Lina Becker

Mesozooplankton Community Dynamics in Response to Suppressed Upwelling

A Case Study Comparing the Gulf of Panama and the Gulf of Chiriquí Using Automated Image Analysis

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2025

Lina Becker

Mesozooplankton Community Dynamics in Response to Suppressed Upwelling

A Case Study Comparing the Gulf of Panama and the Gulf of Chiriquí Using Automated Image Analysis

Master's in Marine and Coastal Systems

Work performed under the supervision of:

Dr. Isabel Maria de Paiva Pinto Mendes

Dr. Ralf Schiebel

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2025

Declaração de autoria do trabalho

Declaro ser o(a) autor(a) deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Statement of authorship

I declare to be the author of this work, which is original and unpublished. Authors and works consulted are duly cited in the text and included in the reference list.

Lina Becker

© Lina Becker

A Universidade do Algarve reserva para si o direito, em conformidade com o disposto no Código do Direito de Autor e dos Direitos Conexos, de arquivar, reproduzir e publicar a obra, independentemente do meio utilizado, bem como de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição para fins meramente educacionais ou de investigação e não comerciais, conquanto seja dado o devido crédito ao autor e editor respetivos.

The University of Algarve reserves the right, in accordance with the provisions of the Copyright Law and Related Rights, to archive, reproduce and publish the work, regardless of the used medium, as well as to disseminate it through scientific repositories and to allow its copy and distribution for purely educational, research and non-commercial purposes, warranting that due credit is given to the respective author and publisher.

Dedication and acknowledgments

I would like to express my deepest gratitude to all those who have supported and accompanied me throughout the course of this Master's journey.

First and foremost, I am sincerely grateful to Dr. Ralf Schiebel (MPI for Chemistry, Mainz) for his excellent supervision, guidance, and inspiration throughout this project. I also thank Dr. Isabel Maria de Paiva Pinto Mendes (University of Algarve) for her valuable support and organizational assistance, which greatly facilitated my work.

My warm thanks go to the crew of the S/Y Eugen Seibold for their dedication and enthusiasm during the field campaign, without which this research would not have been possible. I would like to thank Anne Kiefer (Mainz) for her kind assistance in the laboratory, and to the Smithsonian Tropical Research Institute (STRI, Panama) team, Dr. Aaron O'Dea, Dr. Carmen Perez, Andrew Sellers, and Steve Paton, for their friendly collaboration and expertise.

I also wish to acknowledge the scientific and technical support provided by colleagues at the IGB Berlin Stechlinsee, in particular Dr. Jens Christian Nejstgaard, Dr. Stella A. Berger, and Christian Dilewski, whose expertise and encouragement made the laboratory work both productive and enjoyable. Special thanks go to Kananat Siangsano for the excellent collaboration on programming, AI expertise, and Python scripts, and to Katerina Symiakaki for her tireless technical support in coding.

I am especially thankful to my fellow Master's colleagues at the University of Algarve, who made this program a truly rewarding academic and personal experience.

Finally, and most importantly, I owe my deepest gratitude to my parents and friends for their unwavering mental support, patience, and encouragement.

Abstract

This thesis investigates mesozooplankton community dynamics in response to anomalous oceanographic conditions, focusing on the 2025 collapse of the seasonal upwelling in the Gulf of Panama. Using a comparative approach with the neighboring Gulf of Chiriquí, the study integrates semi-automated FlowCam imaging with supervised machine learning classifiers to process over 750,000 plankton images. This methodology enabled reliable abundance, and biomass estimates for dominant taxa, particularly copepods, at resolutions and scales unattainable through traditional microscopy.

Hydrographic measurements revealed strongly stratified conditions in the Gulf of Panama, with suppressed nutrient input into the euphotic zone. In contrast, the Gulf of Chiriquí maintained relatively stable conditions. Ecological consequences were profound: the Gulf of Panama, typically characterized by high productivity due to upwelling, exhibited markedly reduced mesozooplankton abundance and biomass. Conversely, the Gulf of Chiriquí showed unexpectedly higher biomass across most depths, highlighting its role as a potential analogue for future oligotrophic states under climate-driven variability.

Copepods remained the numerically and functionally dominant group, underscoring their resilience and central role in trophic transfer. However, size-structured analyses revealed that biomass patterns were not solely linked to abundance but strongly influenced by ontogenetic stage distribution. The disproportionate contribution of larger size fractions to carbon stocks emphasizes the importance of monitoring size structure in addition to abundance.

Methodologically, the study demonstrates the value of hybrid workflows combining automated imaging with targeted manual curation. Despite challenges such as misclassification of transparent taxa and debris, semi-automated imaging proves scalable and essential for long-term ecological monitoring.

Ultimately, the findings provide critical baseline data for understanding mesozooplankton dynamics under anomalous conditions. They highlight both resilience and vulnerability: while copepods persist, the productivity of short food chains may increasingly be at risk in a warming and more variable ocean.

Keywords: mesozooplankton, copepods, upwelling, FlowCam, automated imaging, Tropical Eastern Pacific

Resumo

Esta dissertação analisa a dinâmica das comunidades de mesozooplâncton em resposta a condições oceanográficas anómalas, tomando como estudo de caso a falha do afloramento sazonal no Golfo do Panamá em 2025 e a comparação com o Golfo de Chiriquí (Pacífico Oriental Tropical). O trabalho utiliza uma abordagem metodológica inovadora, combinando análises hidrográficas detalhadas com técnicas de imagem automatizada (FlowCam®) e classificadores de aprendizagem automática, gerando um conjunto de mais de 750.000 imagens de plâncton processadas e classificadas. Esta estratégia permitiu estimativas robustas de abundância, composição e biomassa do mesozooplâncton, especialmente dos copépodes, superando limitações associadas a métodos tradicionais de microscopia.

Do ponto de vista oceanográfico, os dados de temperatura, salinidade e clorofila-a revelaram contrastes marcantes entre os dois sistemas. O Golfo do Panamá, geralmente caracterizado por intensa produtividade primária impulsionada por ventos alísios e transporte de Ekman, não apresentou o afloramento esperado em 2025. Perfis verticais mostraram uma coluna d'água fortemente estratificada, ausência de ressurgência de águas frias e profundas, e concentrações reduzidas de nutrientes na zona eufótica. Em contrapartida, o Golfo de Chiriquí manteve condições oligotróficas estáveis, com níveis de biomassa fitoplanctônica modestos, mas relativamente consistentes.

Os resultados ecológicos dessa anomalia foram profundos. No Golfo do Panamá, a ausência de afloramento reduziu drasticamente a abundância de mesozooplâncton e a biomassa de copépodes, comprometendo a base trófica de cadeias alimentares curtas. Já o Golfo de Chiriquí, tradicionalmente considerado oligotrófico, apresentou valores surpreendentemente mais elevados de abundância e biomassa de mesozooplâncton em quase todas as profundidades amostradas. Essa inversão das expectativas sugere que, em cenários de colapso de afloramentos, sistemas oligotróficos podem temporariamente sustentar comunidades planctónicas mais robustas do que sistemas historicamente produtivos.

A composição taxonómica foi dominada por copépodes, que representaram entre 66 % e 79 % dos indivíduos amostrados. Esse resultado reforça o papel central desse grupo na transferência de energia entre fitoplâncton e níveis tróficos superiores, incluindo peixes, mamíferos marinhos e organismos gelatinosos. Análises de diversidade indicaram que a fração de menor tamanho (<500 μm) concentrou maior riqueza de táxons e diversidade de Shannon, particularmente no

Golfo de Chiriquí, enquanto a fração >1000 µm foi dominada por poucos grupos de grande porte, resultando em menor diversidade.

As análises de biomassa revelaram ainda a importância de considerar a estrutura de tamanhos. Embora a abundância fosse frequentemente maior nas frações menores, a biomassa estava concentrada principalmente na fração de 500–1000 µm, evidenciando a contribuição desproporcional de indivíduos de maior porte para os reservatórios de carbono. Essa observação ressalta que mudanças no tamanho médio dos copépodes podem ter implicações significativas para o fluxo de carbono no oceano e para a eficiência da bomba biológica.

Do ponto de vista metodológico, este trabalho destaca a relevância da integração entre o processamento de imagens automatizado e a curadoria manual de dados. O uso do FlowCam® aliado a classificadores treinados de forma interativa permitiu alcançar uma precisão superior a 90 % para os principais grupos, apesar das dificuldades em distinguir organismos transparentes ou fragmentados, como cnidários e cladócera, frequentemente confundidos com detritos. O método mostrou ser escalável, reprodutível e apto a fornecer séries temporais de elevada resolução necessárias para monitorizar ecossistemas dinâmicos sob a influência da variabilidade climática.

A comparação entre amostras fracionadas por tamanho e amostras combinadas demonstrou diferenças metodológicas importantes. Amostras combinadas tenderam a superestimar a biomassa em relação à soma das frações, especialmente nas camadas superficiais, sugerindo que a recombinação artificial pode alterar a distribuição de tamanhos e introduzir vieses. Este resultado aponta para a necessidade de haver uma padronização dos protocolos de monitorização do mesozooplâncton, sobretudo em estudos comparativos de longo prazo.

Na discussão ecológica, a inversão de padrões entre os dois golfos ganha especial relevância. Historicamente, o Golfo do Panamá apresenta elevadas concentrações de diatomáceas e grandes populações de copépodes durante a estação seca. Contudo, em 2025, a combinação de ventos alísios enfraquecidos e o deslocamento da Zona de Convergência Intertropical resultaram na completa supressão do afloramento. Isso levou à manutenção de águas superficiais quentes, estratificação intensa e escassez de nutrientes, comprometendo a base produtiva do sistema. No Golfo de Chiriquí, por sua vez, a ausência de afloramento permitiu relativa estabilidade, e as comunidades planctónicas responderam de forma mais robusta, acumulando biomassa inclusive nas camadas subsuperficiais (200–300 m).

Esses resultados têm implicações amplas para a compreensão da resiliência dos ecossistemas tropicais face às alterações climáticas. O colapso do afloramento no Golfo do Panamá representa um cenário plausível de futuro, em que a variabilidade atmosférica e oceânica pode comprometer eventos sazonais críticos para a produtividade marinha. A comparação com o Golfo de Chiriquí fornece um referencial valioso sobre como sistemas oligotróficos podem operar sob restrição de nutrientes, apontando para adaptações tróficas que favorecem a reciclagem de matéria orgânica e o uso de vias detritívoras.

Do ponto de vista da gestão, os resultados destacam riscos para a segurança alimentar e económica de comunidades costeiras dependentes de cadeias alimentares curtas e altamente sazonais. A redução da biomassa de copépodes compromete a base de recursos pesqueiros, como anchovas e sardinhas, essenciais para a subsistência e comércio local. Além disso, a ausência do arrefecimento associada ao afloramento aumenta a probabilidade de stresse térmico nos recifes de coral, potencializando episódios de branqueamento.

Em síntese, esta dissertação contribui de forma significativa para o entendimento das respostas do mesozooplâncton a eventos anómalos de oceanografia física, ao mesmo tempo em que valida metodologias inovadoras de monitorização. Ao demonstrar que comunidades planctónicas podem reorganizarse de forma inesperada perante o colapso dos afloramentos, os resultados enfatizam a importância de dados de alta resolução para antecipar os impactos ecológicos e socioeconómicos.

Conclui-se que, embora os copépodes mantenham o seu papel central e demonstrem alguma resiliência, a produtividade das cadeias alimentares marinhas tropicais podem estar cada vez mais vulneráveis num oceano mais quente e mais variável. Assim, os resultados deste estudo fornecem uma base essencial para compreender a dinâmica do mesozooplâncton no Pacífico Oriental Tropical e para o desenvolvimento de estratégias de gestão sustentável dos recursos marinhos em cenários de alteração climática.

Palavras-chave: mesozooplâncton, copépodes, afloramento, FlowCam, imagem automatizada, Pacífico Oriental Tropical

Table of Contents

List of Figures

List of Tables

Abbreviation list

AI – Artificial Intelligence

CTD – Conductivity, Temperature, Depth

DCM – Deep chlorophyll maximum

ENSO – El Niño–Southern Oscillation

GOC – Gulf of Chiriquí

GOP – Gulf of Panama

ITCZ – Intertropical Convergence Zone

MLD – Mixed layer depth

N^2 – Brunt–Väisälä frequency

PCoA – Principal Coordinates Analysis

R/V – Research Vessel

SST – Sea surface temperature

TEP – Tropical Eastern Pacific

$\sigma\theta$ – Potential density anomaly

PSSdb – Pelagic Size Structure database

1. Introduction

The health and resilience of marine ecosystems are vital for humanity. Oceans provide food security, livelihoods, and climate regulation for billions of people, while supporting biodiversity and global biogeochemical cycles(Intergovernmental Panel on Climate Change (IPCC), 2023b). Central to these processes are planktonic organisms, which form the foundation of marine food webs and drive key ecological functions, such as carbon sequestration and nutrient cycling (Wanek et al., 2025). However, rapid environmental change, including climate-driven disruptions to oceanographic processes, now threatens the resilience and productivity of these critical systems (O'Dea et al., 2025). Understanding how plankton communities respond to such disruptions is essential for safeguarding marine resources and ecosystem services on which society depends.

1.1 Mesozooplankton and the role of copepods

Mesozooplankton (0.2–20 mm) represent a diverse and ecologically central group within the plankton community, playing a pivotal role in energy transfer and nutrient cycling (Banse, 1995). They serve as primary consumers in pelagic food webs, transferring energy from phytoplankton to higher trophic levels such as fish, gelatinous mesozooplankton, and marine mammals (Frangoulis et al., 2005; Kiko et al., 2020; Turner, 2004). Among them, copepods are particularly important due to their global abundance, ecological versatility, and high turnover rates. Their widespread presence and biomass dominance in plankton net samples make them a key component of energy flow and nutrient cycling in marine systems (Lampitt & Gamble, 1982; Turner, 2004). In addition to these bottom-up processes, copepods can exert strong top-down control, as their selective grazing on bloom-forming species such as *Phaeocystis* has been shown to regulate primary production and shape broader biogeochemical fluxes and ecosystem stability (Nejstgaard et al., 2007). Equally important, copepods often feed selectively on microzooplankton, particularly ciliates, which are among the fastest grazers of small phytoplankton in warm oceans. This trophic link can have strong cascading effects on phytoplankton dynamics, as demonstrated in mesocosm and bottle incubation studies (Nejstgaard et al., 1997, 2001).

The ecological relevance of copepods stems from both their feeding strategies and their contribution to biogeochemical processes. By grazing on phytoplankton, protozoa, and detritus, copepods participate in nutrient recycling and are vital agents of vertical material flux through the production of fecal pellets and their diel vertical migration (Frangoulis et al., 2005). These processes contribute to the biological carbon pump, sequestering organic matter in deeper ocean layers and playing a role in long-term carbon storage. Their size structure, life stages, and feeding modes can be highly variable and responsive to environmental gradients, making them sensitive indicators of ecosystem change (Boyd et al., 1980; Lampitt & Gamble, 1982).

Despite the ecological significance of copepods, several knowledge gaps remain. Previous research has established the fundamental relationship between wind-driven upwelling, nutrient supply, and the productivity of marine food webs, particularly in temperate regions where upwelling delivers nutrients to the euphotic zone and fuels phytoplankton growth (Fernández-Álamo & Färber-Lorda, 2006; Pennington et al., 2006). However, tropical upwelling systems are less studied, particularly regarding how episodic or anomalous upwelling events, such as upwelling failure, affect mesozooplankton communities and ecosystem functioning. The response of mesozooplankton communities to abrupt changes in upwelling regimes is not well understood, and traditional plankton monitoring methods have struggled to keep pace with the need for rapid, high-resolution data in these dynamic environments. Recent advances in automated imaging and isotopic techniques present new opportunities for detailed plankton community assessment, and their application to tropical, climatically variable systems is still emerging.

1.2 Tropical Eastern Pacific (TEP)

The Tropical Eastern Pacific (TEP) (Figure 1) is a region of strong oceanographic and ecological contrasts, making it a valuable natural laboratory for studying mesozooplankton dynamics (Fiedler & Lavín, 2006). In particular, the Gulf of Panama (GOP) and the Gulf of Chiriquí (GOC), located along the Pacific coast of Panama, represent two ecologically distinct systems.

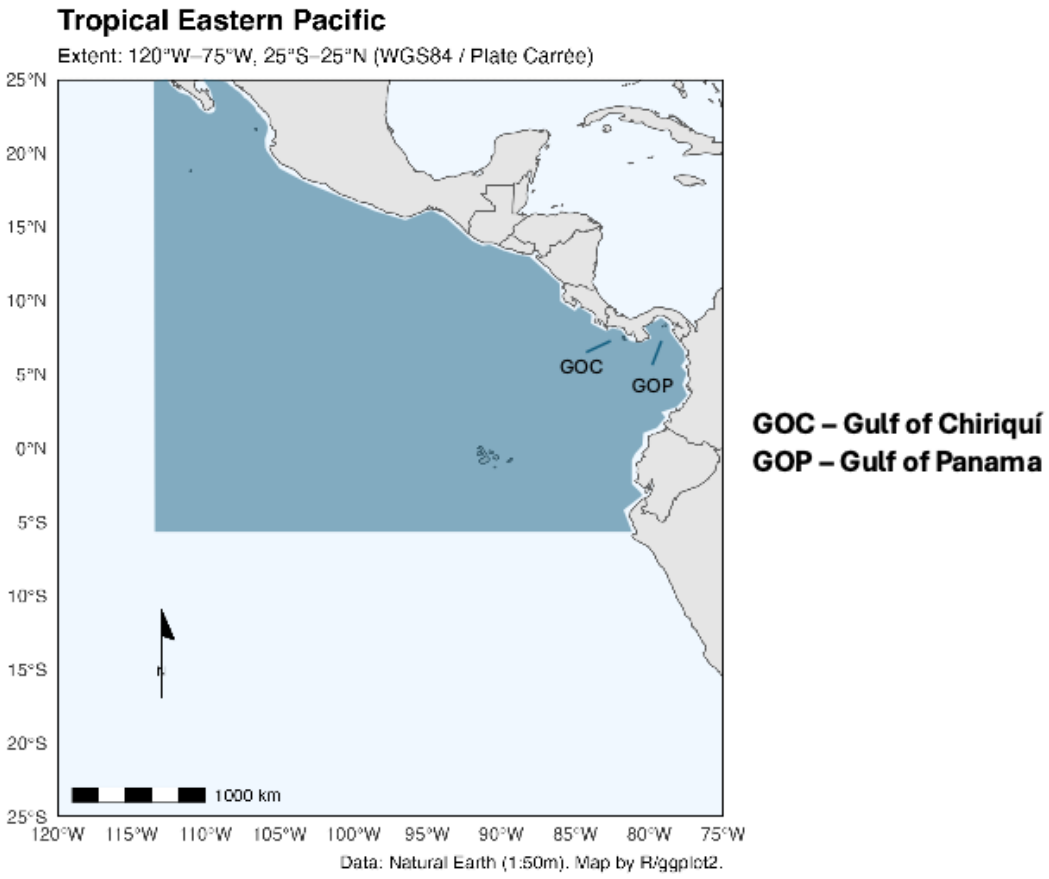


Figure 1: Map of the Tropical Eastern Pacifi, showing the coastal regions of Central and South America. Data source: Natural Earth. Projection: Plate Carrée (WGS84).

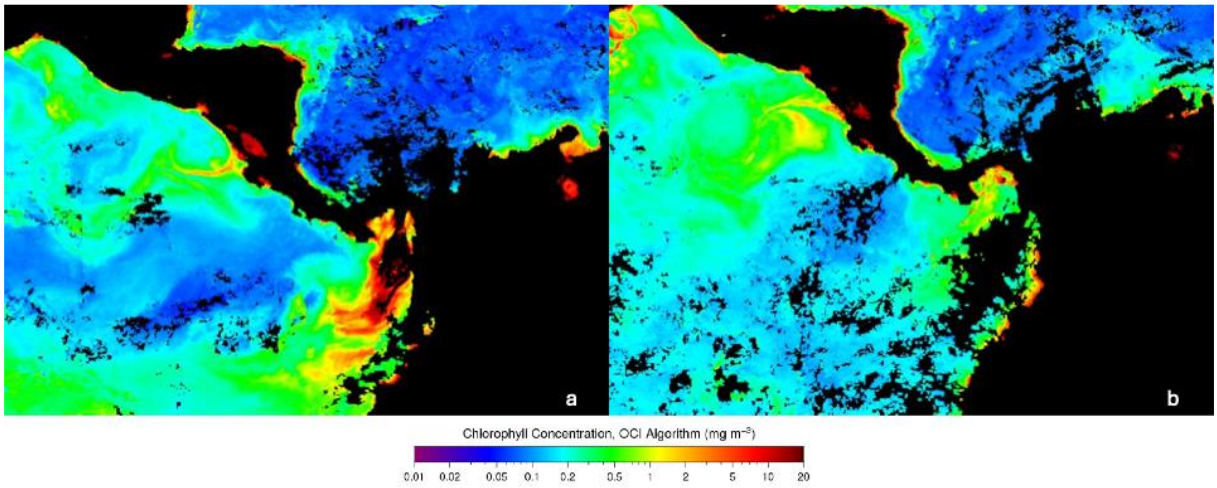
1.2.1 The Gulf of Panama (GOP)

The GOP represents a prime example of a nutrient-rich marine environment typically shaped by seasonal wind-driven upwelling (Glynn & Maté, 1997; Maté, 2005). Situated along the southern coast of Panama, the GOP is directly influenced by the interplay of regional wind patterns and ocean currents. Its geographic proximity to the equatorial current systems and northeasterly trade winds makes it a significant site for marine productivity and ecological studies. This region serves as a natural laboratory for understanding the mechanisms of upwelling, nutrient cycling, and the functioning of productive marine ecosystems (D'Croz & O'Dea, 2007; Perez et al., 2024; Rodriguez-Ruano et al., 2023).

Upwelling in the Gulf of Panama occurs during the dry season, typically from December to April, when the prevailing northeasterly trade winds intensify. These winds drive warm surface waters offshore through Ekman transport, allowing cooler, nutrient-rich deep waters to rise to

the surface (D'Croz & O'Dea, 2007; Perez et al., 2024). This process restores the euphotic zone with nutrients, including the macronutrients nitrate and phosphate, which are critical for phytoplankton growth. The influx of nutrients fuels significant phytoplankton blooms, primarily composed of fast-growing and highly efficient primary producers such as diatoms sustaining the base of the marine food web (MacKenzie et al., 2019).

One of the key indicators of this seasonal productivity is chlorophyll-a concentration, which serves as a proxy for phytoplankton biomass. During upwelling periods, chlorophyll-a levels in the Gulf of Panama increase, reflecting the abundance of phytoplankton (D'Croz & O'Dea, 2007; Perez et al., 2024)(Figure 2). The elevated productivity has been extensively recorded in past studies and highlight the Gulf's role as a hotspot for biological activity. The chlorophyll-a peak signals the onset of productive upwelling period that reverberate through the entire marine ecosystem (D'Croz & O'dea, 2009)(Figure 2).



Mesozooplankton diversity decreases during upwelling and a small number of mesozooplankton opportunists dominates the system, particularly copepods, which act as primary or secondary consumers (Calbet & Saiz, 2005). As such, copepods play a critical role in transferring energy from phytoplankton to higher trophic levels, including fish, marine mammals, and gelatinous mesozooplankton. These trophic interactions are essential for maintaining the ecological balance of the Gulf and contribute to its status as one of the most productive regions in the TEP. The thriving mesozooplankton populations directly sustain economically important fish stocks, such as anchovies and sardines, which are key to local fisheries (Frangoulis et al., 2005).

The intensity and consistency of upwelling in the Gulf of Panama are subject to significant temporal variability. Seasonal wind patterns regulate the strength of upwelling, and larger climatic events like El Niño and La Niña also play a pivotal role. During El Niño years, the weakening of trade winds limit upwelling, leading to a sharp decline in nutrient availability and primary productivity, and cascade through the food web, affecting mesozooplankton populations, fish stocks, and ultimately local fisheries. Conversely, during La Niña conditions upwelling and biological productivity may be enhanced. This variability underscores the need for continuous monitoring to understand the Gulf's resilience to natural and anthropogenic stressors (MacKenzie et al., 2019; Perez et al., 2024).

The Gulf of Panama's upwelling system in tropical waters is contrasting with other major extratropical upwelling regions like the California Current or the Benguela Current. While these systems are driven by similar physical processes, the GOP's location within the TEP allows for a distinctive interplay of tropical biodiversity and upwelling-driven productivity. This combination makes the GOP an important site for studying the ecological and biogeochemical impacts of upwelling in a tropical context (D'Croz & O'Dea, 2007; MacKenzie et al., 2019).

1.2.2 The Gulf of Chiriquí (GOC)

The GOC stands as a natural counterpoint to the nutrient-rich GOP, offering a unique opportunity to study oligotrophic systems within the TEP (Perez et al., 2024; Rodriguez-Ruano et al., 2023). While the Gulf of Panama normally exemplifies the high productivity and ecological dynamics of upwelling systems, the Gulf of Chiriquí provides an invaluable baseline for understanding marine environments characterized by persistently low nutrient availability. This sharp contrast makes the two regions complementary in their ecological significance and research potential (D'Croz & O'Dea, 2007).

As an oligotrophic system, the Gulf of Chiriquí allows exploration of the adaptations and processes that enable marine ecosystems to function under resource-scarce conditions (D'Croz & O'Dea, 2007; MacKenzie et al., 2019; Perez et al., 2024; Rodriguez-Ruano et al., 2023). Main reason for these oligotrophic conditions is the absence of consistent wind-driven upwelling despite its proximity to the Gulf of Panama, primarily due to differences in wind exposure caused by the mid-American mountain chain (Aybar & Tuñon, 1997; Maté, 2005). The Gulf of Chiriquí's coastline is more sheltered from the prevailing trade winds, unlike the Gulf of Panama, where winds align with a gap in the Cordillera (gap winds) to promote

upwelling. Ocean currents and warmer surface waters also affect the conditions needed for surface water column stratification (D'Croz & O'dea, 2009; Perez et al., 2024). Alternative nutrient sources are local riverine inputs, which deliver organic matter and nutrients to coastal areas, potentially fueling short-lived phytoplankton blooms or sporadic upwelling events like the input of small pockets of cool deep water into the surface mixed layer (D'Croz & O'Dea, 2007). These ecosystems rely heavily on nutrient recycling and detrital pathways, offering insights into how energy flows and nutrients are retained in systems where external inputs are limited (D'Croz & O'Dea, 2007; Frangoulis et al., 2005). In contrast, upwelling conditions as in the Gulf of Panama rely on consistent advection of nutrient-rich water to sustain their high biomass and productivity, leading to a more linear transfer of energy through the trophic hierarchy (D'Croz & O'Dea, 2007).

The Gulf of Chiriquí serves as a case study for oligotrophic dynamics in tropical marine ecosystems, which are becoming increasingly relevant as climate change alters nutrient availability and primary productivity in many parts of the ocean. By studying the Gulf of Chiriquí, a better understanding of how marine communities respond to future shifts in nutrient regimes may be facilitated, including potential declines in upwelling intensity in traditionally productive regions or climatic systems such as El Niño-Southern Oscillation (ENSO), which may amplify or diminish local productivity (Rodriguez-Ruano et al., 2023). Copepods, adapt to resource-scarce conditions by employing efficient feeding strategies or relying on detrital pathways (Buesa, 2019).

The oligotrophic conditions in the Gulf of Chiriquí offer a more stable habitat for coral reefs and associated fish species. The Gulf also provides a valuable reference point for comparing the resilience and adaptability of marine ecosystems to nutrient scarcity (D'Croz & O'Dea, 2007). Through its contrast with high- productivity systems and its role in advancing the understanding of oligotrophic ecosystems, the Gulf of Chiriquí holds significant importance for marine ecological and biogeochemical analyses. It not only highlights the diversity of marine ecosystem dynamics within the TEP but also provides essential context for studying the broader impacts of environmental change on oceanic nutrient cycles and productivity.

2. Materials and Methods

2.1 Sampling

Mesozooplankton for the present study were collected during a field campaign conducted from 5 to 9 March 2025 aboard the research sailing yacht *Eugen Seibold*. Sampling was performed at two offshore stations in the Gulf of Panama (Station 209) and the Gulf of Chiriquí (Station 210) (Figure 3). Detailed sampling information, including coordinates, dates, gear types, and times, is summarized in Table 1.

Table 1: Sampling details - March 2025 field campaign aboard the R/V Eugen Seibold

Station	**Code**	**Location (lat, lon)**	**Region**	**Date**	**Gear & depth**	**Time (UTC-5)**
209	ES25C03_209	7°N, 79.5°W	Gulf of Panama (GOP)	06.03.2025	Multicast 100 m	06:30
					Multicast 500 m	07:51
210	ES25C03_210	7.17°N, 82.5°W	Gulf of Chiriquí (GOC)	07.03.2025	Multicast 500 m	06:38
					Multicast 100 m	09:58

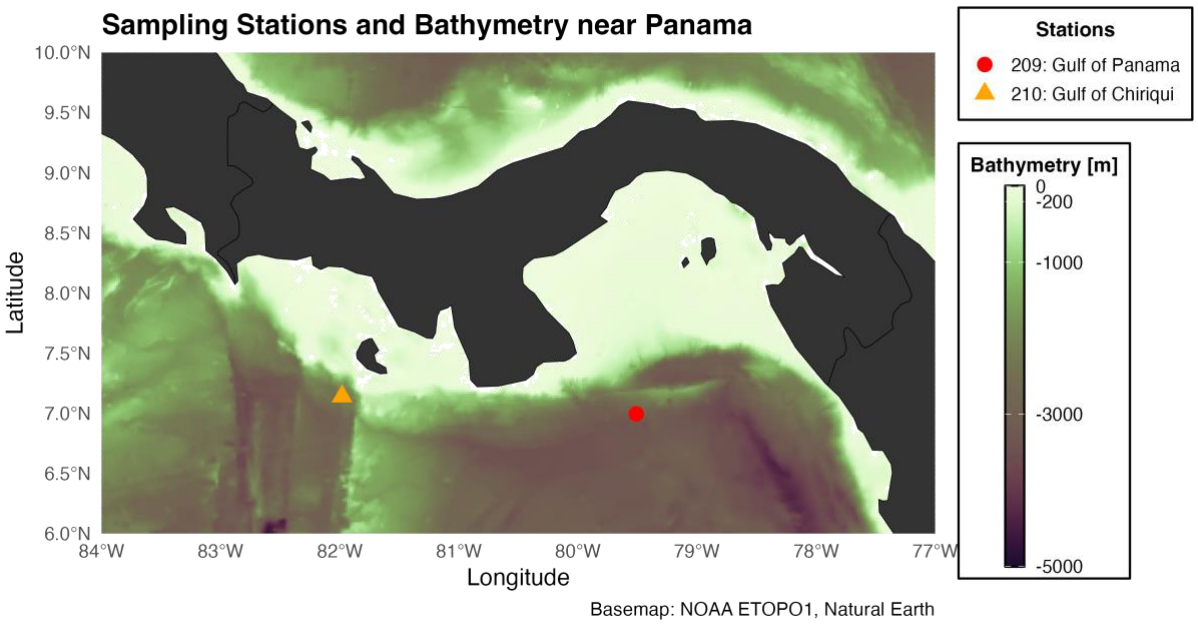


Figure 2: Map of sampling stations in the Gulf of Panama (Station 209, red circle) and the Gulf of Chiriquí (Station 210, yellow triangle). Bathymetry is shown in meters, with shallower areas in light green and deeper regions in purple.

At each site, a multinet system with an opening of 0.25 m^2 and equipped with 100-µm mesh was deployed twice to sample the mesozooplankton community from two depth intervals: one from 100 to 0 m, opening at 100, 80, 60, 40, and 20 m (towing distance 20m), and the one from 500 to 0 m, opening at 500, 400, 300, 200, and 100 m (towing distance 100m). In parallel, CTD-rosette water sampler casts equipped with 5-L Niskin bottles were used to obtain vertical profiles of temperature, salinity, and to collect water samples for nutrient analyses and other geochemical parameters. Mesozooplankton samples were preserved in formalin (Formol) buffered with hexamethylenetetramine and stored under cold (4 °C) conditions. For downstream analysis, the collected material was size-fractionated by sieving into three categories: >1000 µm, 500 - 1000 µm, and <500 µm.

For statistical analyses of hydrographic properties, the upper layer was defined as the water column from the surface down to the thermocline, here identified as the depth at which the vertical temperature gradient first exceeds 0.1 °C m^{-1}. This threshold is commonly used in the literature to delineate the transition between the mixed layer influenced by atmospheric processes and the stratified thermocline (de Boyer Montégut et al., 2004; Sprintall & Roemmich, 1999). The subsurface layer was defined as depths below the thermocline,

representing the more stable and homogeneous deep-water mass below the main thermocline. These thresholds were selected following conventional oceanographic practice, as described by de Boyer Montégut et al. (2004) and Sprintall & Roemmich (1999) and are consistent with the pronounced vertical gradients observed in the CTD profiles collected during this study.

2.2 Hydrographic measurements and CTD processing

Raw conductivity-temperature-depth (CTD) profiles were imported and processed in R (v4.5.1) (R Core Team, 2025) using the *tidyverse*, *oce*, and *janitor* packages. Depth (m) was derived directly from pressure (dbar). Temperature (°C) and salinity were despiked with a 5-point running median filter, which preserves vertical structure while reducing sensor noise, and subsequently interpolated onto a 1-m grid.

In situ density was expressed as potential density anomaly ($\sigma\theta$, kg m^{-3}), calculated from temperature, salinity, and pressure. The mixed-layer depth (MLD) was defined as the shallowest depth where $\sigma\theta$ exceeded the 10-m reference value by $\Delta\sigma\theta = 0.03$ kg m^{-3}, following a common threshold for stratification studies. The Brunt-Väisälä frequency (N^2, s^{-2}) was estimated from vertical density gradients, and $\log_{10}$-scaled N^2 profiles were inspected to locate the pycnocline (depth of maximum stability). Thermocline and halocline depths were identified from the maximum vertical gradients of temperature and salinity, respectively. For visualization, figures were restricted to the upper 200 m of the water column, which captures the main seasonal pycnocline in this region. Profiles of temperature, salinity, $\sigma\theta$, and N^2 were plotted against depth with properly formatted units and symbols. In addition, a temperature-salinity (T-S) diagram with $\sigma\theta$ isopycnals was prepared to highlight water mass characteristics. Summary statistics (mean, SD, min-max) for surface (0-20 m) and deep (>100 m) layers, as well as Spearman rank correlations between temperature and salinity, were calculated for each station.

Chlorophyll-a (µg L^{-1}) was analyzed to resolve the vertical distribution of phytoplankton biomass. Profiles were lightly smoothed with a running median for clarity, and the depth of the deep chlorophyll maximum (DCM) was identified as the depth of peak chlorophyll concentration. For each station, temperature and chlorophyll-a profiles were plotted on aligned axes, with the DCM marked by a horizontal line and annotated to facilitate cross-site comparison.

2.3 Instrument configuration and procedure for FlowCAM®

Analyzing mesozooplankton composition and structure remains a labor-intensive and expertise-driven task. Traditional taxonomic identification methods are time-consuming, requiring manual sorting and microscopy, which limits throughput and reproducibility (Pitois et al., 2018). As a solution, automated imaging systems such as the FlowCam offer scalable alternatives for rapid plankton assessment (Luo et al., 2018a). These systems can capture high-resolution images of mesozooplankton samples, enabling both qualitative and quantitative analysis of plankton communities, especially in combination with artificial intelligence for classification (Bochinski et al., 2019; Eerola et al., 2024; Gorsky et al., 2010; Lacoursière-Roussel et al., 2025; Pierella Karlusich et al., 2022).

To implement this scalable approach, mesozooplankton samples were imaged using a FlowCAM Macro system (Fluid Imaging Technologies, Inc.) equipped with customized flow cells (FireflySci, Inc.) and Olympus microscope objectives (Figure 4). All measurements were conducted using VisualSpreadsheet® software version 6.0 (builds 6.0.4.304 and 6.0.4.317) and its auto-imaging function to capture and process individual mesozooplankton images. Prior to imaging, samples were washed through sieves and rinsed with artificial seawater (35 g sea salt per 1 L deionized water) to remove residues of fixatives and to reduce the risk of cell rupture compared to rinsing with deionized water alone. Inspected samples were subsequently, drained over a sieve and washed back into to their original formalin medium following measurement. A Watson Marlow 323 Peristaltic Pump and a 0.48 cm inner diameter and wall thickness of 0.16 cm tubing was used in the set up (Figure 4).

For the samples with particles smaller than 500 µm, a FlowCAM Macro fitted with a custom-built 500-µm flow cell (10 cm length, 10.5 mm width and 500 µm depth) was used. The system employed a 2.5x Olympus MPlanApo objective (NA 0.08), a frame rate of 14.91 fps, and a magnification of 2x. Mesozooplankton density was adjusted by dilution to either 500 mL or 1 L, and 400 mL or 800 mL of the sample were analyzed. In the case of one particularly dense sample, the total volume was increased to 3 L, of which 500 mL was processed to reduce clogging.

For samples containing mesozooplankton larger than 500 µm, a second configuration was employed using a 1000-µm flow cell (10 cm length, 10.5 mm width and 1000 µm depth), a PlanApo Olympus 1.25x/0.04 objective, and a frame rate of 3.98 fps at 1x magnification. These

samples were typically diluted to 150 mL, of which 110 mL was analyzed. Exceptionally dense samples were diluted to 1 L, and 800 mL were processed. In one case, the sample was diluted to 240 mL and 220 mL were measured.

Figure 3: Flowcam setup with tubing and Peristaltic Pump

To assess consistency across size fractions, a third experimental run was performed where fractions <500 µm and >500 µm were recombined, targeting the 100-1000 µm size range. These samples were analyzed using the 1000-µm flow cell under the same configuration as above. Most samples were resuspended in 250 mL artificial seawater, of which 220 mL were processed; denser samples were resuspended in 500 mL or 1 L, and subsampled at proportional volumes (e.g., 440 mL and 880 mL, respectively).

Samples exceeding 1000 µm were processed manually using stereomicroscopy. For these samples, the following procedure was applied: after homogenization the original sample by gentle swirling in tall, graduated cylinders, the sample volume was sequentially divided by pouring off half while retaining the remaining fraction for further analysis. Water was occasionally added during the splitting process to improve volume accuracy. Dense samples

often required considerable dilution to prevent clogging and to ensure optimal imaging conditions; in extreme cases, up to a 1:128 dilution was necessary. Despite these adjustments, occasional technical challenges such as flow cell clogging and a high proportion of small or amorphous particles (subsequently classified as debris) were encountered, particularly in samples dominated by detritus or empty exoskeletal fragments. Organisms were identified using visual guides including Castellani & Edwards, (2017); Grossmann et al., (2014); Guglielmo et al., (1945) and Larink & Westheide, (2011).

2.4 Preprocessing and model training

Following the imaging process using the FlowCAMs, individual images and a corresponding metadata summary (CSV format) were exported from VisualSpreadsheet® for further classification and analysis. During preprocessing, particles smaller than the target size threshold were excluded from further analysis. Image data were then processed using the open-source pipeline described by Symiakaki et al., (2025), which prepares images and metadata for integration with the LabelChecker platform. Prior to classification, particles imaged by the FlowCam but falling below the relevant size thresholds (<100 µm for the <500 µm fraction and <500 µm for the 500-1000 µm fraction) were removed to avoid including artifacts and non-target material. Classification within LabelChecker was carried out by manually sorting images into taxonomic categories including Appendicularia, Chaetognatha, Chordata, Cladocera, Cnidaria, Copepoda, Crustacea, Debris, Bacillariophyceae, Dinoflagellata, Echinodermata, Planktonic eggs, Nematoda, Phoronida, and Polychaeta. The taxonomic framework was guided in part by Brugnoli Olivera et al., (2023), who conducted a regional study of mesozooplankton in Coiba National Park (Panama).

The full pipeline described by Symiakaki et al., (2025) was followed to develop a machine learning model for automated classification. The model was trained in four successive rounds. In each round, training was performed for five epochs, after which manually corrected labels from LabelChecker were incorporated into the training set. This iterative process gradually improved model performance until an accuracy above 90 % was reached. In the final step, the optimized model was applied to classify the complete image set derived from our field samples. The workflow by Siangsano, (2025) and the accompanying Python scripts are included in the Appendix.

2.5 Data analysis and biomass estimation

Absolute abundances (individuals per sample) and relative abundances (percent contribution of each taxon to the total assemblage) were calculated for major mesozooplankton groups, stratified by station (Stations 209 and 210), water depth, and two size fractions (<500 µm and 500-1000 µm). To account for differences in the volume of water filtered, abundances were standardized to individuals per cubic meter (ind. m^{-3}). The sampled volume was calculated from the net opening area (0.25 m^2) multiplied by the vertical haul distance. For the deeper hauls (500-200 m), a vertical distance of 100 m was sampled, corresponding to 25 m^3 of water, whereas for the upper hauls (100-20 m) the sampled distance was 20 m, corresponding to 5 m^3 of water. In addition, all calculations incorporated the respective dilution factors (see section 2.3) applied prior to FlowCAM analysis, ensuring that the reported abundances reflect the original undiluted sample volumes.

All analyses were performed in R (R Core Team, 2025) using the packages *vegan, ggplot, scales, tidyverse, patchwork, janitor, oce, broom, ggrepel* and *dplyr* for ordination and environmental fitting, alongside base non-parametric functions for diversity comparisons. Alpha diversity was assessed using richness (taxon count), Shannon diversity ($H' = -\Sigma\, p_i \ln p_i$), Simpson diversity ($1 - \Sigma\, p_i^2$), and Pielou's evenness ($H'/\ln S$), where p_i denotes the proportional abundance of taxon i. Differences in these metrics across stations, size fractions, and depth intervals were tested. Where finer pairwise comparisons were warranted (e.g., Shannon diversity between stations within a particular size fraction), Wilcoxon rank-sum tests were applied. Statistical significance of test results was evaluated using p-values ($p < 0.05$), with values between 0.05 and 0.1 treated as marginal trends and interpreted cautiously.

Individual carbon biomass of copepods was estimated with the standard allometric form $C = a \times L^b$ used for copepods (Deneke & Maier G., 2019; Dumont et al., 1975; Uye, 1982). We applied $C = 6.0 \times L^{2.5}$ (µg C ind^{-1}; LLL in mm) where C is carbon mass in µg C per individual and L is body length in millimeters. This relationship represents a compromise across major copepod groups (e.g., calanoids, cyclopoids, harpacticoids) and is suitable for mixed or taxonomically unresolved assemblages; the exponent of 2.5 captures the typical allometric scaling observed in these groups while the coefficient 6.0 reflects a calibrated midpoint between temperate and tropical taxa, as synthesized in foundational studies (e.g., de Azevedo et al., 2012; Dumont et al., 1975; León, 2009). To evaluate whether size-fractionation improves FlowCAM image resolution and quantitative performance, we processed samples both as two

fractions (<500 µm on the small flowcell and 500-1000 µm on the large flowcell) and as a single combined sample. Length measurements were obtained from FlowCAM-derived particle metrics, converted to equivalent body lengths, and biomass was calculated per individual before aggregation to sample-level summaries.

Multivariate patterns in community composition were examined by first applying a Hellinger transformation to the abundance matrix to reduce the influence of dominant taxa and render the data suitable for Euclidean-based ordination (Legendre & Gallagher, 2001; Legendre & Legendre, 2012). Bray-Curtis dissimilarities were then computed on the transformed data, and a Principal Coordinates Analysis (PCoA) was conducted to visualize compositional variation among samples. The effects of station and depth on overall community structure were formally tested using permutational multivariate analysis of variance (PERMANOVA) with 999 permutations. Relationships between mesozooplankton community composition (beta diversity) and environmental variables were assessed by fitting CTD-derived vectors (e.g., temperature, salinity) to the PCoA of Hellinger-transformed abundances based on Bray-Curtis dissimilarities (Envfit) with only vectors showing significant correlations ($p < 0.05$ by permutation) retained and projected onto ordination space.

3. Oceanographic conditions

3.1 Oceanographic conditions during the March 2025 field campaign

During the March 2025 field campaign, oceanographic data were collected at two stations, Station 209 in the Gulf of Panama, and Station 210 in Gulf of Chiriquí, using sequential shallow and deep CTD casts to resolve the vertical structure of the water column. Temperature and salinity profiles (Figure 5) reveal strong stratification, with distinct surface and subsurface water masses. In the upper 20 m, mean temperatures were 27.1 ± 0.01 °C at Station 209, and 29.2 ± 0.5 °C at Station 210, declining to 11.3 ± 1.9 °C and 11.4 ± 2.0 °C, respectively, below 100 m water depth. Observed temperature extremes ranged from a near-surface maximum of 30.0 °C at Station 210 to deep minima of 8.0 °C at Station 209 and 8.2 °C at Station 210.

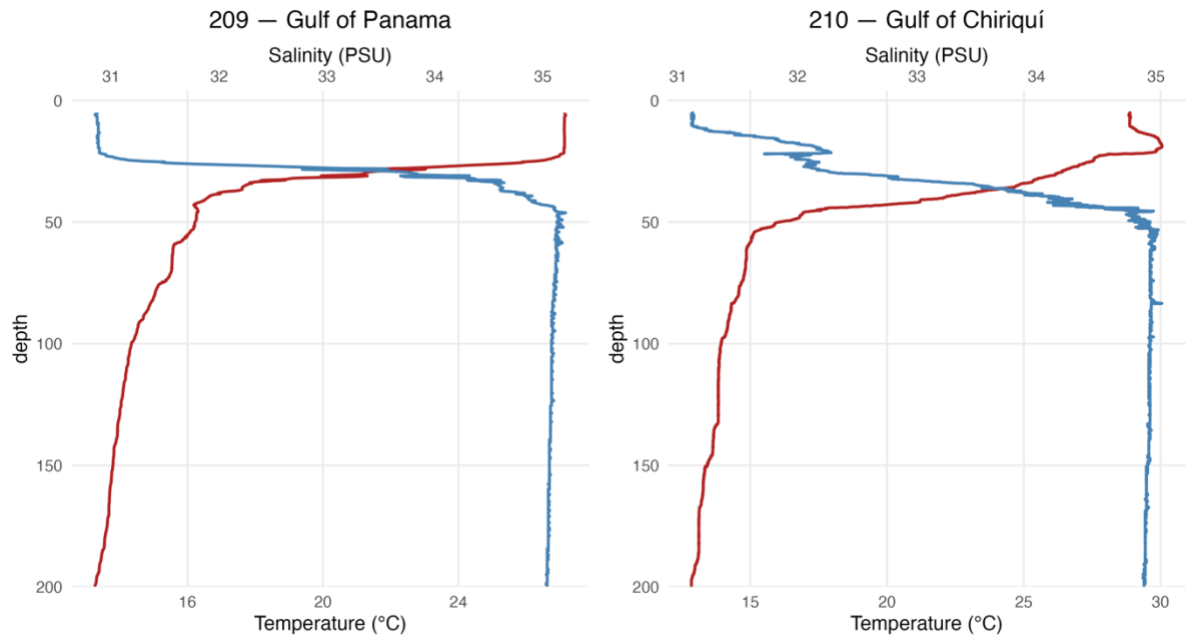


Figure 4: Vertical profiles of salinity and temperature from CTD casts at Stations 209 (Gulf of Panama) and 210 (Gulf of Chiriquí), down to 200 m depth (downcast).

Salinity in the upper 20 m averaged 30.9 ± 0.01 at Station 209 and 31.3 ± 0.4 at Station 210, with relatively uniform subsurface values of 34.9 and 34.8, respectively. Vertical structure differed between the sites: at Station 209, the mixed layer extended to ~22 m, with the thermocline and halocline developing gradually below, spanning approximately 28-60 m, and

the main pycnocline centered around 60-70 m. At Station 210, the mixed layer was shallower (~13 m), while the thermocline and halocline were deeper, spanning roughly 30-60 m, and the pycnocline extended to greater depths (100-150 m), indicating stronger and more persistent stratification over a broader depth range. The T-S diagram (Figure 6) supports this interpretation, showing that both stations shared similar water mass characteristics, with relatively fresh and warm surface waters overlying more saline subsurface waters. Station 210 displayed a broader range of density values, consistent with the deeper and more gradual pycnocline observed in the vertical profiles (Figure 7).

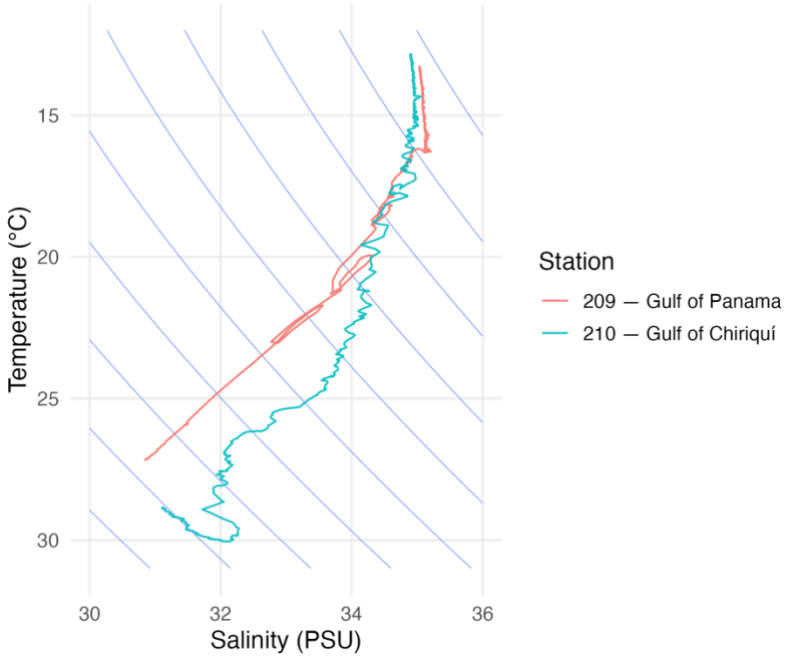


Figure 5: Temperature-salinity (T-S) diagram for Stations 209 -Gulf od Panama (red) and 210 -Gulf of Chiriquí (blue), illustrating the water mass properties along the vertical profiles. Isopycnals (blue lines) indicate density structure

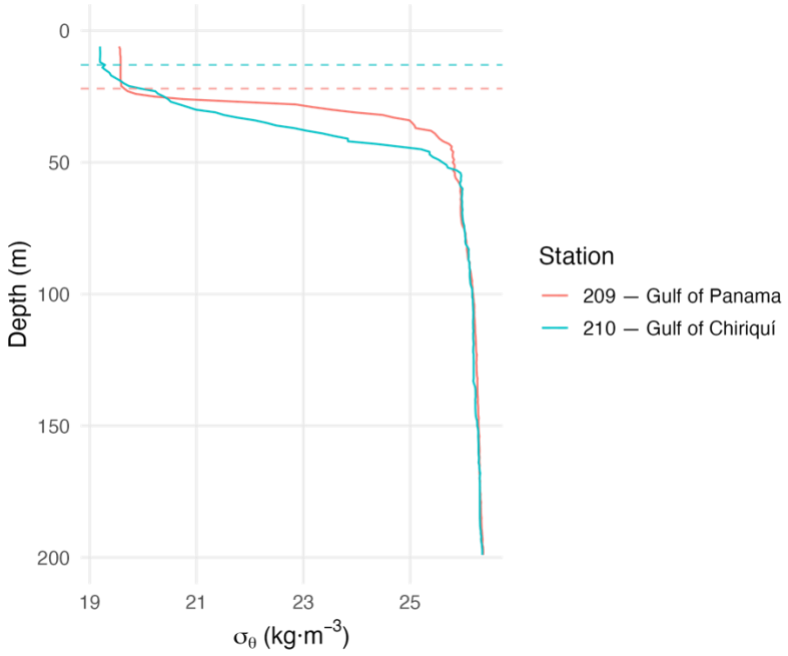


Figure 6: Density (σt) profiles at Stations 209 -Gulf od Panama (red) and 210 -Gulf of Chiriquí (blue), showing the vertical stability of the upper 200 m. Mixed layer depths (dashed lines)

Chlorophyll-a concentrations were strongly coupled to seawater density. At Station 209 (GOP), a pronounced deep chlorophyll maximum (DCM) of up to 29.0 µg Chl-a L^{-1} occurred at 28.1 m, roughly coinciding with the base of the mixed layer and the onset of the thermocline (Figure 8). At Station 210 (GOC), much less Chl-a was present in the DCM, reaching only up to 3.0 µg L^{-1} at 35 m despite comparable hydrography (Figure 6). Surface chlorophyll was low (<1 µg L^{-1}) at both sites, and below the thermocline Chl-a concentration rapidly declined. This contrast suggests that while hydrographic structure was broadly similar between stations, phytoplankton biomass differed by nearly an order of magnitude, pointing to regional differences in production or ecological dynamics.

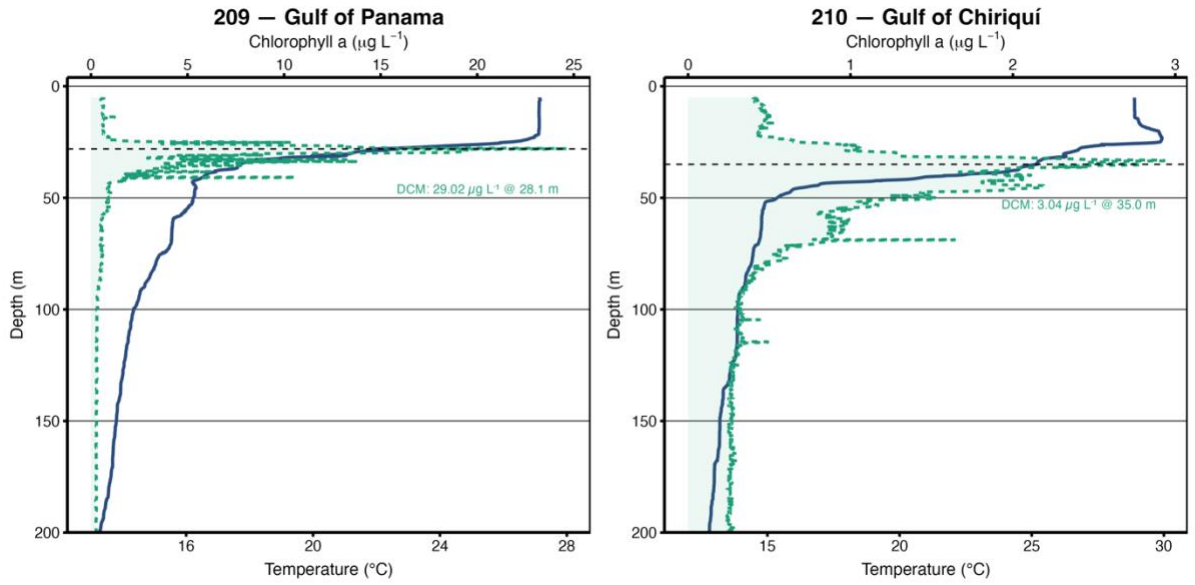


Figure 7: Depth profiles of temperature (solid lines) and chlorophyll-a fluorescence (green dotted lines) at Stations 209 - Gulf od Panama (left) and 210 - Gulf of Chiriquí (right), from the surface to 200 m.

3.2 Anomalous Oceanographic Conditions During the 2025 Field Campaign

While mesozooplankton responses to seasonal upwelling in classic scenarios has been rather well described (Margalef, 1978), little is known about their dynamics during years when upwelling fails altogether, an increasingly likely scenario under future climate change. Addressing this gap is not only important for ecosystem functioning, but also for understanding the resilience of fisheries, carbon sequestration, and ecosystem services that are critical to human societies.

During the 2025 field campaign, these contrasts were expected to be evident; however, anomalous oceanographic conditions were encountered. The seasonal upwelling in the Gulf of Panama, typically a reliable and ecologically crucial event, failed for the first time in at least four decades (O'Dea et al., 2025)(Figure 9). Normally, this phenomenon begins in late January and persists for more than two months, cooling surface waters to ~19 °C at its peak. In stark contrast, the 2025 upwelling was delayed by 42 days, lasted only 12 days, and reached a minimum temperature of just 23.3 °C, well above historical values. The cumulative number of "cold days" (<25 °C) was lower than in any year since satellite records began, underscoring the exceptional character of the event.

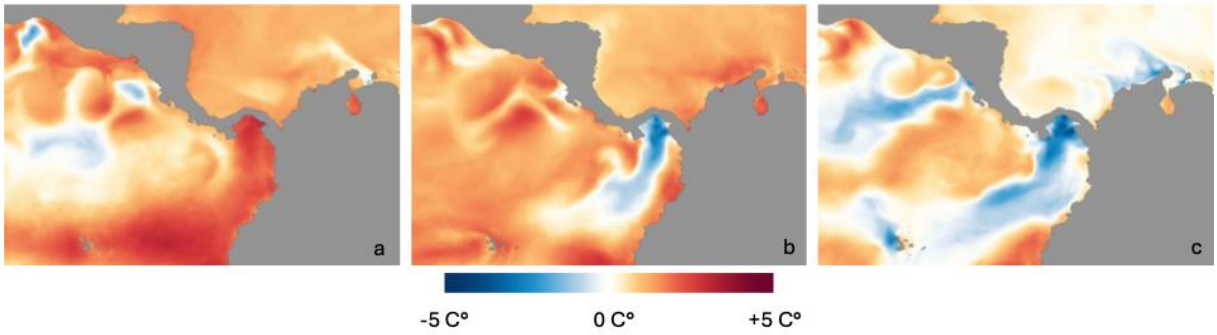


Figure 8: Sea surfaces temperature (SST) in the Tropical eastern Pazific (TEP) during the first week of March 2025 (a), 2024 (b) and 2023 (c) (National Oceanic and Atmospheric Administration, 2025).

This breakdown was closely linked to unusual atmospheric forcing. While northerly winds reached normal intensities when present, their frequency was reduced by nearly three-quarters relative to the historical mean, and periods of wind relaxation were longer. The result was a drastic reduction in cumulative upwelling-favorable wind stress. Regional reanalysis confirmed anomalously weak offshore winds throughout the first quarter of 2025, providing a plausible mechanism for the collapse of upwelling (O'Dea et al., 2025).

Hydrographic profiles further highlighted the anomaly: instead of the cool, nutrient-rich surface layer that usually develops during boreal winter, the water column in March 2025 was strongly stratified, with a warm mixed layer overlying a sharp thermocline at 20-30 m water depth. No evidence of vertical mixing or nutrient uplift into the euphotic zone was detected, consistent with suppressed productivity.

The ecological consequences of this disruption are potentially severe. Reduced nutrient supply is expected to limit phytoplankton growth, weakening the foundation of the marine food web and threatening the productivity of fisheries that depend on seasonal pulses of plankton. In addition, the absence of the cooling effect of upwelled waters increases the likelihood of thermal stress on coral reefs, potentially exacerbating bleaching events (O'Dea et al., 2025).

Against this backdrop, this thesis aims to address critical questions: How do mesozooplankton communities respond to unprecedented disruptions in upwelling-driven productivity? Can modern automated imaging tools provide robust, scalable insights into these dynamics?

To answer these questions, mesozooplankton community composition, abundance, and copepod biomass between the Gulf of Panama and the Gulf of Chiriquí, during the anomalous

2025 dry-season, was analyzed with a semi-automated imaging approach using a FlowCam system coupled with AI classifiers-for reliable detection and quantification of mesozooplankton was conveyed. Cutting-edge imaging technologies advances plankton monitoring methodologies and provides a crucial baseline data for understanding the resilience of marine ecosystems under rapid environmental change, which is vital for a sustainable management and the well-being of coastal communities dependent on ocean resources.

4. Results

4.1. Automated Image Analysis and Model Performance

The FlowCam system imaged between 401 and 254858 particles per sample, resulting in a total of 1103459 images across all stations and size fractions and fraction combinations (Appendix; Table 6). Particle counts varied substantially depending on productivity and detritus load, with the largest datasets recorded in sample from the Gulf of Panama at 0-20 m water depth (Appendix, Table 6).

After preprocessing, a total of 752,478 images were processed and classified using the optimized machine learning model (Model 4) developed for this study (Figure 10).

The classification model was trained and validated over four successive iterations, each involving manual annotation and correction of misclassified images using the LabelChecker platform. Training progress was tracked across five epochs (i.e., five complete passes through the training dataset), with Top-1 accuracy values ranging from 75.1 % to 87.0 % and a maximum Top-5 accuracy of 99.6 %. The minimum validation loss was 0.418 at epoch 4, indicating optimal model performance (Table 2). Application of the final model to the full dataset resulted in an overall performance of ~75 %, i.e., 752,478 classified images, out of which 561,208 were identified as debris, including empty exoskeletons, detrital particles, and other non-target materials. Positively correct assignment rates were achieved for dominant taxa, with copepods (96 %), crustaceans (91 %), dinoflagellates (95 %), polychaetes (93 %), debris (94 %), Bacillariophyceae (100 %), echinoderms (100 %), and eggs (91 %), each exceeding 90 % accuracy (Figure 11). Misclassification rates for these groups ranged from 0 % (Bacillariophyceae, echinoderms) up to 9 % (crustaceans, eggs). In contrast, appendicularians (74 %), chaetognaths (75 %), cladocerans (53. %), and cnidarians (53 %) exhibited lower classification accuracy (Figure 11). This was particularly apparent for cnidarians and cladocerans, which were often misclassified as debris-reflecting challenges in distinguishing transparent or fragmented forms in automated imaging.

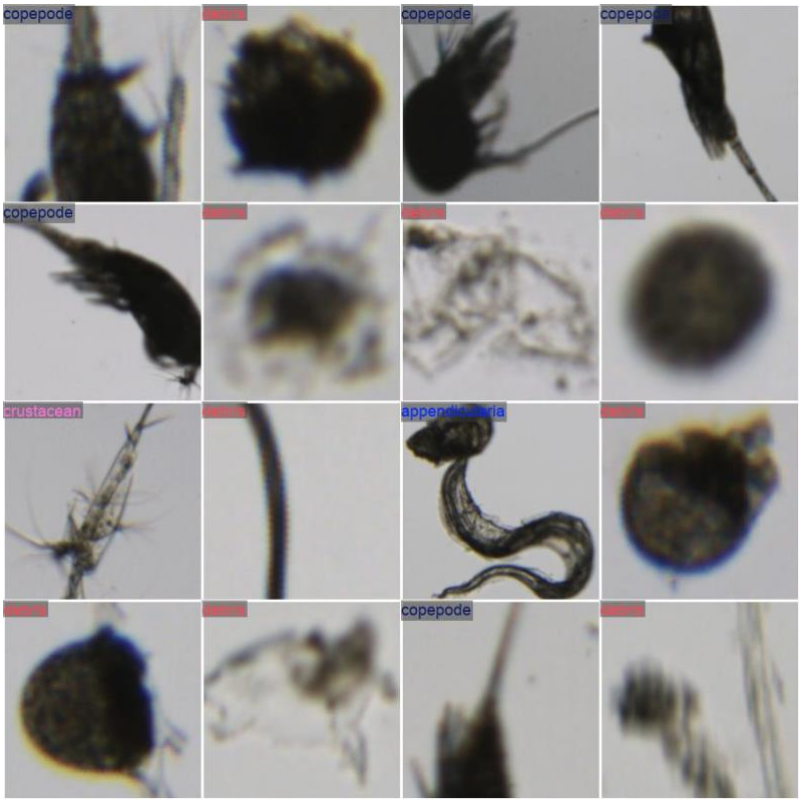


Figure 9: Example FlowCam images of mesozooplankton and particles automatically classified by the machine learning model.

Table 2: Training performance metrics across five epochs, showing Top-1 accuracy, Top-5 accuracy, and validation loss of the classification model.

Epoch	Top-1 Accuracy	Top-5 Accuracy	Validation Loss
1	86.9%	99.3%	0.423
2	85.8%	99.5%	0.445
3	75.1%	97.5%	0.831
4	86.9%	99.5%	0.418
5	82.6%	99.6%	0.511

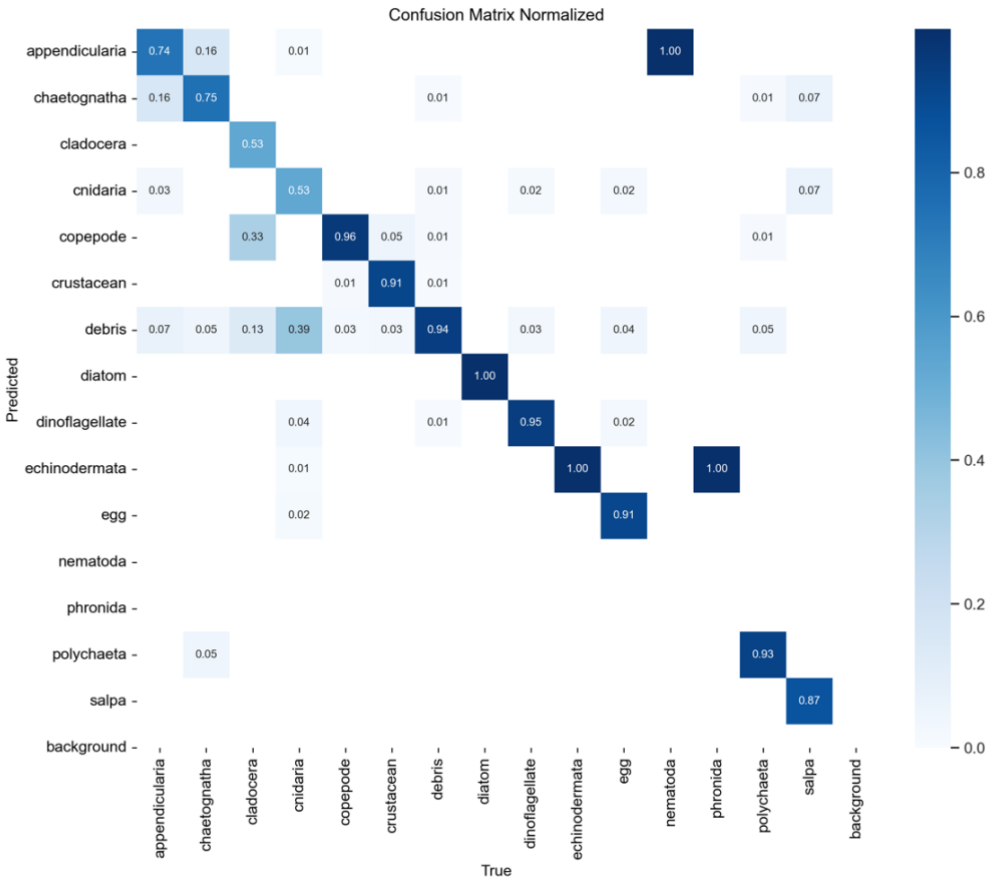


Figure 10: Normalized confusion matrix of the classification model showing predicted versus correct labels for plankton and debris categories. Color intensity represents the proportion of correctly or incorrectly classified instances within each class, with darker blue indicating higher accuracy. Values are row-normalized, showing the proportion of correct classes assigned to each predicted class. Consequently, relationships are not symmetrical: misclassification of class A as class B does not necessarily imply the inverse misclassification at the same rate.

4.2 Mesozooplankton

4.2.1 Mesozooplankton Taxa Across Depths, Stations, and Size Fractions

A total of 14 mesozooplankton taxa were identified across both sampling stations and all size fractions, revealing distinct patterns of abundance, diversity, and community composition with respect to water depth and location (Figure12). The <500 µm fraction consistently exhibited the highest counts in surface waters, with pronounced peaks in the upper 20 m (1225 individuals m^{-3}) and 20-40 m (601 individuals m^{-3}) at Station 209, and at 0-20 m (1585 individuals m^{-3}) and 20-40 m (1549 individuals m^{-3}) at Station 210. In addition, a secondary maximum was detected at 200-300 m at Station 209 (92 individuals m^{-3}), and a minor deep-water peak occurred at 200-300 m at Station 210 (685 individuals m^{-3}).

Summed across all depths and size fractions, abundances were substantially higher in the GOC (Station 210, 3949 individuals m^{-3}) compared to the GOP (Station 209, 1311 individuals m^{-3}), representing a nearly threefold difference.

The largest size fraction (>1000 µm) exhibited a particularly strong contrast between the two gulfs. In the GOP, noteworthy abundances were restricted to the upper water layers (0-20 m: 513; 40 m: 346; 200 m: 12 individuals m^{-3}), whereas in the GOC the large size fraction was strongly represented not only at the surface (20 m: 212; 40 m: 85 individuals m^{-3}) but another peak occurred at 200-300 m (292 individuals m^{-3}). Summed by size fraction, in the GOP 602 individuals m^{-3} occurred in <500 µm, 2649 individuals m^{-3} in 500-1000 µm, and 921 individuals m^{-3} in >1000 µm size fraction, while in the GOC 2649, 1300, and 656 individuals (m^{-3}) occurred, respectively. Thus, overall abundances were markedly higher in Gulf of Chiriquí (Station 210) than in Gulf of Panama (Station 209) (Figure 12).

Relative taxonomic composition confirmed copepods as dominant, accounting for 72.6 % of individuals in size fraction <500 µm, 79 % in 500-1000 µm, and 66.0 % in >1000 µm (Figure 12). Other groups contributed to a variable degree but were never dominant. Crustaceans were most strongly represented in the >1000 µm fraction (12.7 %), while their contribution was lower in <500 µm (10.0 %) (mainly larvae) and minimum in 500-1000 µm (1.4 %). Bacillariophyceae were rare overall, contributing <1 % in fine fractions and being virtually absent from >1000 µm samples (Figure 12).

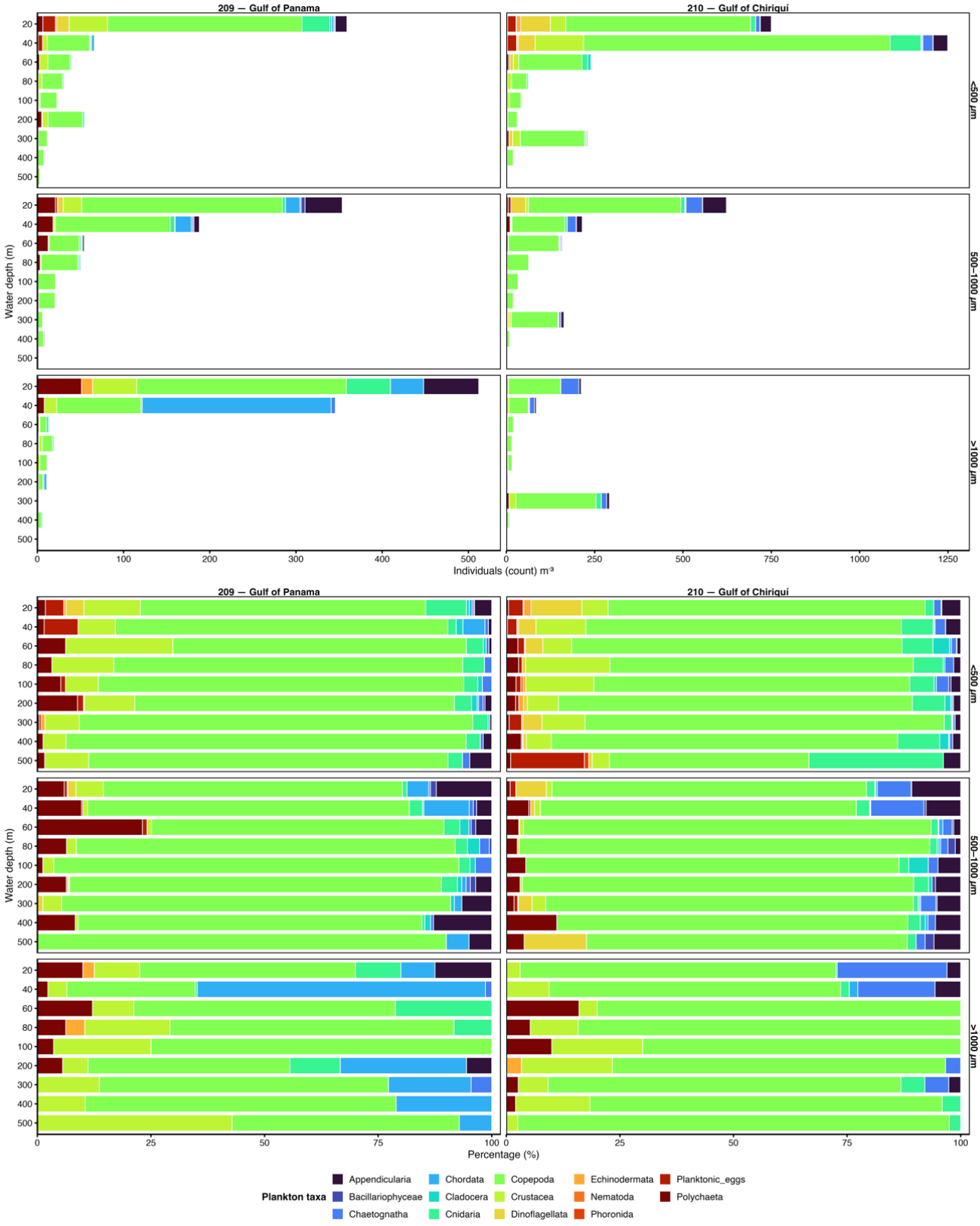


Figure 11: Absolute abundance (upper panels) and relative composition (lower panels) of zooplankton groups at Stations 209 and 210, shown by depth and size fraction (<1000 µm, 500-1000 µm, >500 µm).

Diversity patterns reflect both size fractions and spatial differences. On average, Shannon diversity (H′) and species richness were highest in the <500 µm fraction (H′ = 0.988, Richness = 9.61), intermediate in 500-1000 µm (H′ = 0.815, Richness = 8.39), and lowest in >1000 µm (H′ = 0.894, Richness = 4.33). Station-specific means indicated higher diversity in the GOC for the <500 µm fraction (H′ = 1.086 vs. 0.889), and higher diversity in the GOP for >1000 µm (H′ = 1.069 vs. 0.719). A paired Wilcoxon test across all matched samples showed no significant overall difference in Shannon diversity between stations (p = 0.714). Diversity tended to be highest at thermocline depths including the DCM (40-100 m). Taken together, diversity was highest in the <500 µm fraction (Appendix, Tabel 7).

Multivariate analysis of community structure supported these findings. Principal Coordinates Analysis (PCoA) of Hellinger-transformed abundances (Bray-Curtis) revealed moderate separation by station with partial overlap (Figure 13). The first two axes explained 25.3 % and 14.3 % of the variance (39.6 % cumulatively). While some depth-related clustering occurred particularly among midwater samples, the station effect was more pronounced (Figure 12). PERMANOVA results confirmed a significant effect of station ($F = 3.316$, $R^2 = 0.060$, $p = 0.005$) and a weaker, borderline effect of depth ($F = 1.297$, $R^2 = 0.187$, $p = 0.092$) (Appendix, Tabel 8).

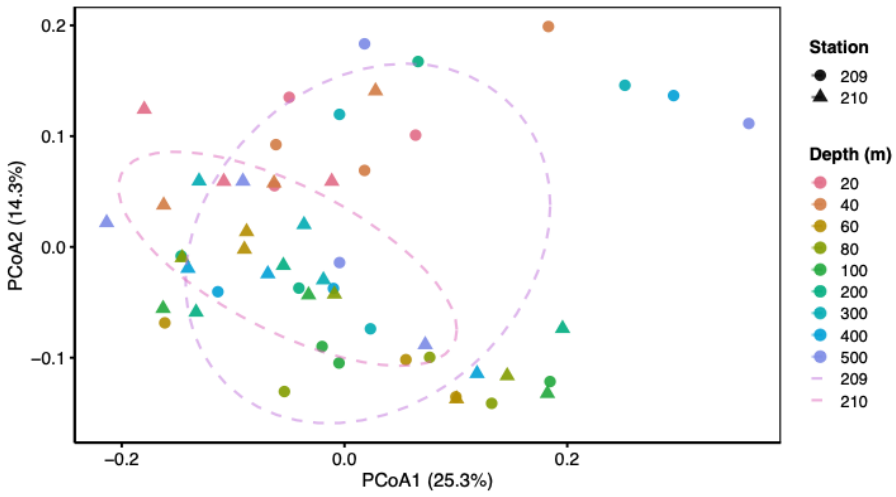


Figure 12: Principal Coordinates Analysis (PCoA) of zooplankton community composition based on Bray–Curtis dissimilarities. Symbols represent stations (circles = Station 209, triangles = Station 210) and colors indicate sampling depth. Dashed ellipses denote station groupings.

4.2.2 Copepod Biomass at Stations and Depths and Size Fraction Comparison (<500 µm and 500-1000 µm vs. Combined Sample)

Direct comparison of the <500 µm and 500 - 1000 µm size fractions show depth-structured differences in both count (individuals m^{-3}) and copepod biomass (m^{-3}) (Figure 14). At both stations, biomass and abundance generally increase toward the sea surface. Across almost all depths, the 500 - 1000 µm fraction exhibits higher biomass than the <500 µm fraction, even when counts are equal or lower.

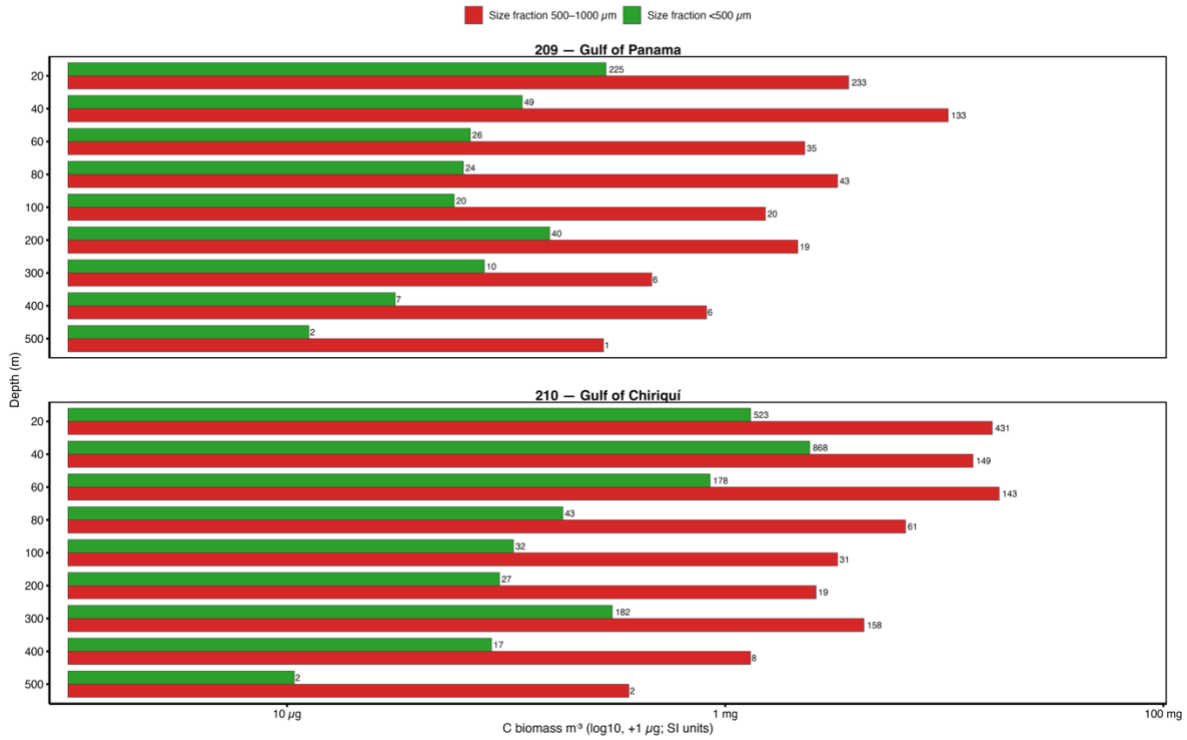


Figure 13: Depth distribution of copepod biomass and abundance at Stations 209 and 210. Bars represent carbon biomass per cubic meter (log_{10}-transformed, +1 offset) for two size fractions: 500-1000 µm; red and <500 µm; green). Numbers next to the bars indicate rounded copepod counts

At Station 209 (GOP) in the deeper layers 500 m, 400 m and 300 m the <500 µm fraction contains more individuals per m^{-3} (e.g., 2 vs 1 at 500 m; 7 vs 6 at 400 m; 10 vs 6 at 300 m, <500 µm vs 500 - 1000 µm), but the 500 - 1000 µm fraction still has higher biomass at the same depths (12 µg vs 277 µg at 500 m; 30 µg vs 821 µg at 400 m; 79µg vs 461 µg at 300 m). Around 200 m, counts remain higher in <500 µm (40 vs 19), yet biomass continues to favor 500 - 1000 µm (157 µg vs 2143 µg). In the upper water column, the pattern tightens: at 100-80 m the 500-1000 µm fraction shows higher counts (e.g., 43 vs 20 at 100 m) and higher biomass, and near

the surface (40-20 m) biomass is clearly larger in 500-1000 µm while counts are comparable or slightly higher there (e.g., 40 m: 133 vs 49; 20 m: 233 vs 225) (Figure 14).

In the Gulf of Chiriquí (Station 210), differences were stronger and more consistent. From 500 m, 400 m, and 300 m, the 500-1000 µm size fraction included the highest biomass (10 µg vs 362 µg at 500 m; 85 µg vs 1303 µg at 400m; 304 µg vs 4300 µg at 300 m, <500 µm vs 500 - 1000 µm), while counts were similar or higher in the <500 µm size fraction (e.g., 182 vs 158 at 300 m). At 200-100 m the <500 µm fraction often shows more individuals (e.g., 27 vs 19 at 200 m; 32 vs 31 at 100 m), yet total biomass remains higher in 500-1000 µm. The contrast is most pronounced in the shallow layers: despite much higher numbers of specimens in <500 µm (e.g., 868 vs 149 at 40 m; 523 vs 431 at 20 m), the 500-1000 µm fraction still delivers higher biomass (2432 µg vs 13520 µg at 40 m; 1396 µg vs 16554 µg at 20 m).

Across the full depth range, a substantially higher total (<500 µm and 500-1000 µm) copepod biomass was present in the GOC than GOP. Integrated over all sampled water depth intervals, biomass in the GOC was ~2.8-fold higher than in the GOP, and the GOC exceeded the GOP at every individual depth (e.g., depth-specific ratios of ~8.5× at 200-300 m, ~7.9× at 40-60 m, and ~4.5× at 0-20 m). At both stations, the 500-1000 µm fraction dominated biomass (typically 85-99 % of the depth-specific total), but the abundance structure differed: at the GOC the <500 µm fraction often carried most of the individuals while contributing far less biomass, for example, at 20-40 m there were 868 vs 149 individuals (<500 µm vs 500-1000 µm), yet biomass remained much higher in the 500-1000 µm size fraction. A similar pattern was observed at 0-20 m (523 vs 431 individuals in 500-1000 µm vs <500 µm). At the GOP, numbers were more balanced and could even favor the <500 µm fraction at some depths (e.g., 100-200 m: 40.3 vs 18.9 individuals), yet biomass still favors the 500-1000 µm fraction. Taken together, the station contrast reflects not only higher overall numbers at the GOC (count total 2 876 vs 899) but, critically, a greater contribution of large-bodied copepods in the 500-1000 µm fraction, which amplified biomass differences between stations, which amplified biomass differences between stations, particularly in the surface and thermocline layers.

Direct comparison of copepod biomass and abundance was made between (i) size-fractionated samples, where the <500 µm and 500-1000 µm fractions were measured separately and their values subsequently summed, and (ii) combined samples, where both fractions were processed together in a single measurement (Figure 15). Across both stations, the combined approach consistently yielded higher values of C-biomass (per m^3) and copepod counts than the sum of

the two fractions, with the largest discrepancies in the surface and thermocline layers (100-300 m).

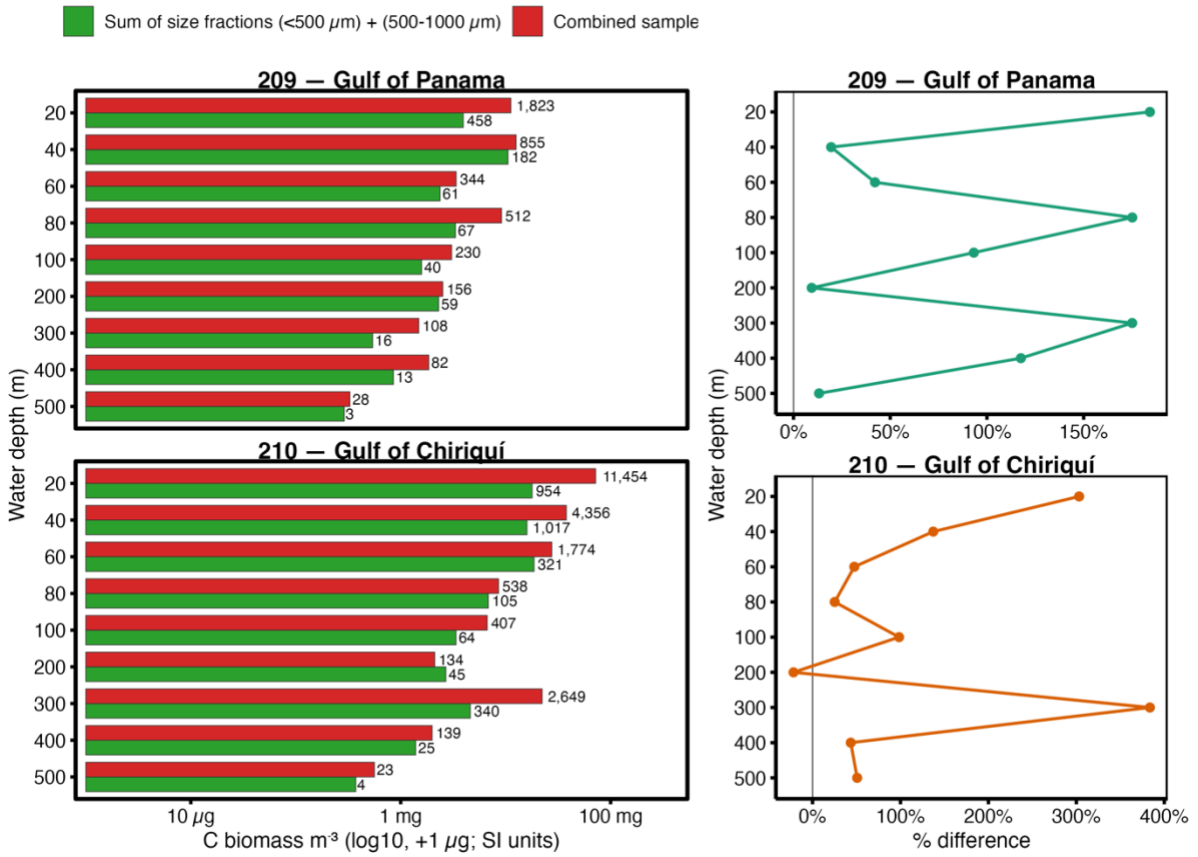


Figure 14: Depth distribution of copepod biomass at Stations 209 and 210, comparing combined samples (red) with the sum of separated size fractions (green). Bars represent carbon biomass per cubic meter (log_{10}-transformed, +1 offset), with numbers showing rounded copepod counts. The panels on the right show the relative difference (%) between combined and separated sampling, where positive values indicate higher biomass estimates in the combined samples.

In the Gulf of Panama (Station 209), combined sample counts exceeded the fraction sum by 6-9-fold in deep layers (e.g., 108 vs 16 at 300-200 m; 82 vs 13 at 400-300 m; 28 vs 3 at 500-400 m), with biomass also higher or similar (e.g., 1485 µg vs 540 µg at 300-200 m; 1852 µg vs 851 µg at 400-300 m; 327 µg vs 289 µg at 500-400 m). At 200-100 m and 60-40 m, combined counts remained higher (e.g., 156 vs 59 at 200-100 m; 344 vs 61 at 60-40 m), while biomass values were comparable (2517 µg vs 2301 µg at 200-100 m; 3372 µg vs 2372 µg at 60-40 m). Between 100 and 60 m, counts were again greater in the combined sample (e.g., 230 vs 67 at 100-80 m; 512 vs 67 at 80-60 m), which also yielded more than double the biomass (3056 µg vs 1582 µg at 100-80 m; 9238 µg vs 3321 µg at 80-60 m). In the upper 40 m, the combined sample dominated strongly in counts (e.g., 1823 vs 458 at 20-0 m; 855 vs 182 at 40-20 m), while biomass showed smaller differences (11212 µg vs 3945 µg at 20-0 m; 12582 µg vs 10533

µg at 40-20 m). The relative-difference panel for biomass indicates increases of ~100-300 % across much of the profile (Figure 15).

In the Gulf of Chiriquí (Station 210), discrepancies were larger and more consistent. Between 500 and 200 m, combined counts exceeded the fraction sum by ~5-8-fold (e.g., 2649 vs 340 at 300-200 m; 139 vs 25 at 400-300 m; 23 vs 4 at 500-400 m), accompanied by higher biomass (22266 µg vs 4604 µg at 300-200 m; 1992 µg vs 1389 µg at 400-300 m; 561 µg vs 372 µg at 500-400 m). At 200-100 m, biomass in the fraction sum exceeded the combined sample (4124 µg vs 2110 µg), though counts were still higher in the combined sample (134 vs 45). Above 200 m, combined counts again dominated (e.g., 11454 vs 954 at 20-0 m; 4356 vs 1017 at 40-20 m; 1774 vs 321 at 60-40 m; 538 vs 105 at 80-60 m; 407 vs 64 at 100-80 m), with biomass differences paralleling these patterns and reaching several-hundred-percent increases in the surface layers.

Taken together, the combined sample produced systematically higher copepod counts and, in most layers, higher biomass than the sum of the <500 µm and 500-1000 µm fractions at both stations. Occasional cases where the fraction sum contained fewer individuals but higher biomass (e.g., Station 210 at 200-100 m) indicate shifts in size composition: the fractions may hold more small copepods, whereas the combined sample captures similar numbers of larger individuals, yielding higher biomass. These contrasts were strongest in the upper 100 m and consistent across both stations.

4.2.3 Relationships between mesozooplankton community composition and oceanographic conditions

The ordination captured 25.3 % (PCoA1) and 14.3 % (PCoA2) of the variance. Permutation tests (envfit) for temperature, salinity, and pH indicated that salinity showed the strongest association with community structure ($r^2 = 0.144$, $p = 0.020$) (Appendix, Table 9). The salinity vector pointed toward positive PCoA1 and negative PCoA2 (loadings: PC1 = 0.191, PC2 = –0.328), suggesting that samples with higher salinity tended to cluster in that sector of the ordination space (Figure 16). However, after adjustment for multiple testing the association was only marginal ($p_{adj} = 0.080$), and the explanatory power remained low. The remaining

variables showed no clear relationships: temperature ($r^2 = 0.081$, $p = 0.110$), conductivity ($r^2 = 0.069$, $p = 0.163$), and pH ($r^2 = 0.038$, $p = 0.344$).

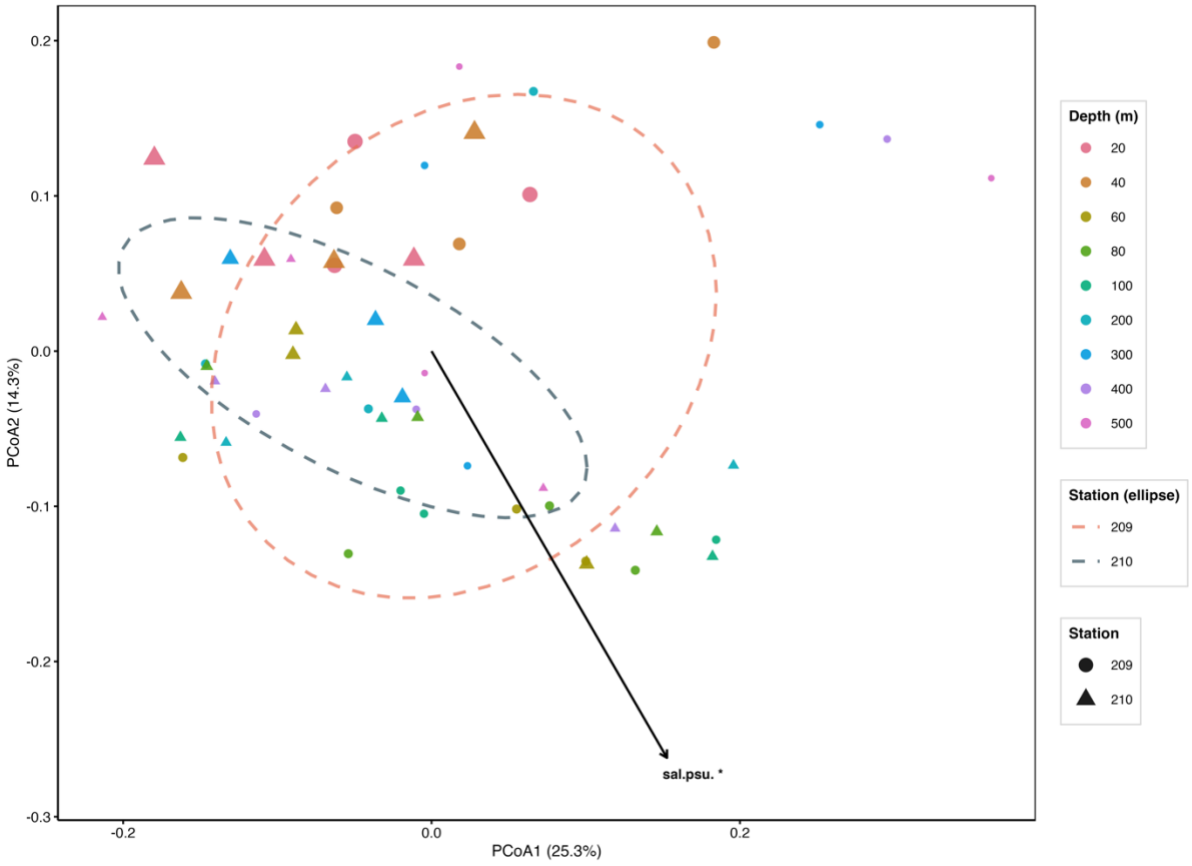


Figure 15: Principal coordinates analysis (PCoA) of zooplankton community composition at Stations 209 and 210 across depths (20-500 m). Point shape indicates station, point color represents depth, and point size is proportional to plankton abundance. Dashed ellipses denote station groupings. The significant environmental vector (salinity, $p < 0.05$) is shown as an arrow, indicating its influence on community structure along the first two PCoA axes (PCoA1: 25.3%, PCoA2: 14.3%).

To separate abundance effects from compositional differences, we also examined rank correlations between total plankton abundance and the same environmental variables (Appendix, Table 10). Total abundance correlated positively and significantly with temperature ($\rho = 0.548$, $p = 1.8\times10^{-5}$)[1] and salinity ($\rho = 0.548$, $p = 1.8\times10^{-5}$), showed a moderate positive association with pH ($\rho = 0.395$, $p = 0.0031$), and no significant relationship with salinity ($\rho = -0.214$, $p = 0.121$). In the PCoA plot, point sizes (scaled to total abundance) were largest at water depth horizons where peak abundances were observed (e.g., 20-40 m at both stations and 200-

[1] ρ denotes Spearman's correlation coefficient, *p* significance level

300 m at station 210), but these abundance gradients did not, by themselves, align as significant compositional drivers indicated by the envfit analysis.

Consistent with our multivariate analyses of community differences, PERMANOVA confirmed a significant station effect ($F = 3.316$, $R^2 = 0.060$, $p = 0.005$) and a weak depth effect ($F = 1.297$, $R^2 = 0.187$, $p = 0.092$). Together, these results indicate that spatial separation between stations is the primary source of β-diversity, while salinity exerts a subtle, directionally consistent influence on community composition along the ordination, and temperature/conductivity primarily covary with abundance rather than species composition at the sampling resolution used here.

5. Discussion

The environment with its physical and biogeochemical conditions provides the framework for the interpretation of mesozooplankton community dynamics and highlight the contrasts between the two systems analyzed here, i.e., the Gulf of Panama and the Gulf of Chiriquí.

5.1 Regional Comparison: Gulf of Panama vs. Gulf of Chiriquí

The most striking ecological outcome of this study is the inversion of expected productivity patterns between the two gulfs, possibly resulting from the failed upwelling in 2025 (O'Dea et al., 2025). The Gulf of Chiriquí generally considered oligotrophic (D'Croz & O'Dea, 2007; MacKenzie et al., 2019), exhibited consistently higher mesozooplankton abundance and copepod biomass than the typically productive Gulf of Panama. Summed abundances reveal nearly threefold higher values in the Gulf of Chiriquí (Figure 12), and copepod biomass was consistently elevated across all depths (Figure 14, 15). Importantly, hydrographic and field observations confirmed that the seasonal upwelling in the Gulf of Panama never developed during 2025, neither before, nor during, or after the sampling campaign, representing a complete collapse of this mechanism (O'Dea et al., 2025).

Under normal conditions, dry-season trade winds drive Ekman transport and nutrient-rich upwelling in the Gulf of Panama, fueling diatom blooms and sustaining high copepod biomass (D'Croz & O'Dea, 2007). In 2025, however, weakened wind forcing and ITCZ displacement suppressed this mechanism, leading to persistently warm, stratified, nutrient-depleted surface waters. Hydrographic profiles from the field campaign (Figures 3 - 7) confirm the absence of typical cold, nutrient-enriched surface layers (O'Dea et al., 2025).

In contrast, the Gulf of Chiriquí, although lacking consistent upwelling, exhibited relatively higher mesozooplankton abundance compared to the Gulf of Panama, including a minor deep mesozooplankton maximum around 200-300 m (Figure 12). This suggests that under stratified conditions, biomass accumulation occurs at subsurface depth, potentially supported by detrital flux or residual subsurface productivity.

Copepods dominated the mesozooplankton assemblages in both systems, comprising 66-79 % of all individuals (Figure 12), consistent with their role as resilient core taxa in tropical ecosystems (Frangoulis et al., 2005; Kiko et al., 2020). Diversity analyses further indicated that

the <500 µm fraction, particularly in the Gulf of Chiriquí, harbored the highest richness and Shannon diversity. By contrast, the >1000 µm fraction was dominated by few large taxa, resulting in reduced diversity.

Notably, mesozooplankton in the GOC (Station 210) showed consistently higher biomass than in the GOP (Station 209) at nearly all depths. Since abundances of the <500 µm fraction were often equal to or higher in the GOP compared to the GOC, the elevated biomass in the GOC must reflect larger average body sizes and later developmental stages. Potential drivers include elevated local production, aggregation at mesoscale features such as fronts or eddies, or recent advection of larger-bodied copepods. This pattern is consistent with ecological theory on life-history strategies: oligotrophic systems such as the GOC tend to favor K-selected taxa, characterized by larger, long-lived individuals that dominate biomass, whereas more eutrophic and variable systems like the GOP often support r-selected, opportunistic species (Margalef, 1978; Schiebel & Hemleben, 2017).

Multivariate analyses confirm these spatial contrasts. Principal Coordinates Analysis (Figure 13) revealed moderate separation between stations with partial overlap, while PERMANOVA identified a significant station effect and only a weak depth effect (Appendix, Table 8). Environmental fitting identified salinity as the only variable with a weak directional effect, while temperature and salinity correlated more strongly with abundance rather than composition (Appendix, Table 10). These results indicate that horizontal separation between gulfs may have been the dominant factor structuring communities under suppressed upwelling.

In summary, mesozooplankton assemblages across the two gulfs were shaped primarily by spatial differences rather than by depth. Both systems exhibited strong surface aggregation at 20–40 m and additional biomass peaks at 200–300 m, with copepods consistently dominating the communities. Size-structuring was also evident, reflecting the contrasting contributions of smaller fractions in the GOP and larger-bodied copepods in the GOC. While diversity patterns varied among fractions and depths, these effects were secondary compared to the pronounced station-level contrasts.

5.2 Ecological Implications

These structural patterns provide the basis for interpreting the ecological implications of the failed upwelling event, illustrating how mesozooplankton communities respond to altered

nutrient supply and stratification. The anomalous conditions documented in 2025 provide an important case study of how tropical plankton communities responded to the failed upwelling event (O'Dea et al., 2025). In the Gulf of Panama, reduced nutrient supply curtailed phytoplankton growth, leading to weakened mesozooplankton biomass. However, the persistence of copepod dominance demonstrates their resilience and capacity to sustain energy transfer even under unfavorable conditions, as, for example, persistence of copepod dominance despite productivity decline (Jiménez-Pérez & Lavaniegos, 2004). Moreover, the detection of a minor subsurface mesozooplankton biomass peak at 200-300 m in the GOC further indicates a downward redistribution of communities, likely reflecting adaptation to detrital pathways and subsurface productivity, consistent with observed subsurface maxima in zooplankton biomass associated with thermocline and OMZ interfaces (Fernández-Álamo & Färber-Lorda, 2006). Taken together, these patterns highlight both the dynamics and adaptability of tropical planktonic ecosystems under disrupted nutrient regimes.

The disproportionate contribution of the 500–1000-µm size fraction to biomass (Figure 15) may indicate an effect of grazing impact and vertical carbon flux. Larger individuals are particularly important because they produce larger fecal pellets that sink more rapidly and are more likely to cross the thermocline, thereby contributing directly to deep-ocean export (Hernández-León, 2009). In addition, zooplankton size and community composition have been shown to regulate both pellet sinking speed and export efficiency, emphasizing that changes in size structure can strongly affect the magnitude of vertical carbon transport (Steinberg & Landry, 2017). Similar findings by Stamieszkin et al., (2015) further demonstrate that body size acts as a master trait controlling fecal pellet flux. Taken together, these studies highlight the need to monitor not only abundance but also size structure when assessing mesozooplankton contributions to carbon cycling.

The long-term implications are significant. If climate change increases the frequency of upwelling failures through shifts in ITCZ dynamics, trade wind variability, and ENSO modulation, the Gulf of Panama may repeatedly experience conditions like those observed in 2025. Such a trend would threaten the productivity of short food chains that support fisheries of economic importance (D'Croz & O'Dea, 2007; O'Dea et al., 2025). In this context, the Gulf of Chiriquí may represent a natural analogue for future oligotrophic states, where communities depend on detrital flux, specialized grazing strategies, and relatively stable conditions for coral

reef systems (Buesa, 2019). Thus, the findings of this study contribute to broader discussions of resilience and vulnerability in tropical marine ecosystems under climate-driven change.

5.3 Methodological Significance

Beyond ecological insights, this study underscores the methodological potential of semi-automated imaging for mesozooplankton monitoring. Systems such as ZooScan deliver reproducible, size-based data at scales previously unattainable through manual microscopy, enabling rapid estimation of size spectra and biomass across large datasets (Gorsky et al., 2010). Autonomous in-situ imaging systems like the APICS, for example, installed at the Western Channel Observatory offer up to a ~100-fold improvement in sampling frequency, capturing rapid plankton dynamics in ways conventional methods cannot (Clark et al., 2025). Semi-automated instruments like the Plankton Imager further advance this field by continuously processing large volumes of water using machine learning classifiers, enhancing both throughput and accuracy (Scott et al., 2021). The development of global, interoperable datasets such as the Pelagic Size Structure database (PSSdb) demonstrates how imaging-derived metrics can support large-scale, comparative ecological studies (Dugenne et al., 2024). Nevertheless, challenges remain-such as image recognition accuracy and the complexity of cluttered datasets-highlighting the need for ongoing methodological refinement (Eerola et al., 2024). These capabilities, combined with routine fractionation underscore that counting-only approaches may severely understate ecologically relevant patterns when size structures shift. The persistent biomass advantage observed in the GOC further illustrates how sensitive, size-informed methodologies can detect station-scale ecological differentiation hidden to abundance-only metrics. Finally, by capturing the anomalous non-upwelling in the year 2025, this dataset offers a valuable baseline for future comparisons when typical upwelling dynamics resume.

5.4 Methodological Reflections: Automated Image Analysis

Automated image analysis has increasingly become an essential tool for plankton research, reducing reliance on time-consuming manual taxonomy and enabling high-throughput sample processing. Similar to our findings, recent advances show that convolutional neural networks can deliver operational, near-real-time plankton recognition, moving the field towards scalable ecological monitoring (Kraft et al., 2022). In this context, the integration of semi-automated imaging through FlowCam with a supervised machine learning classifier substantially improved the efficiency of mesozooplankton community analysis. By processing more than

750,000 images, and exceeding the practical capacity of manual microscopy by far, the analyses included a large number of cases for statistically robust results for dominant taxa such as copepods, Bacillariophyceae, and polychaetes. These methodological advances are consistent with recent efforts to scale plankton monitoring in dynamic marine systems (Luo et al., 2018b; Pierella Karlusich et al., 2022). Importantly, the approach also facilitated abundance assessments and size-based analyses, estimates of community biomass, and ecological interpretation.

The overall performance of the classifier summarized in Table 2 and illustrated in the normalized confusion matrix (Figure 11) shows high accuracy for dominant groups but reduced performance for transparent taxa. Transparent and fragile taxa, including appendicularians and cnidarians, exhibited reduced classification accuracy due to their morphological similarity with debris. Detrital fragments and exoskeletons further inflated the “debris” category, diminishing precision for rare taxa.

Another critical methodological consideration is the handling of size fractions. Direct comparison of separated versus merged samples revealed systematic discrepancies in copepod abundance and biomass estimates, particularly in the surface and mid-depth layers (Figure 15). While merged samples tended to yield higher values, this may reflect an overestimation due to misclassification of small particles as zooplankton, rather than a true increase. By contrast, separated fractions, benefiting from higher image resolution in the individual FlowCams, produced more conservative values and clearer separation of taxa, suggesting that fractionation improves consistency for ecological comparisons.

The rationale for fractionation was further supported by the complementary information provided by the two size classes: in the <500 µm fraction, numerically dominant small forms were more effectively resolved, while the 500–1000 µm fraction accounted for biomass-dominant larger individuals. This distinction is clearly visible in the depth-distributed biomass profiles (Figure 9), where the 500-1000 µm fraction consistently contributed most of the carbon standing stock despite often lower individual numbers. This illustrates a pronounced size-biomass decoupling: biomass increases disproportionately with body size, as the allometric conversion ($C = 6 \cdot L^{2.5}$) amplifies even modest shifts toward larger specimens, i.e. later ontogenetic stages. Consequently, changes in size structure, rather than abundance alone, exert strong control over mesozooplankton carbon stocks (Carlotti et al., 2015; Huskin et al., 2006).

These differences in size fractions were particularly amplified in the upper water column, where accumulation of ontogenetic stages and active growth during productive time-periods lead to small increases in mean body size but disproportionately larger biomass (Le Borgne et al., 2003). This vertical pattern is consistent with the observed distribution of ontogenetic stages and supports the conclusion that surface communities are especially sensitive to size-structure dynamics, as zooplankton aggregate near productivity maxima and act as “gatekeepers” in vertical carbon fluxes (Forest et al., 2012; Frangoulis et al., 2005).

Methodological caveats include boundary effects at the 500 µm sieve-size fraction and may cause slender or damaged individuals to pass into either fraction depending on orientation, and introducing statistical noise, but not accounting for the systematic dominance of the coarser fractions (see general concerns over filtration accuracy in sampling protocols (Karlson et al., 2010)). Subsampling and counting precision are additional sources of variability, as small errors in body-length estimates or underrepresentation of rare large individuals propagate strongly in biomass calculations. Since the carbon conversion relies on pooled coefficients, sensitivity testing (e.g., varying the conversion factor ±10-20 %) would help confirm the robustness of station and fraction patterns.

Together, these findings highlight both the promise and current limitations of semi-automated imaging for ecological field studies. Hybrid workflows combining automation with targeted manual curation remain essential, particularly for rare and fragile taxa. Future improvements should expand training datasets, refine debris classification, and test in-situ imaging systems to further enhance scalability.

5.5 Future Directions

Several pathways for development emerge from this work. Technically, expanding classifier training to better encompass underrepresented taxa and refining debris handling will improve robustness. Concrete follow-ups should include quantifying count-weighted mean body length and biomass share by fraction for each depth and station, using symmetric percent-difference metrics to summarize depth-wise contrasts, and validating FlowCam-derived size distributions against microscopy subsets to confirm upper-tail representation. Ecologically, long-term monitoring across multiple years is required to contrast anomalous seasons with “normal”

upwelling. As plankton organisms may undertake daily vertical migration that alter their distribution (Bandara et al., 2021), it is possible that if samples were taken e.g. in late afternoon, as opposed to in the morning in this study, we would have found a deeper distribution of at least the largest copepods. Thus, in future studies sampling during night and day can be evaluated.

From an applied perspective, integrating semi-automated plankton monitoring into long-term observing networks could provide early-warning indicators of ecosystem stress, support fisheries management, and enhance climate services. Together, these directions highlight both the ecological and methodological relevance of this study for understanding and anticipating mesozooplankton dynamics in a changing ocean. In sum, this study demonstrates how methodological choices, anomalous hydrographic conditions, and regional contrasts shape mesozooplankton communities, setting the stage for the concluding synthesis of their broader ecological and climatic significance.

6. Conclusion

This thesis demonstrates how combining semi-automated imaging and field sampling can provide novel insights into mesozooplankton community dynamics during anomalous oceanographic conditions. By integrating FlowCam imaging with a machine learning classifier, more than 750,000 images were processed with strong accuracy for dominant taxa, delivering abundance and biomass estimates at scales beyond the reach of traditional microscopy. These methodological advances highlight the value of automated size-based imaging as a scalable tool for ecological monitoring, while also exposing persistent challenges in the treatment of transparent taxa, debris (mis-) classification, and fractionation biases. Hybrid workflows that combine automation with targeted manual curation remain essential for achieving robust ecological assessments.

Ecologically, the study captured a rare collapse of the seasonal upwelling in the Gulf of Panama during early 2025, a phenomenon absent before, during, and after the field campaign. This failure inverted the expected productivity gradient between the Gulf of Panama and the Gulf of Chiriquí, with the latter exhibiting higher mesozooplankton abundance and biomass across nearly all depths. Copepods remained the numerically and functionally dominant group, underscoring their resilience under changing conditions and their central role in sustaining energy transfer. The deep biomass maximum observed in the Gulf of Chiriquí highlights that mesozooplankton carbon stocks, and thus their role in export flux, depend not only on abundance but also on shifts in size structure.

From a methodological perspective, the results confirm that fractionation is indispensable for disentangling size-biomass relationships, as different fractions may disproportionately determine biomass and carbon standing stock. This reinforces the importance of monitoring size structure, not just abundance, when assessing mesozooplankton roles in carbon cycling and trophic transfer (Steinberg & Landry, 2017). Future methodological improvements should expand classifier training datasets, refine debris handling, and incorporate sensitivity analyses of biomass conversion, while ecological advances will require multi-year monitoring to contrast anomalous seasons with “normal” upwelling.

In a broader context, this thesis contributes both methodological and ecological knowledge to tropical plankton research. By documenting upwelling failure in the Gulf of Panama, it offers baseline data against which future events may be compared, and provides a proof of concept

for integrating semi-automated imaging into long-term observing systems. As climate change alters wind regimes and ITCZ dynamics (Intergovernmental Panel on Climate Change (IPCC), 2023a), and possibly upwelling dynamics, such scalable approaches will be critical for detecting ecosystem stress, supporting fisheries management, and anticipating changes in carbon cycling. Ultimately, the findings underscore the dual message of resilience and vulnerability: while copepods persist as robust core taxa, the productivity of short food chains and the services they sustain may be increasingly at risk in a warming and more variable ocean.

7. References

Aybar, F., & Tuñon, I. (1997). Diversidad biológica costera y marina del Parque Nacional Coiba, Panamá. *Medio Marino*, 21. http://hdl.handle.net/1834/8048

Bandara, K., Varpe, Ø., Wijewardene, L., Tverberg, V., & Eiane, K. (2021). Two hundred years of zooplankton vertical migration research. *Biological Reviews*, *96*(4), 1547–1589. https://doi.org/10.1111/BRV.12715

Banse, K. (1995). Zooplankton: Pivotal role in the control of ocean production: I. Biomass and production. *ICES Journal of Marine Science*, *52*(3–4), 265–277. https://doi.org/10.1016/1054-3139(95)80043-3

Bochinski, E., Bacha, G., Eiselein, V., Walles, T. J. W., Nejstgaard, J. C., & Sikora, T. (2019). Deep Active Learning for In Situ Plankton Classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11188 LNCS*, 5–15. https://doi.org/10.1007/978-3-030-05792-3_1

Boyd, C. M., Smith, S. L., & Cowles, T. J. (1980). Grazing patterns of copepods in the upwelling system off Peru. *Limnology and Oceanography*, *25*(4), 583–596. https://doi.org/10.4319/lo.1980.25.4.0583

Brugnoli Olivera, E., Molina, L., Till, I., Camarena, M., Morales-Ramírez, A., & Díaz-Ferguson, E. (2023). Mesozooplankton and oceanographic conditions in the North zone of Coiba National Park (Panamá, Central America). *Regional Studies in Marine Science*, *66*. https://doi.org/10.1016/j.rsma.2023.103136

Buesa, R. J. (2019). Plankton based energy transfer rates in four Cuban coastal lagoons. *Estuarine, Coastal and Shelf Science*, *216*, 118–127. https://doi.org/10.1016/j.ecss.2017.10.002

Calbet, A., & Saiz, E. (2005). The ciliate-copepod link in marine ecosystems. *Aquatic Microbial Ecology*, *38*(2), 157–167. https://doi.org/10.3354/AME038157

Carlotti, F., Jouandet, M. P., Nowaczyk, A., Harmelin-Vivien, M., Lefèvre, D., Richard, P., Zhu, Y., & Zhou, M. (2015). Mesozooplankton structure and functioning during the onset

of the Kerguelen phytoplankton bloom during the KEOPS2 survey. *Biogeosciences*, *12*(14), 4543–4563. https://doi.org/10.5194/bg-12-4543-2015

Castellani, C., & Edwards, M. (2017). Marine Plankton: A Practical Guide to Ecology, Methodology, and Taxonomy. In *Limnology and Oceanography Bulletin* (Vol. 26, Issue 4). John Wiley & Sons, Ltd. https://doi.org/10.1002/LOB.10199

Clark, J. R., Fileman, E. S., Fishwick, J., Rühl, S., & Widdicombe, C. E. (2025). The Western Channel Observatory Automated Plankton Imaging And Classification System. *Oceanography*, *38*(1), 29–31. https://doi.org/10.5670/OCEANOG.2025E106

D'Croz, L., & O'Dea, A. (2007). Variability in upwelling along the Pacific shelf of Panama and implications for the distribution of nutrients and chlorophyll. *Estuarine, Coastal and Shelf Science*, *73*(1–2), 325–340. https://doi.org/10.1016/j.ecss.2007.01.013

D'Croz, L., & O'dea, A. (2009). Nutrient and Chlorophyll Dynamics in Pacifi c Central America (Panama). *Proceedings of the Smithsonian Marine Science Symposium*, 335–344. https://www.researchgate.net/publication/230636356

de Azevedo, F., Dias, J. D., Braghin, L. de S. M., & Bonecker, C. C. (2012). Regressões peso-comprimento das espécies de microcrustáceos em uma planície de inundação tropical. *Acta Limnologica Brasiliensia*, *24*(1), 1–11. https://doi.org/10.1590/S2179-975X2012005000021

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, *109*(12), 1–20. https://doi.org/10.1029/2004JC002378

Deneke, R., & Maier G. (2019). Standard-Biomassefaktoren für das Metazooplankton und Empfehlungen zur vereinfachten Biomassebestimmung. In *Zusammenstellung im MS Excel-Format. aktuelle Version 1.0.2*. Download unter: https://www.phytoloss.de.

Dugenne, M., Corrales-Ugalde, M., Luo, J. Y., Kiko, R., O'Brien, T. D., Irisson, J. O., Lombard, F., Stemmann, L., Stock, C., Anderson, C. R., Babin, M., Bhairy, N., Bonnet, S., Carlotti, F., Cornils, A., Crockford, E. T., Daniel, P., Desnos, C., Drago, L., … Vilain, M. (2024). First release of the Pelagic Size Structure database: global datasets of marine

size spectra obtained from plankton imaging devices. *Earth System Science Data*, *16*(6), 2971–2999. https://doi.org/10.5194/essd-16-2971-2024

Dumont, H. J., Van de Velde, I., & Dumont, S. (1975). The dry weight estimate of biomass in a selection of Cladocera, Copepoda and Rotifera from the plankton, periphyton and benthos of continental waters. *Oecologia*, *19*(1), 75–97. https://doi.org/10.1007/BF00377592

Eerola, T., Batrakhanov, D., Barazandeh, N. V., Kraft, K., Haraguchi, L., Lensu, L., Suikkanen, S., Seppälä, J., Tamminen, T., & Kälviäinen, H. (2024). Survey of automatic plankton image recognition: challenges, existing solutions and future perspectives. *Artificial Intelligence Review*, *57*(5). https://doi.org/10.1007/s10462-024-10745-y

Fernández-Álamo, M. A., & Färber-Lorda, J. (2006). Zooplankton and the oceanography of the eastern tropical Pacific: A review. *Progress in Oceanography*, *69*(2–4), 318–359. https://doi.org/10.1016/j.pocean.2006.03.003

Fiedler, P. C., & Lavín, M. F. (2006). Introduction: A review of eastern tropical Pacific oceanography. In *Progress in Oceanography* (Vol. 69, Issues 2–4, pp. 94–100). https://doi.org/10.1016/j.pocean.2006.03.006

Forest, A., Stemmann, L., Picheral, M., Burdorf, L., Robert, D., Fortier, L., & Babin, M. (2012). Size distribution of particles and zooplankton across the shelf-basin system in southeast Beaufort Sea: Combined results from an Underwater Vision Profiler and vertical net tows. *Biogeosciences*, *9*(4), 1301–1320. https://doi.org/10.5194/BG-9-1301-2012

Frangoulis, C., Christou, E. D., & Hecq, J. H. (2005). Comparison of Marine Copepod Outfluxes: Nature, Rate, Fate and Role in the Carbon and Nitrogen Cycles. *Advances in Marine Biology*, *47*, 254–309.

Glynn, P., & Maté, J. L. (1997). Field guide to the Pacific coral reef of panama. *Proc. 8th Int. Coral Reef Sympossium*, *1*, 145–166.

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

(Vol. 32, Issue 3, pp. 285–303). Oxford University Press. https://doi.org/10.1093/plankt/fbp124

Grossmann, M. M., Collins, A. G., & Lindsay, D. J. (2014). Description of the eudoxid stages of Lensia havock and Lensia leloupi (Cnidaria: Siphonophora: Calycophorae), with a review of all known Lensia eudoxid bracts. *Systematics and Biodiversity*, *12*(2), 163–180. https://doi.org/10.1080/14772000.2014.902867

Guglielmo, L., Granata, A., & Guglielmo, R. (1945). *Class Malacostraca; Order Euphausiacea*. 30–35. www.sea-entomologia.org/IDE@

Hernández-León, S. (2009). Top-down effects and carbon flux in the ocean: A hypothesis. *Journal of Marine Systems*, *78*(4), 576–581. https://doi.org/10.1016/J.JMARSYS.2009.01.001

Huskin, I., López, E., Viesca, L., & Anadón, R. (2006). Seasonal variation of mesozooplankton biomass, abundance and copepod grazing in the central Cantabrian Sea (southern Bay of Biscay). *Scientia Marina*, *70*(S1), 119–130. https://doi.org/10.3989/SCIMAR.2006.70S1119

Intergovernmental Panel on Climate Change (IPCC). (2023a). Climate Change 2021 – The Physical Science Basis. *Climate Change 2021 – The Physical Science Basis*. https://doi.org/10.1017/9781009157896

Intergovernmental Panel on Climate Change (IPCC). (2023b). Climate Change 2022 – Impacts, Adaptation and Vulnerability. *Climate Change 2022 – Impacts, Adaptation and Vulnerability*. https://doi.org/10.1017/9781009325844

Jiménez-Pérez, L. C., & Lavaniegos, B. E. (2004). Changes in dominance of copepods off Baja California during the 1997-1999 El Niño and La Niña. *Marine Ecology Progress Series*, *277*, 147–165. https://doi.org/10.3354/meps277147

Karlson, B., Godhe, A., Cusack, C., & Brennan, E. (2010). Intergovernmental Oceanographic Commission Manuals and Guides 55 MICROSCOPIC AND MOLECULAR METHODS FOR QUANTITATIVE PHYTOPLANKTON ANALYSIS. In *Intergovernmental Oceanographic Commission Manuals and Guides*.

Kiko, R., Brandt, P., Christiansen, S., Faustmann, J., Kriest, I., Rodrigues, E., Schütte, F., & Hauss, H. (2020). Zooplankton-Mediated Fluxes in the Eastern Tropical North Atlantic. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.00358

Kraft, K., Velhonoja, O., Eerola, T., Suikkanen, S., Tamminen, T., Haraguchi, L., Ylöstalo, P., Kielosto, S., Johansson, M., Lensu, L., Kälviäinen, H., Haario, H., & Seppälä, J. (2022). Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks. *Frontiers in Marine Science*, *9*, 867695. https://doi.org/10.3389/FMARS.2022.867695/BIBTEX

Lacoursière-Roussel, A., McLean, L., Aubry, C., Maps, F., Finnis, S., Arseneau, J., Milne, R., Macdonald, T., & Guyondet, T. (2025). Contrasting the efficiency of imaging systems for mesozooplankton indicators across Pacific and Atlantic coastal ecosystems. *Ecological Informatics*, *91*, 103372. https://doi.org/10.1016/j.ecoinf.2025.103372

Lampitt, R. S., & Gamble, J. C. (1982). Diet and respiration of the small planktonic marine copepod Oithona nana. *Marine Biology 1982 66:2*, *66*(2), 185–190. https://doi.org/10.1007/BF00397192

Larink, O., & Westheide, W. (2011). *Coastal plankton : Photo Guide for European Seas* (2nd ed., Issue 2). Verlag Dr. Friedrich Pfeil.

Le Borgne, R., Champalbert, G., & Gaudy, R. (2003). Mesozooplankton biomass and composition in the equatorial Pacific along 180. *Journal of Geophysical Research: Oceans*, *108*(12). https://doi.org/10.1029/2000jc000745

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*(2), 271–280. https://doi.org/10.1007/S004420100716/METRICS

Legendre, P., & Legendre, L. (2012). Numerical Ecology. *Developments in Environmental Modelling*, *24*, 337–424. http://www.sciencedirect.com/science/article/pii/B9780444538680500083

León, H. (2009). *Zooplankton biomass estimation from digitized images [Recurso electrónico] : a comparison between subtropical and Antarctic organisms*.

Luo, J. Y., Irisson, J. O., Graham, B., Guigand, C., Sarafraz, A., Mader, C., & Cowen, R. K. (2018a). Automated plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, *16*(12), 814–827. https://doi.org/10.1002/LOM3.10285

Luo, J. Y., Irisson, J. O., Graham, B., Guigand, C., Sarafraz, A., Mader, C., & Cowen, R. K. (2018b). Automated plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, *16*(12), 814–827. https://doi.org/10.1002/lom3.10285

MacKenzie, K. M., Robertson, D. R., Adams, J. N., Altieri, A. H., & Turner, B. L. (2019). Structure and nutrient transfer in a tropical pelagic upwelling food web: From isoscapes to the whole ecosystem. *Progress in Oceanography*, *178*. https://doi.org/10.1016/j.pocean.2019.102145

Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica Acta*, *1*(4), 493–509. https://archimer.ifremer.fr/doc/00123/23403/

Maté, J. L. (2005). Análisis de la situación de la pesca en los golfos de chiriquí y de montijo. *The Nature Conservancy. Panamá*, 1–68.

National Oceanic and Atmospheric Administration. (2025, August 1). *NOAA View Global Data Explorer*. Https://Www.Nnvl.Noaa.Gov/View/Globaldata.Html.

Nejstgaard, J. C., Gismervik, I., & Solberg, P. T. (1997). Feeding and reproduction by Calanus finmarchicus, and microzooplankton grazing during mesocosm blooms of diatoms and the coccolithophore Emiliania huxleyi. *Marine Ecology Progress Series*, *147*(1–3), 197–217. https://doi.org/10.3354/MEPS147197

Nejstgaard, J. C., Naustvoll, L. J., & Sazhin, A. (2001). Correcting for underestimation of microzooplankton grazing in bottle incubation experiments with mesozooplankton. *Marine Ecology Progress Series*, *221*, 59–75. https://doi.org/10.3354/MEPS221059

Nejstgaard, J. C., Tang, K. W., Steinke, M., Dutz, J., Koski, M., Antajan, E., & Long, J. D. (2007). Zooplankton grazing on Phaeocystis : a quantitative review and future challenges. *Biogeochemistry*, *83*(1/3), 147–172. https://doi.org/10.1007/S10533-007-9098-Y

O'Dea, A., Sellers, A. J., Pérez-Medina, C., Pardo Díaz, J., Guzmán Bloise, A., Pöhlker, C., Chiliński, M. T., Aardema, H. M., Cybulski, J. D., Heins, L., Paton, S. R., Slagter, H. A., Schiebel, R., & Haug, G. H. (2025). Unprecedented suppression of Panama's Pacific upwelling in 2025. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(36), e2512056122. https://doi.org/10.1073/PNAS.2512056122/SUPPL_FILE/PNAS.2512056122.SAPP.PDF

Pennington, J. T., Mahoney, K. L., Kuwahara, V. S., Kolber, D. D., Calienes, R., & Chavez, F. P. (2006). Primary production in the eastern tropical Pacific: A review. *Progress in Oceanography*, *69*(2–4), 285–317. https://doi.org/10.1016/J.POCEAN.2006.03.012

Perez, C., Sellers, A. J., Slagter, H., Schiebel, R., & O`Dea, A. (2024). *Chlorophyll Blooms in the Oligotrophic Gulf of Chiriqui, Panama*.

Pierella Karlusich, J. J., Lombard, F., Irisson, J. O., Bowler, C., & Foster, R. A. (2022). Coupling Imaging and Omics in Plankton Surveys: State-of-the-Art, Challenges, and Future Directions. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.878803

Pitois, S. G., Tilbury, J., Bouch, P., Close, H., Barnett, S., & 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*(JAN). https://doi.org/10.3389/fmars.2018.00005

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/>.

Rodriguez-Ruano, V., Toth, L. T., Enochs, I. C., Randall, C. J., & Aronson, R. B. (2023). Upwelling, climate change, and the shifting geography of coral reef development. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-28489-0

Schiebel, R., & Hemleben, C. (2017). Planktic foraminifers in the modern ocean. *Planktic Foraminifers in the Modern Ocean*, 1–358. https://doi.org/10.1007/978-3-662-50297-6

Scott, J., Pitois, S., Close, H., Almeida, N., Culverhouse, P., Tilbury, J., & 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. https://doi.org/10.1093/plankt/fbab018

Siangsano, K. (2025). *IFCB LabelChecker*. https://github.com/KananatSSN/IFCB_LabelChecker.

Sprintall, J., & Roemmich, D. (1999). Characterizing the structure of the surface layer in the Pacific Ocean. *Journal of Geophysical Research: Oceans*, *104*(C10), 23297–23311. https://doi.org/10.1029/1999jc900179

Stamieszkin, K., Pershing, A. J., Record, N. R., Pilskaln, C. H., Dam, H. G., & Feinberg, L. R. (2015). Size as the master trait in modeled copepod fecal pellet carbon flux. *Limnology and Oceanography*, *60*(6), 2090–2107. https://doi.org/10.1002/LNO.10156

Steinberg, D. K., & Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science*, *9*(1), 413–444. https://doi.org/10.1146/annurev-marine-010814-015924

Symiakaki, K., Walles, T. J. W., Park, C., Berger, S. A., Nejstgaard, J. C., Yalçın, G., & Siangsano, K. (2025). *Pipeline for FlowCam data processing with modular open-source software and optional machine 1 learning classification 2*. https://doi.org/10.31219/osf.io/qv2gj_v1

Turner, J. (2004). The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zoological Studies*, *43*(2), 255–266. http://www.sinica.edu.tw/zool/zoolstud/43.2/255.pdf

Uye, S. ichi. (1982). Length-weight relationships of important zooplankton from the Inland Sea of Japan. *Journal of the Oceanographical Society of Japan*, *38*(3), 149–158. https://doi.org/10.1007/BF02110286

Wanek, E., Esteban-Cantillo, O. J., & Bourgeois-Gironde, S. (2025). Valuing marine plankton: a review of ecosystem services and disservices and an expert assessment of the potential of area-based protection. *Frontiers in Marine Science*, *12*. https://doi.org/10.3389/fmars.2025.1607996

8. Appendices

Data Basis

Table 3: Mesozooplankton Community composition

Size_fraction	station	trigger_depth	debris	copepods	polychaeta	chaetognath	cnidaria	cladocera	crustacean	egg	appendicular	chordata	diatom	phronida	echinoderm	nematoda	dinoflagellate
Pan500	209	100	304	98	6	3	4	1	9	1	0	0	0	0	0	0	0
Pan500	209	80	349	120	5	3	8	0	21	0	0	0	0	0	0	0	0
Pan500	209	60	344	130	13	1	8	1	48	0	1	0	0	0	0	0	0
Pan500	209	40	496	245	5	3	6	5	28	25	3	16	0	0	0	0	0
Pan500	209	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pan500	209	500	551	61	1	1	3	0	8	0	4	0	0	0	0	0	0
Pan500	209	400	1061	175	3	1	6	0	10	0	4	0	0	0	0	0	0
Pan500	209	300	769	259	1	0	10	1	23	1	1	0	0	0	3	0	0
Pan500	209	200	4324	1006	126	14	54	18	159	19	21	5	6	1	0	0	1
Pan500	210	100	610	163	5	6	13	1	35	3	5	0	1	0	1	1	0
Pan500	210	80	636	216	9	6	21	1	60	3	5	0	0	0	0	0	3
Pan500	210	60	1231	893	31	14	84	45	78	18	10	5	1	1	1	0	48
Pan500	210	40	4500	4339	15	143	438	11	685	130	206	14	4	4	18	4	238
Pan500	210	20	5298	2616	18	60	72	0	216	120	156	0	6	6	60	0	420
Pan500	210	500	989	58	1	0	39	0	5	21	5	0	0	1	0	0	1
Pan500	210	400	1325	433	19	4	53	11	31	1	10	1	0	0	1	0	4
Pan500	210	300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pan500	210	200	1648	670	18	3	61	11	59	6	14	3	0	0	9	0	8
Pan1000	209	100	190	101	1	4	3	1	3	0	0	0	0	0	0	0	0
Pan1000	209	80	326	213	16	5	7	7	5	0	0	0	1	0	0	0	0
Pan1000	209	60	944	175	63	1	10	5	3	3	10	0	3	0	0	0	0
Pan1000	209	40	2569	665	91	8	27	3	10	3	31	94	7	0	0	0	0
Pan1000	209	20	14259	1164	105	7	18	0	107	11	216	84	23	0	2	0	32
Pan1000	209	500	205	25	0	0	0	0	0	0	1	1	0	0	0	0	0
Pan1000	209	400	573	161	18	1	1	3	1	0	27	0	0	0	0	0	0
Pan1000	209	300	685	142	0	0	0	1	7	0	11	3	0	0	0	0	2
Pan1000	209	200	1045	473	37	5	20	5	1	1	20	5	7	0	1	0	0
Pan1000	210	100	402	157	8	4	4	8	0	0	10	0	0	0	0	0	0
Pan1000	210	80	349	307	8	5	5	1	0	0	4	1	5	0	0	0	1
Pan1000	210	60	560	713	22	16	12	1	7	0	12	7	3	0	0	0	1
Pan1000	210	40	1212	747	53	125	31	1	14	4	82	1	5	0	0	0	10
Pan1000	210	20	5303	2156	26	232	58	4	39	39	336	12	3	0	1	0	207
Pan1000	210	500	559	49	3	1	1	0	0	0	4	0	1	0	0	0	10
Pan1000	210	400	533	190	27	4	7	3	0	0	14	1	0	0	0	0	0
Pan1000	210	300	5553	3299	69	136	42	13	123	33	211	10	11	0	7	0	122
Pan1000	210	200	799	465	16	0	18	4	3	0	30	0	4	0	0	0	0
Panmerge	209	100	11325	1152	19	2	10	2	170	0	49	3	0	0	0	0	8
Panmerge	209	80	17568	2559	99	10	9	8	472	1	74	2	2	0	3	0	15
Panmerge	209	60	15910	1719	130	9	9	11	470	7	47	2	6	0	3	0	23
Panmerge	209	40	22997	4274	203	20	19	13	486	92	101	174	33	0	2	2	31
Panmerge	209	20	514557	18227	557	102	136	11	2875	614	2227	591	80	0	68	0	898
Panmerge	209	500	18500	688	22	5	0	0	91	2	60	7	1	0	0	0	16
Panmerge	209	400	30140	2058	64	10	8	23	194	6	124	1	0	0	0	0	17
Panmerge	209	300	20736	2691	26	5	7	6	241	0	85	5	0	0	0	0	10
Panmerge	209	200	42415	3898	110	9	15	13	415	19	273	14	14	0	4	0	49
Panmerge	210	100	21634	2034	82	7	10	9	200	2	97	0	2	0	1	0	43
Panmerge	210	80	19285	2691	120	10	9	10	333	8	99	1	6	0	0	0	46
Panmerge	210	60	26017	8870	159	18	41	10	795	80	216	26	26	0	13	0	135
Panmerge	210	40	55060	21781	86	118	63	5	2720	228	1800	22	67	0	47	0	810
Panmerge	210	20	91061	50398	117	249	71	17	3744	474	2705	21	92	2	60	0	5788
Panmerge	210	500	25197	512	20	3	16	1	54	4	97	1	0	0	1	0	41
Panmerge	210	400	26796	3066	82	18	19	10	194	5	194	2	5	0	0	0	16
Panmerge	210	300	49285	29136	275	289	79	10	2661	350	1257	32	107	1	14	0	1260
Panmerge	210	200	17947	2945	108	6	22	6	245	2	126	1	2	0	2	0	26
Panmicrosco	209	100	0	42	2	0	0	0	12	0	0	0	0	0	0	0	0
Panmicrosco	209	80	0	60	6	0	8	0	18	0	0	0	0	0	4	0	0
Panmicrosco	209	60	0	38	8	0	14	0	6	0	0	0	0	0	0	0	0
Panmicrosco	209	40	0	488	40	24	8	0	72	0	0	1096	0	0	0	0	0
Panmicrosco	209	20	0	1216	256	0	256	0	256	0	320	192	0	0	64	0	0
Panmicrosco	209	500	0	7	0	0	0	0	6	0	0	1	0	0	0	0	0
Panmicrosco	209	400	0	104	0	0	0	0	16	0	0	32	0	0	0	0	0
Panmicrosco	209	300	0	28	0	2	0	0	6	0	0	8	0	0	0	0	0
Panmicrosco	209	200	0	128	16	0	32	0	16	0	16	80	0	0	0	0	0
Panmicrosco	210	100	0	56	8	0	0	0	16	0	0	0	0	0	0	0	0
Panmicrosco	210	80	0	64	4	0	0	0	8	0	0	0	0	0	0	0	0
Panmicrosco	210	60	0	80	16	0	0	0	4	0	0	0	0	0	0	0	0
Panmicrosco	210	40	0	272	0	72	8	0	40	0	24	8	0	0	0	0	0
Panmicrosco	210	20	0	736	0	256	3	0	32	0	32	0	0	0	0	0	0
Panmicrosco	210	500	0	76	0	0	2	0	2	0	0	0	0	0	0	0	0
Panmicrosco	210	400	0	152	4	0	8	0	32	0	0	0	0	0	0	0	0
Panmicrosco	210	300	0	5664	192	384	384	0	480	0	192	0	0	0	0	0	0
Panmicrosco	210	200	0	88	0	4	0	0	24	0	0	0	0	0	4	0	0

Table 4: CTD data at station 209 and 210 at different trigger depths

Station	record	pres.dbar.	temp.degC.	pH	cond.mS.cm.	PAR.uE.	rawO2.mV.	T_iS.degC.	Chl_a.ug.L.	Battery.V.	CAP25..mS.cm.	date	time	pH_Tc	sal.psu.	sigma.kg.m3.	O_PhPa.	sat...	DO_mg.mg.L.	DO_ml.ml.L.	sound.m.s.	Chl_a.int.	trigger_depth	timestamp
209	7925	503.44	8.02	7.58	36.232	0	4.6	8.03	0.3	10.43	54.86	2025-03-06	15:50:04	7.61	34.76	29.37	5.4	2.55	0.24	0.17	1490.54	678	500	(25-03-06 15:50:04)
209	11517	402.7	9.708	7.61	37.835	0	2.51	9.73	0.29	10.39	54.5	2025-03-06	15:58:38	7.64	34.84	28.69	2.76	1.31	0.12	0.08	1495.22	666	400	(25-03-06 15:58:38)
209	13136	301.95	11.622	7.64	39.71	0	2.83	11.64	0.28	10.35	54.21	2025-03-06	16:02:30	7.67	34.94	27.97	3.03	1.43	0.12	0.09	1500.44	650	300	(25-03-06 16:02:30)
209	14754	201.22	13.156	7.69	41.252	0	10.29	13.17	0.27	10.34	54.05	2025-03-06	16:06:22	7.72	35.04	27.29	11.08	5.25	0.44	0.31	1504.09	638	200	(25-03-06 16:06:22)
209	2670	100.66	14.045	7.69	42.121	0.02	18.31	14.06	0.26	10.48	53.94	2025-03-06	12:31:33	7.71	35.09	26.69	19.1	9.05	0.75	0.52	1505.39	630	100	(25-03-06 12:31:33)
209	5273	80.48	14.234	7.71	42.309	0.09	20.77	14.26	0.29	10.43	53.91	2025-03-06	12:37:45	7.73	35.1	26.57	21.53	10.21	0.84	0.59	1505.68	662	80	(25-03-06 12:37:45)
209	5618	60.43	14.529	7.72	42.592	0.28	24.46	14.56	0.39	10.43	53.87	2025-03-06	12:38:35	7.75	35.1	26.42	25.18	11.94	0.98	0.68	1506.3	798	60	(25-03-06 12:38:35)
209	6047	40.26	15.284	7.74	43.361	1.04	30.76	15.3	0.59	10.43	53.81	2025-03-06	12:39:36	7.77	35.13	26.19	31.47	14.94	1.2	0.84	1508.39	1050	40	(25-03-06 12:39:36)
209	6512	20.17	16.074	7.77	43.985	5.34	33.18	16.12	1.09	10.43	53.54	2025-03-06	12:40:43	7.79	35	25.82	33.68	16	1.27	0.89	1510.35	1702	20	(25-03-06 12:40:43)
209	7127	0	24.878	8.19	5	541.68	183.33	25.37	2.42	10.42	0	2025-03-06	12:42:11	8.19	0.01	-2.91	185.36	89.31	7.31	5.12	1496.39	3430	0	(25-03-06 12:42:11)
210	16459	503.48	8.25	1.54	36.35							2025-03-07	12:43:44		34.66								500	(25-03-07 12:43:44)
210	20855	402.62	9	1.53	37.03							2025-03-07	13:01:26		34.69								400	(25-03-07 13:01:26)
210	24572	301.86	11.37	1.53	39.32							2025-03-07	13:16:25		34.8								300	(25-03-07 13:16:25)
210	5323	201.24	12.768	7.68	40.847	0	8.09	12.79	0.27	10.48	54.07	2025-03-07	14:19:25	7.7	35.02	27.35	8.67	4.1	0.35	0.24	1502.76	638	200	(25-03-07 14:19:25)
210	7411	100.6	13.907	7.74	41.969	0.53	17.54	13.94	0.36	10.45	53.93	2025-03-07	14:24:24	7.77	35.07	26.71	18.29	8.67	0.72	0.5	1504.92	758	100	(25-03-07 14:24:24)
210	7792	80.49	14.377	7.76	42.43	1.67	21.44	14.4	0.53	10.43	53.87	2025-03-07	14:25:18	7.78	35.08	26.53	22.23	10.54	0.87	0.61	1506.12	974	80	(25-03-07 14:25:18)
210	8517	60.34	14.78	7.77	42.804	6.63	22.27	14.82	1.08	10.43	53.8	2025-03-07	14:27:02	7.79	35.07	26.34	22.92	10.87	0.89	0.62	1507.06	1686	60	(25-03-07 14:27:02)
210	8924	40.26	20.681	7.88	47.716	35.61	86.15	20.4	2.21	10.43	52.22	2025-03-07	14:28:00	7.89	34.26	24.2	88.42	42.24	3.11	2.18	1523.2	3162	40	(25-03-07 14:28:00)
210	9255	20.09	29.905	8.24	54.449	296	196.73	29.79	0.64	10.43	49.58	2025-03-07	14:28:48	8.21	32.43	19.92	200.52	97.55	6.17	4.32	1543.08	1118	20	(25-03-07 14:28:48)
210	9907	-0.05	27.915	8.24	1.551	2970.47	181.65	28.19	7.63	10.42	1.46	2025-03-07	14:30:21	8.22	0.73	-3.19	183.65	89.01	6.89	4.83	1504.98	10226	0	(25-03-07 14:30:21)

Table 5: Calculated copepod biomass – copepode_biomass_real represents the standardized values per m³, while copepode_biomass represents the sum from the entire sample

Size_fraction	station	trigger_depth	copepode_biomass	individual_count	copepode_biomass_real	real_count
Pan500	209	100	22802437,68	78	57,00609422	19,5
Pan500	209	80	25138190,55	96	62,84547638	24
Pan500	209	60	27075723,53	104	67,68930882	26
Pan500	209	40	46990321,53	196	117,4758038	49
Pan500	209	20	62650395,25	496	284,774524	225,46
Pan500	209	500	23178157,18	49	11,58907859	2,45
Pan500	209	400	60344039,64	140	30,17201982	7
Pan500	209	300	157185975,1	207	78,59298756	10,35
Pan500	209	200	314349893,7	805	157,1749468	40,25
Pan500	210	100	42825261,38	130	107,0631535	32,5
Pan500	210	80	72256399,61	173	180,640999	43,25
Pan500	210	60	340929647,7	714	852,3241192	178,5
Pan500	210	40	972714799,6	3471	2431,787	867,75
Pan500	210	20	108815363,7	436	1305,784364	523,2
Pan500	210	500	19584319,77	46	9,792159884	2,3
Pan500	210	400	169573187,8	346	84,78659392	17,3
Pan500	210	300	608352704,8	3637	304,1763524	181,852
Pan500	210	200	184744132,4	536	92,37206616	26,8
Pan1000	209	100	559019308,1	74	1524,598113	20,18181818
Pan1000	209	80	1194791339	156	3258,521834	42,54545454
Pan1000	209	60	845506810,7	128	2305,927666	34,9090909
Pan1000	209	40	3819167562	488	10415,91153	133,0909091
Pan1000	209	20	805115207,4	512	3659,61458	232,7272728
Pan1000	209	500	508568274,8	18	277,4008772	0,981818182
Pan1000	209	400	1505574291	118	821,2223408	6,436363636
Pan1000	209	300	845727406,7	104	461,305858	5,672727272
Pan1000	209	200	3929554437	347	2143,39333	18,92727273
Pan1000	210	100	1194659668	115	3258,16273	31,36363636
Pan1000	210	80	2441061640	225	6657,440836	61,36363636
Pan1000	210	60	6530783006	523	17811,22638	142,6363636
Pan1000	210	40	4957407475	548	13520,2022	149,4545455
Pan1000	210	20	7587441906	1976	16554,4187	431,1272728
Pan1000	210	500	664550412,9	36	362,4820434	1,963636364
Pan1000	210	400	2390474493	139	1303,895178	7,58181818
Pan1000	210	300	7882966681	2903	4299,800008	158,3454546
Pan1000	210	200	4766159694	341	2599,72347	18,6
Panmerge	209	100	1344620346	1014	3055,955332	230,4545454
Panmerge	209	80	4020753434	2252	9138,075986	511,8181818
Panmerge	209	60	1484372343	1513	3373,573508	343,8636364
Panmerge	209	40	5536014619	3761	12581,85141	854,7727272
Panmerge	209	20	986664200,6	1604	11212,09319	1822,727273
Panmerge	209	500	719810843	605	327,1867468	27,5
Panmerge	209	400	4074723471	1811	1852,147032	82,3181818
Panmerge	209	300	3267427711	2368	1485,194414	107,6363636
Panmerge	209	200	5536792612	3430	2516,723914	155,9090909
Panmerge	210	100	2937489088	1790	6676,111562	406,8181818
Panmerge	210	80	3768018935	2368	8563,679398	538,1818182
Panmerge	210	60	12112735340	7806	27528,94395	1774,090909
Panmerge	210	40	16653637362	19167	37849,17582	4356,136364
Panmerge	210	20	31680833728	50398	72001,89484	11454,09091
Panmerge	210	500	1234559262	512	561,1633008	23,27272727
Panmerge	210	400	4383291549	3066	1992,40525	139,3636364
Panmerge	210	300	24492707489	29136	22266,09772	2648,727273
Panmerge	210	200	4641772941	2945	2109,896792	133,8636364
Panmicroscope	209	100				8,4
Panmicroscope	209	80				12
Panmicroscope	209	60				7,6
Panmicroscope	209	40				97,6
Panmicroscope	209	20				243,2
Panmicroscope	209	500				0,28
Panmicroscope	209	400				4,16
Panmicroscope	209	300				1,12
Panmicroscope	209	200				5,12
Panmicroscope	210	100				11,2
Panmicroscope	210	80				12,8
Panmicroscope	210	60				16
Panmicroscope	210	40				54,4
Panmicroscope	210	20				147,2
Panmicroscope	210	500				3,04
Panmicroscope	210	400				6,08
Panmicroscope	210	300				226,56
Panmicroscope	210	200				3,52

Table 6: Particle count from the FlowCam samples

Size_fraction	station	trigger_depth	particles_count
Panmerge	209	100	12058
Panmerge	209	80	19812
Panmerge	209	60	17127
Panmerge	209	40	26730
Panmerge	209	20	48823
Panmerge	209	500	17803
Panmerge	209	400	30384
Panmerge	209	300	22452
Panmerge	209	200	70980
Panmerge	210	100	22503
Panmerge	210	80	21163
Panmerge	210	60	34781
Panmerge	210	40	80128
Panmerge	210	20	254858
Panmerge	210	500	26810
Panmerge	210	400	32425
Panmerge	210	300	150150
Panmerge	210	200	22847
Pan500	209	100	1064
Pan500	209	80	602
Pan500	209	60	993
Pan500	209	40	4472
Pan500	209	20	30449
Pan500	209	500	1219
Pan500	209	400	1811
Pan500	209	300	1439
Pan500	209	200	11950
Pan500	210	100	2455
Pan500	210	80	2607
Pan500	210	60	3378
Pan500	210	40	12672
Pan500	210	20	4100
Pan500	210	500	1801
Pan500	210	400	2680
Pan500	210	300	20925
Pan500	210	200	4534
Pan1000	209	100	401
Pan1000	209	80	918
Pan1000	209	60	4483
Pan1000	209	40	5079
Pan1000	209	20	25308
Pan1000	209	500	453
Pan1000	209	400	1674
Pan1000	209	300	1918
Pan1000	209	200	2159
Pan1000	210	100	1081
Pan1000	210	80	759
Pan1000	210	60	1451
Pan1000	210	40	3596
Pan1000	210	20	11655
Pan1000	210	500	1161
Pan1000	210	400	1613
Pan1000	210	300	16398
Pan1000	210	200	2367

Data Analysis

Table 7: Mesozooplankton diversity indices

Station	Depth	ize_fractio	Richness	Shannon	Simpson	Evenness	al_abundar
209	100	<500 µm	7	0,8	0,344	0,411	121,25
209	80	<500 µm	5	0,796	0,388	0,495	156,25
209	60	<500 µm	7	1,013	0,522	0,521	201,25
209	40	<500 µm	9	1,047	0,45	0,477	335
209	500	<500 µm	6	0,802	0,362	0,448	77,5
209	400	<500 µm	6	0,533	0,221	0,298	198,75
209	300	<500 µm	8	0,565	0,243	0,272	298,75
209	200	<500 µm	12	1,103	0,483	0,444	1430
210	100	<500 µm	11	1,115	0,49	0,465	233,75
210	80	<500 µm	9	1,095	0,514	0,499	323,75
210	60	<500 µm	13	1,123	0,459	0,438	1227,5
210	40	<500 µm	14	1,155	0,497	0,438	6246,25
210	20	<500 µm	11	1,158	0,494	0,483	3750
210	500	<500 µm	8	1,392	0,69	0,669	131
210	400	<500 µm	11	0,955	0,406	0,398	567,5
210	200	<500 µm	11	0,926	0,382	0,386	860
209	100	500–1000	6	0,508	0,202	0,284	113,1818
209	80	500–1000	7	0,714	0,298	0,367	255
209	60	500–1000	9	1,102	0,53	0,501	271,3636
209	40	500–1000	10	1,075	0,477	0,467	939,5455
209	20	500–1000	11	1,251	0,542	0,522	1768,182
209	500	500–1000	3	0,394	0,185	0,359	27,27273
209	400	500–1000	7	0,835	0,404	0,429	212,7273
209	300	500–1000	6	0,604	0,26	0,337	165,6364
209	200	500–1000	11	0,803	0,323	0,335	578,1818
210	100	500–1000	6	0,746	0,318	0,416	190,9091
210	80	500–1000	9	0,5	0,182	0,228	339,5455
210	60	500–1000	10	0,528	0,193	0,229	795
210	40	500–1000	11	1,114	0,493	0,464	1074,545
210	20	500–1000	12	1,134	0,499	0,456	3114,545
210	500	500–1000	7	1,043	0,477	0,536	69,54545
210	400	500–1000	7	0,851	0,387	0,437	245,4545
210	300	500–1000	12	0,864	0,339	0,348	4076,136
210	200	500–1000	7	0,608	0,253	0,313	540
209	100	>1000 µm	3	0,665	0,39	0,605	56
209	80	>1000 µm	5	1,12	0,562	0,696	96
209	60	>1000 µm	4	1,121	0,601	0,808	66
209	40	>1000 µm	6	0,95	0,515	0,53	1728
209	20	>1000 µm	7	1,591	0,722	0,818	2560
209	500	>1000 µm	3	0,898	0,561	0,818	14
209	400	>1000 µm	3	0,825	0,476	0,751	152
209	300	>1000 µm	4	1,01	0,541	0,728	44
209	200	>1000 µm	6	1,442	0,704	0,805	288
210	100	>1000 µm	3	0,802	0,46	0,73	80
210	80	>1000 µm	3	0,537	0,277	0,488	76
210	60	>1000 µm	3	0,6	0,333	0,547	100
210	40	>1000 µm	6	1,121	0,547	0,626	424
210	20	>1000 µm	5	0,824	0,457	0,512	1059
210	500	>1000 µm	3	0,233	0,096	0,212	80
210	400	>1000 µm	4	0,703	0,37	0,507	196
210	300	>1000 µm	6	0,877	0,386	0,489	7296
210	200	>1000 µm	4	0,776	0,42	0,56	120

Table 8: Permanova results

Effect	Df	F	R2	p
Station	**1**	**2,64**	**0,081**	**0,019**
Depth	**8**	**1,34**	**0,318**	**0,091**

Table 9: ENVFIT Results

Variable	r2	p_value	p_adj	signif	PC1	PC2	angle_deg
sal.psu.	0,041562	0,063	0,252	.	-0,01973	-0,45109	-92,5047
pH	0,004935	0,331	0,572		0,03857	-0,26223	-81,6326
cond.mS.cm.	0,001623	0,56	0,572		-0,19918	0,024717	172,9263
temp.degC.	0,001469	0,572	0,572		-0,17977	0,077532	156,6696

Table 10: Zooplankton envfit correlations

Index	Env	rho	p_value	signif	label
Total_Abundance	cond.mS.cm.	0.547987616099071	1.79405809593455e-05	**	0.55**
Total_Abundance	pH	0.395037039335212	0.00311395868102728	**	0.40**
Total_Abundance	sal.psu.	-0.213843089314226	0.12050168895343		-0.21
Total_Abundance	temp.degC.	0.547987616099071	1.79405809593455e-05	**	0.55**

Model training results

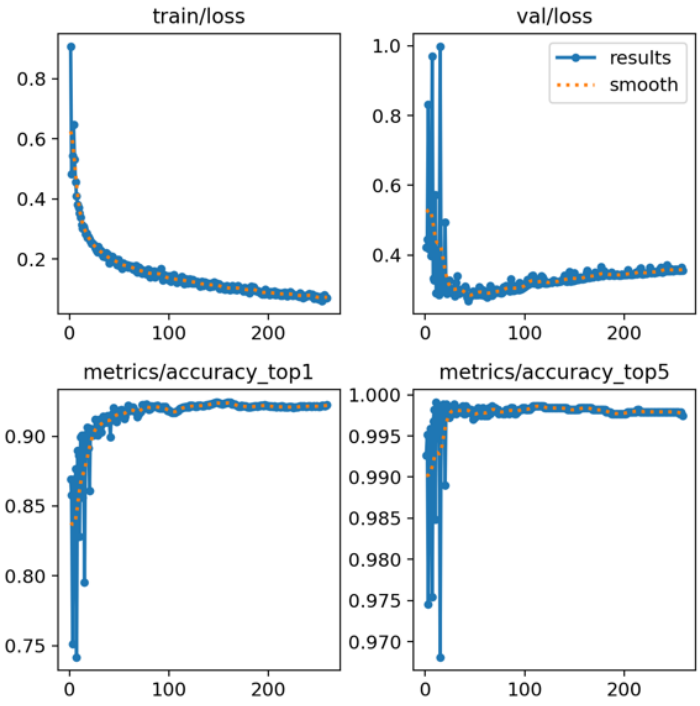


Figure 16: YOLO Model 4 restults

Python skripts

Requirement

```
#Tested environment

#python==3.12.9

numpy==2.2.6

pandas==2.2.3

Pillow==11.2.1

scikit_learn==1.6.1

ultralytics==8.3.134
```

Creating a training dataset

Kananat Siangsano

2025-06-01

```
import os

import shutil

import pandas as pd

import argparse

import glob

from sklearn.model_selection import train_test_split

import numpy as np



def organize_images_with_split(csv_path, source_dir, output_dir, train_ratio=0.7, val_ratio=0.15, test_ratio=0.15, random_state=42):

    """

    Organize images into train/val/test folders based on their labels from a CSV file.

    Rows with empty labels will be skipped.



    Args:

        csv_path: Path to the CSV file containing ImageFilename and LabelTrue columns

        source_dir: Directory containing the source images

        output_dir: Base directory where train/val/test folders will be created

        train_ratio: Proportion of data for training (default: 0.7)

        val_ratio: Proportion of data for validation (default: 0.15)

        test_ratio: Proportion of data for testing (default: 0.15)

        random_state: Random seed for reproducibility (default: 42)

    """

    # Validate split ratios

    if abs(train_ratio + val_ratio + test_ratio - 1.0) > 1e-6:
```

```
        raise ValueError("Train, validation, and test ratios must sum to 1.0")

    # Create output directory structure
    splits = ['train', 'val', 'test']
    for split in splits:
        split_dir = os.path.join(output_dir, split)
        os.makedirs(split_dir, exist_ok=True)

    # Read the CSV file
    df = pd.read_csv(csv_path)

    # Check if required columns exist
    if 'ImageFilename' not in df.columns or 'LabelTrue' not in df.columns:
        raise ValueError("CSV file must contain 'ImageFilename' and 'LabelTrue' columns")

    # Replace empty/NaN labels with a special value and filter out empty labels
    df['LabelTrue'] = df['LabelTrue'].fillna('').astype(str)
    df_filtered = df[df['LabelTrue'].str.strip() != ''].copy()

    if len(df_filtered) == 0:
        print("Warning: No valid labels found in the dataset")
        return

    # Get unique labels and create directories for each split
    unique_labels = df_filtered['LabelTrue'].unique()
    for split in splits:
        for label in unique_labels:
            label_dir = os.path.join(output_dir, split, str(label))
            os.makedirs(label_dir, exist_ok=True)
```

```
    #      print(f"Found      {len(unique_labels)}      unique      labels:
{list(unique_labels)}")


    # Split data for each class to maintain class distribution
    split_stats = {'train': 0, 'val': 0, 'test': 0}
    skipped = 0

    for label in unique_labels:
        # Get all images for this class
        class_data = df_filtered[df_filtered['LabelTrue'] == label].copy()

        if len(class_data) == 0:
            continue

        # If there's only one sample, put it in train
        if len(class_data) == 1:
            train_data = class_data
            val_data = pd.DataFrame()
            test_data = pd.DataFrame()
        elif len(class_data) == 2:
            # If only 2 samples, put one in train and one in test
            train_data = class_data.iloc[:1]
            val_data = pd.DataFrame()
            test_data = class_data.iloc[1:]
        else:
            # First split: separate train from (val + test)
            train_data, temp_data = train_test_split(
                class_data,
                test_size=(val_ratio + test_ratio),
                random_state=random_state,
```

```
            stratify=None
        )

        # Second split: separate val from test
        if len(temp_data) == 1:
            val_data = temp_data
            test_data = pd.DataFrame()
        else:
            val_size = val_ratio / (val_ratio + test_ratio)
            val_data, test_data = train_test_split(
                temp_data,
                test_size=(1 - val_size),
                random_state=random_state,
                stratify=None
            )

    # Copy images to respective split folders
    splits_data = [
        ('train', train_data),
        ('val', val_data),
        ('test', test_data)
    ]

    for split_name, split_data in splits_data:
        if len(split_data) == 0:
            continue

        dest_base_dir = os.path.join(output_dir, split_name, str(label))

        for _, row in split_data.iterrows():
```

```
                image_filename = os.path.basename(str(row['ImageFilename']))
                source_path = os.path.join(source_dir, image_filename)
                dest_path = os.path.join(dest_base_dir, image_filename)

                if os.path.exists(source_path):
                    shutil.copy2(source_path, dest_path)
                    split_stats[split_name] += 1
                else:
                    print(f"Warning: Source image not found: {source_path}")
                    skipped += 1

def organize_images(csv_path, source_dir, output_dir):
    """
    Original function for backward compatibility - now calls the split version with default ratios
    """
    organize_images_with_split(csv_path, source_dir, output_dir)

if __name__ == "__main__":
    # Parse command line arguments for split ratios
    parser = argparse.ArgumentParser(description='Organize images into train/val/test splits')
    parser.add_argument('--input_folder', type=str,
                        help='Input folder path to search for CSV files')
    parser.add_argument('--output_folder', type=str,
                        help='Output folder path for organized dataset')
    parser.add_argument('--train_ratio', type=float, default=0.7, help='Training set ratio (default: 0.7)')
    parser.add_argument('--val_ratio', type=float, default=0.15, help='Validation set ratio (default: 0.15)')
    parser.add_argument('--test_ratio', type=float, default=0.15, help='Test set ratio (default: 0.15)')
```

```
    parser.add_argument('--random_state', type=int, default=42, help='Random
seed for reproducibility (default: 42)')



    args = parser.parse_args()



    # Use command line arguments or defaults

    input_folder_path = args.input_folder

    output_folder_path = args.output_folder



    # Search for files matching the pattern

    csv_files          =          glob.glob(os.path.join(input_folder_path,
'**/LabelChecker_*.csv'), recursive=True)



    # Print the list of matching files

    print(f"Input folder: {input_folder_path}")

    print(f"Output folder: {output_folder_path}")

    print(f"Found {len(csv_files)} LabelChecker files:")

    print(f"Split ratios - Train: {args.train_ratio}, Val: {args.val_ratio},
Test: {args.test_ratio}")

    print(f"Random seed: {args.random_state}")

    print("="*60)



    i = 0

    for file_path in csv_files:

        i += 1

        print(f"\nProcessing {i}/{len(csv_files)}: {file_path}")

        csv_path = file_path

        input_folder_path = os.path.dirname(file_path)

        folder_name = os.path.basename(input_folder_path)

        source_path = os.path.join(input_folder_path, folder_name)



        organize_images_with_split(
```

```
            csv_path,
            source_path,
            output_folder_path,
            train_ratio=args.train_ratio,
            val_ratio=args.val_ratio,
            test_ratio=args.test_ratio,
            random_state=args.random_state
        )
```

Preprocessing of the data

Kananat Siangsano

2025-06-01

```
"""
ROI File Processor CLI
Processes .roi and .adc files to extract images and generate CSV metadata.
"""

import argparse
import sys
from pathlib import Path
import os
import numpy as np
import cv2
import pandas as pd
import uuid



def process_files(input_folder_path, output_folder_path, inplace=False):
    """
```

```
    Process ROI and ADC files to extract images and generate CSV metadata.

    Args:
        input_folder_path (str): Path to input folder
        output_folder_path (str): Path to output folder
        inplace (bool): If True, output to input folder and remove original files
    """
    unique_filenames = list({file.stem for file in Path(input_folder_path).iterdir() if file.is_file()})

    for filename in unique_filenames:
        roi_path = os.path.join(input_folder_path, f"{filename}.roi")
        adc_path = os.path.join(input_folder_path, f"{filename}.adc")
        hdr_path = os.path.join(input_folder_path, f"{filename}.hdr")

        # Check if required files exist
        if not (Path(roi_path).exists() and Path(adc_path).exists()):
            print(f"Skipping {filename}: missing .roi or .adc file")
            continue

        # Create output directory structure
        output_image_folder_path = os.path.join(output_folder_path, filename, filename)
        os.makedirs(output_image_folder_path, exist_ok=True)

        # Set up DataFrame columns
        LC_col_str = "Name,Date,Time,CollageFile,ImageFilename,Id,GroupId,Uuid,SrcImage,SrcX,SrcY,ImageX,ImageY,ImageW,ImageH,Timestamp,ElapsedTime,CalConst,CalImage,AbdArea,AbdDiameter,AbdVolume,AspectRatio,AvgBlue,AvgGreen,AvgRed,BiovolumeCylinder,BiovolumePSpheroid,BiovolumeSphere,Ch1Area,Ch1Peak,Ch1Width,Ch2Area,Ch2Ch1Ratio,Ch2Peak,Ch2Width,Ch3Area,Ch3Peak,Ch3Width,CircleFit,Circularity,CircularityHu,Compactness,ConvexPerimeter,Convexity,EdgeGradient,Elongation,EsdDiameter,EsdVolume,FdDiameter,FeretMaxAngle,FeretMinAngle,FiberCurl,FiberStraightness,FilledArea,FilterScore,GeodesicAspectRatio,GeodesicLength,Geod
```

```
esicThickness,Intensity,Length,Perimeter,Ppc,RatioBlueGreen,RatioRedBlue,Ra
tioRedGreen,Roughness,ScatterArea,ScatterPeak,SigmaIntensity,SphereComplemen
t,SphereCount,SphereUnknown,SphereVolume,SumIntensity,Symmetry,Transparenc
y,Width,BiovolumeHSosik,SurfaceAreaHSosik,Preprocessing,PreprocessingTrue,L
abelPredicted,ProbabilityScore,LabelTrue"

        LC_columns = LC_col_str.split(",")

        LC_df = pd.DataFrame(columns=LC_columns)


        print(f"Processing {filename}...")


        # Read ROI data

        roi_data = np.fromfile(roi_path, dtype="uint8")


        # Process ADC file

        with open(adc_path, 'r') as adc_data:

            for i, adc_line in enumerate(adc_data):

                adc_line = adc_line.split(",")

                output_image_name = f"{filename}_{i:05d}.png"

                output_image_path     =     os.path.join(output_folder_path,
filename, filename, output_image_name)


                # Populate DataFrame

                LC_df.loc[i, 'Name'] = filename

                LC_df.loc[i, 'ImageFilename'] = output_image_name

                LC_df.loc[i, 'Id'] = i

                LC_df.loc[i, 'GroupId'] = i

                LC_df.loc[i, 'ImageW'] = adc_line[15]

                LC_df.loc[i, 'ImageH'] = adc_line[16]

                LC_df.loc[i, 'Uuid'] = uuid.uuid4().hex.upper()


                # Extract and save image

                roi_x = int(adc_line[15])  # ROI width

                roi_y = int(adc_line[16])  # ROI height
```

```
                image_size = roi_x * roi_y


                if image_size > 0:
                    roi_start_bit = int(adc_line[17])
                    roi_end_bit = roi_start_bit + image_size
                    image = roi_data[roi_start_bit:roi_end_bit].reshape((roi_y, roi_x))
                    cv2.imwrite(output_image_path, image)


        # Save CSV file
        LC_file_name = f"LabelChecker_{filename}.csv"
        LC_file_path = os.path.join(output_folder_path, filename, LC_file_name)
        LC_df.to_csv(LC_file_path, index=False)


        # Remove original files if inplace option is enabled
        if inplace:
            print(f"Removing original files for {filename}...")
            try:
                os.remove(roi_path)
                print(f"  Removed {roi_path}")
            except FileNotFoundError:
                pass


            try:
                os.remove(adc_path)
                print(f"  Removed {adc_path}")
            except FileNotFoundError:
                pass


            try:
```

```
                os.remove(hdr_path)
                print(f"  Removed {hdr_path}")
            except FileNotFoundError:
                pass

        print(f"Completed processing {filename}")


def main():
    parser = argparse.ArgumentParser(description='Process ROI and ADC files to extract images and generate CSV metadata.')

    parser.add_argument('input_folder',
                        help='Path to the input folder containing .roi, .adc, and .hdr files')

    parser.add_argument('-o', '--output',
                        default=None,
                        help='Path to the output folder (default: same as input folder)')

    parser.add_argument('--inplace',
                        action='store_true',
                        help='Process files in place (output to input folder and remove original files)')

    parser.add_argument('-v', '--verbose',
                        action='store_true',
                        help='Enable verbose output')

    args = parser.parse_args()
```

```
    # Validate input folder
    input_folder_path = Path(args.input_folder)
    if not input_folder_path.exists():
        print(f"Error: Input folder '{input_folder_path}' does not exist.")
        sys.exit(1)

    if not input_folder_path.is_dir():
        print(f"Error: '{input_folder_path}' is not a directory.")
        sys.exit(1)

    # Determine output folder
    if args.output:
        output_folder_path = Path(args.output)
    else:
        output_folder_path = input_folder_path

    # If inplace is specified, force output to be same as input
    if args.inplace:
        output_folder_path = input_folder_path
        if args.output and Path(args.output) != input_folder_path:
            print("Warning: --inplace specified, ignoring output folder argument")

    # Create output folder if it doesn't exist
    output_folder_path.mkdir(parents=True, exist_ok=True)

    if args.verbose:
        print(f"Input folder: {input_folder_path}")
        print(f"Output folder: {output_folder_path}")
        print(f"In-place processing: {args.inplace}")
```

```
    # Process files
    try:
        process_files(str(input_folder_path),       str(output_folder_path),
args.inplace)
        print("Processing completed successfully!")
    except Exception as e:
        print(f"Error during processing: {e}")
        sys.exit(1)



if __name__ == "__main__":
    main()
```

Model Prediction

Kananat Siangsano

2025-06-01

```
import os
import pandas as pd
import glob
import argparse
from ultralytics import YOLO #8.3.44
import numpy as np


def predict_LC(csv_path, model_path, threshold=0.5):
    """
    Use YOLO model to predict labels for images with empty LabelTrue in CSV
file.
    Updates the CSV file with predictions and probability scores.
```

```
    Args:
        csv_path: Path to the CSV file containing ImageFilename and LabelTrue
columns
        model_path: Path to the YOLO model file
        threshold: Minimum confidence threshold for predictions (default:
0.5)
    """
    # Read the CSV file
    df = pd.read_csv(csv_path)

    # This is to make pandas stop crying
    df['LabelTrue'] = df['LabelTrue'].fillna('')
    df['ProbabilityScore'] = df['ProbabilityScore'].astype(float)
    df['LabelPredicted'] = df['LabelPredicted'].astype(str)
    df['Preprocessing'] = df['Preprocessing'].astype(str)

    # Load YOLO model
    model = YOLO(model_path)

    # Determine image folder path
    folder_path = os.path.dirname(csv_path)
    folder_name = os.path.basename(folder_path)
    image_folder = os.path.join(folder_path, folder_name)

    # Count images to process
    images_to_process = len(df[(df['LabelTrue'] == '') & (df['Preprocessing']
!= 'small')])
    processed_count = 0
    error_count = 0
    prediction_count = 0

    print(f"     → Image folder: {os.path.basename(image_folder)}")
```

```
    print(f"     → Images to process: {images_to_process}")
    print(f"     → Confidence threshold: {threshold}")

    # Process images
    for index, row in df.iterrows():
        if index % 500 == 0 and index > 0:
            print(f"     → Processed {index}/{len(df)} rows...")

        if row['LabelTrue'] == '' and row['Preprocessing'] != 'small':
            processed_count += 1
            try:
                image_filename = row['ImageFilename']
                image_path = os.path.join(image_folder, image_filename)

                results = model(image_path, verbose=False)
                probability_score = float(results[0].probs.top1conf.item())
# Get confidence
                label_predicted                                          =
str(results[0].names[results[0].probs.top1])  # Get class name

                if probability_score >= threshold:
                    df.loc[index,           'ProbabilityScore']          =
round(probability_score, 2)
                    df.loc[index, 'LabelPredicted'] = label_predicted
                    prediction_count += 1

            except Exception as e:
                print(f"     → ERROR: Failed to process {image_filename}")
                error_count += 1
        else:
            df.loc[index, 'ProbabilityScore'] = np.nan
            df.loc[index, 'LabelPredicted'] = ''
```

```
    # Save updated CSV
    df.to_csv(csv_path, index=False, na_rep='')

    print(f"     → COMPLETED: {prediction_count} predictions made")
    print(f"     → Errors: {error_count} images skipped")

    return processed_count, prediction_count, error_count

def process_csv_files(folder_path, model_path, threshold=0.5):
    """
    Process all LabelChecker_*.csv files in the given folder and subfolders.

    Args:
        folder_path: Root directory to search for CSV files
        model_path: Path to the YOLO model file
        threshold: Minimum confidence threshold for predictions
    """
    # Search for files matching the pattern
    csv_files = glob.glob(os.path.join(folder_path, '**/LabelChecker_*.csv'), recursive=True)

    if not csv_files:
        print(f"No LabelChecker_*.csv files found in {folder_path}")
        return

    print(f"Found {len(csv_files)} LabelChecker CSV files:")
    print("="*60)

    total_processed = 0
    total_predictions = 0
```

```
    total_errors = 0

    for i, file_path in enumerate(csv_files, 1):
        csv_path = file_path
        relative_path = os.path.relpath(csv_path, folder_path)

        print(f"{i:3d}. {relative_path}")

        try:
            processed, predictions, errors = predict_LC(csv_path, model_path, threshold)
            total_processed += processed
            total_predictions += predictions
            total_errors += errors

        except Exception as e:
            print(f"     → ERROR: Failed to process file - {str(e)}")

    # Print summary
    print("="*60)
    print(f"PREDICTION SUMMARY:")
    print(f"Total CSV files processed: {len(csv_files)}")
    print(f"Total images processed: {total_processed}")
    print(f"Total predictions made: {total_predictions}")
    if total_errors > 0:
        print(f"Total errors: {total_errors}")
    print(f"Model used: {os.path.basename(model_path)}")
    print(f"Confidence threshold: {threshold}")

if __name__ == "__main__":
```

```
    parser = argparse.ArgumentParser(description='Use YOLO model to predict
labels for images in LabelChecker CSV files')

    parser.add_argument('--input_folder', type=str,

                        help='Input folder path to search for CSV files')

    parser.add_argument('--model_path', type=str,

                        help='Path to YOLO model file')

    parser.add_argument('--threshold', type=float, default=0.8,

                        help='Confidence threshold for predictions (default:
0.8)')


    args = parser.parse_args()


    # Validate input folder exists

    if not os.path.exists(args.input_folder):

        print(f"Error: Input folder does not exist: {args.input_folder}")

        exit(1)


    # Validate model file exists

    if not os.path.exists(args.model_path):

        print(f"Error: Model file does not exist: {args.model_path}")

        exit(1)


    print(f"Input folder: {args.input_folder}")

    print(f"Model path: {args.model_path}")

    print(f"Confidence threshold: {args.threshold}")

    print()


    process_csv_files(args.input_folder, args.model_path, args.threshold)
```

YAML data for the YOLO model

Kananat Siangsano

2025-06-01

```yaml
task: classify
mode: train
model: yolo11m-cls.pt
data: E:\Lina\training data
epochs: 500
time: null
patience: 100
batch: 32
imgsz: 224
save: true
save_period: -1
cache: false
device: '0'
workers: 8
project: null
name: train2
exist_ok: false
pretrained: true
optimizer: auto
verbose: true
seed: 0
deterministic: true
single_cls: false
rect: false
```

```
cos_lr: false
close_mosaic: 10
resume: false
amp: true
fraction: 1.0
profile: false
freeze: null
multi_scale: false
overlap_mask: true
mask_ratio: 4
dropout: 0.0
val: true
split: val
save_json: false
conf: null
iou: 0.7
max_det: 300
half: false
dnn: false
plots: true
source: null
vid_stride: 1
stream_buffer: false
visualize: false
augment: false
agnostic_nms: false
classes: null
retina_masks: false
embed: null
show: false
```

```
save_frames: false
save_txt: false
save_conf: false
save_crop: false
show_labels: true
show_conf: true
show_boxes: true
line_width: null
format: torchscript
keras: false
optimize: false
int8: false
dynamic: false
simplify: true
opset: null
workspace: null
nms: false
lr0: 0.01
lrf: 0.01
momentum: 0.937
weight_decay: 0.0005
warmup_epochs: 3.0
warmup_momentum: 0.8
warmup_bias_lr: 0.1
box: 7.5
cls: 0.5
dfl: 1.5
pose: 12.0
kobj: 1.0
nbs: 64
```

```
hsv_h: 0.015
hsv_s: 0.7
hsv_v: 0.4
degrees: 0.0
translate: 0.1
scale: 0.5
shear: 0.0
perspective: 0.0
flipud: 0.0
fliplr: 0.5
bgr: 0.0
mosaic: 1.0
mixup: 0.0
cutmix: 0.0
copy_paste: 0.0
copy_paste_mode: flip
auto_augment: randaugment
erasing: 0.4
cfg: null
tracker: botsort.yaml
save_dir: runs\classify\train2
```

R Skripts

Community diversity analysis

Lina Becker

2025-09-16

```
library(tidyverse)
library(vegan)
```

```r
library(ggplot2)
library(viridis)

library(patchwork)
library(writexl)
library(tidyr)

theme_publication <- function(base_size = 18) {
  theme_classic(base_size = base_size) %+replace%
    theme(
      panel.border = element_rect(colour = "black", fill = NA, size = 1.1)
,
      axis.line = element_line(size = 0.9),
      axis.ticks = element_line(size = 0.9),
      legend.background = element_blank(),
      legend.key = element_blank(),
      legend.position = "right",
      legend.title = element_text(size = base_size * 0.9, face = "bold"),
      legend.text = element_text(size = base_size * 0.85),
      strip.background = element_blank(),
      strip.text = element_text(size = base_size * 1.1, face = "bold"),
      plot.title = element_text(size = base_size * 1.3, face = "bold"),
      plot.subtitle = element_text(size = base_size * 0.95, face = "plain"
, margin = margin(b = 8)),
      axis.text = element_text(size = base_size * 0.9, color = "black"),
      axis.title = element_text(size = base_size, face = "plain"),
      panel.grid = element_blank()
    )
}

setwd("/Volumes/Linas_data/masterarbeit_sortiert/Appendix/daten_rechnungen
")
df_wide <- read.csv2("Cal_mesozooplankton_community_m3.csv", sep = ";")

df_clean <- df_wide %>%
  pivot_longer(
    cols = Debris:Dinoflagellata,   # alle Taxa
    names_to = "Taxa",
    values_to = "Count"
  ) %>%
  mutate(
    Count = as.numeric(Count),
    Size_fraction = recode(Size_fraction,
                           "Pan500" = "<500 µm",
                           "Pan1000" = "500–1000 µm",
                           "Panmicroscope" = ">1000 µm"
    )
  ) %>%
  filter(Taxa != "Debris", !is.na(Count), Size_fraction != "Panmerge")

df_clean$Size_fraction <- factor(df_clean$Size_fraction, levels = c("<500
µm", "500–1000 µm", ">1000 µm"))

df_abs <- df_clean %>%
```

```
  group_by(station, trigger_depth, Size_fraction, Taxa) %>%
  summarise(Total = sum(Count, na.rm = TRUE), .groups = "drop")

df_abs_full <- df_abs %>%
  complete(station, trigger_depth, Size_fraction, Taxa, fill = list(Total
= 0))

df_abs_full$Size_fraction <- factor(df_abs_full$Size_fraction, levels = c(
"<500 µm", "500–1000 µm", ">1000 µm"))

df_percent_full <- df_abs_full %>%
  group_by(station, trigger_depth, Size_fraction) %>%
  mutate(Percent = 100 * Total / sum(Total, na.rm=TRUE)) %>%
  ungroup() %>%
  mutate(Percent = ifelse(is.na(Percent), 0, Percent))

df_percent_full$Size_fraction <- factor(df_percent_full$Size_fraction, lev
els = c("<500 µm", "500–1000 µm", ">1000 µm"))

depth_levels <- sort(unique(df_abs_full$trigger_depth))
depth_levels_rev <- rev(depth_levels)

df_percent_full <- df_abs_full %>%
  group_by(station, trigger_depth, Size_fraction) %>%
  mutate(Percent = 100 * Total / sum(Total, na.rm = TRUE)) %>%
  mutate(Percent = Percent / sum(Percent, na.rm = TRUE) * 100) %>%
  ungroup()

df_abs$Size_fraction          <- factor(df_abs$Size_fraction, levels = c("
<500 µm", "500–1000 µm", ">1000 µm"))
df_percent_full$Size_fraction <- factor(df_percent_full$Size_fraction, lev
els = c("<500 µm", "500–1000 µm", ">1000 µm"))

station_labels <- c(
  `209` = "209 — Gulf of Panama",
  `210` = "210 — Gulf of Chiriquí"
)

plot_abs <- ggplot(
  df_abs_full,
  aes(x = Total,
      y = factor(trigger_depth, levels = depth_levels_rev),
      fill = Taxa)
) +
  geom_bar(stat = "identity", color = "white", width = 0.85) +
  facet_grid(Size_fraction ~ station, scales = "free_x",
             labeller = labeller(station = station_labels)) +
  scale_fill_viridis_d(option = "turbo") +
  scale_x_continuous(expand = expansion(mult = c(0, 0.05))) +
  theme_publication(base_size = 16) +
  theme(
    axis.text.x = element_text(size = 13),
    axis.text.y = element_text(size = 13),
    axis.title = element_text(size = 15),
```

```
    strip.text = element_text(size = 14),
    legend.position = "bottom"
  ) +
  labs(
    x = "Individuals (count) m⁻³",
    y = "Water depth (m)",
    fill = "Plankton taxa"
  )

plot_rel <- ggplot(
  df_percent_full,
  aes(x = Percent,
      y = factor(trigger_depth, levels = depth_levels_rev),
      fill = Taxa)
) +
  geom_bar(stat = "identity", color = "white", width = 0.85) +
  facet_grid(Size_fraction ~ station,
             labeller = labeller(station = station_labels)) +
  scale_fill_viridis_d(option = "turbo") +
  scale_x_continuous(expand = expansion(mult = c(0, 0.02))) +
  coord_cartesian(xlim = c(0, 100)) +
  theme_publication(base_size = 16) +
  theme(
    axis.text.x = element_text(size = 13),
    axis.text.y = element_text(size = 13),
    axis.title = element_text(size = 15),
    strip.text = element_text(size = 14),
    legend.position = "bottom"
  ) +
  labs(
    x = "Percentage (%)",
    y = "Water depth (m)",
    fill = "Plankton taxa"
  )

combined_plot <- (plot_abs / plot_rel) + plot_layout(guides = "collect") &
theme(legend.position = "bottom")
combined_plot

ggsave("zooplankton_barplots_stacked.png",
       plot = combined_plot,
       width = 20, height = 25, dpi = 400, units = "in")


df_div_wide <- df_clean %>%
  select(Size_fraction, station, trigger_depth, Taxa, Count) %>%
  pivot_wider(names_from = Taxa, values_from = Count, values_fill = 0)
meta <- df_div_wide %>% select(Size_fraction, station, trigger_depth)
taxa_matrix <- df_div_wide %>% select(-Size_fraction, -station, -trigger_d
epth)


diversity_df <- meta %>%
  mutate(
```

```
    richness = specnumber(taxa_matrix),
    shannon = diversity(taxa_matrix, index = "shannon"),
    simpson = diversity(taxa_matrix, index = "simpson"),
    evenness = ifelse(richness > 0, shannon / log(richness), NA)
  ) %>%
  mutate(across(c(richness, shannon, simpson, evenness), ~ round(., 3)))
diversity_df$Size_fraction <- factor(diversity_df$Size_fraction, levels =
c("<500 µm", "500–1000 µm", ">1000 µm"))

pair_keys <- diversity_df %>%
  group_by(Size_fraction, trigger_depth) %>%
  filter(n() == 2) %>% ungroup() %>%
  select(Size_fraction, trigger_depth)

div_df_paired <- diversity_df %>%
  semi_join(pair_keys, by = c("Size_fraction", "trigger_depth"))
div_df_paired$Size_fraction <- factor(div_df_paired$Size_fraction, levels
= c("<500 µm", "500–1000 µm", ">1000 µm"))

paired_wilcox_test <- function(div_df, index) {
  df_wide <- div_df %>%
    select(Size_fraction, trigger_depth, station, !!sym(index)) %>%
    pivot_wider(names_from = station, values_from = !!sym(index), names_pr
efix = "station_")
  s1 <- grep("station_209", names(df_wide), value = TRUE)
  s2 <- grep("station_210", names(df_wide), value = TRUE)
  if (length(s1) == 0 | length(s2) == 0) return(NULL)
  wilcox.test(df_wide[[s1]], df_wide[[s2]], paired = TRUE)
}
wilc_shannon <- paired_wilcox_test(div_df_paired, "shannon")
print(wilc_shannon)

##
##  Wilcoxon signed rank exact test
##
## data:  df_wide[[s1]] and df_wide[[s2]]
## V = 205, p-value = 0.714
## alternative hypothesis: true location shift is not equal to 0

abundance_summary <- df_clean %>%
  group_by(Size_fraction, station, trigger_depth) %>%
  summarise(total_abundance = sum(Count, na.rm = TRUE), .groups = "drop")

diversity_summary <- diversity_df %>%
  left_join(abundance_summary, by = c("Size_fraction", "station", "trigger
_depth"))

diversity_summary <- diversity_summary %>%
  select(
    Station = station,
    Depth = trigger_depth,
    Size_fraction,
    Richness = richness,
    Shannon = shannon,
```

```
    Simpson = simpson,
    Evenness = evenness,
    Total_abundance = total_abundance
  )
library(writexl)
write_xlsx(diversity_summary, "zooplankton_diversity_indices.xlsx")
write.csv(diversity_summary, "zooplankton_diversity_indices.csv", row.name
s = FALSE)

library(tidyverse)
library(ggpubr)

size_labels <- c("<500 µm", "500–1000 µm", ">1000 µm")
names(size_labels) <- c("<500 µm", "500–1000 µm", ">1000 µm")

depth_colors <- c(
  "20" = "#440154",  "40" = "#482475",  "60" = "#414487",
  "80" = "#31688e", "100" = "#26828e", "200" = "#1fa187",
  "300" = "#35b779", "400" = "#6ece58", "500" = "#fde725"
)

library(forcats)

div_df_paired$Size_fraction_plot <- factor(
  div_df_paired$Size_fraction,
  levels = c("<500 µm", "500–1000 µm", ">1000 µm")
)

wilcox_results <- div_df_paired %>%
  group_by(Size_fraction, Size_fraction_plot) %>%
  summarise(
    p = wilcox.test(
      shannon[station == 209],
      shannon[station == 210],
      paired = TRUE
    )$p.value,
    .groups = "drop"
  ) %>%
  mutate(label = paste0("Wilcoxon p = ", formatC(p, format = "e", digits =
2)))

p <- ggplot(div_df_paired, aes(x = factor(station), y = shannon, group = t
rigger_depth)) +
  geom_boxplot(width = 0.65, outlier.shape = NA, fill = "white", color = "
black", size = 1) +
  geom_line(aes(group = trigger_depth), color = "grey60", linetype = "dott
ed", size = 0.8, alpha = 0.7) +
  geom_point(aes(color = factor(trigger_depth)), size = 2.2, alpha = 0.95)
+
  facet_wrap(~Size_fraction_plot, labeller = as_labeller(size_labels), nro
w = 1, strip.position = "bottom") +
  scale_color_manual(values = depth_colors, name = "Depth (m)") +
  labs(
    x = "Station",
```

```
    y = "Shannon index"
  ) +
  theme_publication(base_size = 17)

p_final <- p +
  geom_text(
    data = wilcox_results,
    aes(
      x = 1.5,
      y = max(div_df_paired$shannon, na.rm = TRUE) + 0.08,
      label = label
    ),
    inherit.aes = FALSE,
    size = 5.3,
    fontface = "italic"
  )

print(p_final)

ggsave("shannon_boxplot_publication.png", plot = p_final, width = 12, heig
ht = 6, dpi = 400)

taxa_matrix_paired <- df_div_wide %>%
  semi_join(pair_keys, by = c("Size_fraction", "trigger_depth")) %>%
  select(-Size_fraction, -station, -trigger_depth)
meta_paired <- df_div_wide %>%
  semi_join(pair_keys, by = c("Size_fraction", "trigger_depth")) %>%
  select(Size_fraction, station, trigger_depth)

row_sums <- rowSums(taxa_matrix_paired)
taxa_matrix_paired <- taxa_matrix_paired[row_sums > 0, ]
meta_paired <- meta_paired[row_sums > 0, ]

taxa_matrix_paired[is.na(taxa_matrix_paired)] <- 0


comm_hell <- decostand(taxa_matrix_paired, method = "hellinger")
set.seed(123)
comm_dist <- vegdist(comm_hell, method = "bray")

library(colorspace)

depth_colors <- qualitative_hcl(11, palette = "Dark 3")

pcoa <- cmdscale(comm_dist, k = 2, eig = TRUE)
eigs <- pcoa$eig
var_expl <- round(100 * eigs[1:2] / sum(eigs[eigs > 0]), 1)
xlab <- paste0("PCoA1 (", var_expl[1], "%)")
ylab <- paste0("PCoA2 (", var_expl[2], "%)")

pcoa_points <- as.data.frame(pcoa$points)
colnames(pcoa_points) <- c("PCoA1", "PCoA2")
pcoa_points <- cbind(pcoa_points, meta_paired)
unique_depths <- sort(unique(pcoa_points$trigger_depth))
```

```r
depth_colors <- qualitative_hcl(length(unique_depths), palette = "Dark 3")
names(depth_colors) <- unique_depths

ggplot(pcoa_points, aes(x = PCoA1, y = PCoA2, color = factor(trigger_depth
), shape = factor(station))) +
  geom_point(size = 4, alpha = 0.9) +
  stat_ellipse(aes(group = factor(station), color = factor(station)),
               linetype = "dashed", level = 0.68, size = 1, alpha = 0.5, s
how.legend = TRUE) +
  scale_color_manual(values = depth_colors, name = "Depth (m)") +
  scale_shape_manual(values = c(16, 17), name = "Station") +
  theme_minimal(base_size = 15) +
  coord_equal() +
  labs(
    x = xlab, y = ylab,
    color = "Depth (m)",
    shape = "Station"
  )

ggplot(pcoa_points, aes(x = PCoA1, y = PCoA2, color = factor(trigger_depth
), shape = factor(station))) +
  geom_point(size = 4, alpha = 0.9) +
  stat_ellipse(aes(group = factor(station), color = factor(station)),
               linetype = "dashed", level = 0.68, size = 1, alpha = 0.5, s
how.legend = TRUE) +
  scale_color_manual(values = depth_colors, name = "Depth (m)") +
  scale_shape_manual(values = c(16, 17), name = "Station") +
  theme_minimal(base_size = 15) +
  coord_equal() +
  labs(
    x = xlab, y = ylab,
    color = "Depth (m)",
    shape = "Station"
  )

library(colorspace)
depth_colors <- qualitative_hcl(11, palette = "Dark 3")

pcoa_plot <-ggplot(pcoa_points, aes(x = PCoA1, y = PCoA2, color = factor(t
rigger_depth), shape = factor(station))) +
  geom_point(size = 4, alpha = 0.9) +
  stat_ellipse(aes(group = factor(station), color = factor(station)),
               linetype = "dashed", level = 0.68, size = 1, alpha = 0.5, s
how.legend = TRUE) +
  scale_color_manual(values = depth_colors, name = "Depth (m)") +
  scale_shape_manual(values = c(16, 17), name = "Station") +
  theme_publication(base_size = 15) +
  coord_equal() +
  labs(
    x = xlab, y = ylab,
    color = "Depth (m)",
    shape = "Station"
  )
```

```r
ggsave("PCoA_zooplankton_publication.png", plot = pcoa_plot, width = 10, h
eight = 7, dpi = 400)
ggsave("PCoA_zooplankton_publication.pdf", plot = pcoa_plot, width = 10, h
eight = 7)

adonis2(comm_hell ~ station, data = meta_paired, permutations = 999)

## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = comm_hell ~ station, data = meta_paired, permutations
= 999)
##          Df SumOfSqs      R2     F Pr(>F)
## Model     1  0.16324 0.05995 3.316  0.005 **
## Residual 52  2.55980 0.94005
## Total    53  2.72303 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

adonis2(comm_hell ~ factor(trigger_depth), data = meta_paired, permutation
s = 999)

## Permutation test for adonis under reduced model
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = comm_hell ~ factor(trigger_depth), data = meta_paired
, permutations = 999)
##          Df SumOfSqs      R2      F Pr(>F)
## Model     8  0.51027 0.18739 1.2972  0.092 .
## Residual 45  2.21276 0.81261
## Total    53  2.72303 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

res_station <- adonis2(comm_hell ~ station, data = meta_paired, permutatio
ns = 999)
res_depth   <- adonis2(comm_hell ~ factor(trigger_depth), data = meta_pair
ed, permutations = 999)

permanova_station <- data.frame(
  Effect = "Station",
  Df = res_station$Df[1],
  F = signif(res_station$F[1], 3),
  R2 = signif(res_station$R2[1], 3),
  p = signif(res_station$`Pr(>F)`[1], 3)
)

permanova_depth <- data.frame(
  Effect = "Depth",
  Df = res_depth$Df[1],
  F = signif(res_depth$F[1], 3),
  R2 = signif(res_depth$R2[1], 3),
  p = signif(res_depth$`Pr(>F)`[1], 3)
```

```
)

permanova_summary <- rbind(permanova_station, permanova_depth)

write_xlsx(permanova_summary, "permanova_summary.xlsx")
write.csv(permanova_summary, "permanova_summary.csv", row.names = FALSE)
```

Biomass calculations 1

Lina Becker

2025-09-16

```
library(tidyverse)

library(scales)

setwd("/Volumes/Linas_data/masterarbeit_sortiert/Appendix/daten_rechnungen
")
csv_path <- "cal_biovolume_copepods_m2.csv"
raw <- read.csv2(csv_path, check.names = FALSE)

pick_col <- function(df, pattern) {
  hit <- grep(pattern, names(df), ignore.case = TRUE, value = TRUE)
  if (length(hit) == 0) stop(sprintf("Konnte keine Spalte für Muster '%s'
finden.", pattern))
  hit[1]
}

station_col <- pick_col(raw, "station|stn|sta")
depth_col   <- pick_col(raw, "depth|tiefe|m$")
type_col    <- pick_col(raw, "type|pan|fraction|method")
biom_col    <- pick_col(raw, "copepode?_?biomass_?real|biomass_?real")
count_col   <- pick_col(raw, "real_?count|count_?real")

df <- raw %>%
  transmute(
    station  = .data[[station_col]],
    depth_m  = .data[[depth_col]],
    type_raw = .data[[type_col]],
    biomass  = as.numeric(.data[[biom_col]]),
    n_real   = as.numeric(.data[[count_col]])
  ) %>%
  mutate(
    type_norm = tolower(trimws(as.character(type_raw))),
    type = case_when(
      str_detect(type_norm, "pan\\s*500") | str_detect(type_norm, "^<\\s*?
500") ~ "Pan500 (<500 µm)",
      str_detect(type_norm, "pan\\s*1000") | str_detect(type_norm, "500\\s
*[--]\\s*1000") ~ "Pan1000 (500–1000 µm)",
      TRUE ~ NA_character_
    )
  ) %>%
  filter(!is.na(type)) %>%
```

```r
  select(station, depth_m, type, biomass, n_real)

depth_levels <- df %>%
  distinct(depth_m) %>%
  mutate(d = suppressWarnings(as.numeric(depth_m))) %>%
  arrange(desc(d)) %>%
  pull(depth_m)

agg <- df %>%
  group_by(station, depth_m, type) %>%
  summarise(
    biomass = sum(biomass, na.rm = TRUE),
    n_real  = sum(n_real,  na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    station = factor(station),
    depth_m = factor(depth_m, levels = depth_levels),
    biomass_offset = biomass + 1
  )

station_labels <- c(
  `209` = "209 — Gulf of Panama",
  `210` = "210 — Gulf of Chiriquí"
)

p <- ggplot(agg, aes(x = depth_m, y = biomass_offset, fill = type)) +
  geom_col(position = position_dodge(width = 0.8), width = 0.8,
           colour = "grey20", linewidth = 0.2) +
  geom_text(
    aes(label = scales::comma(round(n_real))),
    position = position_dodge(width = 0.8),
    hjust = -0.2, size = 3
  ) +
  facet_wrap(~ station, ncol = 1, labeller = labeller(station = station_la
bels)) +
  scale_y_continuous(
    trans = "log10",
    labels = scales::label_number(scale = 1e-6, scale_cut = scales::cut_si
("g")),
    expand = expansion(mult = c(0.02, 0.18))
  )+
  scale_fill_manual(
    values = c("Pan500 (<500 µm)" = "#2ca02c",
               "Pan1000 (500–1000 µm)" = "#d62728"),
    labels = c(
      "Pan500 (<500 µm)"    = "Size fraction <500 µm",
      "Pan1000 (500–1000 µm)" = "Size fraction 500–1000 µm"
    ),
    name = NULL
  ) +
  labs(
    x = "Depth (m)",
    y = "C biomass m⁻³ (log10, +1 µg; SI units)"
```

```
  ) +
  theme_minimal(base_size = 12) +
  theme(
    legend.position = "top",
    panel.grid.major.x = element_blank(),
    plot.margin = margin(10, 40, 10, 10),
    panel.spacing = unit(2, "lines")
  )+
  coord_flip()

print(p)

ggsave("pan500_vs_pan1000_biomass_counts.png", p, width = 16, height = 10,
dpi = 400)
```

Biomass calculations 2

Lina Becker

2025-09-16

```
library(tidyverse)

library(scales)

library(patchwork)

setwd("/Volumes/Linas_data/masterarbeit_sortiert/Appendix/daten_rechnungen
")
csv_path <- "cal_biovolume_copepods_m2.csv"
raw <- read.csv2(csv_path, check.names = FALSE)

pick_col <- function(df, pattern) {
  hit <- grep(pattern, names(df), ignore.case = TRUE, value = TRUE)
  if (length(hit) == 0) stop(sprintf("Konnte keine Spalte für Muster '%s'
finden.", pattern))
  hit[1]
}

station_col <- pick_col(raw, "station|stn|sta")
depth_col   <- pick_col(raw, "depth|tiefe|m$")
type_col    <- pick_col(raw, "type|pan|fraction|method")
biom_col    <- pick_col(raw, "copepode?_?biomass_?real|biomass_?real")
count_col   <- pick_col(raw, "real_?count|count_?real")

df <- raw %>%
  transmute(
    station = .data[[station_col]],
    depth_m = .data[[depth_col]],
    type    = .data[[type_col]],
    biomass_µg = as.numeric(.data[[biom_col]]),
    n_real  = as.numeric(.data[[count_col]])
```

```
  )

depth_levels <- df %>%
  distinct(depth_m) %>%
  mutate(d = suppressWarnings(as.numeric(depth_m))) %>%
  arrange(desc(d)) %>% pull(depth_m)

sum501000 <- df %>%
  filter(str_detect(tolower(type), "pan\\s*500") | str_detect(tolower(type
), "pan\\s*1000")) %>%
  group_by(station, depth_m) %>%
  summarise(
    biomass = sum(biomass_µg, na.rm = TRUE),
    n_real  = sum(n_real,  na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(type = "Pan500+Pan1000")

merge_only <- df %>%
  filter(str_detect(tolower(type), "panmerge")) %>%
  group_by(station, depth_m) %>%
  summarise(
    biomass = sum(biomass_µg, na.rm = TRUE),
    n_real  = sum(n_real,  na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(type = "Panmerge")

plot_df <- bind_rows(sum501000, merge_only) %>%
  mutate(
    station = factor(station),
    type    = factor(type, levels = c("Pan500+Pan1000","Panmerge")),
    depth_m = factor(depth_m, levels = depth_levels),
    biomass_offset = biomass + 1
  )

station_labels <- c(
  `209` = "209 — Gulf of Panama",
  `210` = "210 — Gulf of Chiriquí"
)

p_top <- ggplot(plot_df, aes(x = depth_m, y = biomass_offset, fill = type)
) +
  geom_col(position = position_dodge(width = 0.8), width = 0.8, colour = "
grey20", linewidth = 0.2) +
  geom_text(aes(label = scales::comma(round(n_real))),
            position = position_dodge(width = 0.8), hjust = -0.15, size =
3) +
  facet_wrap(~ station, ncol = 1, labeller = labeller(station = station_la
bels)) +
  scale_y_continuous(
    trans = "log10",
    labels = scales::label_number(scale = 1e-6, scale_cut = scales::cut_si
("g")),
```

```
    expand = expansion(mult = c(0.02, 0.18))
  ) +
  scale_fill_manual(
    values = c("Pan500+Pan1000" = "#2ca02c", "Panmerge" = "#d62728"),
    labels = c(
      "Pan500+Pan1000" = "Sum of size fractions (<500 µm) + (500-1000 µm)"
,
      "Panmerge" = "Combined sample"
    ),
    name = NULL
  )+
  labs(
    x = "Water depth (m)",
    y = "C biomass m⁻³ (log10, +1 µg; SI units)"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    legend.position = "top",
    panel.grid.major.x = element_blank(),
    plot.title = element_blank(),
    plot.subtitle = element_blank()
  ) +
  coord_flip(clip="off")

wide <- plot_df %>%
  select(station, depth_m, type, biomass) %>%
  pivot_wider(names_from = type, values_from = biomass) %>%
  mutate(
    pct_diff = 100 * (Panmerge - `Pan500+Pan1000`) / `Pan500+Pan1000`
  ) %>%
  mutate(depth_m = factor(depth_m, levels = depth_levels))

p_bottom <- ggplot(wide, aes(x = depth_m, y = pct_diff, group = station))
+
  geom_hline(yintercept = 0, linewidth = 0.4, colour = "grey40") +
  geom_line(aes(colour = station), linewidth = 0.8) +
  geom_point(aes(colour = station), size = 2) +
  facet_wrap(~ station, ncol = 1, scales = "free_x", labeller = labeller(s
tation = station_labels)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  scale_colour_brewer(palette = "Dark2", guide = "none") +
  labs(
    x = "Water depth (m)",
    y = "% difference"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    panel.grid.major.x = element_blank(),
    plot.title = element_blank(),
    plot.subtitle = element_blank()
  ) +
  coord_flip()
```

```r
combined <- p_top | p_bottom + plot_layout(widths = c(3, 1.6))
combined

ggsave("panmerge_vergleich_mit_diff.png", combined, width = 11, height = 7
, dpi = 300)
```

Chlorophyll plots and calculations

Lina Becker

2025-09-16

```r
required <- c("tidyverse", "janitor", "patchwork")
new_pkgs <- setdiff(required, rownames(installed.packages()))
if (length(new_pkgs)) install.packages(new_pkgs, repos = "https://cloud.r-
project.org")
invisible(lapply(required, library, character.only = TRUE))

paths <- c(
  "/Volumes/Linas_data/masterarbeit_sortiert/Appendix/ctd_daten_original/c
td_chla_profile /250306_ES25C03_209_02_J3061940_4_data-table.csv",
  "/Volumes/Linas_data/masterarbeit_sortiert/Appendix/ctd_daten_original/c
td_chla_profile /250307_ES25C03_210_04_J3071604_2_data-table.csv"
)

out_dir <- "chla_temp_overlap_figs"
dir.create(out_dir, showWarnings = FALSE)

safe_read <- function(path) {
  df <- tryCatch(readr::read_csv(path, show_col_types = FALSE),
                 error = function(e) NULL)
  if (!is.null(df) && ncol(df) > 1) return(df)
  readr::read_delim(path, delim = ";",
                    locale = readr::locale(decimal_mark = ",", grouping_ma
rk = ""),
                    show_col_types = FALSE)
}

read_station <- function(path) {
  raw <- safe_read(path)
  stopifnot(!is.null(raw))
  df  <- janitor::clean_names(raw)
  nm  <- names(df)

  pick <- function(rx_vec) {
    hits <- nm[Reduce(`|`, lapply(rx_vec, function(rx) str_detect(nm, rege
x(rx, ignore_case = TRUE))))]
    if (length(hits)) hits[1] else NA_character_
  }
  chl_col   <- pick(c("\\bchl", "chla", "ug", "µg", "chl_a_ug_l"))
  depth_col <- pick(c("^depth", "depth_m", "depthm", "depth\\.m"))
  temp_col  <- pick(c("^t_is", "t_i_s_deg_c", "\\bdegc\\b", "^temp"))

  need <- c(chl = chl_col, depth = depth_col, temp = temp_col)
```

```
  if (any(is.na(need))) {
    stop("Missing expected columns in ", basename(path),
         "\nHave: ", paste(nm, collapse = ", "),
         "\nNeed something like: Chl_a.ug.L.; depth.m; T_iS.degC")
  }

  file_base <- tools::file_path_sans_ext(basename(path))
  stn_num <- stringr::str_match(file_base, "_(\\d{3})_")[,2]
  if (is.na(stn_num)) stn_num <- stringr::str_extract(file_base, "\\d{3}")
  stn_name <- if (!is.na(stn_num)) paste("Station", stn_num) else file_bas
e

  out <- df %>%
    transmute(
      station = stn_name,
      depth   = suppressWarnings(as.numeric(.data[[depth_col]])),
      chl     = suppressWarnings(as.numeric(.data[[chl_col]])),
      temp    = suppressWarnings(as.numeric(.data[[temp_col]]))
    ) %>%
    filter(!is.na(depth), !is.na(chl), !is.na(temp)) %>%
    arrange(depth) %>%
    filter(dplyr::between(depth, 0, 200))

  if (nrow(out)) {
    imax <- which.max(out$chl)
    out <- out %>% mutate(chlmax_depth = out$depth[imax],
                          chlmax_val   = out$chl[imax])
  }
  out
}

data <- paths %>% lapply(read_station) %>% bind_rows()

stopifnot(nrow(data) > 0)
readr::write_csv(data, file.path(out_dir, "chla_temp_overlap_0-200m_tidy.c
sv"))

runmed_vec <- function(x, k = 5L) if (length(x) < k) x else stats::runmed(
x, k, endrule = "median")

plot_station_overlap <- function(df, smooth = TRUE, k = 5L,
                                 col_temp = "#2B4C7E", col_chl = "#1B9E77"
) {
  st <- unique(df$station); stopifnot(length(st) == 1)

  d <- df %>%
    mutate(temp_sm = if (smooth) runmed_vec(temp, k) else temp,
           chl_sm  = if (smooth) runmed_vec(chl,  k) else chl)

  xmin <- floor(min(d$temp_sm, na.rm = TRUE)); xmax <- ceiling(max(d$temp_
sm, na.rm = TRUE))
  if (!is.finite(xmin) || !is.finite(xmax) || xmin == xmax) {
    xmin <- min(d$temp_sm, na.rm = TRUE) - 1; xmax <- max(d$temp_sm, na.rm
= TRUE) + 1
```

```
  }
  cmin <- 0; cmax <- max(d$chl_sm, na.rm = TRUE)
  if (!is.finite(cmax) || cmax <= 0) cmax <- 1e-3

  scale_chl_to_temp <- function(c) (c - cmin) / (cmax - cmin) * (xmax - xm
in) + xmin
  inv_temp_to_chl   <- function(x) (x - xmin) / (xmax - xmin) * (cmax - cm
in) + cmin

  d <- d %>% mutate(chl_on_temp = scale_chl_to_temp(chl_sm))

  dcm_depth <- unique(d$chlmax_depth)[1]; dcm_val <- unique(d$chlmax_val)[
1]

  base_theme <- theme_minimal(base_size = 12) +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.y = element_line(linewidth = 0.3),
          axis.title.x = element_text(margin = margin(t = 6)),
          axis.title.y = element_text(margin = margin(r = 6)))

  station_titles <- c(
    "Station 209" = "209 — Gulf of Panama",
    "Station 210" = "210 — Gulf of Chiriquí"
  )

  ggplot(d, aes(y = depth)) +
    geom_path(aes(x = temp_sm), linewidth = 1.1, colour = col_temp, na.rm
= TRUE) +
    geom_path(aes(x = chl_on_temp), linewidth = 1.0, linetype = "22",
              colour = col_chl, na.rm = TRUE) +
    geom_ribbon(aes(xmin = pmin(chl_on_temp, xmin), xmax = chl_on_temp),
                fill = col_chl, alpha = 0.08, na.rm = TRUE) +
    geom_hline(yintercept = dcm_depth, linetype = 2, linewidth = 0.5) +
    annotate("text",
             x = xmax, y = dcm_depth + 20,
             label = sprintf("DCM: %.2f \u00B5g L\u207B\u00B9 @ %.1f m", d
cm_val, dcm_depth),
             hjust = 1.05, vjust = 0, size = 3.2, colour = col_chl) +
    scale_y_reverse(limits = c(200, 0), expand = expansion(mult = c(0, 0.0
2))) +
    scale_x_continuous(
      name = "Temperature (°C)",
      limits = c(xmin, xmax),
      sec.axis = sec_axis(~ inv_temp_to_chl(.),
                          name = expression("Chlorophyll a ("*mu*"g L"^-1*
")"))
    ) +
    {
      title_txt <- ifelse(is.na(station_titles[st]), st, station_titles[st
])
      labs(title = title_txt,
           y = "Depth (m")
    } +
    base_theme
```

```
}

by_st <- split(data, data$station)
plots <- lapply(by_st, plot_station_overlap, smooth = TRUE, k = 5,
                col_temp = "#2B4C7E", col_chl = "#1B9E77")

order_names <- tools::file_path_sans_ext(basename(paths))
order_labels <- sapply(paths, function(p) {
  fb <- tools::file_path_sans_ext(basename(p))
  sn <- stringr::str_match(fb, "_(\\d{3})_")[,2]
  if (is.na(sn)) sn <- stringr::str_extract(fb, "\\d{3}")
  if (!is.na(sn)) paste("Station", sn) else fb
})
plots_ordered <- plots[order_labels]

combined_fig <- wrap_plots(plots_ordered, ncol = 2, guides = "collect") &
  theme(legend.position = "none")

png_path <- file.path(out_dir, "stations_temp_chla_overlap_0-200m.png")
ggsave(png_path, combined_fig, width = 12, height = 6, dpi = 300)

pdf_path <- file.path(out_dir, "stations_temp_chla_overlap_0-200m.pdf")
ggsave(pdf_path, combined_fig, width = 12, height = 6, device = cairo_pdf)

message("Saved:\n  - ", png_path, "\n  - ", pdf_path,
        "\nTidy CSV:\n  - ", file.path(out_dir, "chla_temp_overlap_0-200m_
tidy.csv"))
```

CTD Profiles 0 - 200m

Lina Becker

2025-09-16

```
required <- c("tidyverse", "janitor", "oce", "broom")
new_pkgs <- required[!required %in% installed.packages()[, "Package"]]
if (length(new_pkgs)) install.packages(new_pkgs, repos = "https://cloud.r-
project.org")

base_dir <- "/Volumes/Linas_data/masterarbeit_sortiert/Appendix/ctd_daten_
original/ctd_profile"
files <- list.files(path = base_dir, pattern = "_tief\\.csv$", full.names
= TRUE)
if (length(files) == 0) {
  stop("Keine CTD-Dateien gefunden. Pfad/Pattern prüfen: ", base_dir)
}

safe_read <- function(path) {
  df <- tryCatch(
    readr::read_csv(path, show_col_types = FALSE),
    error = function(e) NULL
  )
  if (!is.null(df) && ncol(df) > 1) {
    return(df)
```

```
  }
  message("Fallback-Einlesen mit Semikolon/Dezimal-Komma für ", basename(p
ath))
  df2 <- tryCatch(
    readr::read_delim(
      path,
      delim = ";",
      locale = readr::locale(decimal_mark = ",", grouping_mark = ""),
      show_col_types = FALSE
    ),
    error = function(e) stop("Konnte Datei weder als comma- noch semikolon
-separiert lesen: ", basename(path))
  )
  if (ncol(df2) <= 1) stop("Einlesen ergab nur 1 Spalte für ", basename(pa
th), "; Format ungewöhnlich.")
  df2
}

guess_col <- function(df, patterns, desc) {
  candidates <- names(df)[stringr::str_detect(names(df), stringr::regex(pa
ste(patterns, collapse = "|"), ignore_case = TRUE))]
  if (length(candidates) == 0) {
    stop(sprintf("Konnte keine Spalte für '%s' finden (gesucht nach: %s)."
, desc, paste(patterns, collapse = ", ")))
  }
  if (length(candidates) > 1) {
    message(sprintf("Mehrere mögliche Spalten für '%s' gefunden: %s. Nutze
'%s'.",
                    desc, paste(candidates, collapse = ", "), candidates[1
]))
  }
  candidates[1]
}

read_one <- function(fpath) {
  raw <- safe_read(fpath)
  df <- janitor::clean_names(raw)

  message("Einlese: ", basename(fpath), " — gefundene Spalten: ", paste(na
mes(df), collapse = ", "))

  station <- sub(".*_(\\d+)_tief\\.csv$", "\\1", basename(fpath))
  df <- dplyr::mutate(df, station = station)

  temp_col  <- guess_col(df, c("temp"), "Temperatur")
  depth_col <- guess_col(df, c("depth"), "Tiefe")

  df[[temp_col]]  <- as.numeric(df[[temp_col]])
  df[[depth_col]] <- as.numeric(df[[depth_col]])

  sal_candidates <- names(df)[stringr::str_detect(names(df), stringr::rege
x("(^sal_psu$|^salinity$)", ignore_case = TRUE))]
  if (length(sal_candidates) >= 1) {
    sal_col <- sal_candidates[1]
```

```
    df <- dplyr::rename(df, salinity = !!rlang::sym(sal_col))
    message("Verwende vorhandene Salinität aus Spalte '", sal_col, "' für
Station ", station, ".")
  } else if (any(stringr::str_detect(names(df), stringr::regex("cond", ign
ore_case = TRUE))) &&
             any(stringr::str_detect(names(df), stringr::regex("pres", ign
ore_case = TRUE)))) {
    cond_col <- names(df)[stringr::str_detect(names(df), stringr::regex("c
ond", ignore_case = TRUE))][1]
    pres_col <- names(df)[stringr::str_detect(names(df), stringr::regex("p
res", ignore_case = TRUE))][1]
    cond_vec <- as.numeric(df[[cond_col]]) * 0.1  # mS/cm → S/m
    temp_vec <- as.numeric(df[[temp_col]])
    pres_vec <- as.numeric(df[[pres_col]])
    ctd_obj <- oce::as.ctd(conductivity = cond_vec, temperature = temp_vec
, pressure = pres_vec)
    df <- dplyr::mutate(df, salinity = oce::salinity(ctd_obj))
    message("Salinität für Station ", station, " aus ", cond_col, " & ", p
res_col, " berechnet.")
  } else {
    df <- dplyr::mutate(df, salinity = NA_real_)
    warning("Für Datei ", basename(fpath), " konnte keine Salinität ermitt
elt werden.")
  }

    df <- dplyr::rename(
    df,
    temperature = !!rlang::sym(temp_col),
    depth       = !!rlang::sym(depth_col)
  )

  if (!is.numeric(df$salinity)) {
    has_comma <- any(stringr::str_detect(df$salinity, ","))
    df$salinity <- readr::parse_double(df$salinity,
                                       locale = readr::locale(decimal_mark
= if (has_comma) "," else "."))
  }

  out <- dplyr::select(df, station, depth, temperature, salinity)

  if (!all(c("station", "depth", "temperature", "salinity") %in% names(out
))) {
    stop("Output von ", basename(fpath), " hat nicht erwartete Spalten: ",
paste(names(out), collapse = ", "))
  }

  out
}

combined <- lapply(files, read_one) %>% dplyr::bind_rows()

message("Zusammengeführte Tabelle: Spalten = ", paste(names(combined), col
lapse = ", "))
```

```
## Zusammengeführte Tabelle: Spalten = station, depth, temperature, salini
ty

print(utils::head(combined))

## # A tibble: 6 × 4
##   station depth temperature salinity
##   <chr>   <dbl>       <dbl>    <dbl>
## 1 209      5.01        27.1     30.9
## 2 209      5.12        27.1     30.9
## 3 209      5.24        27.1     30.9
## 4 209      5.37        27.1     30.9
## 5 209      5.47        27.1     30.9
## 6 209      5.63        27.2     30.9

combined <- dplyr::filter(combined, !is.na(depth))

MAX_DEPTH <- 200
combined_zoom <- combined %>% dplyr::filter(depth <= MAX_DEPTH)

station_labels <- c(
  `209` = "209 — Gulf of Panama",
  `210` = "210 — Gulf of Chiriquí"
)

plotdata <- combined_zoom %>%
  tidyr::pivot_longer(
    cols = c(temperature, salinity),
    names_to = "variable",
    values_to = "value"
  ) %>%
  dplyr::mutate(
    variable = dplyr::recode(variable,
                             temperature = "Temperature (°C)",
                             salinity    = "Salinity (PSU)")
  )

base_theme <- ggplot2::theme_minimal(base_size = 13) +
  ggplot2::theme(
    panel.grid.minor = ggplot2::element_blank(),
    strip.placement  = "outside",
    panel.spacing    = grid::unit(0.9, "lines")
  )

make_TS_panel <- function(dat, stn, max_depth = MAX_DEPTH) {
  d <- dat %>% dplyr::filter(station == stn, depth <= max_depth)

  t_min <- min(d$temperature, na.rm = TRUE); t_max <- max(d$temperature, n
a.rm = TRUE)
  s_min <- min(d$salinity,    na.rm = TRUE); s_max <- max(d$salinity,    n
a.rm = TRUE)
  b <- (t_max - t_min) / (s_max - s_min)
  a <- t_min - b * s_min

  ggplot2::ggplot(d) +
```

```
    ggplot2::geom_path(ggplot2::aes(x = temperature, y = depth), colour =
"firebrick", linewidth = 0.9, na.rm = TRUE) +
    ggplot2::geom_path(ggplot2::aes(x = a + b * salinity, y = depth), colo
ur = "steelblue", linewidth = 0.9, na.rm = TRUE) +
    ggplot2::scale_y_reverse(limits = c(max_depth, 0), expand = ggplot2::e
xpansion(mult = c(0, 0.02))) +
    ggplot2::scale_x_continuous(
      name = "Temperature (°C)",
      sec.axis = ggplot2::sec_axis(~ (.-a)/b, name = "Salinity (PSU)")
    ) +
    ggplot2::labs(title = station_labels[[as.character(stn)]]) +
    base_theme +
    ggplot2::theme(
      plot.title = ggplot2::element_text(hjust = 0.5),
      legend.position = "none"
    )
}

stations <- sort(unique(combined_zoom$station))
p_combined <- purrr::map(stations, ~ make_TS_panel(combined_zoom, .x, MAX_
DEPTH)) %>%
  patchwork::wrap_plots(nrow = 1)

p_temp <- ggplot2::ggplot(dplyr::filter(plotdata, variable == "Temperature
(°C)"),
                          ggplot2::aes(x = value, y = depth, color = stati
on)) +
  ggplot2::geom_path(na.rm = TRUE) +
  ggplot2::scale_y_reverse(limits = c(MAX_DEPTH, 0)) +
  ggplot2::labs( title = NULL, x = "Temperature (°C)", y = "Depth (m)", co
lor = "Station") +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

p_sal <- ggplot2::ggplot(dplyr::filter(plotdata, variable == "Salinity (PS
U)"),
                         ggplot2::aes(x = value, y = depth, color = statio
n)) +
  ggplot2::geom_path(na.rm = TRUE) +
  ggplot2::scale_y_reverse(limits = c(MAX_DEPTH, 0)) +
  ggplot2::labs( title = NULL, x = "Salinity (PSU)", y = "Depth (m)", colo
r = "Station") +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

ggplot2::ggsave("ctd_profiles_combined_0-200m.png", p_combined, width = 11
, height = 6, dpi = 300)
ggplot2::ggsave("ctd_temp_profiles_0-200m.png",     p_temp,     width =  6
, height = 5, dpi = 300)
ggplot2::ggsave("ctd_sal_profiles_0-200m.png",      p_sal,      width =  6
, height = 5, dpi = 300)

print(p_combined)
```

```r
print(p_temp)

print(p_sal)

interp_to_grid <- function(df, dz = 1) {
  df <- df %>%
    dplyr::mutate(
      depth       = as.numeric(.data$depth),
      temperature = as.numeric(.data$temperature),
      salinity    = as.numeric(.data$salinity)
    ) %>%
    dplyr::filter(!is.na(.data$depth),
                  !is.na(.data$temperature),
                  !is.na(.data$salinity)) %>%
    dplyr::arrange(.data$depth)

  if (nrow(df) < 5) return(NULL)

  df <- df %>%
    dplyr::group_by(.data$depth) %>%
    dplyr::summarise(
      temperature = mean(.data$temperature, na.rm = TRUE),
      salinity    = mean(.data$salinity,    na.rm = TRUE),
      .groups = "drop"
    )

  zmin <- ceiling(max(0, min(df$depth, na.rm = TRUE)))
  zmax <- floor(min(MAX_DEPTH, max(df$depth, na.rm = TRUE)))
  if (!is.finite(zmin) || !is.finite(zmax) || (zmax - zmin) < 5) return(NU
LL)

  z <- seq(zmin, zmax, by = dz)

  tibble::tibble(
    depth       = z,
    temperature = stats::approx(df$depth, df$temperature, xout = z, rule =
2, ties = "mean")$y,
    salinity    = stats::approx(df$depth, df$salinity,    xout = z, rule =
2, ties = "mean")$y
  )
}

compute_strat_metrics <- function(df_station, lat = 45) {
  g <- interp_to_grid(df_station, dz = 1)
  if (is.null(g)) return(NULL)
  p <- g$depth
  sigma_theta <- oce::swSigmaTheta(
    salinity    = g$salinity,
    temperature = g$temperature,
    pressure    = p
  )
  sigma_theta <- as.numeric(sigma_theta)

  rho    <- 1000 + stats::runmed(sigma_theta, k = 5, endrule = "median")
```

```
  drhodz <- c(NA, diff(rho)) / c(NA, diff(p))
  N2_grid <- -(9.81 / 1025) * drhodz

  runmed <- function(x, k = 5) stats::runmed(x, k = k, endrule = "median")
  T_smooth   <- runmed(g$temperature, k = 5)
  S_smooth   <- runmed(g$salinity,    k = 5)
  sig_smooth <- runmed(sigma_theta,   k = 5)

  dTdz   <- c(NA, diff(T_smooth))   / c(NA, diff(g$depth))
  dSdz   <- c(NA, diff(S_smooth))   / c(NA, diff(g$depth))
  dRdz   <- c(NA, diff(sig_smooth)) / c(NA, diff(g$depth))

  ref_z  <- 10
  ref_ix <- which.min(abs(g$depth - ref_z))
  mld <- if (!length(ref_ix) || is.na(sigma_theta[ref_ix])) NA_real_ else
{
    thr <- sigma_theta[ref_ix] + 0.03
    ix  <- which(sigma_theta >= thr)[1]
    if (is.na(ix)) max(g$depth, na.rm = TRUE) else g$depth[ix]
  }

  thermocline_z <- g$depth[which.max(abs(dTdz))]
  halocline_z   <- g$depth[which.max(abs(dSdz))]
  pycnocline_idx <- suppressWarnings(which.max(N2_grid))
  pycnocline_z   <- if (length(pycnocline_idx) && is.finite(pycnocline_idx
)) g$depth[pycnocline_idx] else NA_real_

  surf <- g %>% dplyr::filter(depth <= min(5, max(depth))) %>% dplyr::summ
arise(
    t = mean(temperature, na.rm = TRUE),
    s = mean(salinity,    na.rm = TRUE),
    sig = mean(sigma_theta[depth <= min(5, max(g$depth))], na.rm = TRUE)
  )
  bott <- g %>% dplyr::filter(depth >= max(depth) - 5) %>% dplyr::summaris
e(
    t = mean(temperature, na.rm = TRUE),
    s = mean(salinity,    na.rm = TRUE),
    sig = mean(sigma_theta[depth >= max(g$depth) - 5], na.rm = TRUE)
  )

  strat_index <- as.numeric(bott$sig - surf$sig)

  tibble::tibble(
    depth = g$depth,
    temperature = g$temperature,
    salinity    = g$salinity,
    sigma_theta = sigma_theta,
    N2          = N2_grid,
    dTdz        = dTdz,
    dSdz        = dSdz,
    dSigdz      = dRdz
  ) %>%
    dplyr::mutate(
      station = unique(df_station$station)[1],
```

```
      mld = mld,
      thermocline_z = thermocline_z,
      halocline_z   = halocline_z,
      pycnocline_z  = pycnocline_z,
      strat_index   = strat_index,
      surf_t = surf$t, surf_s = surf$s, surf_sig = surf$sig,
      bott_t = bott$t, bott_s = bott$s, bott_sig = bott$sig
    )
}

strat_profiles <- combined_zoom %>%
  dplyr::group_by(station) %>%
  dplyr::group_modify(~{
    out <- compute_strat_metrics(.x)
    if (is.null(out)) tibble::tibble() else out
  }) %>%
  dplyr::ungroup()

p_sigma <- ggplot2::ggplot(strat_profiles, ggplot2::aes(x = sigma_theta, y
= depth, color = station)) +
  ggplot2::geom_path(na.rm = TRUE) +
  ggplot2::geom_hline(ggplot2::aes(yintercept = mld, color = station), lin
etype = 2, alpha = 0.6) +
  ggplot2::scale_y_reverse(limits = c(MAX_DEPTH, 0)) +
  ggplot2::labs(titel= NULL,
                x = expression(sigma[theta]~"(kg·m"^-3*")"),
                y = "Depth (m)", color = "Station") +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

p_N2 <- ggplot2::ggplot(strat_profiles, ggplot2::aes(x = N2, y = depth, co
lor = station)) +
  ggplot2::geom_path(na.rm = TRUE) +
  ggplot2::geom_point(ggplot2::aes(y = pycnocline_z), shape = 21, fill = N
A, stroke = 1, na.rm = TRUE) +
  ggplot2::scale_x_log10() +
  ggplot2::scale_y_reverse(limits = c(MAX_DEPTH, 0)) +
  ggplot2::labs(
    title = NULL,
    x = expression(N^2~"(s"^-2*")"),
    y = "Depth (m)",
    color = "Station"
  )  +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

deriv_long <- strat_profiles %>%
  dplyr::select(station, depth, dTdz, dSdz, dSigdz) %>%
  tidyr::pivot_longer(cols = c(dTdz, dSdz, dSigdz), names_to = "var", valu
es_to = "val") %>%
  dplyr::mutate(var = dplyr::recode(var, dTdz = "dT/dz (°C m⁻¹)", dSdz = "
dS/dz (PSU m⁻¹)", dSigdz = "dσθ/dz (kg m⁻4)"))

p_deriv <- ggplot2::ggplot(deriv_long, ggplot2::aes(x = val, y = depth, co
```

```
lor = station)) +
  ggplot2::geom_path(na.rm = TRUE) +
  ggplot2::scale_y_reverse(limits = c(MAX_DEPTH, 0)) +
  ggplot2::facet_wrap(~ var, scales = "free_x", ncol = 3) +
  ggplot2::labs(
    title = NULL,
    x = NULL,
    y = "Depth (m)",
    color = "Station"
  ) +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

Sg <- seq(floor(min(combined_zoom$salinity,    na.rm = TRUE)),
          ceiling(max(combined_zoom$salinity, na.rm = TRUE)), by = 0.1)
Tg <- seq(floor(min(combined_zoom$temperature,    na.rm = TRUE)),
          ceiling(max(combined_zoom$temperature, na.rm = TRUE)), by = 0.2)
grid <- expand.grid(S = Sg, T = Tg)
grid$sigma_theta <- oce::swSigmaTheta(salinity = grid$S,
                                      temperature = grid$T,
                                      pressure = 0)

p_TS <- ggplot2::ggplot() +
  ggplot2::stat_contour(data = grid, ggplot2::aes(x = S, y = T, z = sigma_
theta),
                        bins = 12, linewidth = 0.3, alpha = 0.5) +
  ggplot2::geom_path(data = combined_zoom, ggplot2::aes(x = salinity, y =
temperature, color = station,
                                                        group = interactio
n(station)), alpha = 0.9) +
  ggplot2::scale_y_reverse() +
  ggplot2::labs(
    x = "Salinity (PSU)",
    y = "Temperature (°C)",
    color = "Station"
  ) +
  ggplot2::scale_color_discrete(labels = station_labels) +
  base_theme

ggplot2::ggsave("ctd_density_profiles_0-200m.png", p_sigma, width = 6, hei
ght = 5, dpi = 300)
ggplot2::ggsave("ctd_N2_profiles_0-200m.png",      p_N2,    width = 6, hei
ght = 5, dpi = 300)

ggplot2::ggsave("ctd_derivatives_0-200m.png",      p_deriv, width = 9, hei
ght = 4, dpi = 300)
ggplot2::ggsave("ctd_TS_diagram_0-200m.png",       p_TS,    width = 6, hei
ght = 5, dpi = 300)

print(p_sigma)

print(p_N2)

print(p_deriv)
```

```r
print(p_TS)

station_summary <- strat_profiles %>%
  dplyr::group_by(station) %>%
  dplyr::summarise(
    mld_m              = unique(na.omit(mld))[1],
    thermocline_m      = unique(na.omit(thermocline_z))[1],
    halocline_m        = unique(na.omit(halocline_z))[1],
    pycnocline_m       = unique(na.omit(pycnocline_z))[1],
    strat_index_dsigma = unique(na.omit(strat_index))[1],
    surface_temp_C     = unique(na.omit(surf_t))[1],
    surface_sal_PSU    = unique(na.omit(surf_s))[1],
    surface_sigma      = unique(na.omit(surf_sig))[1],
    bottom_temp_C      = unique(na.omit(bott_t))[1],
    bottom_sal_PSU     = unique(na.omit(bott_s))[1],
    bottom_sigma       = unique(na.omit(bott_sig))[1],
    .groups = "drop"
  )

readr::write_csv(station_summary, "ctd_station_stratification_summary_0-20
0m.csv")
readr::write_csv(strat_profiles,   "ctd_stratified_profiles_grid1m_0-200m.
csv")

print(station_summary)

## # A tibble: 2 × 12
##   station mld_m thermocline_m halocline_m pycnocline_m strat_index_dsig
ma
##   <chr>   <dbl>         <dbl>       <dbl>        <dbl>              <db
l>
## 1 209        22            28          28           66
NA
## 2 210        13            44          31          140
NA
## # ℹ 6 more variables: surface_temp_C <dbl>, surface_sal_PSU <dbl>,
## #   surface_sigma <dbl>, bottom_temp_C <dbl>, bottom_sal_PSU <dbl>,
## #   bottom_sigma <dbl>

upper_layer <- dplyr::filter(combined, depth <= 20)
deep_layer  <- dplyr::filter(combined, depth > 100)

summary_stats <- function(df, layer_name) {
  df %>%
    dplyr::group_by(station) %>%
    dplyr::summarize(
      temp_mean = mean(temperature, na.rm = TRUE),
      temp_sd   = stats::sd(temperature, na.rm = TRUE),
      sal_mean  = mean(salinity,    na.rm = TRUE),
      sal_sd    = stats::sd(salinity,    na.rm = TRUE),
      n         = dplyr::n(),
      .groups   = 'drop'
    ) %>%
    dplyr::mutate(layer = layer_name)
```

```r
}

stats_upper <- summary_stats(upper_layer, "upper_20m")
stats_deep  <- summary_stats(deep_layer,  "below_100m")

stats_all <- dplyr::bind_rows(stats_upper, stats_deep)
print(stats_all)
```

```
## # A tibble: 4 × 7
##   station temp_mean temp_sd sal_mean  sal_sd     n layer
##   <chr>       <dbl>   <dbl>    <dbl>   <dbl> <int> <chr>
## 1 209          27.1 0.00921     30.9 0.00737   209 upper_20m
## 2 210          29.2 0.450       31.3 0.350     416 upper_20m
## 3 209          11.3 1.92        34.9 0.103    5402 below_100m
## 4 210          11.4 1.97        34.8 0.105    5912 below_100m
```

```r
minmax_stats <- combined %>%
  dplyr::group_by(station) %>%
  dplyr::summarize(
    temp_min = min(temperature, na.rm = TRUE),
    temp_max = max(temperature, na.rm = TRUE),
    sal_min  = min(salinity,    na.rm = TRUE),
    sal_max  = max(salinity,    na.rm = TRUE),
    .groups = 'drop'
  )
print(minmax_stats)
```

```
## # A tibble: 2 × 5
##   station temp_min temp_max sal_min sal_max
##   <chr>      <dbl>    <dbl>   <dbl>   <dbl>
## 1 209         8.02     27.2    30.8    35.2
## 2 210         8.24     30.0    31.1    35.0
```

```r
safe_cor_test <- function(x, y) {
  cc <- stats::complete.cases(x, y)
  x <- x[cc]; y <- y[cc]
  if (length(x) < 3L) return(list(estimate = NA_real_, p.value = NA_real_)
)
  ct <- tryCatch(stats::cor.test(x, y, method = "spearman", exact = FALSE)
,
                 error = function(e) NULL)
  if (is.null(ct)) return(list(estimate = NA_real_, p.value = NA_real_))
  list(estimate = unname(ct$estimate), p.value = ct$p.value)
}

cor_results <- combined %>%
  dplyr::group_by(station) %>%
  dplyr::reframe({
    res <- safe_cor_test(temperature, salinity)
    tibble::tibble(spearman = res$estimate, p_value = res$p.value)
  })
print(cor_results)
```

```
## # A tibble: 2 × 3
##   station spearman   p_value
```

```
##   <chr>      <dbl>     <dbl>
## 1 209        0.614 0
## 2 210        0.407 8.86e-295
```

```
cor_stats <- combined %>%
  dplyr::group_by(station) %>%
  dplyr::group_modify(~{
    cc <- stats::complete.cases(.x$temperature, .x$salinity)
    dat <- .x[cc, , drop = FALSE]
    if (nrow(dat) < 3) return(tibble::tibble())
    broom::tidy(stats::cor.test(dat$temperature, dat$salinity, method = "s
pearman", exact = FALSE))
  })
print(cor_stats)
```

```
## # A tibble: 2 × 6
## # Groups:   station [2]
##   station estimate    statistic   p.value method                    al
ternative
##   <chr>      <dbl>        <dbl>     <dbl> <chr>                     <c
hr>
## 1 209        0.614 20926737572. 0         Spearman's rank correlati… tw
o.sided
## 2 210        0.407 40924671104. 8.86e-295 Spearman's rank correlati… tw
o.sided
```

Envfit CTD and community diversity

Lina Becker

2025-09-16

```
library(tidyverse)

library(vegan)

library(colorspace)

library(ggrepel)

library(ggnewscale)

setwd("/Volumes/Linas_data/masterarbeit_sortiert/Appendix/daten_rechnungen"
)

env_data <- read.csv2("250306_ES25C03_ctd.csv")

env_vars <- c("temp.degC.", "sal.psu.", "pH", "cond.mS.cm.")

env_data <- env_data %>%

  mutate(across(all_of(env_vars), ~as.numeric(str_replace(., ",", "."))))

env_data <- env_data %>%

  mutate(trigger_depth = as.numeric(trigger_depth),
```

```
         station = as.character(Station))


meta_env <- meta_paired %>%
  mutate(trigger_depth = as.numeric(trigger_depth),
         station = as.character(station)) %>%
  left_join(env_data, by = c("station", "trigger_depth"))


envfit_result <- envfit(pcoa$points, meta_env[, env_vars], permutations =
999)


signif_vecs <- as.data.frame(scores(envfit_result, "vectors"))
signif_vecs$var <- rownames(signif_vecs)
names(signif_vecs)[1:2] <- c("PC1", "PC2")
sig_vars <- names(which(envfit_result$vectors$pvals < 0.05))
signif_vecs <- signif_vecs %>% filter(var %in% sig_vars)
signif_vecs$label <- ifelse(envfit_result$vectors$pvals[signif_vecs$var] <
0.05,
                            paste0(signif_vecs$var, " *"), signif_vecs$var)


if (!"Total_Abundance" %in% names(pcoa_points)) {
  abundance_sum <- df_abs %>%
    group_by(station, trigger_depth) %>%
    summarise(Total_Abundance = sum(Total, na.rm = TRUE), .groups = "drop")
  pcoa_points$station <- as.character(pcoa_points$station)
  abundance_sum$station <- as.character(abundance_sum$station)
  pcoa_points$trigger_depth <- as.character(pcoa_points$trigger_depth)
  abundance_sum$trigger_depth <- as.character(abundance_sum$trigger_depth)
  pcoa_points <- left_join(pcoa_points, abundance_sum, by = c("station",
"trigger_depth"))
}


pcoa_points$trigger_depth <- factor(
```

```
  pcoa_points$trigger_depth,

  levels = sort(as.numeric(unique(pcoa_points$trigger_depth)))

)

depth_colors  <-  qualitative_hcl(length(levels(pcoa_points$trigger_depth)),
palette = "Dark 3")

names(depth_colors) <- levels(pcoa_points$trigger_depth)



station_colors <- c("209" = "#E76F51", "210" = "#264653")



var_expl <- round(100 * pcoa$eig[1:2] / sum(pcoa$eig[pcoa$eig > 0]), 1)

xlab <- paste0("PCoA1 (", var_expl[1], "%)")

ylab <- paste0("PCoA2 (", var_expl[2], "%)")



pcoa_plot <- ggplot(pcoa_points, aes(

  x = PCoA1, y = PCoA2

)) +

  geom_point(aes(color   =   trigger_depth,   shape   =   station,   size   =
Total_Abundance), alpha = 0.85) +

  scale_color_manual(values = depth_colors, name = "Depth (m)") +

  scale_shape_manual(values = c(16, 17), name = "Station", labels = c("209",
"210")) +

  scale_size_continuous(range = c(2.5, 8), name = "Plankton abundance") +



  ggnewscale::new_scale_color() +



  stat_ellipse(aes(group = station, color = station),

               linetype = "dashed", level = 0.68, size = 1, alpha = 0.55,
show.legend = TRUE) +

  scale_color_manual(values = station_colors, name = "Station (ellipse)") +



  geom_segment(

    data = signif_vecs,
```

```
    aes(x = 0, y = 0, xend = PC1 * 0.8, yend = PC2 * 0.8),

    arrow = arrow(length = unit(0.32, "cm")), color = "black", size = 1.1,

    inherit.aes = FALSE

  ) +

  geom_text_repel(

    data = signif_vecs,

    aes(x = PC1 * 0.85, y = PC2 * 0.85, label = label),

    size = 5.2, fontface = "bold", inherit.aes = FALSE, segment.color =
"grey40"

  ) +

  labs(

    title = "PCoA of zooplankton communities with environmental vectors",

    subtitle = "PCoA points: stations & depths, point size = plankton
abundance\nArrows: significant environmental gradients (*p < 0.05)",

    x = xlab, y = ylab

  ) +

  guides(

    size = guide_legend(order = 3),

    color = guide_legend(order = 1),

    shape = guide_legend(order = 2)

  ) +

  theme_minimal()


print(pcoa_plot)

ggsave("PCoA_envfit_publication.png", plot = pcoa_plot, width = 18, height =
14, dpi = 400)


library(ggnewscale)


pcoa_plot <- ggplot(pcoa_points, aes(

  x = PCoA1, y = PCoA2

)) +
```

```
  geom_point(aes(color = trigger_depth, shape = station, size =
Total_Abundance), alpha = 0.89) +

  scale_color_manual(

    values = depth_colors,

    name = "Depth (m)",

    guide = guide_legend(

      override.aes = list(linetype = 0, shape = 16, size = 5),

      order = 2

    )

  ) +

  scale_shape_manual(

    values = c(16, 17),

    name = "Station",

    labels = c("209", "210"),

  ) +

  scale_size_continuous(

    range = c(3, 8),

    name = "Plankton abundance",

    breaks = c(3000, 6000, 9000),

    labels = c("3,000", "6,000", "9,000"),

    guide = guide_legend(order = 4)

  ) +

  ggnewscale::new_scale_color() +

  stat_ellipse(aes(group = station, color = station), linetype = "dashed",
size = 1.6, alpha = 0.7, level = 0.68, show.legend = TRUE) +

  scale_color_manual(

    values = station_colors,

    name = "Station (ellipse)",

    guide = guide_legend(

      override.aes = list(linetype = "dashed", shape = NA, size = 2.5),

      order = 1
```

```
    )
  ) +
  geom_segment(
    data = signif_vecs,
    aes(x = 0, y = 0, xend = PC1 * 0.8, yend = PC2 * 0.8),
    arrow = arrow(length = unit(0.32, "cm")), color = "black", size = 1.2,
    inherit.aes = FALSE
  ) +
  geom_text_repel(
    data = signif_vecs,
    aes(x = PC1 * 0.85, y = PC2 * 0.85, label = label),
    size = 5.2, fontface = "bold", inherit.aes = FALSE, segment.color =
"grey40"
  ) +
  labs(
    title = "PCoA of zooplankton communities with environmental vectors",
    subtitle = "Points: depth, shape = station, size = abundance\nArrows:
significant environmental gradients (*p < 0.05)",
    x = xlab, y = ylab
  ) +
  guides(
    size = guide_legend(order = 4),
    color = guide_legend(order = 2),
    shape = guide_legend(
      override.aes = list(
        linetype = 0,
        color = "grey20",
        shape = c(16, 17),
        size = 7
      ),
      order = 3
```

```
    )
  ) +
  theme_minimal(base_size = 17) +
  theme(
    legend.title = element_text(face = "bold", size = 16),
    legend.text = element_text(size = 14),
    legend.key.size = unit(1.3, "cm"),
    legend.box.margin = margin(10, 14, 10, 10),
    legend.spacing.y = unit(1.2, "cm"),
    legend.background = element_rect(color = "grey85", fill = "white"),
    legend.position = "right",
    legend.box = "vertical",
    plot.title.position = "plot",
    plot.title = element_text(margin = margin(b = 10)),
    plot.subtitle = element_text(margin = margin(b = 13))
  )



print(pcoa_plot)
ggsave("PCoA_envfit_publication_nicelegend.png",
       plot = pcoa_plot, width = 18, height = 14, dpi = 400, bg = "white",
limitsize = FALSE)
library(dplyr)


ef_r2    <- envfit_result$vectors$r
ef_pval  <- envfit_result$vectors$pvals
ef_scores <- as.data.frame(scores(envfit_result, display = "vectors"))


ef_scores$Variable <- rownames(ef_scores)
names(ef_scores)[1:2] <- c("PC1", "PC2")
```

```
ef_tab <- ef_scores %>%
  mutate(
    r2      = ef_r2[Variable],
    p_value = ef_pval[Variable],
    p_adj   = p.adjust(p_value, method = "fdr"),
    signif  = case_when(
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    ),
    angle_deg = atan2(PC2, PC1) * 180 / pi
  ) %>%
  select(
    Variable, r2, p_value, p_adj, signif, PC1, PC2, angle_deg
  ) %>%
  mutate(
    across(c(r2, p_value, p_adj, PC1, PC2, angle_deg),
           ~formatC(.x, digits = 6, format = "f", decimal.mark = ","))
  )


print(ef_tab)


write.csv2(ef_tab, "envfit_pcoa_summary.csv", row.names = FALSE)
```

Correlations between zooplankton indexes

Lina Becker

2025-09-16

```
library(dplyr)
```

```
library(tidyr)
library(broom)

env_vars <- c("temp.degC.", "sal.psu.", "pH", "cond.mS.cm.")
zoop_vars <- c("Total_Abundance")

abundance_sum <- abundance_sum %>%
  mutate(
    station = as.character(station),
    trigger_depth = as.character(trigger_depth)
  )

env_data2 <- env_data %>%
  mutate(
    station = as.character(Station),
    trigger_depth = as.character(trigger_depth)
  )

meta_paired2 <- meta_paired %>%
  mutate(
    station = as.character(station),
    trigger_depth = as.character(trigger_depth)
  )

data_for_corr <- meta_paired2 %>%
  left_join(abundance_sum, by = c("station", "trigger_depth")) %>%
  left_join(env_data2 %>% select(station, trigger_depth, all_of(env_vars)),
            by = c("station", "trigger_depth"))

corr_list <- list()
```

```
for (zi in zoop_vars) {
  for (ev in env_vars) {
    tmp <- data_for_corr %>%
      filter(!is.na(.data[[zi]]), !is.na(.data[[ev]]))
    if (nrow(tmp) < 3) next
    test <- cor.test(tmp[[zi]], tmp[[ev]], method = "spearman", exact =
FALSE)
    rho <- as.numeric(test$estimate)
    pval <- test$p.value
    corr_list[[length(corr_list) + 1]] <- tibble(
      Index = zi,
      Env = ev,
      rho = rho,
      p_value = pval
    )
  }
}
corr_df <- bind_rows(corr_list)


corr_df <- corr_df %>%
  mutate(
    signif = case_when(
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    ),
    label = paste0(sprintf("%.2f", rho), signif)
  )


cor_table <- corr_df %>%
  select(Index, Env, label) %>%
```

```
  pivot_wider(names_from = Env, values_from = label)


cor_full <- corr_df %>%
  arrange(Index, Env)


print(cor_table)
print(cor_full)


write.csv(cor_full, "zooplankton_env_correlations_full.csv", row.names =
FALSE)
write.csv(cor_table, "zooplankton_env_correlations_wide.csv", row.names =
FALSE)
```