins, C. M., van der Kooij, J., Kröger, S., Kupschus, S., Leech, C., Leonard, K. S., Lynam, C. P., Lyons, B. P., Maes, T., Nicolaus, E. E., Malcolm, S. J., McIlwaine, P., Merchant, N. D., Paltriguera, L., Pearce, D. J., Pitois, S. G., Stebbing, P. D., Townhill, B., Ware, S., Williams, O., and Righton, D. (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. *Frontiers in Marine Science*, 4(AUG):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(6967):661–664.

Benoit-Bird, K. J. and Lawson, G. L. (2016). Ecological Insights from Pelagic Habitats Acquired Using Active Acoustic Techniques. *Annual Review of Marine Science*, 8(1):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 Journal*, 11(6):1331–1344.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Burki, F. and Keeling, P. J. (2014). Rhizaria. *Current Biology*, 24(3):R103–R107.

Cifelli, R. and Sachs, K. N. (1966). Abundance relationships of planktonic Foraminifera and Radiolaria. *Deep-Sea Research and Oceanographic Abstracts*, 13(4):751–753.

Clarke, K. R. and Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Marine Ecology-Progress Series*, 92:205.

Cowen, R. K. and Guigand, C. M. (2008). In situ ichthyoplankton imaging system (ISIIS): System design and preliminary results. *Limnology and Oceanography: Methods*, 6(2):126–132.

Culverhouse, P. F. (2015). An Instrument for Rapid Mesozooplankton Monitoring at Ocean Basin Scale. *Journal of Marine Biology and Aquaculture*, 1(1):1–11.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., David, R., Dell'Anno, A., Dzhembekova, N., Garcés, E., Gasol, J. M., Goela, P., Féral, J. P., Ferrera, I., Forster, R. M., Kurekin, A. A., Rastelli, E., Marinova, V., Miller, P. I., Moncheva, S., Newton, A., Pearman, J. K., Pitois, S. G., Reñé, A., Rodríguez-Ezpeleta, N., Saggiomo, V., Simis, S. G., Stefanova, K., Wilson, C., Martire, M. L., Greco, S., Cochrane, S. K., Mangoni, O., and Borja, A. (2016). Implementing and innovating marine monitoring approaches for assessing marine environmental status. *Frontiers in Marine Science*, 3(NOV):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. *Limnology and Oceanography: Methods*, 3(2):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. *Journal of Plankton Research*, 24(8):797–805.

Eloire, D., Somerfield, P. J., Conway, D. V., 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. *Journal of Plankton Research*, 32(5):657–679.

Giering, S. L., Wells, S. R., Mayers, K. M., Schuster, H., Cornwell, L., Fileman, E. S., Atkinson, A., Cook, K. B., Preece, C., and Mayor, D. J. (2019). Seasonal variation of zooplankton community structure and trophic position in the Celtic Sea: A stable isotope and biovolume spectrum approach. *Progress in Oceanography*, 177(March 2018):101943.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.-B. B., Cawood, A., Pesant, S., García-Comas, C., and Prejger, F. (2010). Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, 32(3):285–303.

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

Harris, R. (2010). The L4 time-series: The first 20 years. *Journal of Plankton Research*, 32(5):577–583.

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

Heath, M. R. (2005). Regional variability in the trophic requirements of shelf sea fisheries in the Northeast Atlantic, 1973-2000. *ICES Journal of Marine Science*, 62(7):1233–1244.

Highfield, J. M., Eloire, D., Conway, D. V., Lindeque, P. K., Attrill, M. J., and Somerfield, P. J. (2010). Seasonal dynamics of meroplankton assemblages at station L4. *Journal of Plankton Research*, 32(5):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. *Progress in Oceanography*, 166(July 2017):30–40.

Irigoien, X., Fernandes, J. A., Grosjean, P., Denis, K., Albaina, A., and Santos, M. (2009). Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment. *Journal of Plankton Research*, 31(1):1–17.

Irigoien, X. and Harris, R. P. (2003). Interannual variability of Calanus helgolandicus in the English Channel. *Fisheries Oceanography*, 12(4-5):317–326.

Johns, D. (2006). The Plankton Ecology of the SEA 8 area. *Strategic Environmental Assessment Programme, UK Dept. of Trade and Industry*, pages 1–44.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Functional Ecology*, 3(1):21.

Legendre, P. and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2):271–280.

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J. B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O., Miloslavich, P., Lara-Lopez, A., Kudela, R. M., Lopes, R. M., Karp-Boss, L., Kiko, R., Jaffe, J. S., Iversen, M. H., Irisson, J. O., Hauss, H., Guidi, L., Gorsky, G., Giering, S. L. C., Gaube, P., Gallager, S., Dubelaar, G., Cowen, R. K., Carlotti, F., Briseño-Avena, C., Berline, L., Benoit-Bird, K. J., Bax, N. J., Batten, S. D., Ayata, S. D., and Appeltans, W. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6(MAR).

Mackas, D. L. and Beaugrand, G. (2010). Comparisons of zooplankton time series. *Journal of Marine Systems*, 79(3-4):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. *Progress in Oceanography*, 137:513–523.

McQuatters-Gollop, A., Johns, D. G., Bresnan, E., Skinner, J., Rombouts, I., Stern, R., Aubert, A., Johansen, M., Bedford, J., and Knights, A. (2017). From microscope

to management: The critical value of plankton taxonomy to marine policy and biodiversity conservation. *Marine Policy*, 83(May):1–10.

MSFD (2009). Marine Strategy Framework Directive Newsletter.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2013). Community ecology package.

Orenstein, E. C., Ratelle, D., Briseño-Avena, C., Carter, M. L., Franks, P. J., Jaffe, J. S., and Roberts, P. L. (2020). The Scripps Plankton Camera system: A framework and platform for in situ microscopy. *Limnology and Oceanography: Methods*, 18(11):681–695.

Owens, N. J., 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 J.W. Dippner and M. Krause. *Journal of Marine Systems*, 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. *Limnology and Oceanography: Methods*, 8(SEPT):462–473.

Pitois, S. G., Bouch, P., Creach, V., Van Der Kooij, J., Kikkawa, T., Kita, J., Ishimatsu, A., Pitois, S. G., Tilbury, J., Bouch, P., Close, H., Barnett, S., and Culverhouse, P. F. (2016). Comparison of zooplankton data collected by a continuous semi-automatic sampler (CALPS) and a traditional vertical ring net. *Journal of Plankton Research*, 38(4):931–943.

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

Pollina, T., Larson, A. G., Lombard, F., Li, H., Colin, S., de Vargas, C., and Prakash, M. (2020). PlanktonScope: affordable modular imaging platform for citizen oceanography. *BioRxiv*.

R Development Core Team (2018). A Language and Environment for Statistical Computing.

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. *Progress in Oceanography*, 68(1):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. *Journal of Plankton Research*, 37(3):519–529.

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.

Sieracki, C. K., Sieracki, M. E., and Yentsch, C. S. (1998). An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, 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. *Progress in Oceanography*, 108:1–42.

Smyth, T., Atkinson, A., Widdicombe, S., Frost, M., Allen, I., Fishwick, J., Queiros, A., Sims, D., and Barange, M. (2015). The Western Channel Observatory. *Progress in Oceanography*, 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(6):e98709.

Southward, A. J., Langmead, O., Hardman-Mountford, N. J., Aiken, J., Boalch, G. T., Dando, P. R., Genner, M. J., Joint, I., Kendall, M. A., Halliday, N. C., Harris, R. P., Leaper, R., Mieszkowska, N., Pingree, R. D., Richardson, A. J., Sims, D. W., Smith, T., Walne, A. W., and Hawkins, S. J. (2004). Long-term oceanographic and ecological research in the western English Channel. *Advances in Marine Biology*, 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 Journal of Marine Science*, 51(4):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-Sea Research Part I: Oceanographic Research Papers*, 49(8):1445–1461.

Steinberg, D. K. and Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science*, 9(1):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. *Artificial intelligence review*, 12(1-3):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. *Progress in Oceanography*, 56(1):7–136.

4

RESOLUTION CHANGES RELATIONSHIPS: OPTIMIZING SAMPLING DESIGN USING SMALL SCALE ZOOPLANKTON DATA

The following Chapter was submitted to Progress in Oceanography and at time of submission is under review, see Published Works. The following is a copy of the manuscript.

4.1 ABSTRACT

Marine research surveys are an integral tool in understanding the marine environment. Recent technological advances have allowed the development of automated or semi-automated methods for the collection of marine data. These devices are often easily implemented on existing surveys and can collect data at finer spatiotemporal resolutions than traditional devices. We used two automated instruments: the Plankton Imager and FerryBox, to collect information on zooplankton, temperature, salinity and chlorophyll in the Celtic Sea. The resulting data were spatiotemporally aligned and merged to decreasing spatial resolutions to explore how distribution patterns and the relationship between variables change across different spatial resolutions. Relative standard deviation was used to describe variability of merged data within grid cells. All variables displayed large, area-wide spatial patterns excluding copepod size which remained consistent across the study area. Copepod biomass and abundance displayed high variations across small spatial scales. Decreasing the sampling resolution changed the description of the data where small spatial changes (those that occur over scales < 3 km) were lost and area wide patterns were emphasized. Furthermore, we found that the choice of resolution can affect both the statistical strength and significance of relationships with high variability at lower resolutions due to the mismatch between the scales of ecological processes and sampling. Determining the optimum sampling resolution to answer a specific question will be dependent upon several factors, mainly the variable measured, season, location and scale of process, which all drive variation. These considerations should be a key element of survey design, helping move towards an integrated approach for an improved understanding of ecosystem processes and gaining a more holistic description of the marine environment.

4.2 INTRODUCTION

Research surveys are fundamental in furthering our understanding of the marine environment. Motivated by providing a holistic, ecosystem approach to monitoring

(Kupschus et al., 2016) or mandated by policy (Vinet and Zhedanov, 2011; Danovaro et al., 2016), technological developments are helping surveys move toward increasingly interdisciplinary approaches (Working Group of International Pelagic Surveys, 2015; Doray et al., 2018). Installing automated technologies, which allow for continuous data collection with little human input, are a straightforward step in achieving this goal. Devices such as the FerryBox, used here, collects both physical and biological variables continuously and reports high frequency data throughout a survey (Petersen and Colijn, 2017). Due to their continuous, automated nature, these data are readily available for (near) real time analysis, or retrospective, post-collection, analysis. These often easy to implement devices can reduce vessel costs (time, fuel or labor) when compared to traditional methods (deployment of nets) or towed imaging devices (e.g., deployment and recovery or reduced vessel speed while towing) and allow for an increased number of variables collected at no or little extra cost. Their use can help surveys stay within financial limitations (Bean et al., 2017; Pitois et al., 2018) and allow for optimized survey design (Kupschus et al., 2016) by easily increasing sampling coverage and intensity (Owen, 2014; Doray et al., 2018). These devices do not require an onboard expert, freeing up vessel space and further reducing costs. They can typically sample in all weather conditions, allowing for data collection in hard to sample locations or reduce the time spent waiting for safe sampling conditions.

In recent decades, automated technologies have become commonplace in multiple marine disciplines. The FerryBox is one of many established options for continuous, automated sampling of physical parameters. Acoustic devices are used globally, commercially and scientifically, for fishing (Mann et al., 2008; Simmonds and MacLennan, 2008), marine mammal research (Johnson and Tyack, 2003; Johnson et al., 2009) and bathymetry (de Moustier, 1986). Other purpose-built devices sample a single component, for example, fish eggs (Checkley et al., 2000) or phytoplankton (Olson et al., 2018). The continuous sampling of zooplankton, globally important in carbon cycles (Steinberg et al., 2002; Steinberg and Landry, 2017), fisheries science (Beaugrand et al., 2003; Heath, 2005; Lauria et al., 2013) and used as climate change indicators (Taylor et al., 2002), provides a unique technological challenge arising

from the difficulties with sampling the entire zooplankton component accurately. Zooplankton includes a wide range of sizes and behaviors and undergoes rapid, temporal and spatial changes, known as plankton patchiness (Mackas et al., 1985; Abraham, 1998), making replicable sampling difficult. These fine scales changes can be seen with traditional net haul data, which provide a 'snapshot' of the zooplankton but have very high replicate tow variability (Wiebe and Wiebe, 1968; Lee and McAlice, 1979; Skjoldal et al., 2013). Capture by netting, and analysis by microscopy, form the gold standard of zooplankton sampling and are principally responsible for our understanding of zooplankton ecology. Their continued use to maintain time series and use as a reliable method is essential to further understanding the zooplankton. Like all sampling devices, nets do suffer some limitations. Deploying plankton nets is time consuming and the collected sample is often preserved using hazardous chemicals for analysis on shore. These challenges have placed pressure on developing cost-effective methods (Danovaro et al., 2016) and in response, technological developments have resulted in a variety of newer devices (Wiebe and Benfield, 2003; Lombard et al., 2019).

The Plankton Imager (PI) is a continuous, automated, imaging device used to sample zooplankton. The PI takes images of all passing particles in seawater pumped onboard a ship (Culverhouse et al., 2016; Pitois et al., 2018; Scott et al., 2021). An initial study evaluating the first generation of the instrument (previously known as Plankton Image Analyzer) against traditional net sampling, found good agreement in the spatial distribution of zooplankton abundances, although noted a portion of fragile organisms (e.g., Appendicularia) were likely to be damaged by the system pump and consequently under-sampled. The study also described the overall lower capture efficiency of the PI with discrepancies mainly resulting from image quality, such as blurred images, which made accurate classification challenging. In response, hardware changes have resolved these issues, resulting in much improved image quality. The PI has since been used to describe temporal changes in the mesozooplankton community Scott et al. (2021). This study found that those fragile species (e.g., Appendicularia) are sampled in sufficient quantity to detect seasonal difference. More recently the application of the PI to ecological indicators has been

tested (Pitois et al., 2021). To date, all published studies have used the PI for point sampling, similar to a deployed ring net, as opposed to continuous sampling. Here we used a new data extraction method to best take advantage of the PI's continuous nature.

The PI has been used alongside the FerryBox routinely during UK fisheries surveys in the Celtic Sea. We use data collected in parallel from these devices to explore small scale changes in the zooplankton in the context of physical parameters and the relationships therein. As automated devices and the ability to collect vast quantities of data become increasingly common place, a new challenge has emerged in that the data collection rate has become faster than the processing rate, resulting in data bottlenecks. It is therefore important to focus collection efforts to gather the correct type of information, at the required locations, times and scales to answer a particular question, balancing research needs with budget limitations. Here, we aim to explore how best to determine the optimal resolution appropriate for the target process or relationship to avoid mismatching between sampling resolution and ecological scales. These can be used to inform future survey design leading to an increasingly holistic survey description.

4.3 MATERIALS AND METHODS

All data were collected in the Celtic Sea from the 3rd of October to the 7th of November 2020 aboard the RV Cefas Endeavour as part of the PELTIC survey (PELagic ecosystems in the Western English Channel and eastern celTIC Seas) (Working Group of International Pelagic Surveys, 2015) (Figure 4.1). All in situ data were collected using the ship's continuous flow system sampling at 4 m below sea level. Zooplankton data were collected using the PI (Pitois et al., 2018). Temperature, salinity and fluorescence were collected using the FerryBox (4H-JENA, Germany). Zooplankton data were sampled at night for consistency and to reduce the effect of vertical migration (Lampert, 1989; Pitois et al., 2018).

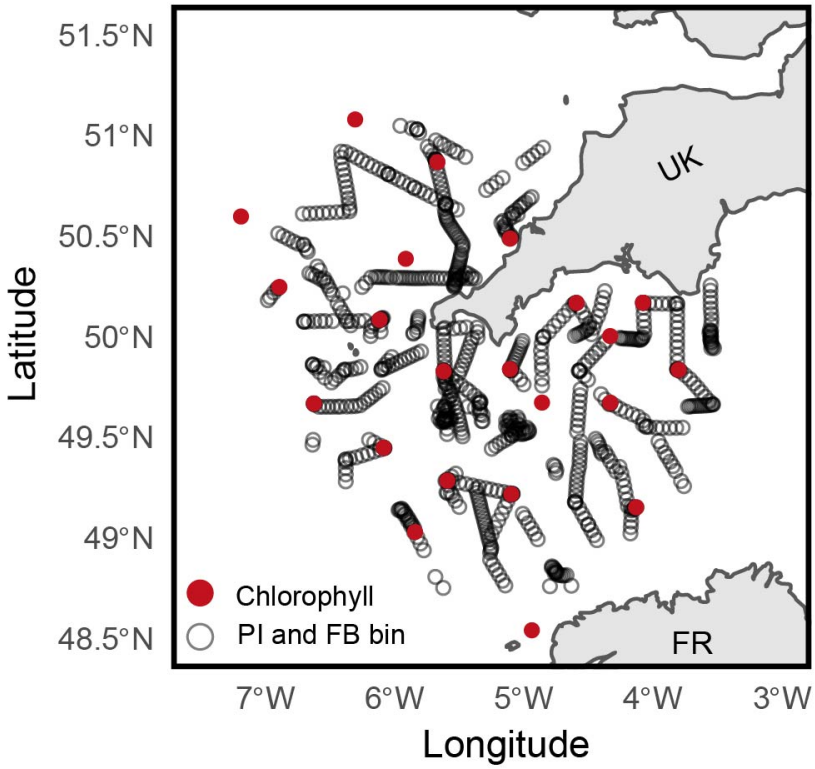

Figure 4.1: Celtic Sea study area and spatial extent of the collected data. Red filled symbols represent in situ discrete chlorophyll samples. Black open symbols represent PI and FerryBox (temperatures, salinity, fluorescence) 10-minute bins.

4.3.1 PLANKTON IMAGER (PI)

The PI was connected to the ship's continuous flow pump, sampling at 22 L min^{-1} with negligible downtime (Figure 4.2). The inlet pipe and internal ships piping have various internal diameters larger than the flow cell which has an internal depth of 12.8 mm, giving a field of view of 10 μm × 20.48 mm. As sea water passes through the flow cell where all passing particles are imaged by a Basler 2048-70kc line scan camera with a scanning rate of 70,000 lines per second. Lines are then stitched together, and regions of interest (ROI) are extracted and saved as images. GPS, time and particle size data (area, length and width) are saved in the metadata of each image. The PI worked continuously throughout the survey. The PI has adjustable minimum and maximum size parameters (min. 100 µm to max. 2 cm). When using this range, the processing rate of the images could not keep up with their collection rate (i.e., the images are captured faster than they can be written to disk. For the survey, the size range was

set from 180 μm to 2 cm. This reduced the number of captured images and allowed for a more manageable dataset for archiving, processing and analysis. Even with this reduced size range, a 1 month survey typically collects 1 tb of data.

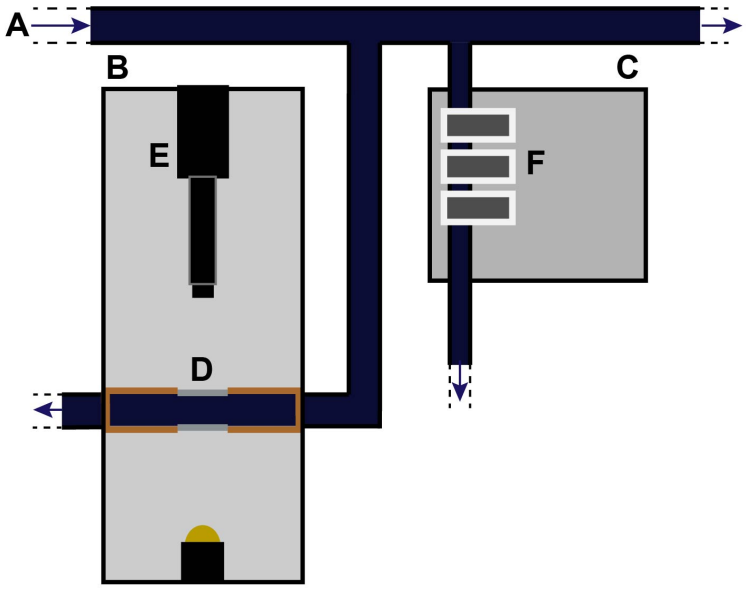

Figure 4.2: Schematic of Plankton Imager (PI) and FerryBox setup aboard the RV Cefas Endeavor. Water is pumped onboard from 4 m below sea level (**A**). This supplies the PI (**B**) and FerryBox (**C**). Within the PI water flows through a flow cell (**D**) where passing particles are imaged by a line scan camera (**E**). Within the FerryBox water passes through a suite of sensors (**F**), here temperature, salinity and fluorescence are used.

Over 70 million images were collected during the survey. In the absence of an accurate classifier for the PI, all images required manual classification. A series of subsets were used to reduce the number of images classified to an achievable quantity. A 0.25°grid was transposed over the study area. Each grid cell typically had multiple transects passing through with the specific number of transects varying based on vessel movements. Data were extracted from the shortest nighttime transect within each 0.25° cell (min = 20 mins, mean = 136 mins, max = 420 mins). The transect time (and therefore water sampled) varied within each grid cell dependent on vessel activities (e.g., steaming between stations or fishing). This extraction resulted in 17 million images for classification. Finally, data were temporally subsampled where 1 in 10 images were extracted from each transect to further reduce the size of the dataset. This process is similar to random subsampling or 'splitting' of a physical sample

and assumes the distribution of organisms within a subsample follows a Poisson distribution (Postel et al., 2000). The resultant 1.7 million images were manually classified to "copepod" or "other" with copepod including the adult and copepodite stages and the latter category comprising all non-copepod zooplankton and detritus. Sorting to only these categories greatly sped up the classification process. The final copepod count per grid cell was multiplied by 10 to resolve for subsampling.

For statistical analysis, the selected transects were divided into 10 minute bins where each bin sampled 0.22 m3 of seawater. This totaled 853 bins (Figure 4.1). The minimum bin size was determined as a compromise between obtaining the smallest possible spatial resolution while sampling a sufficient amount of water to allow for subsampling. Sampling a smaller amount of water (e.g. 1 min and 0.022 m3 of water) may have resulted in unrealistic values when resolving for subsampling. Copepod density was reported as individuals per m3 (indv. m^{-3}). Particle lengths were obtained from image metadata files and used as a proxy for copepod size or total length. Within each 10 minute bin, the geometric mean (geomean) size of all individuals was calculated to take into account their non-normal distribution. This mean value was used to calculate mean copepod wet weight (i.e., individual biomass) following the equation from Pitois et al. (2021):

$$copepodwetweight = 0.299 * copepodprosome^{2.8948} \tag{4.1}$$

This was then upscaled with copepod density to calculate biomass across the bin reported as mg m^{-3}.

4.3.2 *in situ* CHLOROPHYLL MEASUREMENTS

The FerryBox consists of a water inlet connected to the ship's continuous flow (Figure 4.2). It comprises a suite of sensors for measuring physical variables (e.g., temperature, salinity, turbidity, fluorescence and oxygen) and corresponding metadata (GPS, date and time). All data are automatically bin averaged to 1 minute on collection to save storage space. Only temperature (°C), salinity (psu) and fluorescence were used in this study. Discrete chlorophyll samples were taken from

the continuous flow passing through the FerryBox at 22 locations within the study area Figure 4.1. The chlorophyll was extracted using 90% acetone and measured with a Turner fluorometer (Strickland and Parsons, 1972). A linear model was used to convert the FerryBox chlorophyll fluorescence to chlorophyll. The fitted regression model was: [chlorophyll mg m^{-3} = 0.72 * chlorophyll fluorescence + 0.0637]. The regression was statistically significant (R^2 = 0.91, F = 206.1, $p < 0.001$). FerryBox data were spatiotemporally aligned to the 853, 10-minute copepod bins described above. This was achieved by taking the mean of three variables across the bin.

4.3.3 ANALYSIS

At the finest resolution (the 853, 10 minute bins), data were plotted to describe the broader spatial patterns and examine small spatial scale changes in all variables. For copepod biomass, size and density, the change in value between a 10 minute bin and the previous 10 minute bin was examined to see if there was a relationship between distance or time between bins and change in value. For statistical analysis and to investigate how changes in resolution can affect how spatial patterns are described and if small scale changes are omitted or accentuated, the 853 bins were merged to decreasing resolutions. Merging was achieved by taking the mean value of all merged bins for each variable. To explain variation within each cell at each resolution, Relative Standard Deviation (RSD) was used as it expresses the variability of a data set as a percentage relative to its location. RSD is calculated as: RSD = (sample standard deviation / sample mean) × 100.

Four resolutions, 0.1°, 0.25°, 0.5°and 1°were selected. The largest resolution was chosen based on the spatial extents of our study area. A resolution lower than 1° would have resulted in too few cells or a cell that contained the entire data. The selected resolutions were used to visually compare the changes in the description of spatial patterns associated with merging data to coarser resolutions. The relationship between RSD and decreasing resolution was also explored for all variables at resolutions between 0.01°and 0.9°decreasing in 0.01° increments. Here the mean RSD value across all cells was used. For statistical analyses, the same

resolution range (0.01°to 0.9°, by = 0.01°) was used. Spearman's ρ coefficient was used to test for a significant relationship between number of stations per cell and RSD and explore the relationship between copepod biomass and chlorophyll at these resolutions. Copepod biomass, size and density were log-transformed (log10(x+1)) for Figures 4.3, 4.4 and 4.5 to highlight variability.

4.4 RESULTS

4.4.1 SPATIAL DISTRIBUTION

Spatial patterns across the study area for copepod density and biomass were closely aligned (Figure 4.3**A**, 4.3**B**). Higher copepod densities (>8000 indv. m^{-3}) and biomass (>150 mg m^{-3}) were typically found in the middle of the study area with lower values found toward the south (< 2000 indv. m^{-3} and < 50 mg m^{-3}, respectively) (Figure 4.3**A**, 4.3**B**). Density ranged from 45 to 8790 indv. m^{-3} and biomass ranged from < 1 to 155 mg m^{-3}. Copepod size had a more uniform distribution across the study area with no obvious spatial patterns with some localized exceptions of larger copepods found in the northern most extents of the study area (Figure 4.3**C**). Size ranged from 199 to 2590 μm. Large fluctuations in each variable were seen over small spatial scales (between adjacent bins, 5 km), this was less frequent for copepod size and is most evident in the central study area for copepod biomass (Figure 4.3**A** – 4.3**C**).

Small scale changes were not present in chlorophyll concentration, temperature or salinity (Figure 3**D** - 3**F**) with these variables displaying more gradual changes across the area. Temperature was higher toward the east of the study area (Figure 4.3**E**) and salinity was higher toward the south (Figure 4.3**F**). Chlorophyll concentration was consistently low (< 0.6 mg m^{-3}) except for the most south-westerly extents of the study area where the maximum value of 1.5 mg m^{-3} was seen (Figure 4.3**D**).

The large variations across small spatial changes in all copepod variables are better highlighted by Figure 4.4. The change in value between a 10-minute bin and the previous 10-minute bin was explored for density (Figure 4.4**A**), size (Figure 4.4**B**)

and biomass (Figure 4.4**C**). There was no clear relationship between the range and distance from the previous station for any variable at small spatial scales (< 5 km). On the contrary, the highest changes in density and biomass tended to be within 3 km of the previous bin (Figure 4.3**A**, **B**). This was seen most clearly in adjacent datapoints located in the middle of the study area (Figure 4.3**B**, **C**)

4.4.2 DESCRIPTION AT CHANGING RESOLUTIONS

Most resolutions captured the broad spatial patterns evident in the smallest resolution for copepod biomass, density and size (Figure 4.5, Figure 4.3**A** - **C**). For example, the regions of low biomass toward the north and higher biomass toward the southwest of the study area (Figure 4.3**B**) were visible at all resolutions (Figure 4.5). Although, large changes in copepod variables over small spatial scales seen at the smallest resolution were partly lost at a 0.1° resolution and absent entirely at 1° (Figure 4.5). This is true for all copepod variables where a high level of detail was lost by only halving the resolution. For example, the area of low biomass (7° W, 49° N) seen at 0.5° resolution was lost when halving to 1° (4.5, column 1). This loss of small scale detail while capturing broad patterns with decreased resolution was mirrored by copepod density and size.

RSD is shown spatially for the selected resolutions (Figure 4.5) and in increasing 0.01° increments in a scatter plot (Figure 6) for all copepod variables. Using Spearmans ρ, copepod mean density RSD and number of datapoints per cell were consistently significantly related in cells < 0.27° resolution (at 0.26°, Rs = 0.31, p < 0.05, n = 78). For mean biomass RSD, there was consistent significance for cells < 0.4° (at 0.39°, Rs = 0.29, p = 0.05, n = 44). For mean size RSD there was no consistent significant relationship at any resolution. For all three-copepod variables, there were exceptions seen at lower resolutions which may result from an insufficient sample size for the Spearmans ρ test. Copepod sizes were consistently low in RSD (< 30 %) between grid cells both spatially and across resolutions (Figure 4.5). There was an increase in mean RSD, from 9.96 % to 30.55 %, with decreasing resolution (Figure 4.6), although marginal when compared to other variables. Biomass had the highest

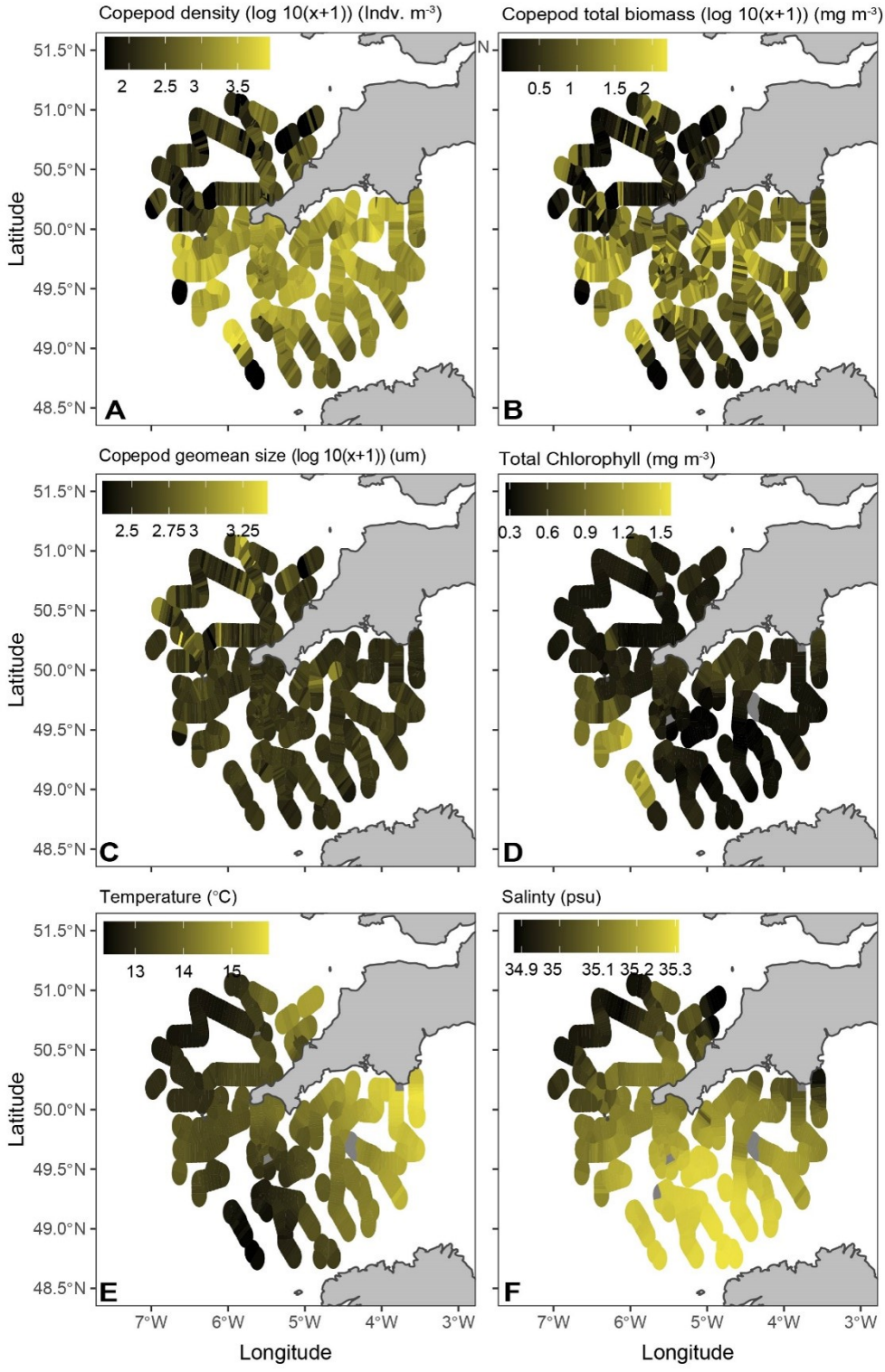

Figure 4.3: Overview of each variable at the finest spatial resolution (10 minute bins, approx. 2.2 m^{-3} seawater) as point data where each point is the bin median latitude and longitude. Point data are highlighted by using Voronoi triangles (with a maximum radius size around the point of 0.1°) which allows for a bigger point size, while avoiding overlap to better highlight small scale changes in the variable. For: (**A**) Copepod density (log10(x+1)) (indv. m^{-3}), (**B**) Copepod total biomass (log10(x+1)) (mg m^{-3}), (**C**) Copepod geomean size (μm), (**D**) Chlorophyll (mg m^{-3}), (**E**) Temperature (°C) and (**F**) Salinity (psu). Color scales are consistent with Figure 4.5 and Figure 4.5 for comparison.

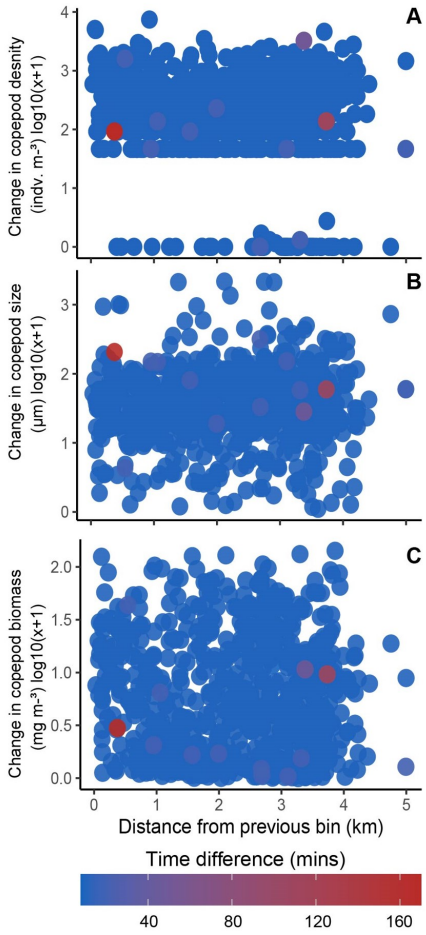

Figure 4.4: Change in parameters value (y-axis) and time (color scales) between two bins as a function of their distance from each other (x-axis) for (**A**) copepod density (indv. m^{-3}), (**B**) copepod geomean size (μm) and (**C**) biomass (mg m^{-3}).

spatial variation in RSD at all resolutions (Figure 4.5). There was a larger increase meal cell biomass, from 54.1 % to 140.57 %, with decreasing resolution (Figure 4.6). Density RSD was more consistent and more closely aligned spatially with biomass than size and a had reduced mean cell RSD.

4.4.3 RELATIONSHIP BETWEEN CHLOROPHYLL AND COPEPOD BIOMASS AT VARYING RESOLUTIONS

At the smallest resolution (10 min bins, Figure. 4.3) there was a weak, significant relationship between chlorophyll and copepod biomass (Rs = 0.3, $p < 0.001$, n = 823). The relationship between chlorophyll and copepod biomass was tested at

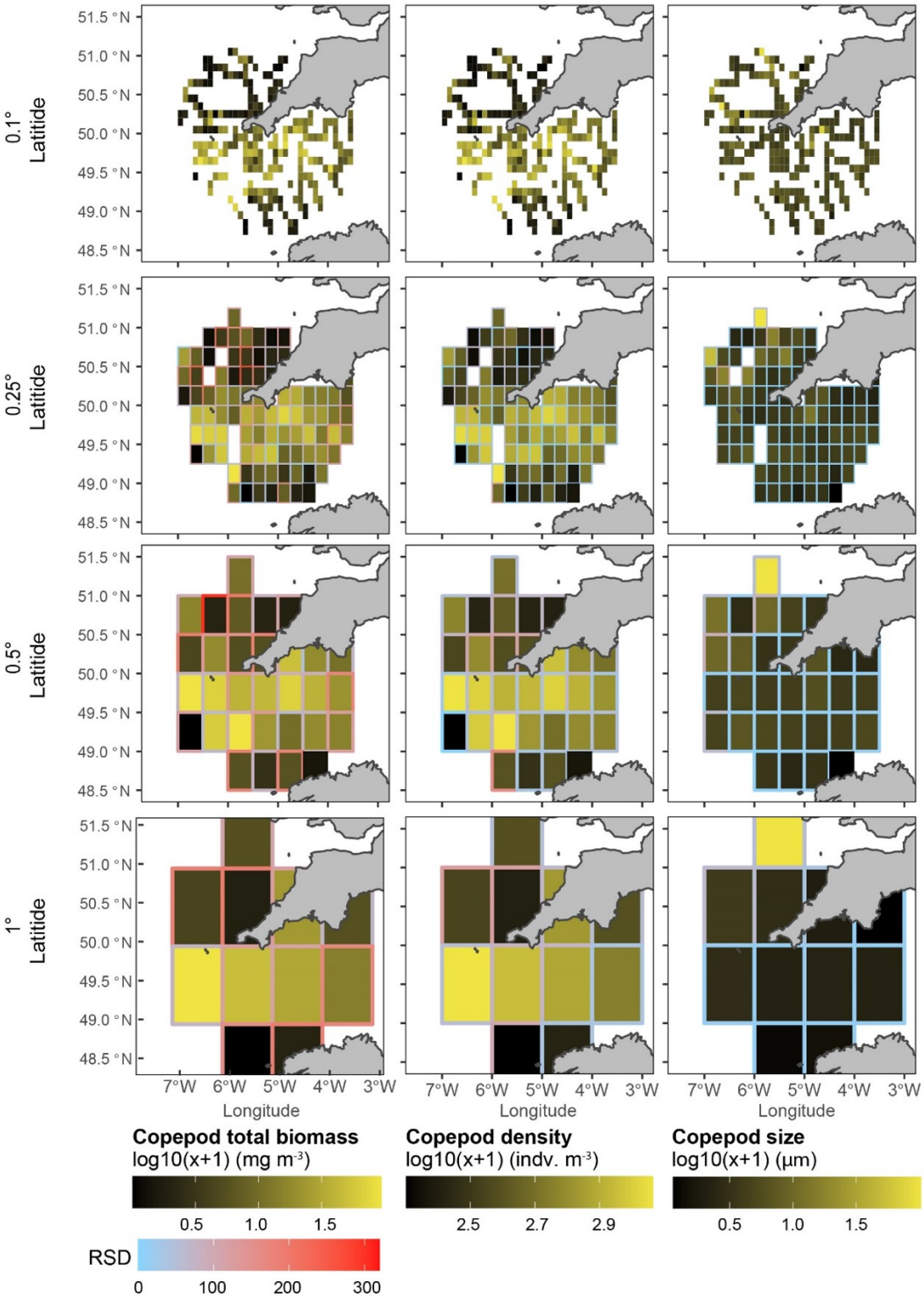

Figure 4.5: 10 minute bins merged to decreasing resolution for (column 1) copepod biomass (mg m^{-3}) (log(x+1)), (column 2) copepod density (indv. m^{-3}) (log(x+1)) and (column 3) copepod size (μm)(log(x+1)) for example resolutions (row 1) 0.1°, (row 2) 0.25°, (row 3) 0.5° and (row 4) 1°. Copepod color scales are the same as Figure 4.3 for comparison. The cell border color indicates relative standard deviation (RSD, %) for the cell. Those cells without a border contain less than 3 data points. RSD color scale is the same for Figure 4.5 and Figure 4.8.

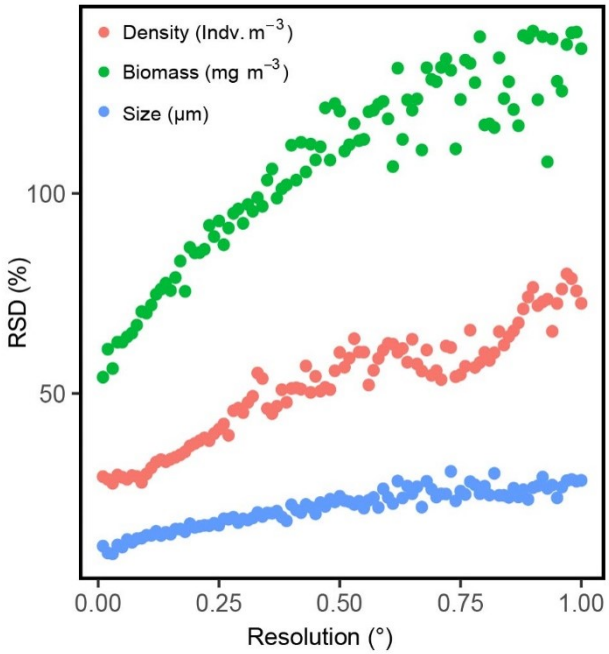

Figure 4.6: Mean cell Relative Standard Deviation (RSD, %) for (green) copepod biomass (mg m^{-3}), (red) copepod density (indv. m^{-3}) and (blue) Size (μm), at all resolutions between 0.05° and 0.9° (increments = 0.01°).

resolutions ranging from 0.05° to 0.9° increasing in steps of 0.01°. The strength of the relationship (Spearmans ρ) and the significance of the relationship are reported in Figure 7. The relationship at the smallest spatial resolution (0.05° x 0.05°) was similar to the ten-minute bin ($p < 0.001$, n = 422). When decreasing resolution from 0.05° to 0.25°, there was little variation in the strength of the relationship and all relationships were significant. For lower resolutions, the strength and significance of the relationship between copepod biomass and chlorophyll became increasingly variable. For example, at a resolution of 0.83° the relationship was not significant and had weak positive correlation (ρ = 0.38, n = 11) while a resolution of 0.84° there was a strong positive correlation and the relationship was significant (ρ = 0.75, n = 11). For resolution lower than 0.9°, there were not enough data points (n < 10) to perform a Spearmans rank analysis (ideally n > 25)

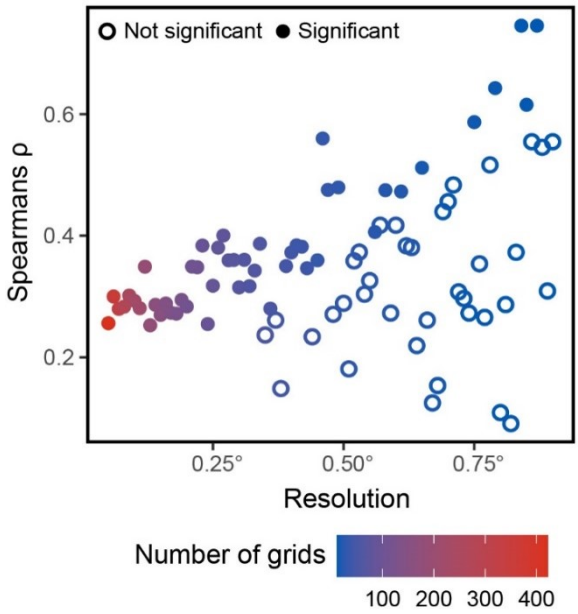

Figure 4.7: The correlation between copepod biomass and chlorophyll using Spearmans ρ against decreasing spatial resolution. Resolutions decrease from 0.05° to 0.9° in increments of 0.01°. The number of grids (grid count) per resolution is indicated by the color scale. The significance of Spearmans correlation (p value < 0.05) is indicated by filled points where non-significant relationships are not filled, and significant relationships are filled.

4.4.4 APPLICATION TO OTHER VARIABLES.

The spatial distribution of chlorophyll concentrations is presented in Figure 4.8 to demonstrate merging of other variables to a decreasing spatial resolution. Chlorophyll concentrations had a broader spatial pattern, where changes occurred over larger distances, than all copepod variables (Figure 4.3**A** - **C**). These patterns are well captured in all selected resolutions (Figure 4.8). There are no small-scale changes in chlorophyll concentration (Figure 4.5) which was reflected in a lower, consistent RSD both spatially and across resolutions (Figure 4.8). The area with the highest chlorophyll concentration, toward the southwest of the study area, also had the highest variation with adjacent cells, which was in turn reflected by a higher RSD (Figure 4.8). Temperature and salinity (Supplementary materials) displayed similar results due to the absence of small spatial changes and broad, slower changes across the study area (Figure 4.8**E**, 4.8**F**).

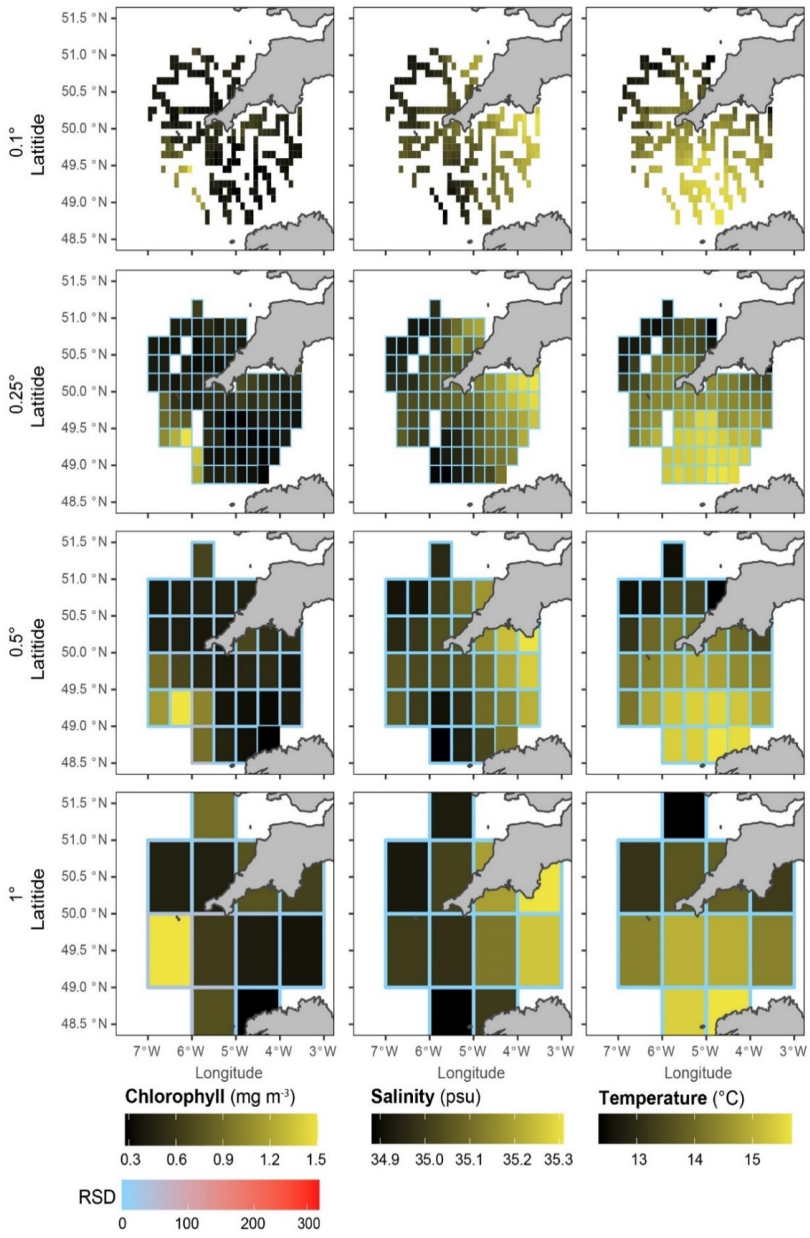

Figure 4.8: 10 minute bins merged to decreasing resolution for (column 1) chlorophyll (mg m^{-3}), (column 2) sea surface temperature (°C) and (column 3) salinity (psu) for example resolutions (row 1) 0.1°, (row 2) 0.25°, (row 3) 0.5° and (row 4) 1°. Variable color scales are the same as Figure 4.3 for comparison. The cell border color indicates relative standard deviation (RSD, %) for the cell. Those cells without a border contain less than 3 data points. RSD color scale is the same for Figure 4.5.

4.5 DISCUSSION

CHANGING SPATIAL RESOLUTION

The use of continuous instruments allowed for data to be obtained at small spatial scales, which in turn captured both wider spatial patterns and small-scale changes in copepod size, abundance and biomass. The small-scale changes in the copepod abundance, indicative of plankton patchiness (Mackas et al., 1985; Abraham, 1998), are not seen in the physical variables where patterns are study area wide. The physical oceanography of the Celtic Sea and Western Approaches, a seasonally stratified area, is well documented (Pingree et al., 1976; Pinnegar et al., 2002; Southward et al., 2004; Smyth et al., 2015). Stratification is known to influence plankton abundances (Fransz et al., 1984; Hure et al., 2022) but the absence of vertical data in this study does not allow for discussion of stratification or its influence on copepod abundance.

Neither surface temperature or salinity appeared correlated with copepod variables. However, the absence of a correlation between zooplankton and physical variables is in line with our understanding that small scale variations in the plankton are driven by a complex series of biological and physical interactions. These were reviewed by Atkinson et al.(2018), using a single point time series (L4 buoy) located in our study area. Average annual densities from the Continuous Plankton Recorder (Richardson, 2008) and reported by Johns (2006) find the majority of copepod families in lower abundance off the North coast of Cornwall. Although our data only cover 1 month, we find a similar spatial distribution, suggesting that the structure of zooplankton communities, within a specific area, remain similar both in time and space. The area wide patterns for copepod densities also match that of a previous study for the region using the PI (Pitois et al., 2018). Despite a lower taxonomic resolution obtained from image identification compared to microscope identification, another study using the PI found that the community structure described the PI is broadly in line with the L4 and CPR (Scott et al., 2021).

As machine learning classifiers for plankton identification from images collected with automated instruments improve in accuracy (The Turing Centre, 2021), it will

be possible to discern zooplankton to increasing taxonomic resolution automatically. Added to the PI, this feature will allow for the removal of the subsampling step that is necessary when manually processing the images. Thus, it will be possible to obtain zooplankton data at higher taxonomic and spatial resolutions, quickly, and at a much lower cost compared to traditional methods. This will be a clear advantage of such systems. Although using the PI has the potential to yield an unprecedented spatial resolution, it cannot replicate the temporal resolution associated with devices such as CPR or longstanding time series such as L4. This is due to the PIs reliance on the vessel on which it's deployed, as it is unrealistic to expect a vessel to survey the same area repeatedly over long periods of time.

4.5.1 OPTIMIZING SURVEY DESIGN

Survey demands often result in ad-hoc, last minute changes reducing assurances of sampling the same spot at a consistent temporal resolution. Thus, a multi-method approach would yield the most complete description of the zooplankton. On the one hand, deploying plankton nets on vessels can help understand the vertical distribution of the plankton, whilst time series are invaluable to understand seasonal and long-term changes (Pitois and Yebra, 2022). On the other hand, understanding the small-scale fluctuations in the plankton, and what drives the high variation between neighboring water parcels, can be better understood using continuous data. Although here, data were subsampled and manually classified which limited the minimum achievable spatial resolution, the findings demonstrate the potential for these instruments to resolve these fine scale interactions driving variation. Furthermore, they demonstrate how the choice of resolution can affect the perceived picture of the plankton as well as relationships between plankton and related variables. Decreasing resolution can result in patterns being emphasized (e.g., chlorophyll) or small-scale changes being lost (e.g., copepod biomass). This demonstrates the 'risk' of a decreased sampling resolution in misrepresenting or incorrectly capturing trends. The variability within cells when merged to a decreasing resolution is not seen by an increased RSD, suggesting RSD is not sensitive to extreme

values if the remainder of the merged cells are consistent. We can expect the changes in the data representation with decreased resolution to be reflected in statistical relationships between variables. Here, we chose to look at copepod biomass and chlorophyll concentrations. Chlorophyll data are readily available as a remote sensing package (Aumont et al., 2015) and many large-scale models rely on these data, inferring prey or carbon from chlorophyll (Landry, 1976; Carlotti and Poggiale, 2010). The relationships between chlorophyll and zooplankton are complex and reported relationships are inconsistent in the literature (Casini et al., 2008; Llope et al., 2012; Schultes et al., 2013; Giering et al., 2019). This variation partly stems from the different types of data and spatial temporal scales used between authors (Pitois et al., 2021). In our study, we find high variation in the strength and statistical significance of the relationship resulting only from changing spatial sampling resolution. Although all correlations are positive, we find both inconsistency in the significance and strength of the correlation at lower resolutions. It is likely that even finer resolution data, achieved through removal of subsampling, will yield the most accurate description of these relationships.

Sampling to the finest possible resolution may not be necessary or relevant to the survey's aim, but rather the choice of resolution, whether in space or time, should match the process studied. A sampling resolution too fine could incur unnecessary costs (in data storage and processing) and not be needed to accurately capture large-scale ecological patterns. For example, a coarser resolution than presented here (2° cells), has been used to successfully capture changes in copepod abundances over time as well investigate their relationship with various physical variables (Bedford et al., 2020). Conversely, too coarse a resolution may miss ecological processes that occur on scales finer than the selected sampling resolution. For example, collecting samples at a specific location once a year (temporal resolution) will not allow to capture seasonal variability.

For our descriptive study we find a spatial resolution of 0.25° to be a good compromise between capturing small scales changes and broader spatial patterns for copepod abundance and biomass. This resolution was the larger end of those resolutions that had consistency in the statistical relationship between copepod

biomass and chlorophyll. Additionally, this resolution can be easily matched to existing products for modelling. For example, a remote sensing ecosystem model output has spatial resolution 0.25° x 0.25° (Aumont et al., 2015). Although each study will likely demand a different resolution dependent on the scale of the processes involved. Let's take, for example, a survey designed to study the timing and location of a specific fish spawning and the ecological processes affecting this (assuming both phyto- and zoo- plankton are variables measured). Prior knowledge will be used to select the overall location and timing of this survey as well as the parameters to measure. At this point, resolution of the measured processes and variables should be taken into consideration. If the survey occurs during the winter months when there is little activity in the plankton ecosystem, sampling these components at very fine temporal and / or spatial resolution is unlikely to be necessary. If, however, that survey occurs during the phytoplankton bloom, a time of fast change within the plankton ecosystem, then a finer resolution that matches the scale of these processes will need to be selected to accurately capture the changes. Similarly, sampling intensity can be adjusted during the course of the survey when and if changes are noticed.

Surveys tend to be designed to collect chosen parameters at preselected locations, usually as many as possible as can be covered by the survey based on time and budget available. In future, as automated tools become common place, optimizing survey design will need to combine different instruments that collect information complementary to each other. For example, automated devices, such as the PI to collect surface data, alongside plankton nets to collect vertical data, would allow a more comprehensive description of the ecosystem studied. Automated tools, and the resolution they yield, may also help to quantify the variability associated with replicate tows resultant from plankton patchiness and be used to better understand its drivers (Wiebe and Wiebe, 1968; Lee and McAlice, 1979; Skjoldal et al., 2013). In theory, this could be achieved with the data presented here, by using a linear regression on the relationship between RSD and Resolution (Figure 4.6). Although, it would be specific to this area and season, a 'survey snapshot' (Huret et al., 2018).

The small-scale resolution presented here, and the future potential for even finer resolution of biological parameters, achieved by reducing or eliminating subsampling, has only become recently possible with devices such as the PI. For this study, the time constraints associated with manually sorting images and the choice to subset the data to the study area meant only a small portion entire survey data were used (2.4 %). This constraint does not apply for other variables, such as physical parameters, where no or little sample processing is required, and all data are instantly available. Recent improvement in classification algorithms (The Turing Centre, 2021) will bring the PI up to speed. Machine learning can eliminate the need for manual classification and thus the PI will be capable of providing continuous data at very fine resolutions (meters and minutes) where data do not need subsampling. Before these solutions can be implemented there are however several challenges that must be overcome, mainly related to the inability of the PI (and other similar systems) to process data as fast as they can collect it. These devices entail a phenomenal data collection rate. For example, if we were to use the full-size range the PI can image, we would collect up to 1 tb of data in less than 10 minutes. This currently not feasible as the technology or protocols to write images this fast does not yet exist. These devices clearly entail a phenomenal data collection rate. While the survey data totaled 2 tb for this study, we have made provision for 10 tb of data for the same survey to take place in 2023. The cost of storage and of compute is reducing but these must also form part of survey planning.

4.6 CONCLUSION

We demonstrate the importance of sampling resolution, in the context of pelagic studies, and how it affects relationships between selected measured parameters and their perceived resulting picture. The increasing use of automated and semi-automated technologies allows us to sample at a much finer resolution than previously possible across much larger spatial scales. This is especially true for zooplankton where the PI has the potential to provide unprecedented fine spatial data at moderate taxonomic resolution. Sampling resolution for each measured process

will therefore need to be considered as part of optimum survey design. This is to ensure that sampling matches the resolution of the measured process at a specific place and time, and that only necessary data is collected to remain within the survey budgetary constraints. Integrating data collected from various instruments (both traditional and novel) will help to optimize sampling resolution for an improved understanding of ecosystem processes and ultimately, a more holistic view of marine ecosystems in all dimensions.

4.7 BIBLIOGRAPHY

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, 391(6667):577–580.

Atkinson, A., Polimene, L., Fileman, E. S., Widdicombe, C. E., McEvoy, A. J., Smyth, T. J., Djeghri, N., Sailley, S. F., and Cornwell, L. E. (2018). Comment. What drives plankton seasonality in a stratifying shelf sea? Some competing and complementary theories. *Limnology and Oceanography*, 63(6):2877–2884.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015). PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, 8(8):2465–2513.

Bean, T. P., Greenwood, N., Beckett, R., Biermann, L., Bignell, J. P., Brant, J. L., Copp, G. H., Devlin, M. J., Dye, S., Feist, S. W., Fernand, L., Foden, D., Hyder, K., Jenkins, C. M., van der Kooij, J., Kröger, S., Kupschus, S., Leech, C., Leonard, K. S., Lynam, C. P., Lyons, B. P., Maes, T., Nicolaus, E. E., Malcolm, S. J., McIlwaine, P., Merchant, N. D., Paltriguera, L., Pearce, D. J., Pitois, S. G., Stebbing, P. D., Townhill, B., Ware, S., Williams, O., and Righton, D. (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. *Frontiers in Marine Science*, 4(AUG):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(6967):661–664.

Bedford, J., Ostle, C., Johns, D. G., Atkinson, A., Best, M., Bresnan, E., Machairopoulou, M., Graves, C. A., Devlin, M., Milligan, A., Pitois, S., Mellor, A., Tett, P., and McQuatters-Gollop, A. (2020). Lifeform indicators reveal large-scale shifts in plankton across the North-West European shelf. *Global Change Biology*, (February):1–16.

Casini, M., Lövgren, J., Hjelm, J., Cardinale, M., Molinero, J. C., and Kornilovs, G. (2008). Multi-level trophic cascades in a heavily exploited open marine ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, 275(1644):1793–1801.

Checkley, D. M., Dotson, R. C., and Griffith, D. A. (2000). Continuous, underway sampling of eggs of Pacific sardine (Sardinops sagax) and northern anchovy (Engraulis mordax) in spring 1996 and 1997 off southern and central California. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 47(5-6):1139–1155.

Culverhouse, P. F., Williams, R., Gallienne, C., Tilbury, J., and Wall-Palmer, D. (2016). Ocean-Scale Monitoring of Mesozooplankton on Atlantic Meridional Transect 21. *Journal of Marine Biology and Aquaculture*, 2(1):1–13.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., David, R., Dell'Anno, A., Dzhembekova, N., Garcés, E., Gasol, J. M., Goela, P., Féral, J. P., Ferrera, I., Forster, R. M., Kurekin, A. A., Rastelli, E., Marinova, V., Miller, P. I., Moncheva, S., Newton, A., Pearman, J. K., Pitois, S. G., Reñé, A., Rodríguez-Ezpeleta, N., Saggiomo, V., Simis, S. G., Stefanova, K., Wilson, C., Martire, M. L., Greco, S., Cochrane, S. K., Mangoni, O., and Borja, A. (2016). Implementing and innovating marine monitoring approaches for assessing marine environmental status. *Frontiers in Marine Science*, 3(NOV):213.

de Moustier, C. (1986). Beyond bathymetry: Mapping acoustic backscattering from the deep seafloor with sea beam. *Journal of the Acoustical Society of America*, 79(2):316–331.

Doray, M., Petitgas, P., Romagnan, J. B., Huret, M., Duhamel, E., Dupuy, C., Spitz, J., Authier, M., Sanchez, F., Berger, L., Dorémus, G., Bourriau, P., Grellier, P., and Massé, J. (2018). The PELGAS survey: Ship-based integrated monitoring of the Bay of Biscay pelagic ecosystem. *Progress in Oceanography*, 166(October 2017):15–29.

Fransz, H. G., Miquel, J. C., and Gonzalez, S. R. (1984). Mesozooplankton composition, biomass and vertical distribution, and copepod production in the stratified central north sea. *Netherlands Journal of Sea Research*, 18(1):82–96.

Giering, S. L., Wells, S. R., Mayers, K. M., Schuster, H., Cornwell, L., Fileman, E. S., Atkinson, A., Cook, K. B., Preece, C., and Mayor, D. J. (2019). Seasonal variation of zooplankton community structure and trophic position in the Celtic Sea: A stable isotope and biovolume spectrum approach. *Progress in Oceanography*, 177(March 2018):101943.

Heath, M. R. (2005). Regional variability in the trophic requirements of shelf sea fisheries in the Northeast Atlantic, 1973-2000. *ICES Journal of Marine Science*, 62(7):1233–1244.

Hure, M., Batistić, M., and Garić, R. (2022). Copepod Diel Vertical Distribution in the Open Southern Adriatic Sea (NE Mediterranean) under Two Different Environmental Conditions.

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. *Progress in Oceanography*, 166(July 2017):30–40.

Johns, D. (2006). The Plankton Ecology of the SEA 8 area. Technical report, Sir Alistair Hardy Foundation for Ocean Science, Plymouth.

Johnson, M., De Soto, N. A., and Madsen, P. T. (2009). Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: A review. *Marine Ecology Progress Series*, 395:55–73.

Johnson, M. P. and Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28(1):3–12.

Kupschus, S., Schratzberger, M., and Righton, D. (2016). Practical implementation of ecosystem monitoring for the ecosystem approach to management. *Journal of Applied Ecology*, 53(4):1236–1247.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Functional Ecology*, 3(1):21.

Lauria, V., Attrill, M. J., Brown, A., Edwards, M., and Votier, S. C. (2013). Regional variation in the impact of climate change: Evidence that bottom-up regulation from plankton to seabirds is weak in parts of the Northeast Atlantic. *Marine Ecology Progress Series*, 488:11–22.

Lee, W. Y. and McAlice, B. J. (1979). Sampling variability of marine zooplankton in a tidal estuary. *Estuarine and Coastal Marine Science*, 8(6):565–582.

Llope, M., Licandro, P., Chan, K. S., and Stenseth, N. C. (2012). Spatial variability of the plankton trophic interaction in the North Sea: A new feature after the early 1970s. *Global Change Biology*, 18(1):106–117.

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J. B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O., Miloslavich, P., Lara-Lopez, A., Kudela, R. M., Lopes, R. M., Karp-Boss, L., Kiko, R., Jaffe, J. S., Iversen, M. H., Irisson, J. O., Hauss, H., Guidi, L., Gorsky, G., Giering, S. L. C., Gaube, P., Gallager, S., Dubelaar, G., Cowen, R. K., Carlotti, F., Briseño-Avena, C., Berline, L., Benoit-Bird, K. J., Bax, N. J., Batten, S. D., Ayata, S. D., and Appeltans, W. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6(MAR).

Mackas, D. L., Denman, K. L., and Abbott, M. R. (1985). Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science*, 37(2):653–674.

Mann, D. A., Hawkins, A. D., and Jech, J. M. (2008). Active and Passive Acoustics to Locate and Study Fish. In *Fish Bioacoustics*, pages 279–309. Springer.

Olson, R. J., Zettler, E. R., and DuRand, M. D. (2018). Phytoplankton analysis using flow cytometry. In *Handbook of methods in aquatic microbial ecology*, pages 175–186. CRC Press.

Owen, K. R. (2014). Flow Cytometric Investigation of the Size Spectrum of North Sea Phytoplankton Communities. (May):272.

Petersen, W. and Colijn, F. (2017). FerryBox Whitebook.

Pingree, R. D., Holligan, P. M., Mardell, G. T., and Head, R. N. (1976). The influence of physical stability on spring, summer and autumn phytoplankton blooms in the celtic sea. *Journal of the Marine Biological Association of the United Kingdom*, 56(4):845–873.

Pinnegar, J. K., Jennings, S., O'Brien, C. M., and Polunin, N. V. (2002). Long-term changes in the trophic level of the Celtic Sea fish community and fish market price distribution. *Journal of Applied Ecology*, 39(3):377–390.

Pitois, S. and Yebra, L. (2022). Contribution of marine zooplankton time series to the United Nations Decade of Ocean Science for Sustainable Development. *ICES Journal of Marine Science*, 79(3):722–726.

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). *Ecological Indicators*, 123:107307.

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

Postel, L., Fock, H., and Hagen, W. (2000). Biomass and abundance. In *ICES Zooplankton Methodology Manual*, pages 83–192. Elsevier.

Richardson, A. J. (2008). In hot water: Zooplankton and climate change. *ICES Journal of Marine Science*, 65(3):279–295.

Schultes, S., Sourisseau, M., Le Masson, E., Lunven, M., and Marié, L. (2013). Influence of physical forcing on mesozooplankton communities at the Ushant tidal front. *Journal of Marine Systems*, 109-110(SUPPL.):S191–S202.

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.

Simmonds, J. and MacLennan, D. N. (2008). *Fisheries acoustics: theory and practice*. John Wiley & Sons.

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. *Progress in Oceanography*, 108:1–42.

Smyth, T., Atkinson, A., Widdicombe, S., Frost, M., Allen, I., Fishwick, J., Queiros, A., Sims, D., and Barange, M. (2015). The Western Channel Observatory. *Progress in Oceanography*, 137:335–341.

Southward, A. J., Langmead, O., Hardman-Mountford, N. J., Aiken, J., Boalch, G. T., Dando, P. R., Genner, M. J., Joint, I., Kendall, M. A., Halliday, N. C., Harris, R. P., Leaper, R., Mieszkowska, N., Pingree, R. D., Richardson, A. J., Sims, D. W., Smith, T., Walne, A. W., and Hawkins, S. J. (2004). Long-term oceanographic and ecological research in the western English Channel. *Advances in Marine Biology*, 47:1–105.

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-Sea Research Part I: Oceanographic Research Papers*, 49(8):1445–1461.

Steinberg, D. K. and Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science*, 9(1):413–444.

Strickland, J. D. and Parsons, T. R. (1972). 1972. A practical handbook of seawater analysis. 2nd ed. *Bull. Fish. Res. Bd.*, Can:167.

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

The Turing Centre (2021). DSG Report 2021. Technical report.

Vinet, L. and Zhedanov, A. (2011). A 'missing' family of classical orthogonal polynomials. *Journal of Physics A: Mathematical and Theoretical*, 44(8):22–142.

Wiebe, P. H. and Benfield, M. C. (2003). From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography*, 56(1):7–136.

Wiebe, P. H. and Wiebe, P. H. (1968). Plankton Patchiness: Effects on repeated net tows. *Limnology and Oceanography*, 13(2):315–321.

Working Group of International Pelagic Surveys (2015). Manual for International Pelagic Surveys (IPS). Technical report, ICES, Copenhagen.

5

HARMONISING CONTINUOUS ZOOPLANKTON AND FISHERIES DATA TO REEVALUATE THEIR RELATIONSHIPS

The following chapter aims to compare a new method with an existing published article and thus has large overlap in the methodology, see Published Works.

5.1 ABSTRACT

Zooplankton, specifically copepods, are the principal prey for the majority of commercial pelagic fish. Their relationship with fish is difficult to statistically describe which stems from the complex range in behaviours, spatial distributions and interactions between the zooplankton and fish. We used continuous zooplankton and fish data collected concurrently in the Celtic Sea to investigate these relationships at fine spatiotemporal resolutions. Zooplankton data were collected using the Plankton Imager and all images were sorted using a machine learning classifier. Fish data were collected using acoustics. Data were harmonised to a minimum spatial resolution of 1 nautical mile (1.85 km). There was good overlap of the datasets. High variability in zooplankton size and abundance was seen over short spatial and temporal scales, indicative of plankton patchiness. Fish spatial distributions varied between species. Horse Mackerel, Anchovy and Sardine were found throughout the study area. The following speices where found only in specific areas: Boarfish (south of Cornwall and west of wales); Herring (north of Cornwall and west of Wales) and Sprat (north west of France).

Anchovy and Horse Mackerel had a negligible correlation with copepod abundance and size. Boarfish and Sardine had a weak positive correlation with copepod abundance and size. An unexpected, negative correlation was found between Sprat and Herring and copepod abundance and size, contrasting the literature. The correlation with Sprat was likely due to a mismatch between data coverage and Sprat behaviour, where a known spawning and feeding area (Lyme Bay) was absent from the zooplankton data. Sampling at different spatiotemporal resolutions did not have significant effect on the strength of correlations for those relationships tested. This demonstrates that sampling to the finest spatiotemporal resolution may not always be appropriate. Instead time series and larger spatiotemporal resolutions may provide further insight into these relationships.

5.2 INTRODUCTION

Zooplankton occupy a critical position within pelagic food webs, providing a crucial carbon link between primary producing phytoplankton and larger planktivores. They are the principal prey for a range of planktivorous fish species (Confer and Blades, 1975; Beaugrand et al., 2003), seabirds (Pakhomov and McQuaid, 1996; Lauria et al., 2012) and marine mammals (Sims and Quayle, 1998; Sims, 1999). They are sensitive indicators of climate change (Taylor et al., 2002) with climate change induced changes in the plankton demonstrated to cascade up the trophic chain (Lauria et al., 2012; Pitois et al., 2012). The difficulty in understanding these relationships in part originates from the wide ranging scales seen by both prey and predator, such as plankton patchiness (Mackas et al., 1985; Abraham, 1998) or behavioural patterns in fish such as feeding. As a result, modelling the relationship between zooplankton and fish in 'end to end' ecosystem models or individual fisheries models for stock assessment remains a major challenge (Travers et al., 2007; Carlotti and Poggiale, 2010).

The ability to continuously infer fish biomass from acoustic data to a fine spatial resolution is an established technique (Working Group of International Pelagic Surveys, 2015; Van Der Kooij et al., 2016). Until recently, achieving a similar sampling frequency for mesozooplankton has been impossible. Technological advances (for a review see Lombard et al., 2019) have allowed zooplankton sampling at an unprecedented temporal and spatial resolution and in turn brought about new data challenges. Devices such as the Plankton Imager (PI), used here (Pitois et al., 2021; Scott et al., 2021), Video Plankton Recorder (Davis et al., 1992) and Shadowed Image Particle Profiling and Evaluation Recorder (SIPPER) (Samson et al., 2001) have the capacity to collect millions of images over a short space of time. The challenge is in identifying these images (Culverhouse et al., 2006; Benfield et al., 2007; MacLeod et al., 2010) and understanding how to best use these types of data. Sorting these images manually requires similar methods to those associated with traditional sampling (e.g. splitting a sample to an achievable size) and can provide ecologically consistent data for analysis (Gorsky et al., 2010; Scott et al., 2021). However, these methods do not

take full advantage of the high volume of available data and share similar problems to traditional methods associated with subsampling (Ohman and Lavaniegos, 2002). Machine learning provides a solution for sampling all images quickly and at a reduced cost (Grosjean et al., 2004; Fernandes et al., 2009; Faillettaz et al., 2016).

A recently developed classification algorithm, developed in collaboration with the Alan Turing Institute (for a detailed description of the algorithm see Section 2.6.2, (The Turing Centre, 2021) allows for application of a machine learning classifier to Plankton Imager (PI) images. The PI is a continuous particle imager which take images of all particles in a flow-through system connected to the ship's water supply (for a more in depth description, see Section 2.3). The PI is capable of collecting > 70 million images in a month long survey. In the previous chapters, subsampling was used to reduce the dataset to a number of images achievable by manual classification. Using a machine learning classifier allows us to move away from subsampling or point sampling, thus making the most of the entire set of images.

Here we aim to follow up on the work by Pitois et al. (2021) on linking a 2-dimensional ecological indicator (Copepod Mean Size and Total Abundance - CMSTA) to higher tropic levels, using pelagic fish biomass. (CMSTA) indicator is derived from the HELCOM Mean Size and Total Stock (MSTS) indicator. It is a two-dimensional indicator describing the relationship between mean zooplankton (as copepod) size and total abundance (Pitois et al., 2021; Gorokhova et al., 2013). These pelagic fish are prey for larger piscivorous fish (Trenkel et al., 2005), marine mammals and seabirds (Kaschner et al., 2006) and have significant economic value as commercial fisheries (Pinnegar et al., 2002; Rochet et al., 2010). Pitois et al. (2021) found that Herring was the only fish positively correlated with copepod mean size, being found where copepods were larger but not necessary more abundant. No other statistically significant correlations were found between any copepod variable and any pelagic fish. The data produced by using the machine learning algorithm presents a unique opportunity to apply the same method of analysis using much larger and finer resolved data. This will allow for the relationships between fish and copepods to be re-evaluated and at the same time explore the benefits of using continuous data.

5.3 METHODS

All data were collected in the Celtic Sea from the 3rd of October to the 7th of November 2020 aboard the RV Cefas Endeavour as part of the PELTIC Survey (PELagic ecosystems in the Western English Channel and eastern celTIC seas) (Working Group of International Pelagic Surveys, 2015) (Figure 5.2). Plankton data were recorded 24/7 and collected using the Plankton Imager (PI) (Pitois et al., 2018). Fish biomass was inferred from acoustic data. Although the acoustic data are continuous, only daylight hours were processed.

5.3.1 PLANKTON IMAGER

All zooplankton data were collected using the PI. For detailed methods on the PI operational usage see Section 2. During the Peltic Survey a total of 71 million images were collected. Here, the raw data were subset to align with fish data and to reduce the data to an achievable, analysable quantity in terms of compute time. In line with fisheries data, only daylight hours were processed. This was achieved using the suncalc R package (Thieurmel et al., 2019) which provided sunrise and sunset for each survey day resulting in easy separation of the day data. The speed of the classifier also required further reducing the data size, as it was not possible to sample all daylight images. Data were filtered to a maximum ten minute bin size of 44,616 images (see Section 2.3.5 for details on raw data storage). This value was the third quartile of the data (Figure. 5.1) and used to avoid periods of high particulate matter or phytoplankton content. The target images were converted using Raw2RGB (Section 2.4.1) as required by the Turing classifier (Section 2.6.2). The classifier was then run using a looped python script to classify the images. Finally, the classifier output for each 10 minute bin was merged to produce the final dataset.

Images were classified to two categories: copepod and non-copepods, where the latter also included detritus. We assumed the error in copepod classification by the Turing classifier is consistent across all data. The accuracy of the classifier was 97%,

this is detailed in Section 2.6.2. Abundance, size and biomass were calculated as per Pitois et al. (2021). For a detailed calculation of these variables see Section 2.4.3.

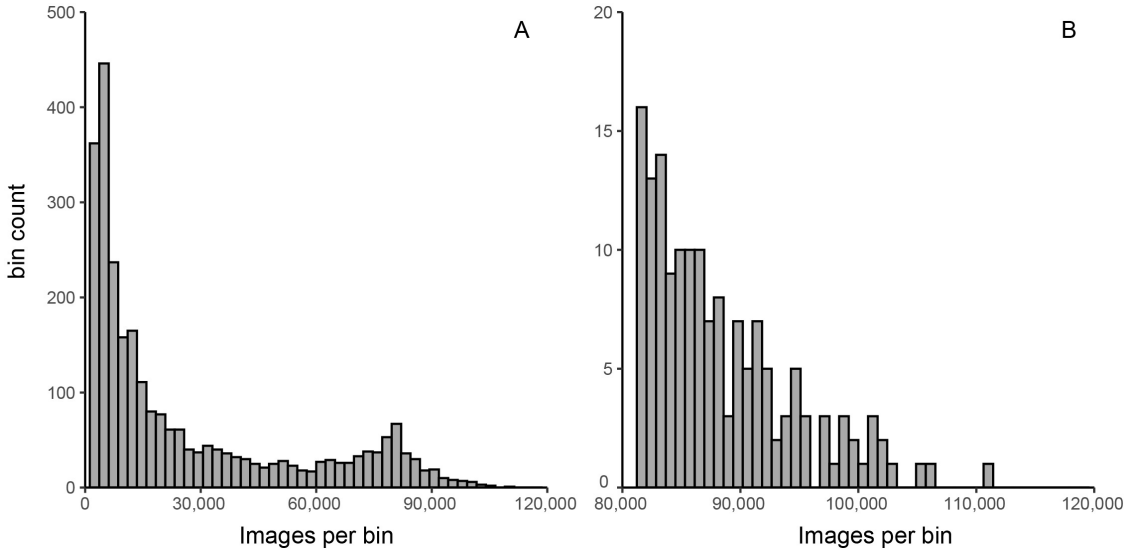

Figure 5.1: Number of images per 10 minute bin for the 2020 Peltic Survey. Data were captured over 1 month. All data are shown in panel **A**. Only bins with > 80,000 images are shown in panel **B** as these are indistinguishable in panel **A**. The Y-Axis changes between panels.

5.3.2 PELAGIC FISH BIOMASS ESTIMATES

The exact same method for extracting fish acoustic data were used as in Pitois et al. (2021). See *Published Works*. The following is a quote from the publication:

> *Acoustic data were collected along transects during the day, using a Simrad EK60 scientific echosounder, with the split-beam transducers mounted on the vessel's drop keel at a depth of 3.2 m below the vessel's hull or 8.2 m sub surface. Three operating frequencies were used during the survey (38, 120 and 200 kHz) for trace recognition purposes, with 38 kHz data used to generate the abundance estimate for clupeids (and other fish with swimbladder) and 200 kHz for Horse Mackerel (Van Der Kooij et al., 2016). All frequencies were calibrated at the start of the survey.*
>
> *A pelagic midwater trawl with a vertical opening of c. 12 m was used to collect information on species and size composition and provide biological samples, and was fitted with a 20 mm codend liner to ensure*

the retention of small and juvenile fish. Trawl monitoring, trawl door type and dimensions and rigging are described in (Working Group of International Pelagic Surveys, 2015). As the trawls were deployed to obtain qualitative rather than quantitative information, no fixed trawl duration was employed during the survey although deployment was generally 30 min, with haul targeted on schools. All components of the catch from the trawl hauls were sorted and weighed; fish and other taxa were identified to species level. Length frequency and individual length-weight data were collected for all species of the catch. Total length measurements of Sprat Sprattus Sprattus, Sardine Sardina pilchardus, anchovy Engraulis encrasicolus, boarfish Capros acer and Herring Clupea harengus were to be taken to the nearest 0.5 cm below, Horse Mackerel Trachurus trachurus were measured to the whole cm below. Where possible the total catch component of the haul per species was measured. When this was not, a representative sub-sample was taken, species identified, and lengths obtained to provide a true (length) representation of the species.

Biomass estimates for pelagic fish species followed routine methods (Working Group of International Pelagic Surveys, 2015). The acoustic recordings of Nautical Area Scattering Coefficient (NASC, m^2 *nmi^{-2}) for each nautical mile along the transects were partitioned by species based on school characteristics and trawl catches. To determine the underlying spatial distribution of pelagic fish species, statistical models (Generalized Additive Models, GAMs) were employed using physical covariates as predictors (i.e. latitude/longitude/depth/distance from coast). Analysis of covariance between predictors using the Variance Inflation Factor indicated that depth and distance from coast were strongly correlated (VIF > 2). Similarly, depth and longitude where strongly correlated so models were built for fish species using latitude/longitude/distance from coast only (all VIF < 2).*

5.3.3 SPATIOTEMPORAL ALIGNMENT AND ANALYSIS

The main challenge was designing a program to spatiotemporally align the copepod and fish data while allowing the flexibility to treat space and time as independent variables. This was required as the ship often repeated transects or crossed a previous track. Therefore, it was not possible to simply merge all datapoints within a chosen spatial subset (e.g., a grid cell of 0.25° by 0.25° as done in Section 4) as there would have been instances where the data range over 1 month, for example, when the ship was steaming back to port crossing a transect from the start of the survey. Although, this method was used to describe overall spatial trends of the data across the survey month it was not used in any statistical analyses. The program for spatially aligning the data was set up to accept user chosen spatial resolution (in degrees) and temporal resolution (in minutes). The program has no upper or lower bounds on these parameters and is only bound by the compute speed of the machine or the maximum data resolution. For this study, the minimum resolution was set by the fisheries data, where data were resolved to 1 nautical miles (1.8 km).

Following alignment, the program was used to experiment with how varying spatial and temporal resolution can alter statistical relationships. The program was looped through permutations of varying spatial (0.1, 0.2, 0.25, 0.4, 0.5, 0.75, 1 deg) and temporal resolutions (20, 30, 40, 50, 60 minutes) to find the strongest relationship between copepod size and abundance and fisheries biomass. The strength of the relationship was determined using Spearmans ρ. After determining the best spatiotemporal fit that specific resolution was used to align the data and produce 2D plots to examine the influence of copepod size and abundance on fish biomass. The code is published in Appendix B).

5.4 RESULTS

5.4.1 DATA ALIGNMENT

The incident of aligned fish and copepod data depended upon the chosen spatiotemporal resolution and increased with decreased resolution. Figure 5.2 shows the spatial data alignment at an example 0.1 by 0.1 degree resolution. Fish data were more frequently available than copepod data (Figure 5.2). At this example resolution, fish data were present in 247 (fish only) + 238 (copepods and fish) grid cells whereas copepod data were present in 74 (copepods only) + 247 (copepods and fish) grid cells (Figure 5.2). Grid cells without either data are not shown.

In Lyme Bay area, the PI was being used for experiments with the size data and data rates to determine a compromise between storage space and size range (see Section 2.5.1). Based on experience from previous years the PI was turned off in the Bristol Channel to avoid flooding the storage drive with images of detritus. The detritus prevalence is particularly high due to freshwater input and estuarine mixing.

5.4.2 SPATIAL DISTRIBUTION

Figure 5.3 describes the spatial distributions of copepod abundance and size and fish biomass. All present data were merged independently of the temporal scale and therefore do not account for multiple vessel entries into the same grid cell. There are no grey cells present for copepod abundance or sizes as when the PI was recording (Figure 5.3), copepods were found in all 0.25° cells.

The spatial distribution of copepod abundance followed no obvious basin wide pattern and there was high variation over small spatial scales across the study area (Figure 5.3). There was marginal consistency in lower abundance off the coast of Wales and the north coast of France. Copepod length had a more uniform distribution with the exception of coastal waters off the south coast of the Cornwall where an area of larger copepods (geomean length > 400 μm) was present. There was no obvious relationship between size and abundance.

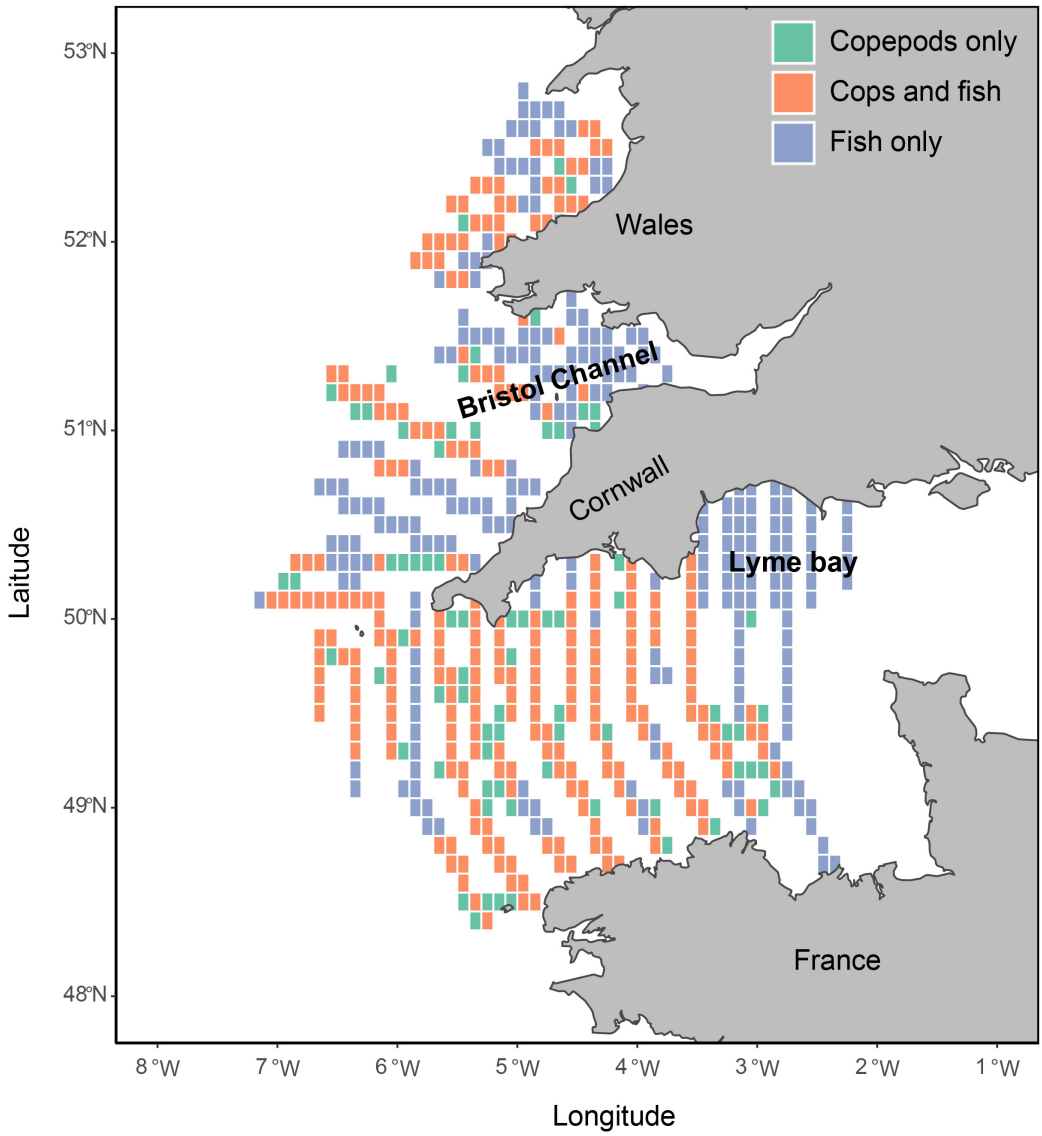

Figure 5.2: Spatial alignment of variables at an example 0.1° to show those grid cells where copepod data, fish data or copepod and fish data were present. Where cops = copepod. Note there is high variability within grid cells at this resolution.

There was high variation between basin wide fish spatial distributions (Figure 5.3). Anchovy, Horse Mackerel and Sardine were found throughout the study area (Figure 5.3). Sardine was found in highest abundance below the extents of Cornish peninsula whereas anchovy had a more uniform distribution. Neither anchovy or sardine were found in high abundance in the most seaward, south westerly cells. Sprat were only found toward the north of the study area and in Lyme Bay. Boarfish were found in the

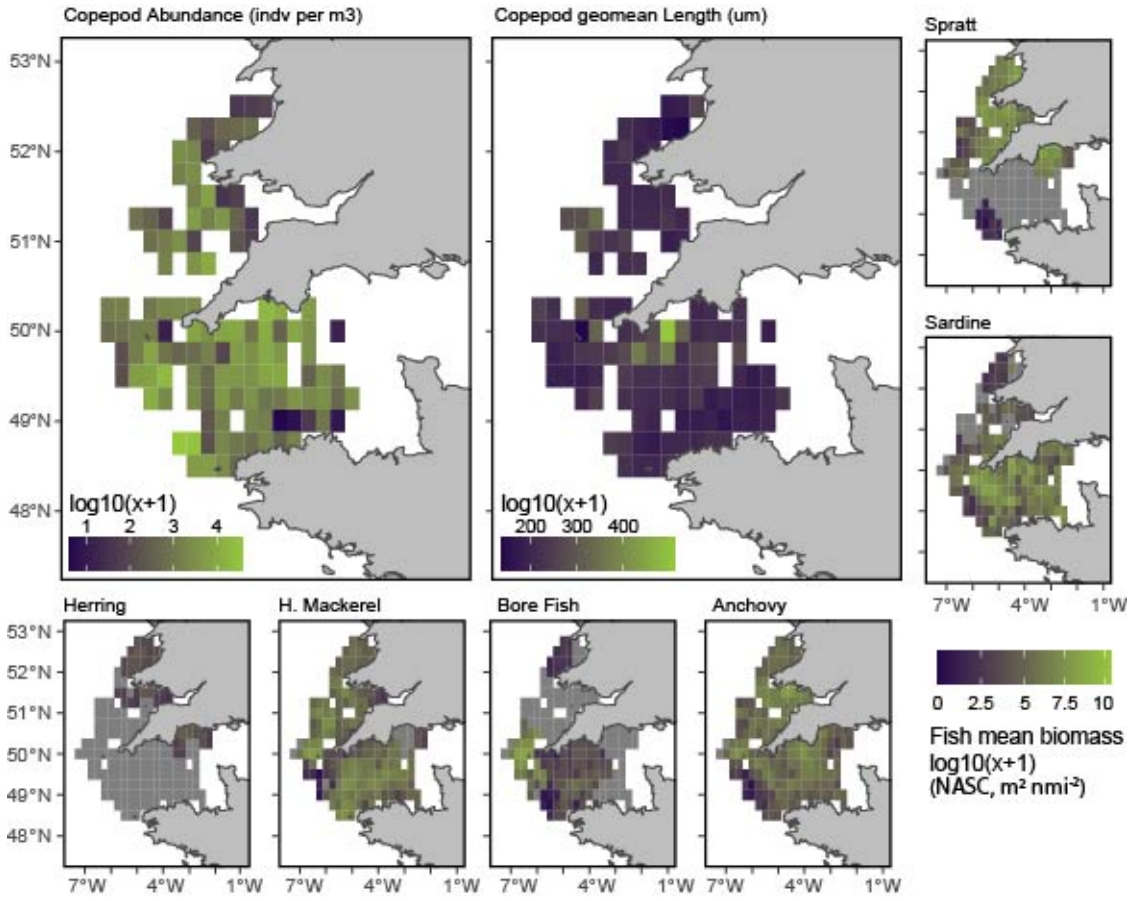

Figure 5.3: Spatial distribution of all copepod and fish variables across the study area where data were averaged over 0.25° grid cells. All data except for copepod geomean length were logged (log(x+1)). The temporal component of our data was ignored. Grey cells indicate 0 for fish data. The fish biomass legend (bottom right) is common to all fish.

bottom half of the study area, below Cornwall and in high abundance at the most westerly, seaward sites. Herring had the lowest overall biomass of all fish species present and was only found of the west coast of Wales. With this exception, the remainder of the fish were found with similar overall biomass across the study area. There appears no obvious spatial correlation between any of the fish species or either copepod abundance or size when using only spatial data merging. There is reduced overlap of copepod data with Sprat and Herring (Figure 5.3).

5.4.3 RELATIONSHIP BETWEEN VARIABLES

Anchovy and Horse Mackerel showed negligible correlation with copepod abundance (ρ = 0.15 and ρ = -0.1, respectively, Table 5.1) and size (ρ = 0.18 and ρ = -0.12,

respectively, Table 5.1). Horse Mackerel was the only fish with a non-significant correlation with either copepod size or abundance (Table 5.1). Boarfish had a weak positive correlation with copepod abundance (ρ = 0.37, p < 0.001, Table 5.1, Figure 5.4). For all fish with a moderate or stronger (ρ > 0.25) negative or positive correlation with copepod abundance, the correlation was about 50% of the ρ for size. Sardine had a strong positive correlation (ρ = 0.47, p < 0.001, Table 5.1, Figure 5.5) with copepod abundance whilst Sprat and Herring both had a negative correlation (ρ = -0.4 and -0.37, respectively, p for both < 0.001, Table 5.1). There was high variability between fish in which spatiotemporal scale yielded the strongest correlations, although the actual correlation only varied marginally across resolutions (Figure 5.4). Sprat had the strongest correlation at the finest spatial resolution (0.1°). For Sardine and Horse Mackerel the strongest correlation was found at 1 °. The temporal resolution yielded the strongest correlation for all fish was either 40 or 60 minutes with the exception of Horse Mackerel. The strongest correction between Horse Mackerel and copepod biomass was at 20 minutes (Figure 5.4).

Every species was strongest correlated with size at a different resolution than abundance except for Sardine where 60 minutes and 1 degree yielded both the strongest correlations (Figure 5.4 and 5.5).

Fish	Abundance (ρ)	Abundance (p)	Size (ρ)	Size (p)
Anchovy	0.15	**0.04**	0.18	**<0.001**
Boarfish	0.37	**< 0.001**	0.10	**0.01**
Herring	-0.53	**<0.001**	-0.28	**<0.001**
Horse Mackerel	-0.10	0.08	-0.12	0.12
Sardine	0.47	**<0.001**	0.18	**0.03**
Sprat	-0.53	**<0.001**	-0.26	**<0.001**

Table 5.1: Spearmans ρ for relationship between fish and copepod size (column 4) and abundance (column 2). Corresponding p-values are shown in the next column and those significant (p < 0.05) are show in bold. These are the strongest correlations from various spatiotemporal scales (see Figure 5.4 and Figure 5.5).

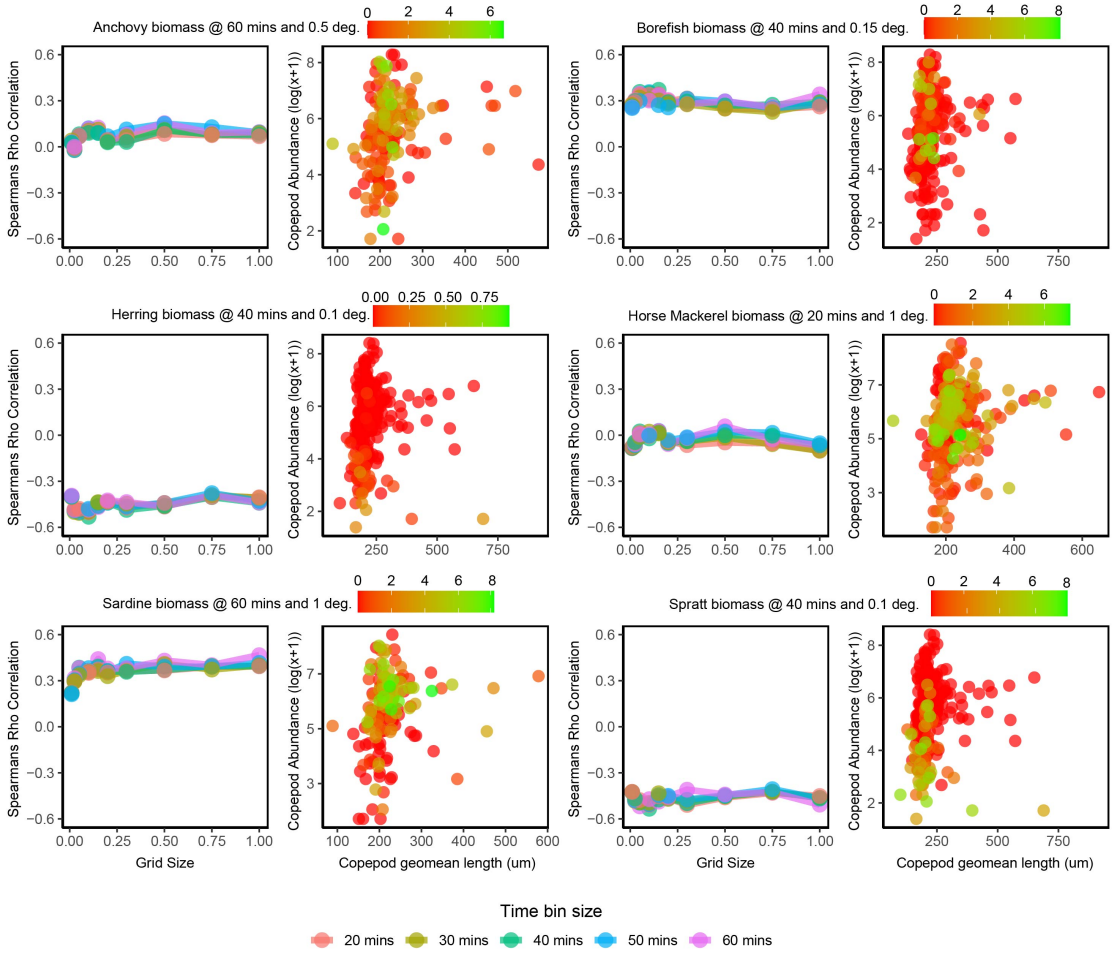

Figure 5.4: Copepod abundance and fish biomass dual plots for each fish species. The left most plot for each species shows the varying strength of correlation (Spearmans ρ) between copepod abundance and fish biomass with changing spatiotemporal resolution. The resolution with the strongest positive or negative correlation is shown atop the figures (e.g., Sardine @ 60 minutes and 1 degrees). This resolution was used to create the species scatter plot (right panel per species). This plot shows a scatter graph where points represent a grid cell and are coloured by the fish biomass (NASC, m^2 nmi^{-2}) (log10(x+1)). The raw data were averaged within grid cells then correlated. The legend colour scale varies for each fish. Points are stacked from highest fish biomass to lowest.

5.5 DISCUSSION

Plankton patchiness (Mackas et al., 1985; Abraham, 1998) was evident in spatial results. Similarly, to the findings of Section 4, fine spatial changes in copepod abundance, indicative of plankton patchiness, were lost with decreasing spatial resolution. Sampling all images allowed for analysis for the variation in plankton (plankton patchiness) over temporal scales within a single grid cell. We find that

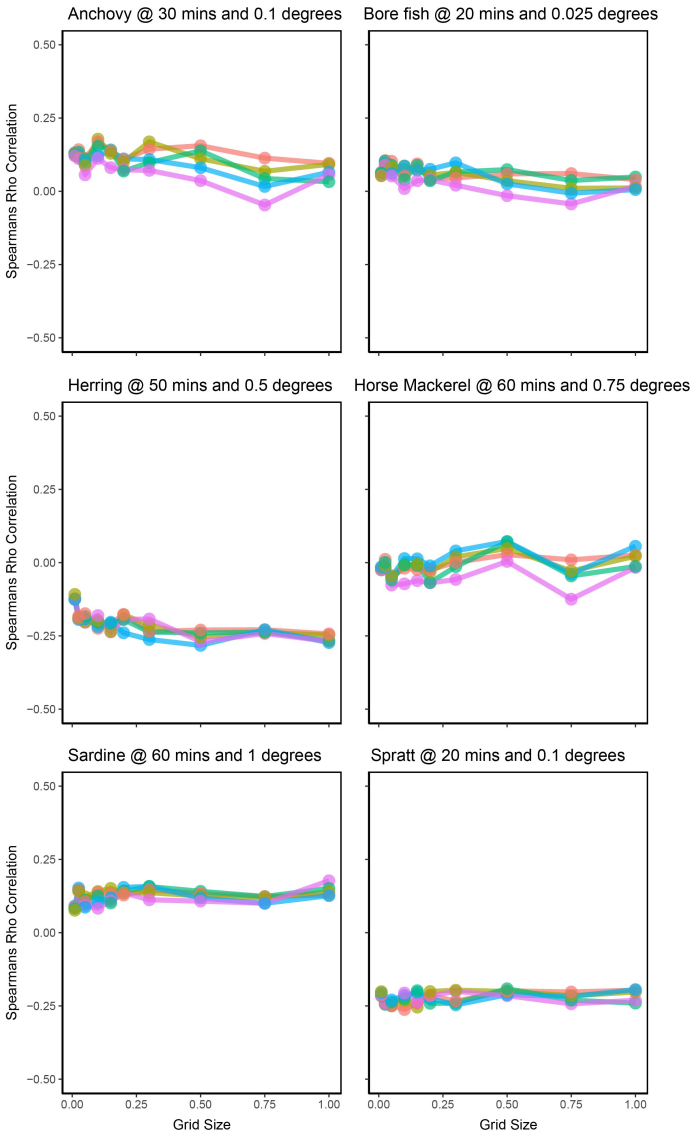

Figure 5.5: Copepod size and fish biomass correlation plot for each fish species. Each plot shows each species varying strength of correlation (Spearmans ρ) between copepod size and fish abundance with changing spatiotemporal resolution. he resolution with the strongest positive or negative correlation is shown atop the figures (e.g., Anchovy @ 20 minutes and 0.1 degrees). For scatter plots see Figure 5.4.

variation in copepod abundance was not necessarily a function of time, with large changes in copepod abundances over small time scales, which aligns with high variability seen between replicate new tows (Wiebe and Wiebe, 1968).

We had good overlap between fish and plankton data. This allowed for investigation of relationships between fish and copepods at different spatiotemporal resolutions. The program written, (Appendix B) to align the data, accepted spatial

resolutions as small as meters and temporal resolutions as small as seconds, but available computer power limited the minimal spatial and temporal resolutions to 0.01° and 20 minutes. In future, as technology progresses to allow faster processing speed, these smaller scales can be resolved. The spatial resolution was also limited by the binned fish data (1 nautical mile), although these can be resolved to a finer resolution for future studies. For all spatiotemporal combinations, correlations between fish and copepods variables did not change significantly across combinations. This contrasts with findings from Section 4 where changes in spatial resolution had a considerable effect on the statistical correlation between chlorophyll and copepod biomass. This suggests that the ecological processes that govern the relationships between copepods and fish occur over greater spatial scales than those between copepods and phytoplankton (Section 4, Figure 4.7) and the complexity of the factors driving chlorophyll, zooplankton and fish are difficult to capture through examining a single 'snapshot' (Huret et al., 2018).

Independent of correlation type (positive, negative or neutral), we find all relationships between fish species (excluding Mackerel) and both copepod size and abundance, to be significant. This contradicts the results from Pitois et al. (2021) who found a correlation between Herring biomass and copepod size, but none between other fish species and copepod variables (both size and abundance). We used the same data collection methods (PI for copepods and acoustics for fish) and analytical techniques (Spearman's ρ). The only difference resides in the number of zooplankton images sampled and therefore the number of data points available for statistical testing. The strength and consistency of significant correlations (many reported relationships had $p < 0.001$) may arise from the number of stations and choice of statistical analysis as Spearman's ρ is sensitive to dataset size with a preferable minimum of 15 data pairs. If more data had been available for Pitois et al. (2021) they may have been more of a chance in finding significant relationships. This may be the reason for the higher probability values, and thus confidence in our findings. Plankton patchiness, not visible by point sampling used by Pitois et al. (2021), may also distort the relationships we investigated here. In Section 4 we find that large, fine scale changes in the copepod abundance are lost when merging data to a fixed spatial

resolution. The loss of these extreme values when average, which could be areas with high fish and plankton density, may be in part responsible for some of the negative correlations we find here.

The positive correlation between copepod size and Herring presented in Pitois et al. (2021, $\rho = 0.51$, $p = 0.01$) is supported by the literature (Flinkman et al., 1998; Corten, 2000). Conversely, we find Herring to be negatively correlated to size and abundance, although less so with abundance. We also find Sprat and Horse Mackerel to be both negatively correlated with size and abundance. It is difficult to speculate as to why we find the negative correlations, as copepods are known to be the primary food source for most pelagic fish (Confer and Blades, 1975; Garrido et al., 2008; Reid et al., 2001). Prey size is also known to be important in driving in prey selection as well as controlling food quality and availability (Pitois and Fox, 2006; Barton et al., 2013; Van Deurs et al., 2015). Sardine, Sprat and Horse Mackerel have similar feeding apparatus meaning they have the ability to preferentially switch between filter feeding to particulate feeding based on prey availability (Garrido et al., 2008). Another explanation may be a temporal mismatch between our data and peak fish feeding. Although this is unlikely as species such as Sprat and Herring are known to be feeding in autumn to over winter (Patel et al., 2022; Capuzzo et al., 2022). Another potential explanation is the hypothesis that these species are not limited by food (Pitois et al., 2021) and are therefore not changing their behaviour to seek food. There is food abundant enough in all locations meaning other factors are driving their spatial distribution. The weak negative correlations between fish biomass and copepod size may also arise from the the relatively uniform spatial distribution of copepod size across the study area (also demonstrated with reduced spatial resolution in Pitois et al., 2021, and Section 4) resulting from a similar distribution of species across the study area (Johns, 2006; Scott et al., 2021). These negative correlations with size, suggesting that these species preferentially eat smaller plankton, is at odds with the published literature. An unlikely hypothesis is that solely top-down control, known to be a governing factor for zooplankton (Lynam et al., 2017; Reid et al., 2000), is dominating the processes and the reasons for our finding few and smaller copepods is due to preferential predation from higher species. In reality a suite of processes

govern zooplankton dynamics, review by Atkinson et al.. The most relevant are the the loophole hypothesis, whereby physiochemical changes favour some taxa (Irigoien et al., 2005), mortality-controlled copepod phenology (Irigoien and Harris, 2003; Maud et al., 2015) and zooplankton feeding traits (Sailley et al., 2015). These processes could be combined with top-down or bottom-up effects as well as fish behaviours such as migration or speeding only in specific areas and result in the weak spatial corrections presented here.

Unfortunately, the possibility exists that the stronger, negative correlations, particularly those seen in Sprat and Herring, are coincidental artefacts and not a true representation of ecological relationships. These correlations may result from a spatial misalignment between fish behaviours and our data, a limitation of a 'survey snapshot' where variables collected concurrently are too variable and longer time series may be more informative (Huret et al., 2018). For example, fish are feeding in an area where we have no data. Conversely, it may mean that for those areas with a negative correlation fish are moving between target feeding or spawning areas. Lyme Bay (Figure 5.2) is home to a consistent, resident population of Sprat (ICES, 2021). We have no data for Lyme Bay and the negative correlation we find between Sprat and copepod abundance elsewhere in the study area suggests that Sprat are transient in these areas, moving between areas such as Lyme Bay for spawning and feeding. On future surveys, now the PI is established, these areas will be incorporated in the survey.

Although these findings contribute to the hypothesis that fish behaviour, such as spawning and nursery grounds, are more important in governing spatial distributions than prey availability (Pitois et al., 2021), the data provided no new insight into the smaller scale processes of fish (e.g. local migration, changing in feeding behaviors or shoaling). The reason for not capturing the Lyme Bay region and thus the misalignment with Sprat was simply because experiments with the PI size parameters were being run to determine the most appropriate compromise between storage and target size (see Section 2.5.1). In the future, as the PI becomes fully operational and automated, a more complete record will be available for each survey. This will allow for areas of interest to other disciplines (e.g. fish spawning areas) or hydrodynamic

features (e.g. fronts or upwelling) can be retrospectively subsampled. This can be combined with the conclusions from the Section 4 regarding choice of spatial resolution as analysing all data might not be necessary. The first consideration should be the hypothesis-driven study area, followed by the choice of sampling resolution. Here, where the choice of resolution did not have significant impact on the statistical relationships it would be sensible to sample to the lower resolution possible.

5.6 CONCLUSION

We find that choice of resolution does not effect the correlation or significance between fish biomass and copepod abundance or size. We find unexpected correlations between Sprat and Herring and copepod abundance and size. These relationships, if true could lead to new insight into and call for a rethink about what governs the interactions between pelagic fish and their prey. The tools and programs used for this chapter will form the basis for data analysis with the PI and mergers with other continuous instruments going forward. Any continuous data can be used as long as a latitude, longitude and 'datetime' are provided. Optimisation of the code will quickly allow users to specify the spatial and/or temporal resolutions they require to fit their question. These data when used as a time series could provide a powerful data for trends and relationship analysis. The relationships also pose questions about the need for collecting continuous data when looking at small scale processes (e.g. single fish feeding) for a population wide question. Combining longer time series with larger spatial or temporal scales and very fine frequency data may provide a better approach to deciphering these relationships.

5.7 BIBLIOGRAPHY

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, 391(6667):577–580.

Atkinson, A., Polimene, L., Fileman, E. S., Widdicombe, C. E., McEvoy, A. J., Smyth, T. J., Djeghri, N., Sailley, S. F., and Cornwell, L. E. (2018). Comment. What drives plankton seasonality in a stratifying shelf sea? Some competing and complementary theories. *Limnology and Oceanography*, 63(6):2877–2884.

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., Kiørboe, T., and Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, 16(4):522–534.

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(6967):661–664.

Benfield, M. C., Grosjean, P., Culverhouse, P. F., Irigoien, X., Sieracki, M. E., Lopez-Urrutia, A., Dam, H. G., Hu, Q., Davis, C. S., Hanson, A., Pilskaln, C. H., Riseman, E. M., Schultz, H., Utgoff, P. E., and Gorsky, G. (2007). RAPID: Research on Automated Plankton Identification. *Oceanography*, 20(SPL.ISS. 2):172–218.

Capuzzo, E., Wright, S., Bouch, P., Collingridge, K., Creach, V., Pitois, S., Stephens, D., and Kooij, J. v. d. (2022). Variability in structure and carbon content of plankton communities in autumn in the waters south-west of the UK. *Progress in Oceanography*, 204(October 2021):102805.

Carlotti, F. and Poggiale, J. C. (2010). Towards methodological approaches to implement the zooplankton component in "end to end" food-web models. *Progress in Oceanography*, 84(1-2):20–38.

Confer, J. L. and Blades, P. I. (1975). Omnivorous zooplankton and planktivorous fish. *Limnology and Oceanography*, 20(4):571–579.

Corten, A. (2000). A possible adaptation of herring feeding migrations to a change in timing of the Calanus finmarchicus season in the eastern North Sea. *ICES Journal of Marine Science*, 57(4):1261–1270.

Culverhouse, P. F., Williams, R., Simpson, B., Gallienne, C., Reguera, B., Cabrini, M., Fonda-Umani, S., Parisini, T., Pellegrino, F. A., Pazos, Y., Wang, H., Escalera, L., Moroño, A., Hensey, M., Silke, J., Pellegrini, A., Thomas, D., James, D., Longa, M. A., Kennedy, S., and Del Punta, G. (2006). HAB Buoy: A new instrument for in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, 28(2):245–250.

Davis, C. S., GALLAGER, S. M., Berman, M. S., Haury, L. R., and Strickler, J. R. (1992). The Video Plankton Recorder (VPR): Design and initial results. *Arch. Hydrobiol. Beih.*, 36:67–81.

Faillettaz, R., Picheral, M., Luo, J. Y., Guigand, C., Cowen, R. K., and Irisson, J. O. (2016). Imperfect automatic image classification successfully describes plankton distribution patterns. *Methods in Oceanography*, 15-16:60–77.

Fernandes, J. A., Irigoien, X., Boyra, G., Lozano, J. A., and Inza, I. (2009). Optimizing the number of classes in automated zooplankton classification. *Journal of Plankton Research*, 31(1):19–29.

Flinkman, J., Aro, E., Vuorinen, I., and Viitasalo, M. (1998). Changes in northern baltic zooplankton and herring nutrition from 1980s to 1990s: Top-down and bottom-up processes at work. *Marine Ecology Progress Series*, 165:127–136.

Garrido, S., Ben-Hamadou, R., Oliveira, P. B., Cunha, M. E., Chícharo, M. A., and van der Lingen, C. D. (2008). Diet and feeding intensity of sardine Sardina pilchardus: correlation with satellite-derived chlorophyll data. *Marine Ecology Progress Series*, 354:245–256.

Gorokhova, E., Lehtiniemi, M., Lesutiene, J., Strake, S., Uusitalo, L., Demereckiene, N., and Amid, C. (2013). Zooplankton mean size and total abundance. *HELCOM Core Indicator Report*.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.-B. B., Cawood, A., Pesant, S., García-Comas, C., and Prejger, F. (2010). Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, 32(3):285–303.

Grosjean, P., Picheral, M., Warembourg, C., and Gorsky, G. (2004). Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science*, 61(4):518–525.

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. *Progress in Oceanography*, 166(July 2017):30–40.

ICES (2021). Inter-benchmark to revise the advice framework for the Sprat stock in 7.de based on the most recent changes to data-limited short-lived species assessments (IBPSprat). *ICES Scientific reports*, 3(23):42 pp.

Irigoien, X., Flynn, K. J., and Harris, R. P. (2005). Phytoplankton blooms: a 'loophole'in microzooplankton grazing impact? *Journal of Plankton Research*, 27(4):313–321.

Irigoien, X. and Harris, R. P. (2003). Interannual variability of Calanus helgolandicus in the English Channel. *Fisheries Oceanography*, 12(4-5):317–326.

Johns, D. (2006). The Plankton Ecology of the SEA 8 area. *Strategic Environmental Assessment Programme, UK Dept. of Trade and Industry*, pages 1–44.

Kaschner, K., Watson, R., Trites, A. W., and Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, 316:285–310.

Lauria, V., Attrill, M. J., Pinnegar, J. K., Brown, A., Edwards, M., and Votier, S. C. (2012). Influence of Climate Change and Trophic Coupling across Four Trophic Levels in the Celtic Sea. *PLoS ONE*, 7(10).

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J. B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O.,

Miloslavich, P., Lara-Lopez, A., Kudela, R. M., Lopes, R. M., Karp-Boss, L., Kiko, R., Jaffe, J. S., Iversen, M. H., Irisson, J. O., Hauss, H., Guidi, L., Gorsky, G., Giering, S. L. C., Gaube, P., Gallager, S., Dubelaar, G., Cowen, R. K., Carlotti, F., Briseño-Avena, C., Berline, L., Benoit-Bird, K. J., Bax, N. J., Batten, S. D., Ayata, S. D., and Appeltans, W. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6(MAR).

Lynam, C. P., Llope, M., Möllmann, C., Helaouët, P., Bayliss-Brown, G. A., and Stenseth, N. C. (2017). Interaction between top-down and bottom-up control in marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 114(8):1952–1957.

Mackas, D. L., Denman, K. L., and Abbott, M. R. (1985). Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science*, 37(2):653–674.

MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312):154–155.

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. *Progress in Oceanography*, 137:513–523.

Ohman, M. D. and Lavaniegos, B. E. (2002). Comparative zooplankton sampling efficiency of a ring net and bongo net with comments on pooling of subsamples. Technical report.

Pakhomov, E. A. and McQuaid, C. D. (1996). Distribution of surface zooplankton and seabirds across the Southern Ocean. *Polar Biology*, 16(4):271–286.

Patel, R., Roy, S., Capuzzo, E., and Kooij, J. V. D. (2022). Seasonality of diet overlap among small pelagic fish in the waters southwest of the UK. pages 1–40.

Pinnegar, J. K., Jennings, S., O'Brien, C. M., and Polunin, N. V. (2002). Long-term changes in the trophic level of the Celtic Sea fish community and fish market price distribution. *Journal of Applied Ecology*, 39(3):377–390.

Pitois, S. G. and Fox, C. J. (2006). Long-term changes in zooplankton biomass concentration and mean size over the Northwest European shelf inferred from Continuous Plankton Recorder data. *ICES Journal of Marine Science*, 63(5):785–798.

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). *Ecological Indicators*, 123:107307.

Pitois, S. G., Lynam, C. P., Jansen, T., Halliday, N., and Edwards, M. (2012). Bottom-up effects of climate on fish populations: Data from the continuous plankton recorder. *Marine Ecology Progress Series*, 456:169–186.

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

Reid, P. C., Battle, E. J. V., Batten, S. D., and Brander, K. M. (2000). Impacts of fisheries on plankton community structure. *ICES Journal of Marine Science*, 57(3):495–502.

Reid, P. C., De Fatima Borges, M., and Svendsen, E. (2001). A regime shift in the north sea circa 1988 linked to changes in the north sea horse mackerel fishery. *Fisheries Research*, 50(1-2):163–171.

Rochet, M. J., Trenkel, V. M., Carpentier, A., Coppin, F., de Sola, L. G., Léauté, J. P., Mahé, J. C., Maiorano, P., Mannini, A., Murenu, M., Piet, G. J., Politou, C. Y., Reale, B., Spedicato, M. T., Tserpes, G., and Bertrand, J. A. (2010). Do changes in environmental and fishing pressures impact marine communities? An empirical assessment. *Journal of Applied Ecology*, 47(4):741–750.

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. *Journal of Plankton Research*, 37(3):519–529.

Samson, S., Hopkins, T., Remsen, A., Langebrake, L., Sutton, T., and Patten, J. (2001). A system for high-resolution zooplankton imaging. *IEEE Journal of Oceanic Engineering*, 26(4):671–676.

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.

Sims, D. W. (1999). Threshold foraging behaviour of basking sharks on zooplankton: Life on an energetic knife-edge? *Proceedings of the Royal Society B: Biological Sciences*, 266(1427):1437–1443.

Sims, D. W. and Quayle, V. A. (1998). Selective foraging behaviour of basking sharks on zooplankton in a small-scale front. *Nature*, 393(6684):460–464.

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

The Turing Centre (2021). DSG Report 2021. Technical report.

Thieurmel, B., Elmarhraoui, A., and Thieurmel, M. B. (2019). Package "suncalc".

Travers, M., Shin, Y. J., Jennings, S., and Cury, P. (2007). Towards end-to-end models for investigating the effects of climate and fishing in marine ecosystems. *Progress in Oceanography*, 75(4):751–770.

Trenkel, V. M., Pinnegar, J. K., Dawson, W. A., Du Buit, M. H., and Tidd, A. N. (2005). Spatial and temporal structure of predator-prey relationships in the Celtic Sea fish community. *Marine Ecology Progress Series*, 299:257–268.

Van Der Kooij, J., Fässler, S. M., Stephens, D., Readdy, L., Scott, B. E., and Roel, B. A. (2016). Opportunistically recorded acoustic data support Northeast Atlantic mackerel expansion theory. *ICES Journal of Marine Science*, 73(4):1115–1126.

Van Deurs, M., Jorgensen, C., and Fiksen, O. (2015). Effects of copepod size on fish growth: A model based on data for North Sea sandeel. *Marine Ecology Progress Series*, 520:235–243.

Wiebe, P. H. and Wiebe, P. H. (1968). Plankton Patchiness: Effects on Repeated Net Tows. *Limnology and Oceanography*, 13(2):315–321.

Working Group of International Pelagic Surveys (2015). Manual for International Pelagic Surveys (IPS). Technical report, ICES, Copenhagen.

6

SYNTHESIS

6.1 ABSTRACT

In the following chapter the key results are discussed in the context of the original objectives. The limitations of the thesis are briefly reviewed. Limitations identified with the instrument prior the PhD research which have been addressed during the course of the PhD research are also reviewed. Finally, the wider implications and recommendations for future research are discussed.

6.2 SUMMARY OF KEY RESULTS

This study presents zooplankton data at a unprecedented spatiotemporal resolution and is the first study to match these data with other continuous variables at these resolutions while maintaining the potential to discern species to a moderate (e.g. family-level) taxonomic resolution. The analyses on these data captured seasonal changes in the zooplankton community using a more traditional point-sampling method. This demonstrates the importance of sample resolution when examining fine-scale processes such as the relationship between chlorophyll and zooplankton

and found an unexpected negative correlations with pelagic fish at finer resolutions. These data have been obtained using the Plankton Imager which (at the time of writing) is unparalleled in its ability to capture zooplankton data at these fine resolutions to a moderate degree of taxonomic resolution without the need for deployment where data collection is automated. Obtaining this resolution was only possible due to the machine learning classifier developing during the PhD reserach. While the programming, training and successful testing of the classifier was achieved through a collaborative project with data scientists, images from the PhD data were used to build the classifier. This demonstrates the exciting potential of coupling high frequency data and machine learning classifiers may have on our future understanding plankton ecology.

The data analysis chapters (Chapters 3, 4 & 5) essentially increase in spatiotemporal resolution from the last chapter and follows the natural progression of the instrument development and the increasing power of the analytics tools available during the PhD. The first chapter (Chapter 3) uses the PI in a very similar way to a traditional net with low spatial resolution and high taxonomic resolution. The second chapter starts to take advantage of the continuous data but no machine learning classifier was available. Data were therefore spatially subset and temporally subsampled yielding a higher spatial resolution than Chapter 3 but at the cost of taxonomic resolution. Only copepods were classified to speed up the classification process. Finally, Chapter 5 takes advantage of the new classifier and samples all data yielding the finest spatiotemporal resolution. To ensure image classification accuracy, images were only classified to the copepod level.

The thesis successfully answers the objectives set out in Chapter 1.5. The key results from each chapter and outlined in the context of the 6 objectives. Objective 6 (*To summarise the major findings from the thesis to provide context and suggestions for future studies*) is excluded from the list below as it is addressed in this chapter

1. Through a literature review provide a context for the PI amongst other zooplankton samplers and demonstrate its potential value to zooplankton research.

- **The Plankton Imager is a unique instrument that has the potential to describe zooplankton at an unprecedented spatiotemporal scale**. Through a brief review of common approaches and available devices in Chapter 1 the place for the PI among existing devices, and its unique properties, are described. Currently, it is difficult to compare it any other available device. It harnesses the advantages associated with imaging zooplankton as opposed to their physical collection (e.g. speed, cost, automated) while negating the effects of the more common towed approaches (e.g. changes in vessel behaviour, wake). It makes use of existing ship's infrastructure and requires little modification for installation. Finally, it is automated, meaning it requires little human effort during operation.
- **The key findings answering objectives 3 to 5 also demonstrate the PIs value to zooplankton research.**

2. Contribute to ongoing development of the system and its methodology for full operational deployment.

   - **A comprehensive method for the PI is detailed including the hardware, software and standard operating procedure.** The method described in Chapter 2 provides a reference for the thesis for future applications of the PI. Although the hardware is not applicable to the newer PI-10 the operating procedure and data handling methods will remain largely the same.
   - **The PI-10 is now a commercial, ready to purchase instrument**, Figure 2.8. The PhD reserach has contributed to the development and ecological demonstration of the instrument. The major contributions are detailed in Appendix C.

3. Evaluate the PIs ability for community and ecosystem approach to fisheries.

   - **Interannual and seasonal changes were identified in mesozooplankton community of the Celtic Sea.** These data were obtained using point

locations and subsampling in the Celtic sea across several years, totalling 107 stations. A moderate (family-level) taxonomic resolution was possible with PI images which was detailed enough to describe changes in the community.

- **Variation was higher between years than seasons** This was due to the large variation in the community between years in autumn attributed to a mismatch between the survey time and the breaking down of summer stratification (Southward et al., 2004; Harris, 2010). The spring community, where the water column is more stable during the survey period, was more consistent between years.
- **The seasonality of taxa adhered to those presented in the literature** (Johns, 2006; Eloire et al., 2010; Highfield et al., 2010; Giering et al., 2019).

4. Develop an analytic method to best use the PIs ability to collect continuous, high frequency zooplankton data.

    - **Temporal subsampling yielded small scale zooplankton data.** Sampling 1 in every 10 images and manually classifying those images described both area wide patterns and large, small scale changes in copepods biomass, size and abundance, indicative of plankton patchiness (Mackas et al., 1985; Abraham, 1998).
    - **Reducing sampling resolution by merging adjacent data, changes the spatial description of the data** where small spatial changes (those that occur in scales < 3 km) were lost and area wide patterns were emphasized.
    - **The choice of resolution can affect both the statistical strength and significance of relationships.** The relationship between copepod biomass and chlorophyll varied significantly with spatial resolution. This variation is also demonstrated by conflicting relationships between these variables within the literature (Casini et al., 2008; Llope et al., 2012; Schultes et al., 2013; Giering et al., 2019; Pitois et al., 2021).
    - **There was a negative correlation with Sprat and Herring with copepod size and abundance and positive correlation with Anchovy and Sardine.**

These data were compared to pelagic fish data, another continuous variables collected using acoustics. Comparison of these variables revealed an unexpected negative correlation of copepod size and abundance with Sprat and Herring where previous results in the literature find the contrary (Pitois et al., 2021; Flinkman et al., 1998; Corten, 2000).

- **Choice of resolution, both spatial and temporal, had little effect on these correlations.** Contrary to those results presented in Chapter 4 changing spatial or temporal resolution had little influence on the strength or correlation. This demonstrates the importance of sampling scale versus process size where fine spatial scales may not always be the best approach and longer temporal scales (e.g. months or years) and time series would be better suited.

5. To help develop and test a machine learning algorithm on image data obtained from the PI to yield unprecedented fine scale data.

    - **Successful deployment of a machine learning algorithm yielded copepod data at very fine spatial scales.** Using the developed Turing Classify all images collected during a 1 month survey were analysed. Data can potentially be resolved to meters and minutes.
    - **Development will continue into the future.** The training set will continue development as more and more images are collected. If (when) the PI becomes a more common place to a collaborative approach between institutions will yield the best training set.

6.3 STUDY AND INSTRUMENT LIMITATIONS

The data chapters (Chapters 3, 4 & 5) use the PI in the 'best way' possible at the time, meaning data were resolved to the finest spatial or temporal resolution using the tools available. As aforementioned, in line with improvements to the instrument and our ability to resolve data to finer spatiotemporal scales increases with time. Each chapter uses a different data 'type': Chapter 3 uses point, subsampled data; Chapter 4 uses

near-continuous temporally subsampled data and Chapter 5 uses continuous data without any subsampling. This improvement in spatial resolution with each chapter (at the cost of taxonomic resolution) partially negates, or avoids all-together, some of the limitations in the previous chapter. The reduced taxonomic resolution seen in the latter data analysis chapters is only a temporary limitation as the capacity to resolve images to increased taxonomic resolution with machine learning classifiers is quickly increasing.

In Chapter 3 (Zooplankton community seasonal and annual changes) data are sub sampled, point data which is analogous to using ring nets and sub sampling for microscopy. Thus, the data are subject to the same limitations of subsampling (see Chapter 1.2.1). Mainly, the under or over representation of rarer species. Images were manually classified to the highest possible taxonomic resolution which takes time. As a result the data had a significantly reduced spatial coverage and resolution compared with the latter chapters. Chapter 4 still uses subsampled data but here data are temporally sub sampled. The limited spatial resolution is significantly reduced using this approach (Figure 3.1 to Figure 4.1), although at the cost of taxonomic resolution. Sampling to copepods only avoids the over or under representation of rarer species but removes any capacity for community-type studies and analysis such as those performed in Chapter 3. Finally, Chapter 5 builds on the work done in the previous chapter but negates subsampling all together by employing the Turing Classifier (Chapter 2.6.2). This removes completely the limitations associated with subsampling and provides fine spatial zooplankton data but does not provide any community data. Achieving subsampling-free, fine spatial data and an increased taxonomic resolution is nearly achievable and is discussed below.

Two (Chapters 4 & 5) of the three data chapters provide only a 'snapshot' of the zooplankton. They are single cruise analyses that only cover a single season. The limitations of the 'survey snapshot' (Huret et al., 2018) are addressed in the respective chapter discussions. This limitation primarily arises from the instrument being in a developmental stage throughout the PhD and will be avoidable in the future when the PI is resident aboard on the RV Endeavour or other research vessels. Additionally, the late-PhD arrival of tools required to perform these non-subsample analysis and

the time taken required to run them on the data archive was outside the scope of the PhD timeline. The PI will now be routinely used in the spring and autumn surveys (Table 2.1) and will allow for future time series studies at finer spatial resolution, and in combination with improved classifiers, higher taxonomic resolution.

FIXED SAMPLING DEPTH

The PIs reliance on the vessels continuous flow restricts vertical sampling to 4 meters below sea level. Instead of a limitation, this may be considered a characteristic of the instrument, in-line with other plankton instruments where no device can sample the whole plankton and each instruments samples a portion of the plankton. To obtain vertical data, the PI can be used in combination with other instruments, see below (section 6.5).

6.3.1 LIMITATIONS ADDRESSED WITH THE THESIS

The completion of the instrument coincided with the end of the PhD. A month before submission, the finished, commercial purchasable product, the PI-10, was tested (Chapter 2.8). In the final months of the PhD, a machine learning classifier that allowed for classification all images with acceptable error (Chapter 2.6.2) was developed. These developments successful complete objective 2 (Chapter 1.5) and resolve many of the more major limitations present in the thesis.

The foremost limitation, now addressed, is the inability to sample all images. It is now possible to sample all images and avoid subsetting or subsampling the data which is required for manual classification. In turn, this avoids the issues associated with subsampling (see Chapter 1.2.1) but also allows for data to be gained at an unprecedented spatial and temporal resolution. This potential is initially explored in Chapter 5 although limitations associated with device development (e.g. testing data rate capacities) resulted in a mismatch between data and ecological processes (e.g., the absence of PI data in Sprat feeding zones, Figure 5.3). These limitations will be absent in future surveys as the data rates (/ limits) are now known and the instrument is virtually automated. The impetus is now on expanding capacity to store more data

and refine classifiers to resolve zooplankton to even higher taxonomic resolution, discussed below.

These two developments cover many of the limitations present in the course of the thesis. In addition, over the course of the PhD, feedback between myself (from experience as the end-user both at sea and during data analysis) and the instrument designers also resulted in tweaks to the software and hardware as well as establishing protocols in response to limitations (for some examples, see Appendix C).

6.4 WIDER IMPLICATIONS

The PI provides opportunity for collection of a new type of plankton data as well as avoiding some of the complexities associated with zooplankton sampling. Thus, the wider implications of the thesis are two fold. Firstly is the demonstration of PI, and imaging devices in general, as a cost-effective alternative method for collecting zooplankton data. Secondly is the ecological insight that can be obtained from these tools that have the capacity to sample zooplankton data at new resolutions.

The PI was developed in response to the limitations associated with traditional net sampling (Chapter 1.2.1) (Agnarsson and Kuntner, 2007; Bean et al., 2017; Pitois et al., 2018; Danovaro et al., 2016) coupled with mandated monitoring through policy (McQuatters-Gollop et al., 2017; Bedford et al., 2018) and increasingly complex ecosystem (Mitra et al., 2014) and biogeochemical modelling (Steinberg et al., 2002; Steinberg and Landry, 2017). The thesis begins to demonstrate the capacity of one of these newer approaches to answering existing ecological questions at reduced time and financial costs. The thesis, and the results therein, form part of a large collection of work where imaging instruments are becoming increasingly commonplace in monitoring strategies and providing a viable, and often preferential, alternative to traditional devices. In Chapter 4 the PI is able to describe changes in zooplankton community similarly to those methods that analysis the specimens via light microscopy (Johns, 2006; Eloire et al., 2010; Highfield et al., 2010). There are too many recent examples to cite, but a simple Google Scholar result for "zooplankton

imaging" for articles post-2021 returns 1000s of hits, with the diversity of applications for imaging devices also growing.

While there a range of imaging devices currently available, the PI remains fairly unique in its high throughput capacity while retaining the ability to discern moderate taxonomic resolution (e.g. copepods to family level). It is difficult to find a device similar in the literature. The recently developed 'Planktoscope' (Pollina et al., 2020) uses a similar method but samples at 0.1 ml/min, an order of magnitude less than the PI. Now the PI is established the future implications, in terms of ecological studies can start to be explored. Although, the PI, like all plankton sampling devices, is constrained and thus biased, towards a certain portion of the zooplankton (Owens et al., 2013). It may not be the most appropriate tool for specific ecological questions and must be considered as an option among many other devices where the researcher must choose the most appropriate device to the question (Skjoldal et al., 2013).

The wider ecological implications of the thesis are the exciting result of trying to develop cost effective devices. The combination of imaging and machine learning and its application to plankton science is growing (Crisci et al., 2012; Irisson et al., 2022). This coupling allows for describing the zooplankton at an ever increasing spatiotemporal resolution. Chapter 4 is an example of obtaining small scale zooplankton data although manually classified. This is followed by Chapter 5 which builds on the previous chapter and presents even finer data obtained by using a machine learning algorithm. This progression between chapters is itself a demonstration in the benefits of using a machine learning approach compared with manual taxonomy. Machine learning classifiers have the capacity to deal with the large number of images collected by imaging devices (Culverhouse et al., 2006; Benfield et al., 2007; MacLeod et al., 2010) which would be impossible to analyse manually. Similarly to zooplankton imaging, there are numerous articles applying machine learning to zooplankton. In terms of the wider implications, these chapters form the foundation for a range of studies that can better harness continuous zooplankton data.

6.5 FUTURE RESEARCH

Although improvements to the instrument hardware, software and data analysis pipelines will continue development, the PI (or now the 'PI-10') is now a finished tool. The hardware and software are developed to a point that ensure the instrument is automated. This means a dedicated crew member is no longer required to accompany the instrument. Dedicated team members incur substantial costs (with the notable exception of PhD students) and removing this requirement allows for a much more flexible deployment regime of the instrument both in terms of surveys and research vessels. There are currently plans to put the existing PI-10 on more Cefas surveys aboard the RV Cefas Endeavour and deploy anther PI on a different research vessel. Finally, as a finished instrument, periods of development or modification to the instrument are no longer needed. These have previously resulted in a missed survey (Table 2.1). In future, the instrument will be more consistently used at sea which will allow for formation of new time series. This frames the context for much of the future work suggestions detailed below.

Due to the novel nature of the PI data there were many potential research avenues that might have been pursued during the PhD. For example, the PIs ability to capture Radiolaria, discussed in Chapter 1 and seen in the results of Chapter 3, could have been used to better understand the spatial distribution of these species. There are likely many other examples of studies that can be achieved using the existing PI data archive. Some future example applications of the PI, now it is established are detailed:

- **Fine spatial changes in zooplankton over hydrodynamic features (e.g. fronts).** The PI could be used to explore fine changes in the zooplankton community, biomass or sizes over relatively small hydrodynamic features. The width of ocean fronts may be as small as 100 m (Belkin and Cornillon, 2007) and are important governing features of zooplankton dynamics (Graham and Hamner, 2001; Genin et al., 2005) and the PI is well placed to resolve data to these resolutions.

- **Use in hard to reach areas where deployment of nets or towed devices is not safe.** Deployment of any devices can be made impossible by sea conditions. For example, Drakes Passage, in the Southern Ocean is hydrodynamically complex (Grelowski et al., 1986; López et al., 1999) and can result in unfavourable sea states for deployments. Several studies exist documenting the zooplankton using nets with low sampling resolution For example Vedenin et al., 2019. If the PI were to be installed on an annual survey that crosses these seas a time series could be built with relative safety, cost and effort.

- **Formation of fine scale, moderate taxonomic resolution time series.** The PI lends itself to the formation of time series in much the same way as the Continuous Plankton Recorder (Richardson et al., 2006). The PI will be used routinely and left aboard ship collecting data. These time series are powerful tools in assessing long term changes in the zooplankton (for a review see Mackas and Beaugrand, 2010), for example they have been used to identify changes in the zooplankton in response to climate change (Richardson, 2008; Taylor et al., 2002).

- **Interpolate with vertical depth data.** The PI data is solely horizontal. Zooplankton undergo significant diurnal vertical migrations (Gliwicz, 1986; Bandara et al., 2021) which are missed by our data. Future studies could integrate vertical data, captured by ring nets or other imaging devices, to provide a 3D description of the zooplankton.

- **Data analysis pipeline.** Although not strictly a research output, the analyses within the thesis demanded a unique approach to harmonising several sources of continuous data. The code written for Chapters 4 & 5 could be the base for which a developer builds software for harmonising continuous data with ease. This pipeline can be built on in future and perhaps automated. Bringing these sources of data together over several years has the capacity to provide a detailed, holistic description of ecosystem and describe changes over time.

6.6 CONCLUSION

The thesis meets the aims set out in the introductory chapter and forms a foundation for future studies using continuous plankton data of a similar nature. The examples of future work listed above are only a few examples of where devices such as the PI can provide a unique perspective on existing questions. Ecologically speaking, the thesis demonstrates imaging devices have the capacity to sufficiently describe temporal changes in zooplankton assemblages similar to a ring nets. Relationships previous described in studies with lower spatial resolution are re-investigated at finer spatiotemporal scales with varied results (compared to the literature), demonstrating the complexity of these relationships, the importance of sampling scale and time series. On completion of the thesis, work begins on deploying the PI routinely, the formation of time series and the potential deployment of other PIs on other research vessels.

6.7 BIBLIOGRAPHY

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, 391(6667):577–580.

Agnarsson, I. and Kuntner, M. (2007). Taxonomy in a changing world: Seeking solutions for a science in crisis. *Systematic Biology*, 56(3):531–539.

Bandara, K., Varpe, O., Wijewardene, L., Tverberg, V., and Eiane, K. (2021). Two hundred years of zooplankton vertical migration research. *Biological Reviews*, 96(4):1547–1589.

Bean, T. P., Greenwood, N., Beckett, R., Biermann, L., Bignell, J. P., Brant, J. L., Copp, G. H., Devlin, M. J., Dye, S., Feist, S. W., Fernand, L., Foden, D., Hyder, K., Jenkins, C. M., van der Kooij, J., Kröger, S., Kupschus, S., Leech, C., Leonard, K. S., Lynam, C. P., Lyons, B. P., Maes, T., Nicolaus, E. E., Malcolm, S. J., McIlwaine, P., Merchant, N. D., Paltriguera, L., Pearce, D. J., Pitois, S. G., Stebbing, P. D., Townhill, B., Ware, S., Williams, O., and Righton, D. (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. *Frontiers in Marine Science*, 4(AUG):263.

Bedford, J., Johns, D., Greenstreet, S., and McQuatters-Gollop, A. (2018). Plankton as prevailing conditions: A surveillance role for plankton indicators within the Marine Strategy Framework Directive. *Marine Policy*, 89(January):109–115.

Belkin, I. M. and Cornillon, P. C. (2007). Fronts in the world ocean's large marine ecosystems. *ICES CM*, 500(130):21.

Benfield, M. C., Grosjean, P., Culverhouse, P. F., Irigoien, X., Sieracki, M. E., Lopez-Urrutia, A., Dam, H. G., Hu, Q., Davis, C. S., Hanson, A., Pilskaln, C. H., Riseman, E. M., Schultz, H., Utgoff, P. E., and Gorsky, G. (2007). RAPID: Research on Automated Plankton Identification. *Oceanography*, 20(SPL.ISS. 2):172–218.

Casini, M., Lövgren, J., Hjelm, J., Cardinale, M., Molinero, J. C., and Kornilovs, G. (2008). Multi-level trophic cascades in a heavily exploited open marine ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, 275(1644):1793–1801.

Corten, A. (2000). A possible adaptation of herring feeding migrations to a change in timing of the Calanus finmarchicus season in the eastern North Sea. *ICES Journal of Marine Science*, 57(4):1261–1270.

Crisci, C., Ghattas, B., and Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, 240:113–122.

Culverhouse, P. F., Williams, R., Simpson, B., Gallienne, C., Reguera, B., Cabrini, M., Fonda-Umani, S., Parisini, T., Pellegrino, F. A., Pazos, Y., Wang, H., Escalera, L., Moroño, A., Hensey, M., Silke, J., Pellegrini, A., Thomas, D., James, D., Longa, M. A., Kennedy, S., and Del Punta, G. (2006). HAB Buoy: A new instrument for in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, 28(2):245–250.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., David, R., Dell'Anno, A., Dzhembekova, N., Garcés, E., Gasol, J. M., Goela, P., Féral, J. P., Ferrera, I., Forster, R. M., Kurekin, A. A., Rastelli, E., Marinova, V., Miller, P. I., Moncheva, S., Newton, A., Pearman, J. K., Pitois, S. G., Reñé, A., Rodríguez-Ezpeleta, N., Saggiomo, V., Simis, S. G., Stefanova, K., Wilson, C., Martire, M. L., Greco, S., Cochrane, S. K., Mangoni, O., and Borja, A. (2016). Implementing and innovating marine monitoring approaches for assessing marine environmental status. *Frontiers in Marine Science*, 3(NOV):213.

Eloire, D., Somerfield, P. J., Conway, D. V., 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. *Journal of Plankton Research*, 32(5):657–679.

Flinkman, J., Aro, E., Vuorinen, I., and Viitasalo, M. (1998). Changes in northern baltic zooplankton and herring nutrition from 1980s to 1990s: Top-down and bottom-up processes at work. *Marine Ecology Progress Series*, 165:127–136.

Genin, A., Jaffe, J. S., Reef, R., Richter, C., and Franks, P. J. S. (2005). Swimming against the flow: a mechanism of zooplankton aggregation. *Science*, 308(5723):860–862.

Giering, S. L., Wells, S. R., Mayers, K. M., Schuster, H., Cornwell, L., Fileman, E. S., Atkinson, A., Cook, K. B., Preece, C., and Mayor, D. J. (2019). Seasonal variation of zooplankton community structure and trophic position in the Celtic Sea: A stable isotope and biovolume spectrum approach. *Progress in Oceanography*, 177(March 2018):101943.

Gliwicz, M. Z. (1986). Predation and the evolution of vertical migration in zooplankton. *Nature*, 320(6064):746–748.

Graham, W. M. and Hamner, W. M. (2001). A physical context for gelatinous zooplankton aggregations: a review. *Jellyfish blooms: ecological and societal importance*, pages 199–212.

Grelowski, A., Majewicz, A., and Pastuszak, M. (1986). Mesoscale hydrodynamic processes in the region of Bransfield Strait and the southern part of Drake Passage during BIOMASS-SIBEX 1983/84. *Polish Polar Research*, 7(4).

Harris, R. (2010). The L4 time-series: The first 20 years. *Journal of Plankton Research*, 32(5):577–583.

Highfield, J. M., Eloire, D., Conway, D. V., Lindeque, P. K., Attrill, M. J., and Somerfield, P. J. (2010). Seasonal dynamics of meroplankton assemblages at station L4. *Journal of Plankton Research*, 32(5):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. *Progress in Oceanography*, 166(July 2017):30–40.

Irisson, J. O., Ayata, S. D., Lindsay, D. J., Karp-Boss, L., and Stemmann, L. (2022). Machine Learning for the Study of Plankton and Marine Snow from Images. *Annual Review of Marine Science*, 14(January):277–301.

Johns, D. (2006). The Plankton Ecology of the SEA 8 area. *Strategic Environmental Assessment Programme, UK Dept. of Trade and Industry*, pages 1–44.

Llope, M., Licandro, P., Chan, K. S., and Stenseth, N. C. (2012). Spatial variability of the plankton trophic interaction in the North Sea: A new feature after the early 1970s. *Global Change Biology*, 18(1):106–117.

López, O., Garcıa, M. A., Gomis, D., Rojas, P., Sospedra, J., and Sánchez-Arcilla, A. (1999). Hydrographic and hydrodynamic characteristics of the eastern basin of the Bransfield Strait (Antarctica). *Deep Sea Research Part I: Oceanographic Research Papers*, 46(10):1755–1778.

Mackas, D. L. and Beaugrand, G. (2010). Comparisons of zooplankton time series. *Journal of Marine Systems*, 79(3-4):286–304.

Mackas, D. L., Denman, K. L., and Abbott, M. R. (1985). Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science*, 37(2):653–674.

MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature*, 467(7312):154–155.

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

Mitra, A., Castellani, C., Gentleman, W. C., Jónasdóttir, S. H., Flynn, K. J., Bode, A., Halsband, C., Kuhn, P., Licandro, P., Agersted, M. D., Calbet, A., Lindeque, P. K., Koppelmann, R., Møller, E. F., Gislason, A., Nielsen, T. G., and St. John, M. (2014). Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. *Progress in Oceanography*, 129(PB):176–199.

Owens, N. J., 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 J.W. Dippner and M. Krause. *Journal of Marine Systems*, 128:240–242.

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). *Ecological Indicators*, 123:107307.

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

Pollina, T., Larson, A. G., Lombard, F., Li, H., Colin, S., de Vargas, C., and Prakash, M. (2020). PlanktonScope: affordable modular imaging platform for citizen oceanography. *BioRxiv*.

Richardson, A. J. (2008). In hot water: Zooplankton and climate change. *ICES Journal of Marine Science*, 65(3):279–295.

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. *Progress in Oceanography*, 68(1):27–74.

Schultes, S., Sourisseau, M., Le Masson, E., Lunven, M., and Marié, L. (2013). Influence of physical forcing on mesozooplankton communities at the Ushant tidal front. *Journal of Marine Systems*, 109-110(SUPPL.):S191–S202.

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. *Progress in Oceanography*, 108:1–42.

Southward, A. J., Langmead, O., Hardman-Mountford, N. J., Aiken, J., Boalch, G. T., Dando, P. R., Genner, M. J., Joint, I., Kendall, M. A., Halliday, N. C., Harris, R. P.,

Leaper, R., Mieszkowska, N., Pingree, R. D., Richardson, A. J., Sims, D. W., Smith, T., Walne, A. W., and Hawkins, S. J. (2004). Long-term oceanographic and ecological research in the western English Channel. *Advances in Marine Biology*, 47:1–105.

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-Sea Research Part I: Oceanographic Research Papers*, 49(8):1445–1461.

Steinberg, D. K. and Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science*, 9(1):413–444.

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

Vedenin, A. A., Musaeva, E. I., Zasko, D. N., and Vereshchaka, A. L. (2019). Zooplankton communities in the Drake Passage through environmental boundaries: a snapshot of 2010, early spring. *PeerJ*, 7:e7994.

A

METADATA EXPLORER FOR PI DATA

A.1 INTRODUCTION

The PI generates millions of images per survey. Understanding where the data are and how much data are in an area is useful knowledge when retrospectively subsampling. A more educated decision can be made about where to sample of if its sensible to sample an area (e.g. if there is an exceptional concentration of images in a specific areas compute time might be unrealistic). The PI Meta data explorer (Figure A.1 was developed in response to this need and provides and interactive tool for exploring the data. This tool will continue to be developed post-phd and was never truly finished. The code below was created in year 2 of the PhD and as such is not optimised or up in line with best-coding practices for R.

The code makes use of the leaflet package, a geographical information system for the interactive exploration of data. The code can accept user defined parameters to create a single bin (Figure A.2, line 8). The code also has a 'cutting factor' which keeps only 1 in every n images. This improves render speed. After running the code an interactive .html page is available. This has the potential to be shared online on a server. The code also generates a start and end time which was originally required by

the PI_Subsample software (Chapter 2.4.1). It also generates a polygon code which can be used to subset the data.

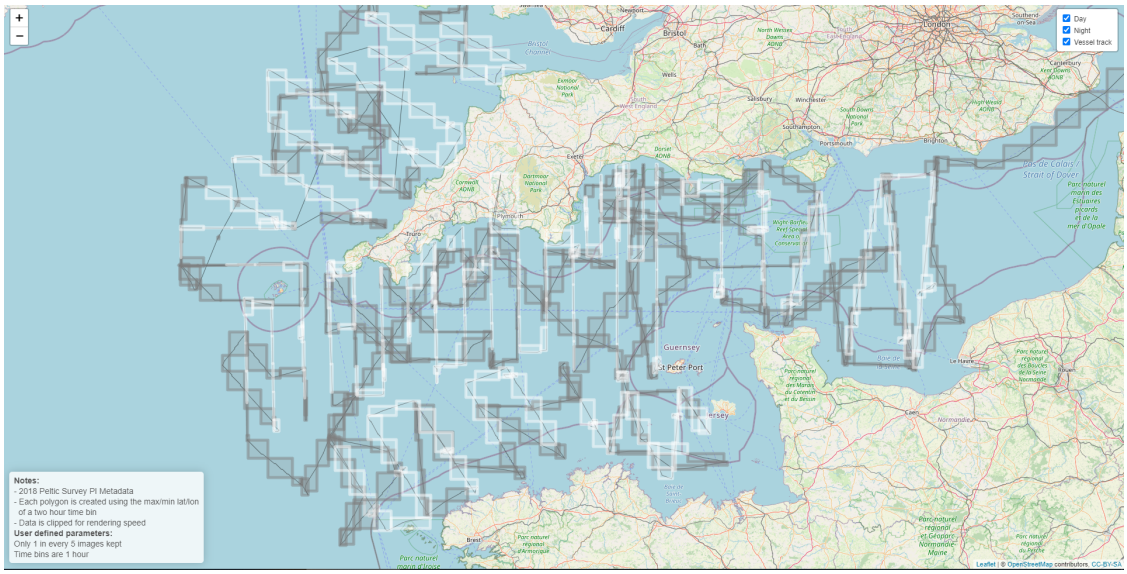

Figure A.1: An overview of the interactive PI Metaexplorer program. This can be launched in any web browser. This example shows data for 2018 Peltic Survey. The notes (bottom left) show the user defined parameters as well as information on how each polygon is created.

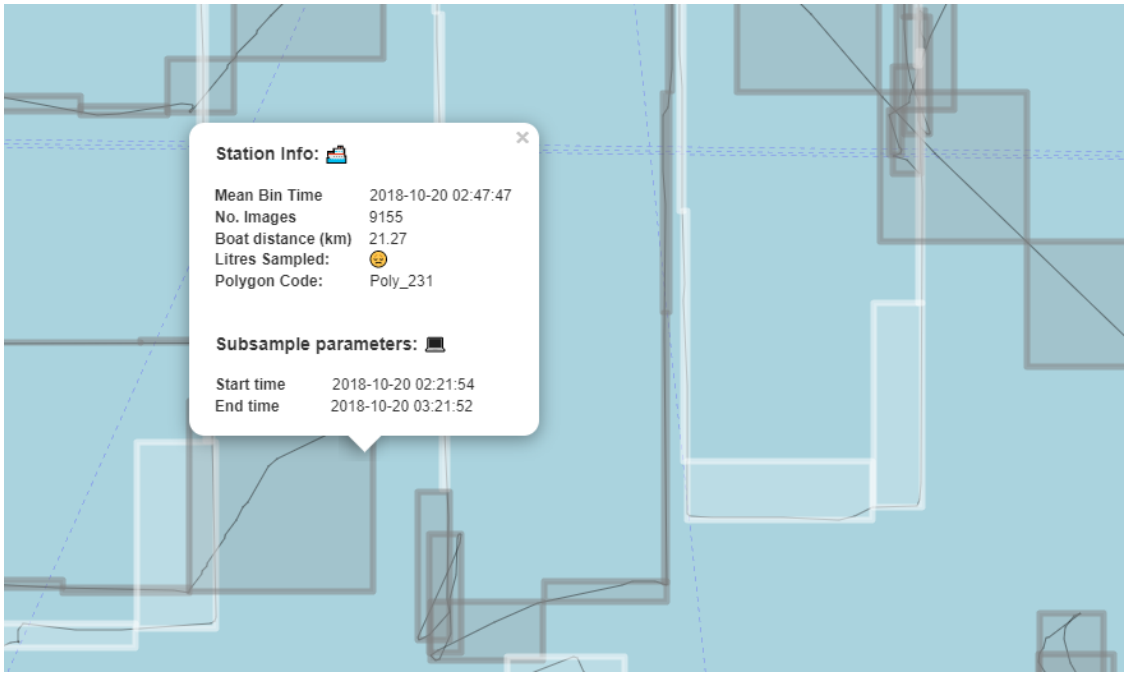

Figure A.2: A selected bin (below the pop up) showing the meta data for the bin. The popup shows the time, distance travelled as well as generating parameters for sub-setting the data.

A.2 USE FOR MSC PROJECT

The code was successfully used to inform educated retrospective subsampling of PI data for an MSc project. The user selected bins to ensure the best spatial coverage. The project investigated the relationship between copepod biomass and remote sensing derived chlorophyll data with consideration of the time lag between peak chlorophyll and biomass.

A.3 CODE

```
# User defined variables

# How many images to keep
# e.g. a cutting factor of 25 would keep 1 in 25 images
cutting_factor = 5

# How much time to split the bins into
user_bin_time  = "1 hour"
# Must be a number and then a peroid
# week
# hour
# month

# ~ # ~ # ~ # Reading in data ####

# Set working directory
setwd("D:\\OneDrive\\OneDrive - University of East Anglia\\PIA_
    ↪ meta_explorer")

# Reading CSV from python script
joined <- read.csv("Metadata_allimages_2018.csv")

# ~ # ~ # ~ # Tidying and truncating data ####

library(dplyr)

# Renaming vars to standard format
joined <- joined %>%
  rename(lat = img_lattitude,
         lon = img_longtitude)

# Clipping the dataset by X, entered before last )
joined_cut      = joined[seq(1, nrow(joined), cutting_factor), ]

# Writing cut dataframe for ships transect
cut_file_name <- paste0("cut_data_by_1in",cutting_factor,".csv")

write.csv(joined_cut, cut_file_name,
          row.names = F)

# Extracting just hour
# Function to select right most characters
substrRight <- function(x, n){
  substr(x, nchar(x)-n+1, nchar(x))
}

# Removing hour and converting time to POSIX
```

```
47  # and extracting date
48  joined_cut <- joined_cut %>%
49    mutate(hour = substrRight(img_time, 8),
50           img_time = as.POSIXct(img_time),
51           date = as.Date(img_time))
52
53  # ~ # ~ # ~ # Binning time by n hours ####
54
55  library(hms)
56  library(dplyr)
57  library(tibbletime)
58
59  # Creating a dummy variable not to loose time
60  joined_cut$bintime_start <- joined_cut$img_time
61
62  # Binning by time
63  binned_time <- joined_cut %>%
64    arrange(bintime_start) %>%
65    as_tbl_time(index = bintime_start) %>%
66    # Time choice below
67    collapse_by(period = user_bin_time,
68                start_date = first(joined_cut$bintime_start),
69                side = "start",
70                # Choosing if to round up
71                clean = F)
72
73
74  # Summarising the needed variables for the plot
75  library(geosphere)
76
77  binned_time <- binned_time %>%
78    group_by(bintime_start) %>%
79    # Creating a lagged variable to calculate distance
80    mutate(lat_prev = lag(lat,1), lon_prev = lag(lon,1) ) %>%
81    # Using Geosphere to calculate a distance between two points
82    mutate(dist = distHaversine(matrix(c(lon_prev, lat_prev), ncol =
        ↪  2),
83                                matrix(c(lon, lat),   ncol = 2)))
                                    ↪ %>%
84    # Summarising variables
85    summarise(dist_m = sum(dist,na.rm=T),
86              count  = length(filename),
87              #
88              avglat = mean(lat),
89              avglon = mean(lon),
90              maxlat = max(lat),
91              minlat = min(lat),
92              maxlon = max(lon),
93              minlon = min(lon),
94              #
95              min_time = min(img_time),
96              max_time = max(img_time),
97              avg_time = mean(img_time),
98              date     = mean(date)) %>%
99    rename(lat = avglat,
100           lon = avglon)
101
102
103 # ~ # ~ # ~ # Calculating sunrise and sunset ####
104
105 library(suncalc)
106
107 # Getting sunlight times based on average date
108 bin_time_sun <- getSunlightTimes(data = binned_time,
109                                  keep = c("sunrise", "sunset"))
110
111 # Pulling those variables neede across
112 # Meta data
113 bin_time_sun$bintime_start  <- binned_time$bintime_start
```

```
114 bin_time_sun$count     <- binned_time$count
115 # Min lats/lons
116 bin_time_sun$maxlat    <- binned_time$maxlat
117 bin_time_sun$minlat    <- binned_time$minlat
118 bin_time_sun$maxlon    <- binned_time$maxlon
119 bin_time_sun$minlon    <- binned_time$minlon
120 # Distance
121 bin_time_sun$dist_m    <- binned_time$dist_m
122 # Mean time
123 bin_time_sun$avg_time  <- binned_time$avg_time
124 bin_time_sun$min_time  <- binned_time$min_time
125 bin_time_sun$max_time  <- binned_time$max_time
126
127 # Calculating if a bin is mainly in day or in the night
128 bin_time_sun <- bin_time_sun %>%
129   mutate(day  = ifelse(avg_time < sunset & avg_time > sunrise, "
        ↪ Day", "Night")) %>%
130   mutate(day  = as.factor(day))
131
132
133 # ~ # ~ # ~ # Naming the polygons for future reference ####
134
135 # Simple name based on first to last time
136 bin_time_sun$poly_name <- paste0("Poly_", 1:as.numeric(count(bin_
        ↪ time_sun)))
137
138 # ~ # ~ # ~ # Saving he grouped data and cleaning the workplace
        ↪ ####
139
140 # Writing the data
141
142 user_time_bin_no_space <- sub("␣", "", user_bin_time)
143
144 time_bin_csv_name <- paste0("timebin_", user_time_bin_no_space,"_
        ↪ cuttingfactor_",cutting_factor,".csv")
145
146 write.csv(bin_time_sun, time_bin_csv_name,
147           row.names = F)
148
149 # Removing all but that nasty csv
150 rm(list=ls()[! ls() %in% c("joined", "cutting_factor","user_bin_
        ↪ time","time_bin_csv_name","cut_file_name")])
151
152 # Reading data back in
153 plotdata <- read.csv(time_bin_csv_name)
154
155 # Changing time to factor
156 plotdata$day <- as.factor(plotdata$day)
157
158
159 # ~ # ~ # ~ # Creating the polygons ####
160
161 library(sf)
162
163 lst <- lapply(1:nrow(plotdata), function(x){
164   # Creating the polygons using the min and max lats and lons
165   # for each hour (need FIX here - better method?)
166   res <- matrix(c(plotdata[x, 'maxlon'], plotdata[x, 'maxlat'],
167                   plotdata[x, 'maxlon'], plotdata[x, 'minlat'],
168                   plotdata[x, 'minlon'], plotdata[x, 'minlat'],
169                   plotdata[x, 'minlon'], plotdata[x, 'maxlat'],
170                   plotdata[x, 'maxlon'], plotdata[x, 'maxlat'])
171                 , ncol =2, byrow = T
172   )
173   st_polygon(list(res))})
174
175 # Extracting the geometry per polygon
176 # Create simple feature geometry list column
177 plotdata$geomtry <- st_sfc(lst)
178 str(plotdata)
```

```
179 data_assf <- st_sf(plotdata)
180
181
182
183 # ~ # ~ # ~ # Splitting the shapefile data frame by day and night ####
184
185 # night
186 night_only <- data_assf[data_assf$day == 'Night',]
187 # day
188 day_only   <- data_assf[data_assf$day == 'Day',]
189
190
191 # ~ # ~ # ~ # Getting the boat transect and changing to sp ####
192
193 # Reading cut file
194 transect <- read.csv(cut_file_name)
195
196 # Changing image time to indexing
197 transect <- transect %>%
198   mutate(img_time_num = as.numeric(as.POSIXct(img_time, tz = "GMT")))
199
200 # ~ # ~ # ~ # Plotting in Leaflet ####
201
202 library(leaflet)
203 library(leaflet.extras)
204 library(htmltools)
205
206 # Title
207 rr <- tags$div(
208   (HTML("<b>Notes:</b><br>",
209         "- 2018 Peltic Survey PI Metadata <br>",
210         "- Each polygon is created using the max/min lat/lon<br>",
211         "  of a two hour time bin <br>",
212         "- Data is clipped for rendering speed <br>",
213         "<b>User defined parameters:</b><br>",
214     "Only 1 in every", as.character(cutting_factor), "images kept", "<br>",
215         "Time bins are", as.character(user_bin_time), "<br>"
216     )))
217
218 map <- leaflet() %>%
219   #
220   addTiles() %>%
221   #
222   addPolylines(data = transect,
223                lng = ~lon,
224                lat = ~lat,
225                group = "Vessel track",
226                weight = 1,
227                color = "black") %>%
228   #
229   addPolygons(data = night_only,
230               group = "Night",
231               color = "Grey",
232               highlightOptions = highlightOptions(color = "red",
233                                                   weight = 4,
234                                                   bringToFront = TRUE),
235               popup = ~paste0("<h3>Station Info: &#128674;</h3>",
236                               "<b>Mean Bin Time           </b>",
237                               avg_time,
238                               "<br>",
239                               "<b>No. Images                 </b>",
```

```
240                                 count*cutting_factor,
241                                 "<br>",
242                                 "<b>Boat␣distance␣(km)␣  &
                                        ↪ nbsp;</b>",
243                                 round(dist_m/1000, digits = 2),
244                                 "<br>",
245                                 "<b>Litres␣Sampled:␣␣  &
                                        ↪ nbsp;    &
                                        ↪ nbsp;</b>",
246                                 "&#128542;␣",
247                                 "<br>",
248                                 "<b>Polygon␣Code:␣␣   
                                        ↪ ;     
                                        ↪   </b>",
249                                 poly_name,
250                                 "<br>",
251                                 "<br>",
252                                 "<h3>Subsample␣parameters:␣&
                                        ↪ #128187;</h3>",
253                                 "<b>Start␣time␣   &
                                        ↪ nbsp;    &
                                        ↪ nbsp; </b>",
254                                 min_time,
255                                 "<br>",
256                                 "<b>End␣time␣    
                                        ↪ ;     
                                        ↪   </b>",
257                                 max_time)) %>%
258     #
259     addPolygons(data = day_only,
260                 group = "Day",
261                 color = "White",
262                 highlightOptions = highlightOptions(color = "red",
263                                                     weight = 4,
264                                                     bringToFront =
                                                            ↪ TRUE),
265                 popup = ~paste0("<h3>Station␣Info:␣&#128674;</h3>",
266                                 "<b>Mean␣Bin␣Time␣   
                                        ↪      &
                                        ↪ nbsp; </b>",
267                                 avg_time,
268                                 "<br>",
269                                 "<b>No.␣Images   &
                                        ↪ nbsp;    &
                                        ↪ nbsp;    &
                                        ↪ nbsp;  ␣</b>",
270                                 count*cutting_factor,
271                                 "<br>",
272                                 "<b>Boat␣distance␣(km)␣  &
                                        ↪ nbsp;</b>",
273                                 round(dist_m/1000, digits = 2),
274                                 "<br>",
275                                 "<b>Litres␣Sampled:␣␣  &
                                        ↪ nbsp;    &
                                        ↪ nbsp;</b>",
276                                 "&#128542;␣",
277                                 "<br>",
278                                 "<b>Polygon␣Code:␣␣   
                                        ↪ ;     
                                        ↪   </b>",
279                                 poly_name,
280                                 "<br>",
281                                 "<br>",
282                                 "<h3>Subsample␣parameters:␣&
                                        ↪ #128187;</h3>",
283                                 "<b>Start␣time␣   &
                                        ↪ nbsp;    &
                                        ↪ nbsp; </b>",
284                                 min_time,
285                                 "<br>",
```

```
286                                   "<b>End time     
                                         ↪ ;     
                                         ↪   </b>",
287                                   max_time)) %>%
288     addLayersControl(
289       overlayGroups = c("Day", "Night", "Vessel track"),
290       options = layersControlOptions(collapsed = FALSE)) %>%
291     addControl(rr, position = "bottomleft")
292
293 map
```

B

HARMONISING CONTINUOUS DATA STREAMS

B.1 INTRODUCTION

This tool is used to harmonise continuous data streams to user defined space and size bins (lines 12 & 13). Data must be binned in order to run statistical analysis even if the bins are small. Although used in thesis for zooplankton data the tool can be used with any continuous data as long as longitude, latitude and datetime field are provided. The only consideration is the minimum temporal bin of the raw data. For example, in Chapter 5 the minimum bin size was 1 nautical mile due to the fisheries acoustic data being binned to that spatial resolution. Finally the only consideration is: the smaller the time / bin sizes the longer the code takes to run.

B.2 USE IN CHAPTERS

The code was used to merge various continuous data the thesis. For Chapter 4 only a spatial merge was performed, this code is not shown but is essentially the same code

excluded lines 73-142. In Chapter 5 data were merged by both space and time and this is the code used.

B.3 CODE

```
library(reshape2)
library(here)
library(tidyverse)
conflict_prefer("filter", "dplyr")

gm_mean = function(x, na.rm=TRUE){
  exp(sum(log(x[x > 0]), na.rm=T) / length(x))
} # geomean function

# changeable paramters

temporal_bin_size = "20 mins"
spatial_grid_size = 0.05# must be x/1 of a degree


#### reading in raw data ####

# Ferrybox
ferrybox_data <- read_csv(here("data","ferrybox", "Peltic_fb_2020_JS.csv")) %>%
  rename(lat = Latitude,
         lon = Longitude) %>%
  dplyr::select(dateTime,
                variable,
                Mean,
                lat,
                lon)

ferrybox_data <- dcast(ferrybox_data, lat + lon + dateTime ~ variable,
                       value.var = "Mean",
                       fun.aggregate = mean)


# Fish
fish_noJUV <- read_csv(here("data",
                            "fish_data",
                            "PELTIC20_nasc_by_species_noJUV_agg_final.csv")) %>%
  rename(lon = Lon_M,
         lat = Lat_M) %>%
  drop_na() %>%
  mutate(dateTime = paste(Date_M, hours_mins),
         dateTime = as.POSIXct(dateTime, format = "%d/%m/%Y %H:%M:%OS",
                               tz = "UTC")) %>%
  select(-c(Interval,
            Date_M,
            Time_M,
            hours_mins,
            NASC.SBF))

# Plankton
cops_only <- read_csv(here("data", "PI_FB_FC_ST", "Copepodnosubsample.csv")) %>%
  drop_na() %>%
  mutate(copepods_logged = log10(Major)+1) %>%
  select(Major,
```

```
54          DateTime,
55          lat,
56          lon) %>%
57   rename(dateTime = DateTime)  %>%
58   mutate(length_um  = Major*10,
59          length_mm  = length_um/1000,
60          taxa_ww_mg = 0.0299*(length_mm^2.8348),
61          taxa_ww_ug = taxa_ww_mg * 1000,
62          bin_min_time = as.POSIXct(plyr::round_any(dateTime, 600,
               ↪ f = floor),
63                                    tz = "UTC", origin = "
                                         ↪ 1960-01-01"),
64          bin_max_time = as.POSIXct(plyr::round_any(dateTime, 600,
               ↪ f = ceiling),
65                                    tz = "UTC", origin = "
                                         ↪ 1960-01-01"))
66
67
68 # Plankton bin metadata
69 # PI_bin_times <- read_csv(here("data", "PI_FB_FC_ST", "bin_counts
      ↪ .csv"))$datetime
70
71 #### creating temporal and spatial bins ####
72
73 ## Time
74
75 temporal_bins <- as.data.frame(
76   seq(round(min(ferrybox_data$dateTime - 24*3600), "hour"),
77       round(max(ferrybox_data$dateTime + 24/3600), "hour"),
78       by = temporal_bin_size)) %>%
79   rename(temporal_bins_start = 1) %>%
80   mutate(temporal_bins_start_fact = as.factor(temporal_bins_start)
        ↪ )
81
82 temporal_bins$temporal_bin_id <- ids::random_id(as.numeric(count(
      ↪ temporal_bins)), 3)
83
84 ## Space
85
86 gridder_consistentXY <- function(dataframe, resolution){
87
88   sing_res <- resolution
89   xCellSizeGrid <- yCellSizeGrid <- sing_res
90
91   # Projection
92   wgs.84 <- CRS("+proj=longlat␣+datum=WGS84␣+no_defs␣+ellps=WGS84␣
        ↪ +towgs84=0,0,0")
93
94   datasetSP <- SpatialPointsDataFrame(coords = dataframe[,c("lon",
        ↪  "lat")],
95                                       data = data.frame("id" = 1:
                                            ↪ nrow(dataframe)),
96                                       proj4string = wgs.84)
97
98
99   extentDatasetSP <-extent(datasetSP)
100
101
102   mincellX <- -8.25
103   mincellY <- 48
104
105   maxcellX <- -2
106   maxcellY <- 53
107
108   xgridcount <- ((maxcellX-mincellX)/xCellSizeGrid)+2
109   ygridcount <- abs((maxcellY-mincellY)/yCellSizeGrid)+2
110
111   grid <- GridTopology(cellcentre.offset = c(mincellX, mincellY),
112                        cellsize = c(xCellSizeGrid, yCellSizeGrid),
113                        cells.dim = c(xgridcount, ygridcount))
```

```
114
115   # Create SpatialGrid object
116   gridSpatial <- SpatialGrid(grid = grid, proj4string = wgs.84)
117
118
119   # Convert to SpatialPixels object
120   gridSpatialPixels <- as(gridSpatial, "SpatialPixels")
121
122   # Convert to SpatialPolygons object
123   gridSpatialPolygons <- as(gridSpatialPixels, "SpatialPolygons")
124
125   # Add 'id' and 'values' to every polygon
126   gridSpatialPolygons$id <- 1:nrow(coordinates(gridSpatialPolygons
        ↪ ))
127   # gridSpatialPolygons$values <- paste("grid_",sing_res , "_",1:
        ↪ nrow(coordinates(gridSpatialPolygons)), sep = "")
128
129   # Get attributes from polygons
130   samplePointsInPolygons2 <- datasetSP %over% gridSpatialPolygons
131
132   # Result
133   datasetResult <- data.frame(dataframe, samplePointsInPolygons2)
134
135   output_df <- sf::st_as_sf(gridSpatialPolygons)
136   output_df <- merge(datasetResult, output_df, by = "id")
137   output_df <- sf::st_as_sf(output_df)
138
139   return(output_df)
140
141
142 }
143
144
145 gridder_consistentXY_gridgeo <- function(dataframe, resolution){
146
147   sing_res <- resolution
148   xCellSizeGrid <- yCellSizeGrid <- sing_res
149
150   # Projection
151   wgs.84 <- CRS("+proj=longlat␣+datum=WGS84␣+no_defs␣+ellps=WGS84␣
        ↪ +towgs84=0,0,0")
152
153   datasetSP <- SpatialPointsDataFrame(coords = dataframe[,c("lon",
        ↪  "lat")],
154                                       data = data.frame("id" = 1:
                                              ↪ nrow(dataframe)),
155                                       proj4string = wgs.84)
156
157
158   extentDatasetSP <-extent(datasetSP)
159
160
161   mincellX <- -8.25
162   mincellY <- 48
163
164   maxcellX <- -2
165   maxcellY <- 53
166
167   xgridcount <- ((maxcellX-mincellX)/xCellSizeGrid)+2
168   ygridcount <- abs((maxcellY-mincellY)/yCellSizeGrid)+2
169
170   grid <- GridTopology(cellcentre.offset = c(mincellX, mincellY),
171                        cellsize = c(xCellSizeGrid, yCellSizeGrid),
172                        cells.dim = c(xgridcount, ygridcount))
173
174   # Create SpatialGrid object
175   gridSpatial <- SpatialGrid(grid = grid, proj4string = wgs.84)
176
177
178   # Convert to SpatialPixels object
```

```
179   gridSpatialPixels <- as(gridSpatial, "SpatialPixels")
180
181   # Convert to SpatialPolygons object
182   gridSpatialPolygons <- as(gridSpatialPixels, "SpatialPolygons")
183
184   # Add 'id' and 'values' to every polygon
185   gridSpatialPolygons$id <- 1:nrow(coordinates(gridSpatialPolygons
          ↪ ))
186   # gridSpatialPolygons$values <- paste("grid_",sing_res , "_",1:
          ↪ nrow(coordinates(gridSpatialPolygons)), sep = "")
187
188   # Get attributes from polygons
189   samplePointsInPolygons2 <- datasetSP %over% gridSpatialPolygons
190
191   # Result
192   datasetResult <- data.frame(dataframe, samplePointsInPolygons2)
193
194   gridded_geo <- sf::st_as_sf(gridSpatialPolygons)
195   # output_df <- merge(datasetResult, output_df, by = "id")
196   # output_df <- sf::st_as_sf(output_df)
197
198   return(gridded_geo)
199
200 }
201
202
203
204
205
206 #### Assigning time bins to ferrybox, copepods and fishing ####
207
208 ## Time
209
210 # Ferrybox
211 ferrybox_data$temporal_bins_start_fact <- cut(ferrybox_data$
        ↪ dateTime,
212                                               breaks = temporal_
                                                    ↪ bins$temporal_
                                                    ↪ bins_start)
213 ferrybox_data <- merge(ferrybox_data, temporal_bins,
214                        by = "temporal_bins_start_fact") %>%
215   select(-temporal_bins_start_fact)
216
217 # Fish data
218 fish_noJUV$temporal_bins_start_fact <- cut(fish_noJUV$dateTime,
219                                            breaks = temporal_bins$
                                                 ↪ temporal_bins_
                                                 ↪ start)
220 fish_noJUV <- merge(fish_noJUV, temporal_bins,
221                     by = "temporal_bins_start_fact") %>%
222   select(-temporal_bins_start_fact)
223
224 # Plankton data
225 cops_only$temporal_bins_start_fact <- cut(cops_only$dateTime,
226                                           breaks = temporal_bins$
                                                ↪ temporal_bins_start
                                                ↪ )
227 cops_only <- merge(cops_only, temporal_bins,
228                    by = "temporal_bins_start_fact") %>%
229   select(-temporal_bins_start_fact)
230
231
232 #### Assigning space bins to ferrybox, copepods and fishing ####
233
234 # Ferrybox
235 ferrybox_data <- gridder_consistentXY(ferrybox_data, spatial_grid_
        ↪ size) %>%
236   mutate(grid_time_id = paste(id, temporal_bin_id, sep =   "_"))
        ↪ %>%
237   select(temporal_bin_id,
```

```
238           everything())
239
240 # Fish data
241 fish_noJUV <- gridder_consistentXY(fish_noJUV, spatial_grid_size)
      ↪ %>%
242   mutate(grid_time_id = paste(id, temporal_bin_id, sep =  "_"))
        ↪ %>%
243   select(temporal_bin_id,
244           everything())
245
246 # Cops data
247 cops_only <- gridder_consistentXY(cops_only, spatial_grid_size)
      ↪ %>%
248   mutate(grid_time_id = paste(id, temporal_bin_id, sep =   "_"))
        ↪ %>%
249   select(temporal_bin_id,
250           everything())
251
252 #### Summarising per grid_time_id
253
254 # Ferrybox
255 ferrybox_grouped <- ferrybox_data %>%
256   group_by(as.factor(grid_time_id)) %>%
257   summarise(mean_lat  = mean(lat, na.rm = T),
258             mean_lon  = mean(lon, na.rm = T),
259             mean_time = mean(dateTime, na.rm = T),
260             #
261             mean_fluor  = mean(FLUORS, na.rm = T),
262             mean_FTU    = mean(FTU, na.rm = T),
263             mean_sal    = mean(SAL, na.rm = T),
264             mean_sst    = mean(SST, na.rm = T),
265             mean_windsp = mean(WINDSP, na.rm = T))%>%
266   sf::st_drop_geometry() %>%
267   rename("grid_time_id" = 1) %>%
268   select(-c(mean_lat,
269             mean_time,
270             mean_lon))
271
272
273 # Fishies
274 fish_noJUV_grouped <- fish_noJUV %>%
275   group_by(as.factor(grid_time_id)) %>%
276   summarise(mean_lat  = mean(lat, na.rm = T),
277             mean_lon  = mean(lon, na.rm = T),
278             mean_time_fish = mean(dateTime, na.rm = T),
279             #
280             mean_depth = mean(Depth, na.rm = T),
281             #
282             SPR_mean = mean(SPR, na.rm = T),
283             PIL_mean = mean(PIL, na.rm = T),
284             ANE_mean = mean(ANE, na.rm = T),
285             HER_mean = mean(HER, na.rm = T),
286             HOM_mean = mean(HOM, na.rm = T),
287             BOF_mean = mean(BOF, na.rm = T),
288             WHB_mean = mean(WHB, na.rm = T))%>%
289   sf::st_drop_geometry() %>%
290   rename("grid_time_id" = 1) %>%
291   select(-c(mean_lat,
292             mean_lon,
293             mean_time_fish
294   ))
295
296 # Plankton
297 # first need to figure out run time per bin
298
299 cops_grouped <- cops_only %>%
300   group_by(as.factor(grid_time_id)) %>%
301   summarise(mean_lat  = mean(lat, na.rm = T),
302             mean_lon  = mean(lon, na.rm = T),
```

```
303             mean_time_cops = mean(dateTime, na.rm = T),
304
305             PI_on_time = abs(difftime(min(bin_min_time),
306                                       max(bin_max_time),
307                                       units = "mins")),
308             PI_h20samp_m3 = as.numeric(22*PI_on_time/1000),
309
310             cop_len_geomean_um = gm_mean(length_um),
311
312             count_count = length(Major),
313             cop_mean_ww_ug = mean(taxa_ww_ug, na.rm = T),
314             cop_tot_ww_ug  = sum(taxa_ww_ug, na.rm = T))%>%
315
316   mutate(copepod_abun = count_count*(1/PI_h20samp_m3),
317          cop_bio_ugM3_total = cop_tot_ww_ug*(1/PI_h20samp_m3),
318          cop_bio_mgM3_total = cop_bio_ugM3_total/1000) %>%
319   rename("grid_time_id" = 1) %>%
320   sf::st_drop_geometry() %>%
321   select(-c(cop_bio_ugM3_total,
322             cop_mean_ww_ug,
323             cop_tot_ww_ug,
324             PI_h20samp_m3,
325             PI_on_time,
326             count_count,
327             mean_lat,
328             mean_lon))
329
330
331
332
333
334 ultimatedata <- merge(cops_grouped, fish_noJUV_grouped, by = "grid
    ↪ _time_id", all = T)
335
336 ultimatedata <- merge(ultimatedata, ferrybox_grouped, by = "grid_
    ↪ time_id", all = T)
337
338 ultimatedata <- ultimatedata %>%
339   filter_if(~is.numeric(.), all_vars(!is.infinite(.)))
340
341
342 # ultimatedata <- ultimatedata %>%
343 #   filter_at(vars(SPR_mean, PIL_mean, ANE_mean, HER_mean, HOM_
    ↪ mean, BOF_mean, WHB_mean), . > 0.0005)
```

C

CONTRIBUTION TO DEVELOPMENT OF THE PI

C.1 INTRODUCTION

Part of the PhD aims were to develop the instrument. Feedback between myself as user of the instrument and designers helped development. List below are some of the more major changes to the PI I initiated.

C.2 ITEMS

- **Max images per minute.** This is likely my most important contribution to the development of the PI. The PI originally had a buffer that was used to store images that were waiting to be written to disk. This happened during areas of high particulate concentrations (sand etc.). If the buffer was full, the next 10 min bin would be affected even if there was a more normal amount of particles. A 'bad' ten minutes could effective spoil a few hours data with this method. Using this max images per min meant that if a patch of high particulate was hit then the next bin would be unaffected as the software would stop trying to

write the previous bins images to disk and instead record the number of images waiting. This resulted in a significant increase in the number of usable bins following a survey.

- **Preview window on PIA_Subsample redesign** (Figure 2.5). This was designed to be more user friendly and show more information when previewing a sample (e.g. number of images in timeframe). Several bugs which crashed the GUI were also removed during this process.

- **ExtHDSpeedTest.exe - hard drive testing software**. This software was written by one of the PI developers to try to quantify the gains that could be made by moving from spinning hard drives to solid states disks. It turns out gains were marginal.

- **Standard Operating Procedure (SOP)**. The SOP for the PI was written for deployment aboard the RV Cefas Endeavour in the final few months of the PhD. This is now used by other PI users.