THREE-DIMENSIONAL MAPPING OF OCEAN VARIABLES AND MARINE ECOSYSTEMS CHARACTERIZATION ON THE PORTUGUESE COAST

Telmo Geraldes Dias

Dissertation submitted in partial fulfilment of the requirements for the Degree of Mestre em Ciência e Sistemas de Informação Geográfica (Master in Geographical Information Systems and Science)

NOVA Information Management School

Three-dimensional mapping of ocean variables
and marine ecosystems characterization
on the Portuguese coast

Dissertation supervised by

PhD Ana Cristina Marinho da Costa

Dissertation co-supervised by

MSc Paulo Jorge Antunes Nunes

September 2022

DECLARATION OF ORIGINALITY

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or not published) are referenced.

This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere.

Lisbon, September 26th, 2022

Assinado por : **Telmo Geraldes Dias**
Num. de Identificação: 12320825
Data: 2022.09.26 17:17:25+01'00'

Telmo Geraldes Dias

ACKNOWLEDGEMENTS

This endeavor would not have been possible without the continuous and prolific support of my supervisor Prof. Ana Cristina Costa. Special thanks to my co-supervisor Cdr. Paulo Antunes Nunes in guiding and assisting the research.

This dissertation marks the end of the master in Geographic Information Systems and Science. I could not have undertaken this journey without the enlightening assistance of the course coordinator Prof. Marco Painho. Special thanks to the teaching staff: Prof. André Oliveira, Prof. Emmanuel Stefanakis, Prof. Fernando Bação, Prof. Fernando Gil, Prof. Filipe Campos, Prof. Hugo Martins, Prof. João David, Prof. Mário Caetano, and Prof. Wolfgang Reinhardt.

I am also grateful for the technical support of the Divison of Marine Chemistry and Pollution of the Hydrographic Institute, namely, PhD Carla Palma and PhD Carlos Borges.

Thanks should also go to Capt. Luís Bessa Pacheco and Cdr. João Delgado Vicente, from the Hydrographic Institute, for the opportunity to attend this master's course.

I would also like to acknowledge the support of Esri staff: Eric Krause, Ivonne Seler, Gonzalo Espinoza, Nerrissa Pinto, and Diamantino Correia.

Lastly, as there is no knowledge without information, and no information without data, I would like to mention that the data used in this research was collected for the project AQUIMAR – Marine Knowledge Supporting Aquaculture (MAR-02.01.01-FEAMP-0107), funded by the MAR 2020 Operational Program.

Three-dimensional mapping of ocean variables and marine ecosystems characterization on the Portuguese coast

ABSTRACT

The oceans are a persistent topic on the international agenda, either because of the vital importance of marine ecosystems for human life or because their preservation and sustainability are in jeopardy. The United Nations Decade of Ocean Science for Sustainable Development (2021-2030) embodies this mission to improve and extend ocean knowledge towards effective policies and actions. One of its major challenges is ocean digitalization.

The aim of this work is to map, in three dimensions, the physical and chemical properties of seawater and to characterize ecological marine units on the Portuguese coast. We used data from project AQUIMAR, collected in-situ along four campaigns, on two areas: B, on the northwestern coast, from Aveiro to Figueira da Foz; and E, on the south coast, from Sagres to Tavira. We selected eight ocean variables: temperature, salinity, dissolved oxygen, nutrients (phosphate, total oxidized nitrogen, and silicate), chlorophyll a, and pH. Digital representations of their distribution in the water column were created by spatial interpolation using the Empirical Bayesian Kriging 3D method, which had reliable performance as an automated interpolator. Marine ecosystems were characterized through K-means cluster analysis, independently for each of the four campaigns. The results were exported as NetCDF files and voxel layers and published on the web, on the ArcGIS Online infrastructure, and on a ncWMS server.

This study improved the knowledge of the spatial distribution and evolution of ocean variables and ecological marine units in the two areas of the Portuguese coast. Based on marine data observations, we produced multidimensional surfaces and shared them in standardized formats, suited for further exploration and analysis by stakeholders. These ocean state base maps are essential for ocean preservation, sustainability, and governance, and for effectively creating digital twins of the ocean.

Mapeamento tridimensional de variáveis oceânicas e caracterização de ecossistemas marinhos na costa Portuguesa

RESUMO

Os oceanos são um tema constante na agenda internacional, quer pela importância vital dos ecossistemas marinhos para a vida humana, quer porque a sua preservação e sustentabilidade se encontram em risco. A Década da Ciência Oceânica para o Desenvolvimento Sustentável das Nações Unidas (2021-2030) materializa esta missão de melhorar e ampliar o conhecimento do mar, conduzindo a políticas e ações eficazes. Um dos seus principais desafios é a digitalização dos oceanos.

O objetivo deste trabalho é mapear, em três dimensões, as propriedades físicas e químicas da água do mar e caracterizar os ecossistemas marinhos na costa portuguesa. Foram utilizados dados do projeto AQUIMAR, recolhidos in situ ao longo de quatro campanhas, em duas áreas: B, na costa noroeste, entre Aveiro e a Figueira da Foz; e, na costa sul, entre Sagres e Tavira. Foram selecionadas oito variáveis oceânicas: temperatura, salinidade, oxigénio dissolvido, nutrientes (fosfato, nitrato+nitrito e silicato), clorofila a e pH. As representações digitais da sua distribuição na coluna de água foram criadas através de interpolação espacial, utilizando o método Empirical Bayesian Kriging 3D, que evidenciou um desempenho fiável, enquanto interpolador automatizado. Os ecossistemas marinhos foram caracterizados independentemente para cada uma das quatro campanhas, através de análise de clusters com recurso ao método K-means. Os resultados foram exportados para ficheiros NetCDF e camadas voxel e publicados na web, na infraestrutura ArcGIS Online e num servidor ncWMS.

Este estudo reforçou o conhecimento acerca da distribuição espacial e evolução das variáveis oceânicas e ecossistemas marinhos nas duas áreas da costa portuguesa. Com base em observações de dados marinhos, foram produzidas superfícies multidimensionais e partilhadas em formatos padronizados, adequados para exploração e análise futuras por parte dos utilizadores. Estes mapas de base do estado do oceano são essenciais para a sua preservação, sustentabilidade e governança, bem como para a criação eficaz de gémeos digitais do oceano.

KEYWORDS

Cluster analysis

Marine ecosystem

Ocean variable

Three-dimensional interpolation

Voxel layer

PALAVRAS-CHAVE

Análise de clusters

Camada voxel

Ecossistema marinho

Interpolação tridimensional

Variável oceânica

LIST OF ACRONYMS AND ABBREVIATIONS

°C	Degree Celsius
µmol	Micromole
2D	Two-Dimensional
3D	Three-Dimensional
ANN	Artificial Neural Network
ASE	Average Standard Error
CDF	Cumulative Distribution Function
CF	Climate and Forecast
CICES	Common International Classification of Ecosystem Services
CIESIN	Center for International Earth Science Information Network
CRPS	Continuous Ranked Probability Score
CSV	Comma-Separated Values
CTD	Conductivity, Temperature and Depth
DCM	Deep Chlorophyll Maximum
DestinE	Destination Earth
DGRM	Directorate-General for Natural Resources, Safety and Maritime Services
DIVA	Data-Interpolating Variational Analysis
DO	Dissolved Oxygen
DTO	Digital Twin of the Ocean
EBK	Empirical Bayesian Kriging
EBK3D	Three-Dimensional Empirical Bayesian Kriging
EBM	Ecosystem-Based Management
EC	European Commission
EEA	European Environment Agency
eEOV	Ecosystem Essential Ocean Variable
EEZ	Exclusive Economic Zone
EIF	Elevation Inflation Factor
EMODnet	European Marine Observation and Data Network
EMU	Ecological Marine Unit
EOV	Essential Ocean Variable
EPA	United States Environmental Protection Agency

ES	Ecosystem Services
ESA	European Space Agency
ESDA	Exploratory and Spatial Data Analysis
Esri	Environmental Systems Research Institute
EU	European Union
GNSS	Global Navigation Satellite System
GOODS	Global Open Oceans and Deep Seabed
GOOS	Global Ocean Observing System
IAPSO	International Association for the Physical Sciences of the Oceans
IDW	Inverse Distance Weighting
IH	Hydrographic Institute
IOC	Intergovernmental Oceanographic Commission
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
L	Liter
LOQ	Limit of Quantification
m	Meter
MBV	Minimum Bounding Volume
ME	Mean Error
MEA	Millennium Ecosystem Assessment
MEOW	Marine Ecoregions of the World
mg	Milligram
MLD	Mixed Layer Depth
MSP	Marine Spatial Planning
N	Nitrogen
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
NRP	Ship of the Portuguese Republic
O	Oxygen
OBA	Object Based Analysis
OGC	Open Geospatial Consortium

OMZ	Oxygen Minimum Zone
OV	Ocean Variable
P	Phosphate
PCA	Principal Component Analysis
PCM	Presidence of the Council of Ministers
PSU	Practical Salinity Unit
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Square Error
RMSSE	Root Mean Square Standardized Error
RVar	Ratio of the Variance
SC	Silhouette Coefficient
SCOR	Scientific Committee on Oceanic Research
SDG	Sustainable Development Goal
Si	Silicate
SME	Mean Standardized Error
SO	Specific Objective
SOM	Self-Organizing Map
SOOS	Southern Ocean Observing System
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SVM	Support Vector Machine
TEEB	The Economics of Ecosystems & Biodiversity
TEOS	Thermodynamic Equation of Seawater
TON	Total Oxidized Nitrogen
UN	United Nations
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
WEF	Word Economic Forum
WMS	Web Map Service
WOA13	World Ocean Atlas 2013
WOR	World Ocean Review

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background

The oceans are often considered the final frontier for science, either because they cover more than half of the Earth's surface (nearly 71%) (Duxbury & Cenedese, 2022), although only accounting for approximately 16% of the world's species (Costello & Chaudhary, 2017b), or because only 23.4% (Seabed 2030, 2022) of the seafloor is currently mapped with high resolution. They have been a persistent topic on the international agenda, and lately, given the great challenges of climate change and sustainability, concerns about their resilience, accessibility and digitization are emerging.

In 2015, the 2030 Agenda for Sustainable Development, founded by the United Nations (UN), included, for the first time, a specific objective related to the oceans: Sustainable Development Goal (SDG) 14 – Conserve and sustainably use the oceans, seas, and marine resources for sustainable development (UN, 2015). Later, in 2017, the UN declared the 2021-2030 period as the Decade of Ocean Science for Sustainable Development (UN, 2021), recognizing the importance of the marine realm for life on Earth, its alarming degradation, and the urgent need for adaptation strategies and policies based on scientific knowledge (UN, 2020). The Ocean Decade includes ten specific challenges that may contribute to a healthy, productive, and predicted ocean (Ryabinin, et al., 2019). Particularly significant for the scope of this study is challenge 8, which aims to create a digital representation of the ocean, "including a dynamic ocean map, which provides free and open access for exploring, discovering, and visualizing past, current, and future ocean conditions in a manner relevant to diverse stakeholders" (UN, 2022). This digital representation is commonly referred to as the digital twin of the ocean (DTO).

The digital twin concept was first introduced in 2002 as a model to improve product lifecycle management (Grieves, 2019). This concept has evolved, and it is now generally defined as a virtual representation of a physical entity capable of replicating its past, present, and future states. A continuous synchronization between the physical and virtual environments can leverage the system and enhance the understanding and prediction of the entity's behavior (Jones, et al., 2020). Digital twins are being applied to multiple domains, including the Earth's and the oceans' systems. At the European level, the major initiative is Destination Earth (DestinE), which aims to build a detailed digital twin of the Earth by 2030 to assess and predict the interaction between human activities and natural phenomena (EC, 2022a). Regarding the DTO, there are also two major initiatives: DITTO, which is a global program of the Ocean Decade (DITTO, 2022), and Iliad, which is an European Union (EU) funded project (Iliad, 2022).

The implementation of the 2030 Agenda, concerning the preservation and digitization of the oceans is a priority at the European (EC, 2022b) (EC, 2022c) and national levels. The National Strategy for the Sea 2021-2030 has translated these priorities into two specific objectives (SO): SO 1 – Combat climate change and pollution and protect and restore the ecosystems, and SO 9 – Encourage reindustrialization and productive capacity and digitize the ocean (PCM, 2021a). Regarding SO 1, there is also an emblematic measure (number one) that stipulates the implementation of a national program for habitats, ecosystems, and marine and coastal ecosystem services mapping (PCM, 2021b).

In essence, the international, European, and national contexts point to the need to preserve and digitize the ocean to achieve sustainability. While the preservation of the ocean is inherent to its sustainability, when it comes to digitization, the reasoning is less straightforward. According to Figure 1, ocean science is a key driver of the policies required to generate societal benefit. The digitalization of the sea can be seen as a strategy for making ocean data more transparent and accessible to involve and enable citizens, businesses, and governments to make better-informed decisions (EC, 2020). At the basis of this scientific discovery process are marine data collection and ocean mapping. Notably, in this process, ocean mapping is important not only for science but also directly for policy making, as it can support sea management, for instance, through marine spatial planning (MSP) tools (Fernandes, et al., 2020). Like the brain cells map the world to facilitate human orientation (O'Keefe, 2014), we need to map the ocean, in all its dimensions, to support and improve our actions.

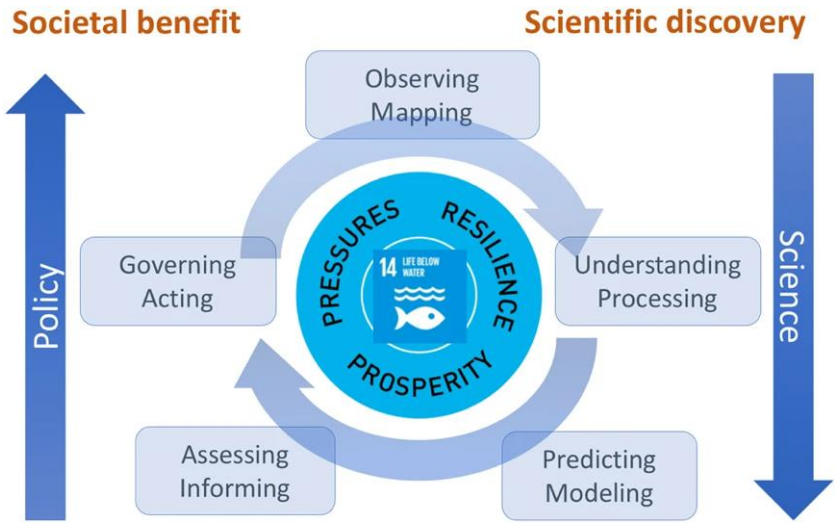


Figure 1. Ocean science as driver of ocean policies.
Source: *(Visbeck, 2018)*.

Three-dimensional (3D) mapping of physical and chemical properties of seawater (e.g., temperature, salinity, dissolved oxygen (DO), nutrients, chlorophyll a, and pH), also referred to

as ocean variables (OV), based on scattered in-situ observations is of major importance, as it produces a digital representation of the water column at a specific time and place. The spatial interpolation of these local measurements, systematically collected by ships or other platforms at sea, creates continuous surfaces, which are prolific tools to produce insights and enhance further analysis by stakeholders.

The spatial distribution of OVs is critical in multiple domains, from climate change to the blue economy, for instance, in the identification of marine ecosystems with similar physical and chemical properties. These ecosystems can be addressed as ecological marine units (EMU) and their benefits for human well-being and sustainability have been long demonstrated (Sayre, et al., 2017b).

Therefore, 3D mapping of OVs and the characterization of EMUs represent a baseline for ocean governance (i.e., a snapshot of the ocean state and behavior at a specific time and place), which is a core subject in the current agenda. From this base model or starting point, persistent monitoring and change detection (e.g., by human impact or climate change) can easily become a reality. This work also represents a precursor to a multifaceted DTO, with features like diagnosis, prognosis, and simulation (what if scenarios) to support and improve decision-making.

1.2. Research aim and objectives

The objects of this research are eight of the physical and chemical properties of seawater, namely, temperature, salinity, DO, nutrients (phosphate (PO_4), total oxidized nitrogen (TON) (NO_X), and silicate (SiO_2)), chlorophyll a, and pH. Due to its sampling techniques, the measurements of these OVs are usually scattered and discontinuous in space and time. For this reason, the processing of these data points can benefit from the use of geostatistical methods, to build continuous surfaces, and data mining techniques, to identify ecosystems with specific physical and chemical properties.

We used data acquired within the scope of the AQUIMAR[1] project. Data was collected on board two hydrographic vessels, by a rosette equipped with a multi-parameter probe and water sampling bottles, along four campaigns, and focused on two areas of the Portuguese coast: one in the northwestern region, approximately between Aveiro and Figueira da Foz, and another in the south region, approximately between Sagres and Tavira.

[1] Marine Knowledge Supporting Aquaculture (https://aquimar.hidrografico.pt [accessed 12 September 2022]).

The research aim (general objective) is to map, in 3D, the physical and chemical properties of seawater and to characterize EMUs on the Portuguese coast. In order to achieve this, we defined three SOs, as described in Table 1.

General Objective	Specific Objectives	
▪ **Map**, in 3D, physical and chemical properties of seawater and **characterize** ecological marine units on the Portuguese coast.	SO 1	▪ **Create** 3D surfaces of the physical and chemical properties of seawater through spatial interpolation.
	SO 2	▪ I**dentify** and **describe** ecological marine units through cluster analysis.
	SO 3	▪ **Publish** the results on a web application.

Table 1. General and specific objectives.

We created 3D surfaces of the eight OVs (SO 1) using an extension of the Empirical Bayesian Kriging (EBK) method (Gribov & Krivoruchko, 2020), and identified and described EMUs (SO 2) through K-means cluster analysis (Zhao, et al., 2020). Finally, the web application was published (SO 3) using the ArcGIS Online[2] infrastructure.

One of the expected outcomes of this research is an improved knowledge of the spatial distribution and evolution of OVs and EMUs in the two areas of the Portuguese coast. Furthermore, the web mapping application will enable users to explore the results and apply them in specific studies.

1.3. Structure

The work is structured into six chapters.

The first chapter (introduction) presents the research theme, justifies its topicality and relevance, and describes the research aim and objectives.

The second chapter (literature review) focuses on the major concepts associated with this work and describes the existing knowledge, methods, and gaps.

The third chapter (data and study areas) details the data used and outlines its geographic extent, provenance, sampling scheme, and structure.

The fourth chapter (methods) explains the methodology adopted, divided into seven stages.

The fifth chapter (results and discussion) presents and discusses the results, divided into five sections.

The last chapter (conclusion) reviews the main results and conclusions and highlights the knowledge contribution, including limitations, recommendations, and proposals for further development.

[2] https://www.arcgis.com [accessed 12 September 2022].

2. LITERATURE REVIEW

This chapter reviews the key concepts related to this work, providing an outline of the existing knowledge, methods, and gaps. First, we present a bibliometric analysis, highlighting the major concepts, their networks of relationships, and their evolution in time. Then we describe the marine and coastal ecosystems (their importance and how they are defined), the ocean variables (their relation to the ecosystems and their global and typical distribution), and the available mapping methods.

2.1. Bibliometric analysis

The assessment of scientific literature concerning this study was initially explored through a bibliometric analysis performed with VOSviewer 1.6.17[3], which is a software tool to construct and visualize bibliometric networks. The search was conducted on the Scopus[4] database, on titles, abstracts, and keywords, on January 3rd, 2022, using the following string "marine ecosystem map"[5]. It returned 1 583 results, published from 1977 to 2021.

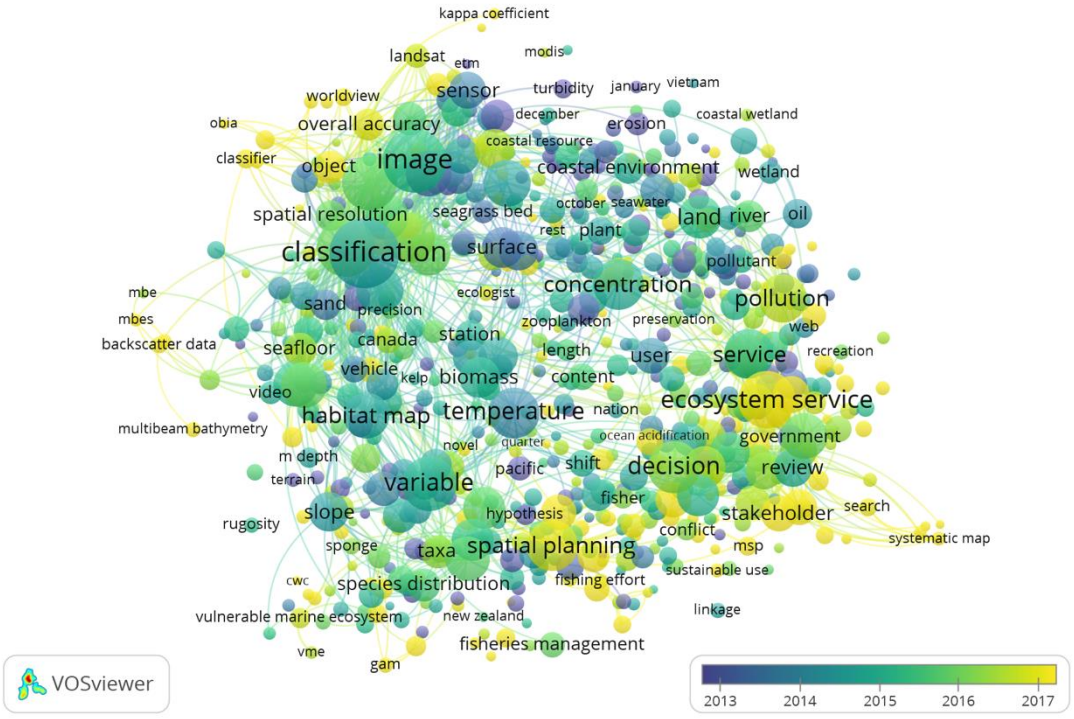


Figure 2. VOSviewer overlay visualization, classified by publication date.

[3] https://www.vosviewer.com [accessed 12 September 2022].
[4] https://www.scopus.com [accessed 12 September 2022].
[5] TITLE-ABS-KEY (marine AND ecosystem AND map).

The distribution of keywords used in the titles and abstracts of the articles can be evaluated from different perspectives. The relative size of the keywords illustrates their incidence; the network visualization shows how the literature is clustered around the keywords; and the overlay visualization, for instance, classified by publication date, reveals the evolution of keywords in time. Thus, we can perceive a more recent focus on ecosystem services (ES), MSP, stakeholders, and pollution (Figure 2).

Although this analysis can be a helpful start, it is not comprehensive. Therefore, we structured the literature review in accordance with the study methodology, focusing on the core topics discussed afterwards (OVs, interpolation and mapping methods, and marine ecosystems), and their main ramifications.

2.2. Marine and coastal ecosystems

Marine and coastal ecosystems are the richest and most valuable ecosystems in the world (Martínez, et al., 2007). They are also the main producers of oxygen (O_2) and the largest absorbers of carbon dioxide (CO_2) (Mandal, et al., 2022). However, despite their importance, they embody the most severely exploited ecosystems in the world (Barbier, 2017).

Coastal regions represent 4% of the earth's land area and 11% of the world's oceans (Barbier, 2017). Nevertheless, they contain about 40% of the world's population (within 100 kilometers of the coast) (CIESIN, 2012) and 90% of the catches from marine fisheries (Barbier, 2017). Coastal zones also reveal higher rates of population growth and urbanization (Neumann, et al., 2015). So, as population density and economic activity in these areas escalate, so does the pressure on coastal ecosystems. The most common environmental concerns are habitat conversion, land cover change, pollutant loads, and the introduction of invasive species. These strains can lead to loss of biodiversity, coral reef bleaching, oxygen depletion (hypoxia), algae bloom, siltation, water quality degradation and new diseases (UN, 2007).

Although marine and coastal ecosystems can be seriously threatened by anthropogenic disturbances (Zamboni, et al., 2021) and climate change (IPCC, 2021), they are also vital in reducing the vulnerability of coastal communities to hazards, namely through wave attenuation, sediment capture, vertical accretion, erosion reduction and the mitigation of storm surge and debris movement (Spalding, et al., 2014). This has particular relevance, as extreme weather events have been perceived, since 2017, as the most probable global risk. Furthermore, in 2021, four of the top seven global risks by likelihood and four of the top seven global risks by impact are related to environmental threats, damage, or crisis (WEF, 2021).

Marine and coastal ecosystems provide a wide range of benefits to human well-being. These gains are usually designated as ES, a concept first introduced in the last quarter of the twentieth century (De La Cruz, 2021). Marine ES were formally defined by the Millennium Ecosystem Assessment (MEA) as "benefits that people obtain from coastal and marine ecosystems" and aggregated into four major categories: provisioning, regulating, cultural and supporting services (MEA, 2005); as described in Figure 3. In recent years, the growing interest in this theme has provided several definitions and classification schemes delivered by other initiatives: The Economics of Ecosystems & Biodiversity (TEEB) (TEEB, 2022), the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (IPBES, 2022) and the Common International Classification of Ecosystem Services (CICES) (CICES, 2022). Nowadays, the CICES scheme, updated in 2018, is the most widely accepted, defining ES as "the contributions that ecosystems make to human well-being" and dividing them into three main categories: provisioning, regulating and maintenance, and cultural[6] (Buonocore, et al., 2021).

ECOSYSTEM SERVICES	Coastal									Marine		
	Estuaries and marshes	Mangroves	Lagoon and salt ponds	Intertidal	Kelp	Rock and shell reefs	Seagrass	Coral reefs	Inner shelf	Outer shelves edges slopes	Seamounts & mid-ocean ridges	Deep sea and central gyres
Biodiversity	X	X	X	X	X	X	X	X	X	X	X	X
Provisioning services												
Food	X	X	X	X	X	X	X	X		X	X	X
Fibre, timber, fuel	X	X	X						X	X		X
Medicines, other resources	X	X	X		X			X	X			
Regulating services												
Biological regulation	X	X	X	X		X		X				
Freshwater storage and retention	X		X									
Hydrological balance	X		X									
Atmospheric and climate regulation	X	X	X	X		X	X	X	X	X		X
Human disease control	X	X	X	X		X	X	X				
Waste processing	X	X	X				X	X				
Flood/storm protection	X	X	X	X	X	X	X	X				
Erosion control	X	X	X				X	X				
Cultural services												
Cultural and amenity	X	X	X	X	X	X	X	X	X			
Recreational	X	X	X	X	X			X				
Aesthetics	X		X	X				X				
Education and research	X	X	X	X	X	X	X	X	X	X	X	X
Supporting services												
Biochemical	X	X			X			X				
Nutrient cycling and fertility	X	X	X	X	X	X		X	X	X	X	X

Figure 3. ES provided by marine and coastal habitats.
Source: *(UNEP, 2006)*.

The importance of marine and coastal ecosystems to the human society is widely recognized, not just for the benefits they provide, but also because they can offer a new framework for sustainable management. Thus, they should be preserved, maintaining their diversity and

[6] The "supporting services" category, originally defined by the MEA, was excluded in the CICES scheme, because it was considered as part of the structures, processes, and functions of ecosystems (Buonocore, et al., 2021).

productivity, rather than undergoing unprecedented changes, at different scales, due to human direct and indirect actions (Buonocore, et al., 2021). Ecosystems may supply a new integrated management approach to assure their own long-term sustainability – ecosystem-based management (EBM) – which recognizes the full spectrum of interactions within an ecosystem, including humans, instead of considering single issues, species, or ES in isolation (O'Higgins, et al., 2021). EBM handles marine and coastal ecosystems in a way that acknowledges their complexity and comprehends their connections. One of the fundamental aspects of EBM is that it is a location-based approach, where an ecosystem represents the place (UNEP, 2011).

Therefore, mapping marine and coastal ecosystems, revealing their locations and spatial patterns, and inferring their state and trends, is an essential task to mainstream them into policy and decision-making initiatives (Buonocore, et al., 2021). To achieve this, we first need to describe how ecosystems can be defined and mapped.

2.2.1. Definition of ecosystems

There are several ecosystem definitions. We chose to highlight the ones best linked to the aim of this work. An ecosystem can be defined as "the complex of living organisms, their physical environment, and all their interrelationships in a particular unit of space" (Britannica, 2022a) or as "the complex of a community of organisms and its environment functioning as an ecological unit" (Merriam-Webster, 2022a). From these ecosystem definitions, we can draw three conclusions: (i) they have unique characteristics (e.g., biological, physical, chemical, etc.); (ii) they exist in a specific location; and (iii) they may be referred to as ecological units.

Ecosystems are areas (or volumes) that can be physically distinguished and geographically located. However, the mapping of EMUs presents some challenges compared to the mapping of terrestrial ecosystems (Sayre, et al., 2017b):

- Dimensionality. Terrestrial ecosystems are usually processed as two-dimensional features, modelled by raster surfaces, while marine ecosystems are essentially three-dimensional entities (water column), bounded by two surfaces (seabed and sea surface). In this sense, they need to be mapped using 3D points (latitude, longitude, depth, and attributes) (Sayre, et al., 2017b);
- Characteristics. The defining characteristics of ecosystems are different on land and at sea. While terrestrial ecosystems may be defined by distinct blends of bioclimate, vegetation, landforms and lithology, marine ecosystems are usually mapped through

different combinations of biological, physical, and chemical properties of seawater[7] (Sayre, et al., 2017b);

- Dynamics. The marine environment is more spatially and temporally dynamic than the terrestrial one (Sayre, et al., 2017b). The constant changes in seawater properties make marine ecosystem mapping a time-related and a never-ending task. It also ensures that systematic mapping and consistent monitoring can easily detect disturbances.

The first major division of the ocean ecological regions are the pelagic and benthic provinces (Webb, 2021). The pelagic zone equals the water column of the entire ocean, and it is vertically structured into five depth regions (Britannica, 2022b): epipelagic, mesopelagic, bathypelagic, abyssalpelagic and hadalpelagic. The benthic zone matches the seabed, including the sediment surface and some sub-surface layers, and it is horizontally divided into five discrete areas (Webb, 2021): littoral or intertidal, sublittoral or shelf, bathyal, abyssal and hadal. The water column can also be structured into light zones depending on the maximum depth reached by each component of the electromagnetic spectrum (Pinet, 2009). This division of the ocean water column and seafloor is illustrated in Figure 4.

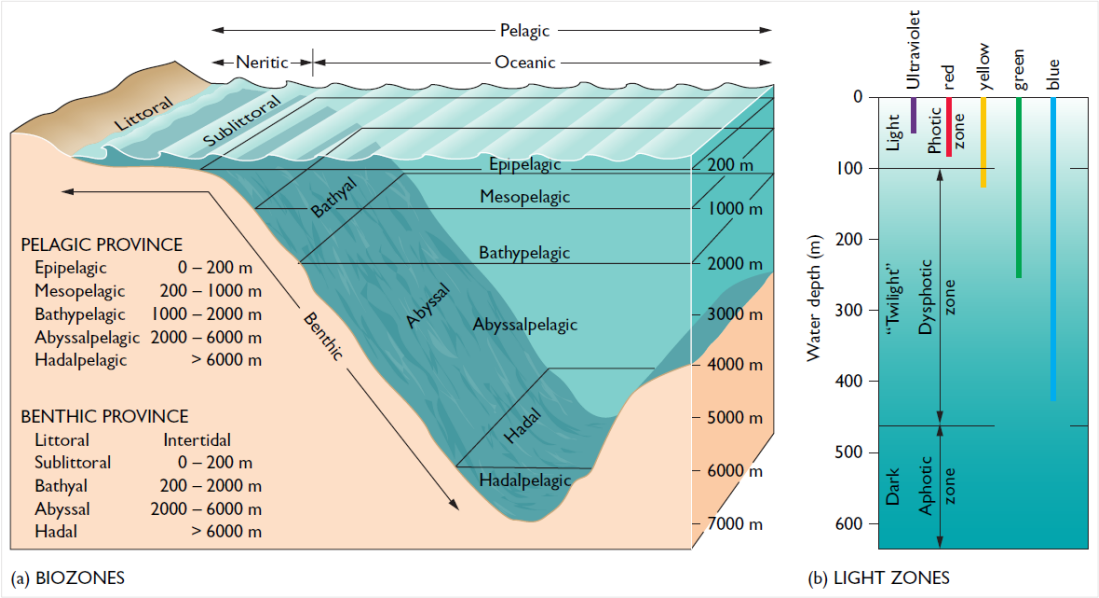


Figure 4. Oceanic divisions and zonation.
Source: *(Pinet, 2009)*.

The oceanic zones have specific characteristics and discrete boundaries, some more perceptible than others. For instance, the epipelagic zone extends from the surface to approximately 200 m

[7] Marine ecosystems mapping can also use bathymetry, similarly, as terrestrial ecosystems mapping uses landforms. However, this characteristic (seafloor morphology) is more commonly used in 2D marine ecosystem mapping, as it is an exclusive attribute of the seabed that does not affect the entire water column.

depth, having enough sunlight to support photosynthesis and being home to iconic marine animals (Oceana, 2022). Next, the mesopelagic zone spreads from the bottom of the epipelagic zone to approximately 1 000 m, where sunlight cannot reach, revealing lower levels of oxygen and higher levels of nutrients (Costello & Breyer, 2017a). These vertical and horizontal differences, between ecosystems, make 3D mapping possible.

2.2.2. Mapping marine ecosystems

Ocean mapping and the subsequent classification of marine regions have been developed for various purposes, based on multiple criteria and methodologies, and have created numerous products at different scales. (Zhao, et al., 2020) provides a summary of some of these initiatives. The mapping of EMUs is usually associated with the identification and characterization of water masses. This concept has been a recurrent subject of research, initially from a qualitative point of view and lately from a quantitative analysis (Tomczak, 1999).

The ocean's water column structure was consistently described in 1942, addressing the general distribution of some physical properties of seawater (temperature, salinity, and density) (Sverdrup, et al., 1942). The following works had evolving complexity, from which we selected the ones that represent breakthroughs or milestones related to the aim of this work:

- In 1986, global water masses, regarding temperature and salinity, were mapped, describing their distribution, boundaries, and formation regions (Emery & Meincke, 1986);
- Also in 1986, large marine ecosystems, located in coastal regions, were mapped based on expert opinion (qualitative methods) and introducing the importance of ecosystem variability and management (Sherman & Alexander, 1986);
- In 2004, water masses on continental shelves were mapped using satellite imagery (sea surface temperature and ocean color – chlorophyll) and based on multivariate cluster analysis (quantitative methods) (Oliver, et al., 2004);
- In 2007, the coastal and shelf areas were classified in marine ecoregions of the world (MEOW), once again, based on expert opinion and introducing a nested system of 12 realms, 62 provinces and 232 ecoregions (Spalding, et al., 2007). This work was further extended to the pelagic provinces of the world (Spalding, et al., 2012);
- In 2008, biogeographic provinces were mapped using satellite imagery (sea surface temperature and ocean color – chlorophyll) and based on multivariate cluster analysis, introducing spatial and temporal province dynamics and the subsequent changes in their boundaries (Oliver & Irwin, 2008). The province dynamics were further developed,

regarding their seasonal changes and the detection of spatial shifts (Reygondeau, et al., 2013), and the use of other environmental variables (Zhao & Costello, 2019);

- In 2009, the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) produced a global biogeographic classification – Global Open Oceans and Deep Seabed (GOODS) – providing a classification both for the pelagic and benthic provinces, based on expert opinion (UNESCO, 2009). The deep-sea classification was further refined by subsequent studies (Watling, et al., 2013);
- In 2017, two separate studies produced global biogeographic (Sutton, 2017) and biochemical (Reygondeau, et al., 2017a) classifications of the mesopelagic zone. These studies followed similar approaches implemented for the epipelagic zone (Longhurst, 2007). The biogeographic classification integrated expert opinion with environmental variables (temperature, salinity, and DO) considered to be significant ecological drivers, through a modified Delphi Method (Linstone & Turoff, 2002). The biochemical classification was generated in 3D and used seven environmental parameters (temperature, salinity, DO concentration, nutrients concentration (nitrate (NO_3), silicate (SiO_2) and phosphate (PO_4)) and particulate organic carbon flux) from the World Ocean Atlas 2013 (WOA13) (Levitus, et al., 2015) and a combination of clustering algorithms;
- Also in 2017, global marine ecosystem mapping was produced in 3D, using six OVs (temperature, salinity, DO, nitrate, phosphate, and silicate) from the WOA13, and based on K-means cluster analysis (Sayre, et al., 2017b). This work also introduced the concept of EMU;
- In 2020, near surface marine ecosystems were mapped globally, based on cluster analysis. This study distinguished seven marine ecosystems and, as it was based on 20 environmental variables, it can be considered one of the most comprehensive global classifications to date. The results were compared with other objective classification schemes, for instance, revealing a good resemblance to the EMUs (Zhao, et al., 2020).

This brief, but focused, description of the main marine ecosystem mapping initiatives through time, allows us to summarize the following:

- Physical, chemical, and biological properties of seawater and their spatial distribution along the water column, are the main drivers for ecosystem mapping;
- Improved availability of this data is producing more objective classifications, evolving from expert opinion to unsupervised data-driven methods (Zhao & Costello, 2019);

- Still, much of the data that is being used comes from averaged records of ocean properties, generated for annual, seasonal and monthly time-scales, at 0,25° to 1° horizontal resolution, and interpolated with specific vertical intervals (Levitus, et al., 2015), instead of using in-situ measurements at a local scale;
- Finally, although ocean properties are three-dimensional, ecosystem mapping has traditionally been applied to horizontal layers, for instance, benthic province (Harris & Whiteway, 2009), mesopelagic zone (Sutton, 2017) or photic zone (Zhao & Costello, 2019). Only in recent years, three-dimensional mapping has been achieved, at global (Reygondeau, et al., 2017a) (Sayre, et al., 2017a) and regional scales (Reygondeau, et al., 2017b).

2.3. Ocean variables

Physical and biochemical properties of seawater, usually called OVs (GOOS, 2021), are critical to understanding global scale events (Stewart, 2008). The spatial distribution of these variables influences water column stratification, for instance, affecting biomass concentration and positioning (MacKenzie & Adamson, 2004). It also defines ocean (thermohaline) circulation (Wunsch, 2002) and regulates interaction between the sea surface and the atmosphere, shaping the climate (Jayne & Marotzke, 1999).

It is vital to measure and map these variables, not only to provide precise forecasts and early warnings, to project climate scenarios and to assess ocean's health (GOOS, 2021), but also, as shown, to accurately map ecosystems. Table 2 describes the essential ocean variables (EOV) defined by the IOC program Global Ocean Observing System (GOOS), based on relevance, feasibility, and cost effectiveness (GOOS, 2021).

Physics	Biochemistry	Biology and Ecosystems
▪ Sea State ▪ Ocean surface stress ▪ Sea ice ▪ Sea surface height ▪ Sea surface temperature ▪ Subsurface temperature ▪ Surface currents ▪ Subsurface currents ▪ Sea surface salinity ▪ Subsurface salinity ▪ Ocean surface heat flux	▪ Oxygen ▪ Nutrients ▪ Inorganic carbon ▪ Transient tracers ▪ Particulate matter ▪ Nitrous oxide ▪ Stable carbon isotopes ▪ Dissolved organic carbon	▪ Phytoplankton biomass and diversity ▪ Zooplankton biomass and diversity ▪ Fish abundance and distribution ▪ Marine turtles, birds, mammals abundance and distribution ▪ Hard coral cover and composition ▪ Seagrass cover and composition ▪ Macroalgal canopy cover and composition ▪ Mangrove cover and composition ▪ Microbe biomass and diversity ▪ Invertebrate abundance and distribution
Cross-disciplinary (including human impact)		
	▪ Ocean color ▪ Marine debris	▪ Ocean sound

Table 2. Essential OVs.
Source: *(GOOS, 2021)*.

OVs can be measured directly (e.g., temperature, salinity, pressure) or used to derive other properties (e.g., heat capacity, thermal expansion, dissolving power, transparency) (Sverdrup, et al., 1942). One of the most important properties for studying ocean circulation, density, is typically computed using an equation of state (Thermodynamic Equation of Seawater (TEOS)) (IOC; SCOR; IAPSO, 2010)) based on temperature, salinity, and pressure (Pawlowicz, 2013).

2.3.1. Ocean variables and ecosystems

EOVs were selected for comprehensive ocean observing purposes and are not solely dedicated to ecosystem mapping. In this sense, some studies, of which we highlight the latest, have tried to identify specific variables for marine ecosystem monitoring, designated as ecosystem EOVs (eEOV):

- In 2015, a study applied to the Exclusive Economic Zone (EEZ) of Australia identified ten eEOVs candidates, among which we highlight chlorophyll, oxygen, and nutrients (Hayes, et al., 2015);
- In 2016, a study focused on the Southern Ocean Observing System (SOOS) (SOOS, 2022) identified eEOVs as biological or ecological quantities, and linked them to some seawater properties like temperature, nutrients, oxygen, and pH (Constable, et al., 2016);
- In 2018, a similar study was developed globally, but also focused on biological and ecological EOVs, and briefly referred to temperature and oxygen (Miloslavich, et al., 2018);
- In 2019, a study of habitat delineation used a set of five variables related with water masses properties: temperature, salinity, mixed layer depth (MLD), chlorophyll and nutrients (nitrate). This approach introduced principal component analysis (PCA) and hierarchical clustering to derive habitats and relate them to phytoplankton communities (Weber, et al., 2019);
- In 2020, a study for deep-ocean monitoring identified a set of ecological variables among five scientific areas: biodiversity; ecosystem functions; impacts and risk assessment; climate change, adaptation, and evolution; and ecosystem conservation. The monitoring of these variables was directly associated with temperature and oxygen, and indirectly with salinity and pH (Danovaro, et al., 2020).

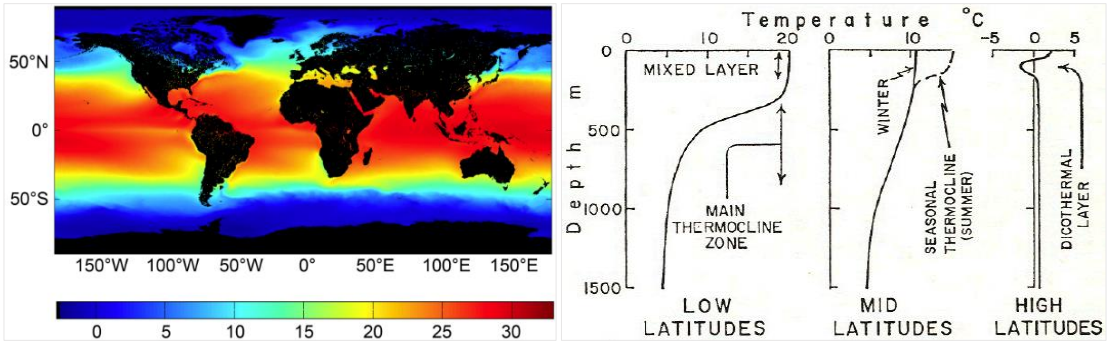
Based on these studies and the ones related to ecosystem mapping, presented before, we can conclude that there is a widespread agreement on some of the variables that might be relevant to ecosystem definition and mapping: temperature, salinity, oxygen, nutrients, chlorophyll, and

pH. The following paragraphs describe the typical horizontal and vertical distributions of these variables in the ocean.

a. Temperature

Temperature is a central variable of the ocean. Its spatial distribution determines the location and characteristics of many marine biomes. Changes in temperature can affect trophic interactions (Durant, et al., 2019) or circulation patterns and nutrients concentration in marine ecosystems (Pastuszak, et al., 1999). Seawater temperature also affects the interaction between the ocean and the atmosphere, influencing global climate (EPA, 2021).

Sea surface temperature (SST) global horizontal distribution depends on latitude. Temperature decreases with latitude, and the isothermal lines, generally, have an east-west orientation, unveiling slight north-south inflexions in coastal areas, due to local events (NOAA, 2020). The vertical distribution is influenced by depth. Temperature decreases with depth and the specific profile shape is determined by latitude and season (Pickard & Emery, 1990). These patterns are illustrated in Figure 5.



(a) (b)

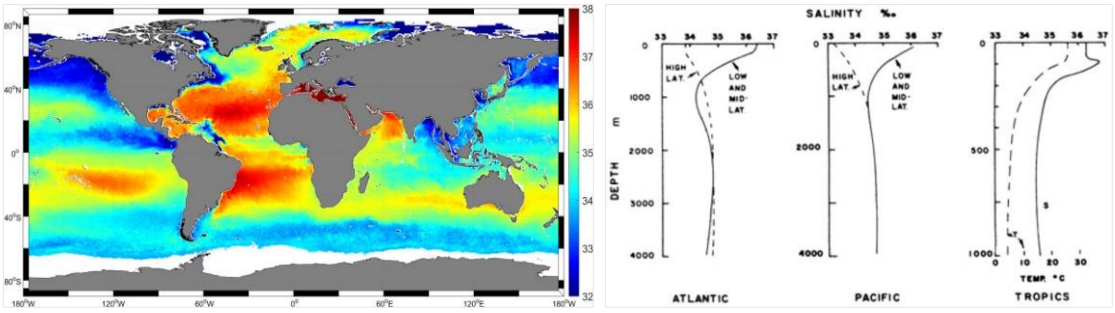
Figure 5. Typical horizontal and vertical distributions of ocean temperature.
Sources: (a) *(NOAA, 2020)*; (b) *(Pickard & Emery, 1990)*.

b. Salinity

Salinity is also considered a major ocean variable, affecting ocean currents and water mass interactions and hydrodynamics (Reygondeau, et al., 2017a). Furthermore, salinity can outline ecosystem structure and species distribution, because of the specific salinity tolerances of aquatic organisms and how their vital biological functions are affected (Smyth & Elliott, 2016).

Sea surface salinity (SSS) global horizontal distribution also reveals a zonal pattern (determined by latitude), although more complex than SST. Salinity is higher between 20 and 30 degrees north and south and decreases towards the equator and the poles, revealing a positive correlation with the evaporation-precipitation rate (Reul, et al., 2020). The vertical distribution is influenced by depth, exhibiting a more complex relation than temperature. In the upper layer

of the ocean, density, which is the main property responsible for the vertical stability of a water mass, is mainly determined by temperature, rather than salinity. As salinity variation is not sufficient to override the temperature effect, there can be either high or low salinity in the warmer upper layer. In low and mid-latitudes, salinity decreases with depth until it reaches a minimum, and then it remains relatively uniform or has a slight increase, while the profile shape depends on latitude and ocean basins (Pickard & Emery, 1990). These patterns are illustrated in Figure 6.



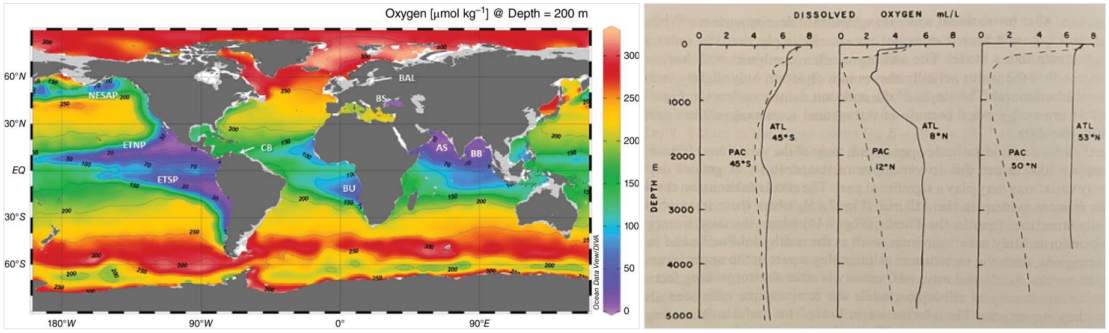
(a) (b)

Figure 6. Typical horizontal and vertical distributions of ocean salinity.
Sources: (a) *(Reul, et al., 2020)*; (b) *(Pickard & Emery, 1990)*.

c. Oxygen

DO in the oceans influences the metabolism of organisms, inducing their distribution (EEA, 2021) and the biochemistry and structure of marine ecosystems (Wilson, et al., 2019). This variable is important for identifying oxygen minimum zones (OMZ), which represent barriers to the migration of non-adapted organisms (Betrand, et al., 2010). For instance, the rising of the OMZ upper limit is reducing the available habitat for some species (Stramma, et al., 2012).

DO global horizontal distribution is complex and derives from an interplay of ocean interaction with the atmosphere, water circulation, and biological productivity, which involves oxygen consumption. The highest DO concentrations are usually located at high-latitudes, while the low ones are found at mid-latitudes, namely on the western coasts (WOR, 2010). The vertical distribution is typically characterized by high values near the surface and the OMZ in the upper 1 000 m, depending on latitude and ocean basins (Pickard & Emery, 1990). These patterns are illustrated in Figure 7.



(a) (b)

Figure 7. Typical horizontal and vertical distributions of ocean DO.
Sources: (a) *(Jürgens & Taylor, 2018)*; (b) *(Pickard & Emery, 1990)*.

d. Nutrients

Nutrient concentration (nitrate (NO_3), silicate (SiO_2) and phosphate (PO_4)) limit primary production within the epipelagic layer (Bristow, 2017), shaping the carbon influx into the mesopelagic (Sigman & Hain, 2012). They can also be used as indicators of the biogeochemical processes occurring and that affect ecosystem structure (Reygondeau, et al., 2017a).

Nutrients global horizontal and vertical distribution mainly depends on ocean productivity (Sigman & Hain, 2012). Typically, there are lower values of nutrients in the upper layers, due to their use by phytoplankton in the euphotic zone, and higher values in deeper waters, due to the decay and sinking of organic matter. This cycle is sustained by physical processes like upwelling (Pickard & Emery, 1990). These patterns are illustrated in Figure 8.

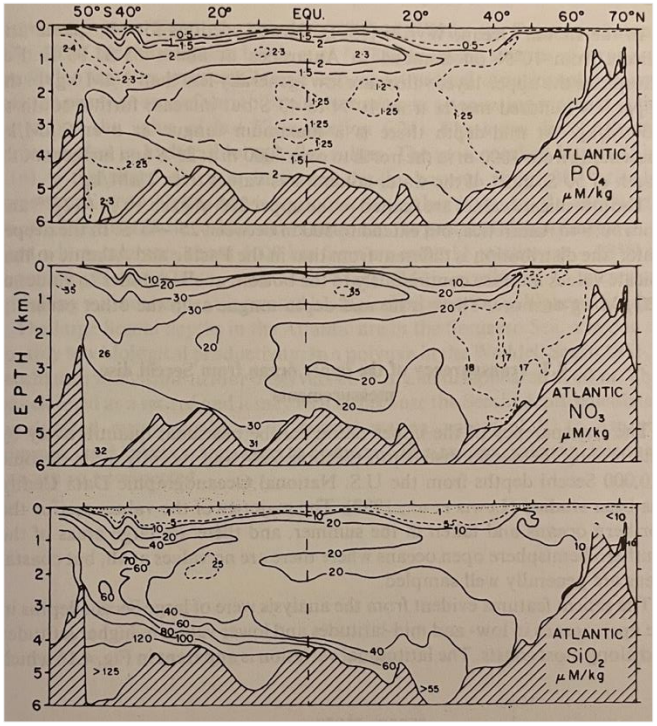
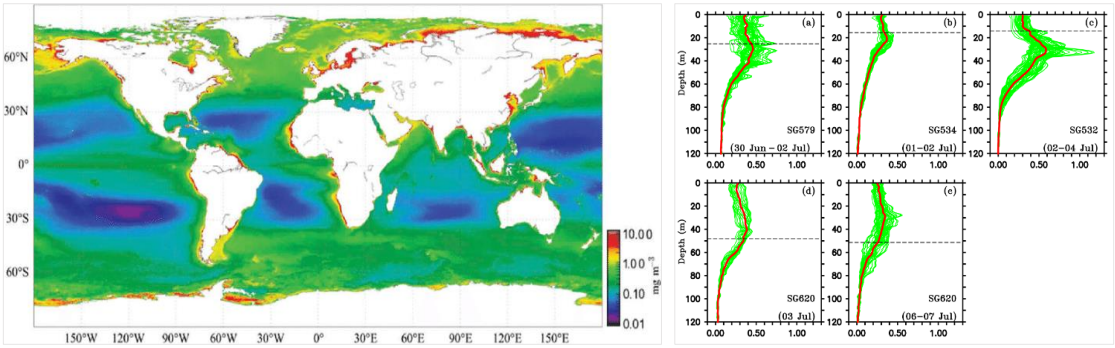


Figure 8. Typical horizontal and vertical distributions of ocean nutrients.
Source: *(Pickard & Emery, 1990)*.

e. Chlorophyll

Chlorophyll is also linked to ocean productivity as it allows organisms to photosynthesize. It can be used to assess phytoplankton concentration (Marrari, et al., 2017) and hence estimate primary production potential.

Chlorophyll global horizontal distribution is related to upwelling events, where cooler and richer water arises to the surface, evidencing higher values at high latitudes, around the equator and along the shores of continents (NASA, 2022). Typically, its vertical distribution is characterized by an increasing gradient from the surface to the limit of the euphotic zone, where it reaches a maximum (deep chlorophyll maximum (DCM)) and starts to decay. These patterns are illustrated in Figure 9.

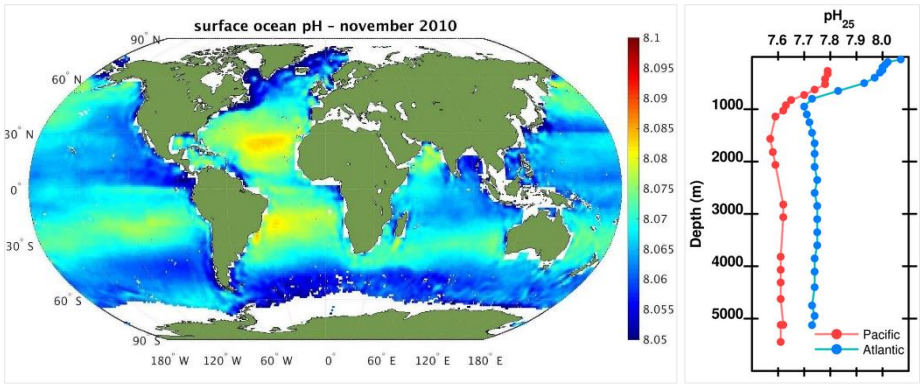


(a) (b)

Figure 9. Typical horizontal and vertical distributions of ocean chlorophyll.
Sources: (a) *(Couto, et al., 2016)*; (b) *(Thushara, et al., 2019)*.

f. pH

Ocean acidification, which reduces seawater pH, is linked with the absorption of carbon dioxide (CO_2) by the ocean. As CO_2 in the atmosphere increases, more of it is dissolved into seawater. The decrease in ocean pH represents a threat to marine ecosystems (Doney, et al., 2020). For instance, this process captures carbonate ions, which are used by many organisms to build shells and skeletons. Furthermore, even slight changes in pH disturb many chemical reactions, including some essential to life (Smithsonian, 2018). Additionally, there is also evidence that ecosystems' own processes can modulate pH variability (Lowe, et al., 2019). The global horizontal and vertical distributions are illustrated in Figure 10.



(a) (b)

Figure 10. Typical horizontal and vertical distributions of ocean pH.
Sources: (a) *(ESA, 2021)*; (b) *(Mackie, et al., 2011)*.

2.3.2. Measuring and mapping ocean variables

OVs can be derived by remote sensing: temperature (Minnett, et al., 2019), salinity (Vinogradova, et al., 2019), oxygen (Kim, et al., 2020), nutrients (Wang, et al., 2018), chlorophyll (ocean color) (Groom, et al., 2019) and pH (Gregor & Gruber, 2021).

However, these methods are usually limited to the sea surface. The measurement of ocean variables along the water column has been conducted locally by ships or, more recently, by unmanned vehicles like drifters (NOAA, 2022). On board ships, these variables are generally determined through multi-parameter stations, where a device with multiple sensors is dropped along the water column, maintaining the same horizontal position (latitude and longitude), and recording values for several depths, resulting in depth profiles for each variable, for each station. In-situ measurements, especially the ones made by ships, can be expensive (Arbic, 2015) and they represent discrete observations of continuous variables. In this sense, spatial interpolation is required to estimate values at unsampled locations and to build a surface (two-dimensional) or a volume (three-dimensional) from the initial network of surveyed points.

Spatial interpolation is based on two basic assumptions. The first is spatial autocorrelation, which derives from Tobler's first law of geography, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In a more refined formulation, this means that "nearby things are more similar than distant things" (Goodchild, 2009). The second is that the values are continuous in space (Wu & Hung, 2016). Both assumptions are true for OVs.

a. Spatial interpolation

Spatial interpolation comprises multiple methods (e.g., inverse distance weighting (IDW), splines, kriging), which can be classified according to a few criteria (e.g., global and local, exact and inexact, deterministic and stochastic, linear and nonlinear, univariate and multivariate) (Li

& Heap, 2008). Each of these methods uses specific parameters whose tuning can produce different surfaces from the same sampled points, even when applying the same method (Wu & Hung, 2016).

One of the main distinctions among spatial interpolation techniques is between deterministic and stochastic or geostatistical methods. A deterministic interpolator (e.g., IDW) only provides the estimated values, while a geostatistical interpolator (e.g., kriging) also provides the prediction errors associated with the estimations (Li & Heap, 2008). In this sense, deterministic models like IDW can be adequate for dense and regularly spaced data (Isaaks & Srivastava, 1989) and helpful for exploratory and spatial data analysis (ESDA). Nevertheless, they are less adequate when the dataset has few samples or clustered points (Isaaks & Srivastava, 1989) or when the spatial autocorrelation structure is not isotropic (Lu & Wong, 2008).

Despite these general principles, to this date, there is no universal rule regarding which is the most adequate interpolation technique (Wu & Hung, 2016). This suitability assessment depends on each situation, for instance, the sampling design and spatial distribution (e.g., sample density, spatial distribution and size) and the data characteristics (distribution, isotropic and anisotropic, variance and range of values) (Li & Heap, 2008) (Kravchenko, 2003). There are several studies that review and compare spatial interpolation methods applied to, e.g., environmental variables (Li & Heap, 2014), climatic and bioclimatic variables (Amiri, et al., 2020) (Attorre, et al., 2007), air quality (Noi & Murray, 2022) (Yuval, et al., 2017) (Wong, et al., 2004), air temperature (Wang, et al., 2017), groundwater contamination (Mirzaei & Sakizadeh, 2016), soil properties (AbdelRahman, et al., 2021) (Göl, et al., 2017), in which the conclusions are usually limited to the specific use case.

IDW and kriging represent the most frequently used methods in environmental sciences (Li & Heap, 2011). Regarding OVs, there is also a specific method, called data-interpolating variational analysis (DIVA) (Troupin, et al., 2012), whose development is funded by the European Marine Observation and Data Network (EMODnet) (EMODnet, 2021) and SeaDataNet (SeaDataNet, 2022b) projects.

Classical kriging has some downsides: (i) kriging weights are highly sensitive to variogram model misspecification (Pilz & Spöck, 2008); (ii) its complexity and specificity to the input data, such as semivariogram modelling and neighborhood definition, is time-consuming and rather subjective (Li & Heap, 2008).

EBK was developed in 2012. Its main objective was to implement a reliable automatic interpolator (Krivoruchko, 2012) that could unravel some of the typical disadvantages of the

classic geostatistical methods: (i) assuming data embodies a stationary process; (ii) describing data with a single model; (iii) not considering model uncertainty in the predictions; (iv) and not resolving coincident data and individual measurement errors (Gribov & Krivoruchko, 2020). EBK has been extensively described (Gribov & Krivoruchko, 2020) and compared to other geostatistical methods (Krivoruchko & Gribov, 2019) (Gupta, et al., 2017), and it has also been used in various research works (Li, et al., 2020) (Kutuzov, et al., 2019) (Nocco, et al., 2019) (Nogueira, et al., 2019) (Samsonova, et al., 2017) (Fabijańczyk, et al., 2017).

EBK differs from other kriging methods by considering the uncertainty provided by the estimation of the semivariogram model. While other methods estimate a single semivariogram model from the sampled locations and use it to predict the values in unsampled locations, assuming that this particular semivariogram is true for the interpolation region, EBK simulates several semivariogram models for each subset. The values in unsampled locations are then predicted based on a semivariogram spectrum generated by selecting individual semivariograms from the semivariogram distributions in the locations' neighborhood (Esri, 2022g).

This process can be summarized in four phases:

- Estimate a semivariogram based on sampled locations (Esri, 2022g);
- Use this semivariogram to simulate new values at sampled locations;
- Estimate a new semivariogram based on simulated values at sampled locations. This semivariogram gets a weight depending on how likely the observed values can be generated from it;
- Repeat phases two and three a specified number of times. In each iteration, the semivariogram estimated in phase one is used to simulate new values at sampled locations, and these simulated values are used to estimate a new semivariogram and its weight.

OVs in the water column have x-y-z coordinates. However, 3D interpolation methods are still being developed and evaluated (Sahlin, et al., 2014). For a long time, 3D interpolation of OVs was accomplished by stacking two-dimensional (2D) horizontal surfaces interpolated for multiple depth levels (Lauvset, et al., 2016) (Troupin, et al., 2012).

EBK for 3D interpolation (EBK3D) uses a similar process to EBK, however some configurations had to be adapted, for instance: the search neighborhood is designed in 3D (sphere, tetrahedron, cube); the difference between the horizontal and vertical changes in data values is accounted by two new parameters (elevation inflation factor (EIF) and trend removal) (Gribov & Krivoruchko, 2020). EBK3D has already been applied to OVs, namely, DO (Milbrandt, et al.,

2021), temperature and salinity (Nobre, et al., 2020), temperature and nitrate (Hennes, 2020), and in other domains (Cuesta, 2020).

b. Voxel layers

The voxel concept was first used in 1976 (Merriam-Webster, 2022b). It is defined as the 3D equivalent of a pixel and it stands for "volume (vo) element (el)" (Encyclopedia, 2016). A voxel represents a value (a single sample or data point) on a regularly spaced 3D grid (WikiGIS, 2011). Voxels are used in 3D object modelling, based on regularly sampled points instead of surfaces, and benefit from their capability of dealing with highly detailed objects as well as objects with poorly defined boundaries (Intel, 2011). They have been used in video games (Hughes, et al., 2014), digital twins (Voxelmaps, 2021) and simulating reality, namely in the visualization and analysis of medical (Ashburner & Friston, 2009) and scientific data, for instance, GIS visibility modelling (Chmielewski & Tompalski, 2017).

2.4. Mapping methods

Ecosystem mapping has evolved from individual expert opinion to unbiased data-driven methods, which are usually divided into two distinct approaches: unsupervised and supervised (Delua, 2021).

A supervised method, which can include, for instance, classification and regression techniques, uses a trained or labelled dataset (Soni, 2018). In the ecosystem mapping context, this means that we would need datasets with delimited values of OVs a priori associated with an ecosystem designation (label or desired output). That is not the case in this study, where we try to derive insights from the data and infer its natural structure, namely the different ecosystem locations and characteristics, without using predetermined categories. Nevertheless, there are some studies that used supervised classification methods in similar tasks, e.g., support vector machine (SVM) to map coastal marine ecosystems through satellite imagery (Pottier, et al., 2021), SVM and random forest (RF) to map aquatic ecosystems through remote sensing images (Martínez-Santos, et al., 2021), and RF and object based analysis (OBA) to map geomorphic and benthic zones of coral reefs (Roelfsema, et al., 2020). Regarding the specific mapping of marine benthic habitats there is also a study that evaluated supervised and unsupervised methods (Calvert, et al., 2014).

Unsupervised methods include clustering, association, and dimensionality reduction (Delua, 2021). Clustering involves dividing data points into a determined number of groups, such that data points inside each group are more similar among themselves than compared to data points

in other groups (Soni, 2018). As the globally accepted marine ecosystem classifications (e.g., MEOW, GOODS) do not provide framed values of OVs for each ecosystem, clustering has been the most widely used method for marine ecosystem mapping (Zhao, et al., 2020) (Reygondeau, et al., 2017a) (Sayre, et al., 2017b).

There are several clustering methods (Harmouch, 2021): hierarchical (Reijden, et al., 2021), partitioning, e.g., K-means (Sayre, et al., 2017b), and artificial neural networks (ANN), e.g., Self-Organizing Map (SOM) (Li, et al., 2018a), each with specific advantages and drawbacks . Regarding marine ecosystem mapping, K-means has been the most widely used method (Zhao, et al., 2020) (Sayre, et al., 2017b). It normally requires three user inputs:

- Number of clusters. This is the most critical parameter and there are several criteria to assess the optimal k value, e.g., gap statistic (Tibshirani, et al., 2001), pseudo-F statistic (Sayre, et al., 2017b), silhouette score (Rousseeuw, 1987). Usually, these statistics are compared for multiple k values, also considering their meaning for the analysis domain (Jain, 2010);
- Cluster initialization. The algorithm is sensitive to the initial cluster centroids, as different seeds can lead to dissimilar final clusters (Jain, 2010). This can be sorted out by calling upon distinctive initialization methods (Celebi, et al., 2013);
- Distance metric. K-means normally uses Euclidean distance, resulting on spherical clusters. Depending on the characteristics of the input data, other distance metrics have been proposed, with the expense of higher complexity and computational cost (Jain, 2010).

Besides these critical configurations and some disadvantages (e.g., sensitivity to outliers, inadequacy to deal with non-convex or different sized clusters) (Bação, 2020b), K-means has been widely used due to its simplicity, versatility, data handling capability and reliable results (Celebi, et al., 2013).

3. DATA AND STUDY AREA

This chapter describes the data used in this work, highlighting its geographic extent, provenance, sampling scheme and structure.

Data was acquired for the project AQUIMAR, which aims to identify suitable locations for aquaculture, from the areas defined in the MSP scheme (DGRM, 2019), based on physical, chemical, and biological characteristics.

AQUIMAR included several scientific campaigns with in-situ data collection. This study focuses on two areas of the Portuguese coast, illustrated in Figure 11: B, on the northwestern coast, approximately between Aveiro and Figueira da Foz, and E, on the south coast, approximately between Sagres and Tavira. These regions were chosen for two key reasons: (i) they are the ones, from project AQUIMAR, with the largest number of samples, which can be an advantage for the subsequent interpolation; and (ii) they represent two distinct oceanographic areas (Leitão, et al., 2019), with distinctive water masses (Bettencourt, et al., 2004), which can enhance the analysis and the results.

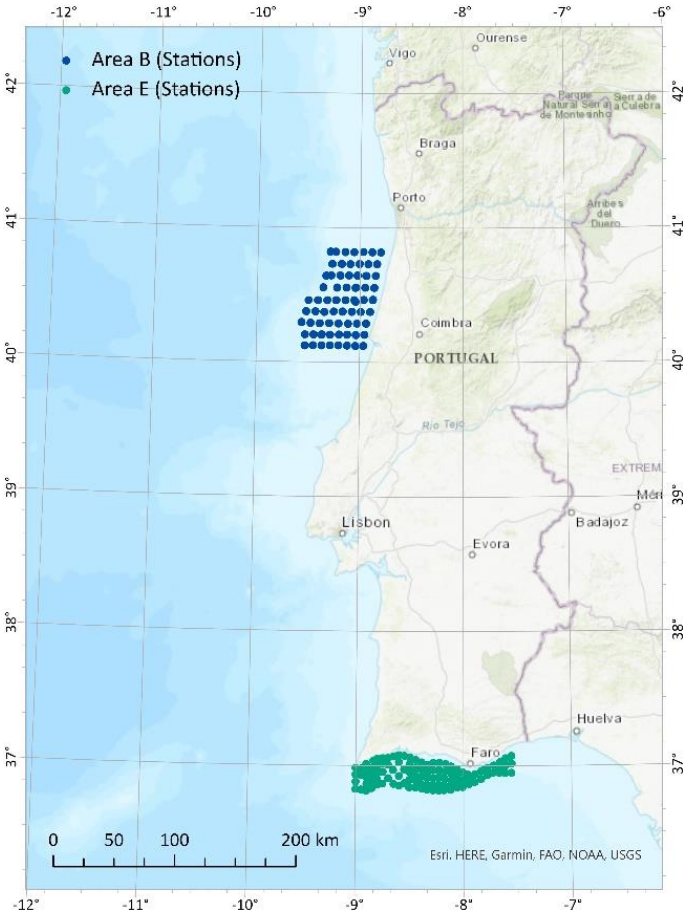


Figure 11. Location of the study areas.

The four specific campaigns conducted in the study areas occurred between October 2018 and March 2020 on board two hydrographic vessels of the Portuguese Navy, NRP *D. Carlos I*[8] and NRP *Almirante Gago Coutinho*[9]. Data was collected using a rosette equipped with a multi-parameter probe (*Ocean Seven 320 Plus WOCE-CTD*[10], featuring pressure, temperature, and conductivity sensors (CTD[11]), and 11 Niskin water sampling bottles of eight liters[12]. Positioning was provided by Global Navigation Satellite System (GNSS) receivers and processed in PT-TM06/ETR89 (EPSG:3763)[13] reference system.

In each CTD station (same horizontal position), temperature was measured directly by the multi-parameter probe and salinity was calculated using TEOS-10 (IOC; SCOR; IAPSO, 2010). Both values were processed and validated from 5 m depth to the seafloor, at 1 m intervals, for each profile.

In each water station, samples for analysis of DO, nutrients (phosphate (PO_4), TON (NO_X) (the sum of nitrate and nitrite) (Capone, et al., 2008), and silicate (SiO_2)), chlorophyll a and pH were collected approximately at the following depths: 5 m, 25 m, 50 m, 75 m, 100 m, 150 m, 200 m, and 5 m above the seabed, depending on the station's depth range (IH, 2020). The analyses were performed at the chemistry laboratories of the Hydrographic Institute (IH), which are accredited according to the ISO/IEC 17025 standard (ISO, 2017) for the analysis of most of these parameters in saline waters. DO and pH were determined, respectively, according to methods 4500-O G and 4500-H+ B (Baird & Bridgewater, 2017). Chlorophyll a was determined according to a modification of Lorenzen's method (Jeffrey & Humphrey, 1975), and TON was determined according to (Strickland & Parsons, 1972). PO_4 and SiO_2 were determined according to (Murphy & Riley, 1962), by procedures adapted to segmented flow analysis.

Table 3 describes the measurement uncertainty (analytical component) of each OV (1-sigma level) and Figure 12 illustrates the sampling scheme of the stations, along the depth profiles.

[8] https://www.marinha.pt/pt/os_meios/hidrograficos/Paginas/nrp-d-carlosI.aspx [accessed 12 September 2022].
[9] https://www.marinha.pt/pt/os_meios/hidrograficos/Paginas/nrp-almirante-gago-coutinho.aspx [accessed 12 September 2022].
[10] https://www.idronaut.it/multiparameter-ctds/oceanographic-ctds/os320plus-oceanographic-ctd [accessed 12 September 2022].
[11] Conductivity, temperature, and depth.
[12] https://www.generaloceanics.com/model-1010-niskin-water-sampler-8l.html [accessed 12 September 2022].
[13] https://www.dgterritorio.gov.pt/geodesia/sistemas-referencia/portugal-continental/PT-TM06-ETRS89 [accessed 12 September 2022].

OV	Range	Uncertainty
Temperature	0 - 35.0 °C	0.0017 %
Salinity	2.0 - 42.0 PSU	0.00425 PSU
Dissolved Oxygen	0.12 - 2.0 mgO_2/L	10.95 %
	2.0 - 15.0 mgO_2/L	7.72 %
Phosphate (PO_4)	0.2 - 0.4 µmol/L	9.05 %
	0.4 - 50 µmol/L	4.19 %
Total Oxidized Nitrogen (NO_X)	1.0 - 2.0 µmol/L	9.84 %
	2.0 - 100 µmol/L	2.88 %
Silicate (SiO_2)	0.3 - 0.6 µmol/L	12.02 %
	0.6 - 30.0 µmol/L	4.93 %
Chlorophyll a	0.10 - 25.0 mg/m^3	5 %
pH	2.0 - 10.0	0.045

Table 3. Measurement uncertainty of the OVs.

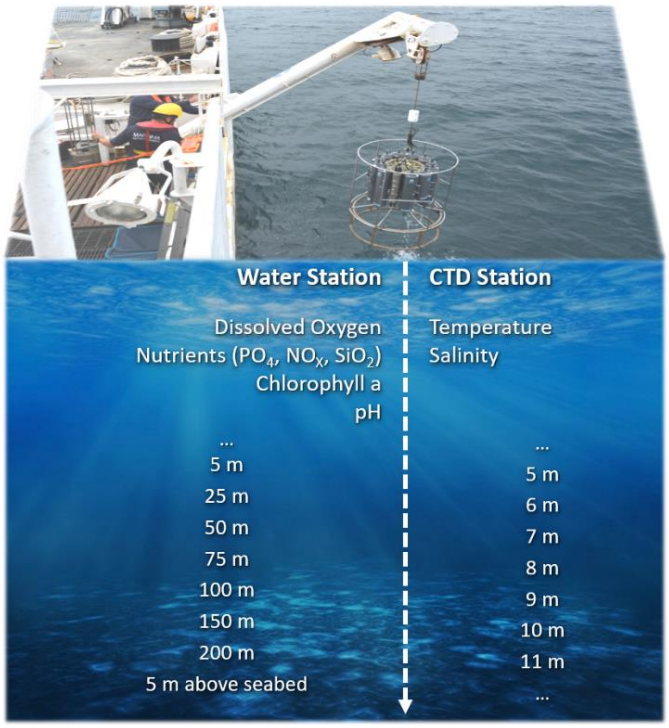


Figure 12. Sampling scheme along the depth profiles.

Table 4 provides a summary of the campaigns, which are further detailed in Annex A. For abbreviation purposes, data from the specific campaigns and areas is referenced as follows: campaign 1, area B (C1B); campaign 1, area E (C1E); campaign 2, area B (C2B), etc. The number of stations and samples in each area varied across the four campaigns. For instance, in C1B, data was collected in 65 stations, for a total of 5 956 sample points, while in C2B, data was collected in 59 stations, for a total of 4 835 sample points. As explained by the sample scheme and detailed in Table 4, there are more CTD samples than water samples, and CTD stations are also more numerous than water stations, especially in region E. Therefore, temperature and salinity have a higher sampling density than the other OVs (DO, nutrients, chlorophyll a, and pH).

For the interpretation of the results, we considered C1 and C3 (early autumn) and C2 and C4 (late winter/early spring) seasonal pairs. Although data collection occurred in analogous seasons, the environmental and oceanographic conditions on both years were not the same.

Campaign	Dates	Vessel	Area	Ref	CTD		Water	
					Stations	Samples	Stations	Samples
1st	2018-10-05	NRP *Alm. Gago Coutinho*	B	C1B	65	5 956	63	264
IHPT-AQUIMAR2018-2	2018-10-27		E	C1E	108	10 704	55	223
2nd	2019-04-16	NRP *D. Carlos I*	B	C2B	59	4 835	59	237
IHPT-AQUIMAR2019-2	2019-05-13		E	C2E	104	9 890	54	218
3rd	2019-10-07	NRP *Alm. Gago Coutinho*	B	C3B	66	5 991	64	265
IHPT-AQUIMAR2019-4	2019-10-31		E	C3E	114	10 158	60	238
4th	2020-02-24	NRP *Alm. Gago Coutinho*	B	C4B	57	4 584	57	222
IHPT-AQUIMAR2020-1	2020-03-20		E	C4E	114	10 407	60	241

Table 4. AQUIMAR scientific campaigns considered in this study.

4. METHODS

This chapter describes the methodology adopted in this study, which was divided into seven stages, illustrated in Figure 13.

Data preparation consisted of its organization and conversion to adequate formats, and the pre-processing of missing, incorrect, or inconsistent values. ESDA included data characterization through descriptive and spatial statistics. 3D interpolation was achieved with EBK3D method. Grid definition involved the determination of the point mesh, with specific horizontal and vertical resolutions as well as 3D boundaries, where the OVs values would be predicted. Cluster analysis was performed using K-means method and only considering seven OVs, as chlorophyll a was excluded. Voxel layers were created based on multidimensional Network Common Data Form (NetCDF) (Unidata, 2022b) files. The web application was built using the ArcGIS Online infrastructure and a StoryMap[14] layout.

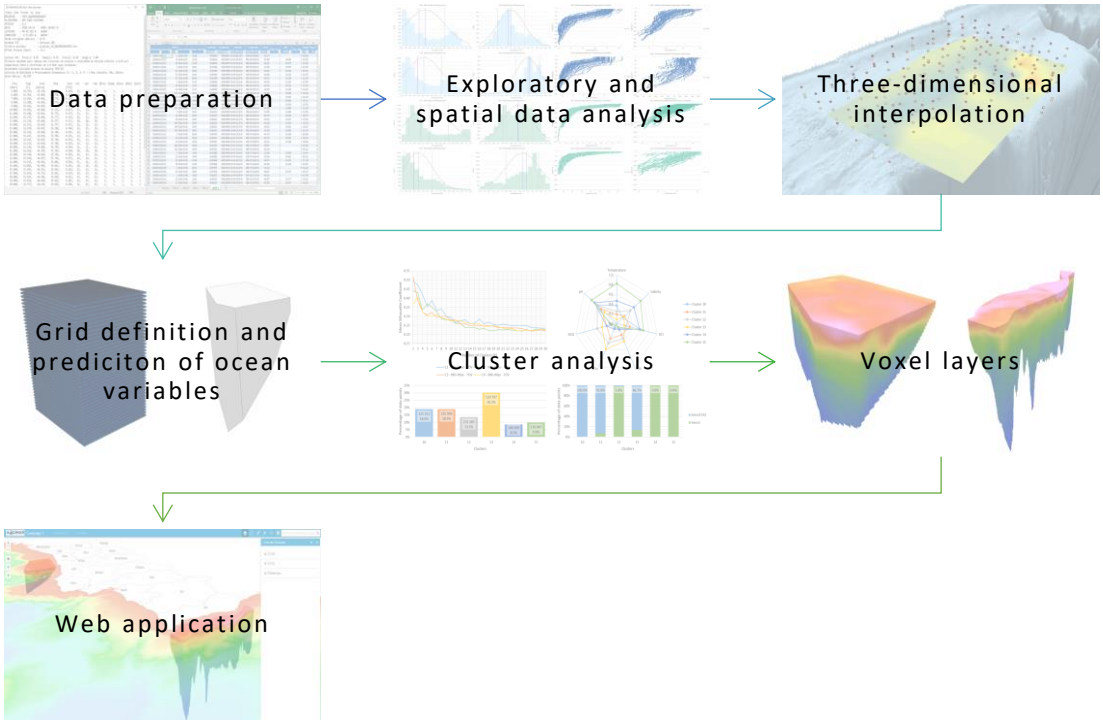


Figure 13. Methodology stages.

4.1. Data preparation

During data preparation, OVs values flagged with quality control codes “3” (probably bad value) and “4” (bad value) (SeaDataNet, 2022a) were deleted, as they were not substantial. OVs values flagged with quality control code “Q” (value bellow the limit of quantification (LOQ))

[14] https://storymaps.arcgis.com [accessed 12 September 2022].

(SeaDataNet, 2022a) were changed to half the LOQ value (Hecht, et al., 2018). As they were frequent in some OVs (e.g., nutrients), this improved the estimation of the OVs values (in comparison with deleting them or changing them to 0) but decreased the accuracy of the estimation of its true variance (Beal, 2001).

Data was organized by campaign and study area in a tabular format and imported to ArcGIS Pro 3.0.1[15].

4.2. Exploratory and spatial data analysis

ESDA is a key stage of the geostatistical process, as it reveals significant data properties, e.g., data distribution, outliers, relationships between variables and spatial patterns (Anselin, 2005). ESDA was conducted in ArcGIS. Each OV was characterized in 3D using descriptive statistics (location, dispersion, symmetry, kurtosis, and association) and charts (histograms and scatter plots).

4.3. Three-dimensional interpolation

The 3D interpolation of each OV, for all the campaigns and study areas, was achieved with the EBK3D method.

EBK3D estimations were assessed using the cross-validation method (Li, et al., 2020) (Yan, et al., 2019) (Esri, 2022d), which removes each observation, one by one, and predicts the associated data value with the remaining observations. It then compares the predicted to the observed values using the following statistics (Esri, 2022a):

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$$

Equation 1

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$

Equation 2

$$\mathrm{SME} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)/\sigma(y_i)$$

Equation 3

[15] https://www.esri.com/en-us/arcgis/products/arcgis-pro [accessed 12 September 2022].

$$\mathrm{RMSSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(y_i - x_i)/\sigma(y_i)]^2}$$

Equation 4

$$\mathrm{ASE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma^2(y_i)}$$

Equation 5

In these equations: ME is the mean error; RMSE is the root mean square error; SME is the mean standardized error; RMSSE is the root mean square standardized error; ASE is the average standard error; x_i is the observed value, y_i is the predicted value and σ is the standard deviation. Additionally, we also computed the ratio of the variance (RVar), to assess EBK3D ability to preserve the variance (Haberlandt, 2007).

$$\mathrm{RVar} = \frac{\mathrm{Var}(y_i)}{\mathrm{Var}(x_i)}$$

Equation 6

EBK produces two additional statistics (Esri, 2022a):

- Percent in 95 % interval: the percentage of points that are in a 95% cross-validation confidence interval;
- Average continuous ranked probability score (CRPS): the quadratic measure of the difference between the predictive cumulative distribution function (CDF) and the empirical CDF of each observation (Gneiting & Raftery, 2007) . While other statistics only compare the data to single-point predictions, this diagnostic compares it to a full distribution.

Table 5 describes the ideal values for EBK3D statistics (Esri, 2022a) (Johnston, et al., 2001) and Table 6 mentions some of the conclusions that can be drawn from the results, concerning the variability of the predictions (Li, et al., 2020) (Krivoruchko, 2011).

Statistic	Ideal value
Mean Error (ME)	Close to 0
Root Mean Square Error (RMSE)	As small as possible
Mean Standardized Error (SME)	Close to 0
Root Mean Square Standardized Error (RMSSE)	Close to 1
Average Standard Error (ASE)	Close to RMSE
Ratio of the variance	Close to 1
Percent in 95% interval	Close to 95
Average CRPS	As small as possible

Table 5. EBK3D statistics ideal values.

Result	Variability of the predictions
ASE $\approx$ RMSE and RMSSE ≈ 1	Correctly assessed
ASE $>$ RMSE and RMSSE < 1	Overestimated
ASE $<$ RMSE and RMSSE > 1	Underestimated

Table 6. EBK3D conditions to assess the variability of the predictions.

The configuration of EBK3D advanced model parameters, described in Table 7, was based on: (i) the manual testing of the available options; (ii) the corresponding use case description (Esri, 2022g); (iii) the cross-validation results; and (iv) the visual aspect of the interpolated surface. To select the appropriate configuration, we used several OVs with different sample schemes. As it could easily turn into a never-ending task, for the visualization assessment we opted for horizontal surfaces, at 25 m depth[16], in both areas.

	Default Model	Adjusted Model
Subset size	100	100
Overlap factor	1	1
Number of simulations	100	100
Transformation type	None	None
Semivariogram type	Power	Exponential
Order of trend removal	None	First[17]

Table 7. EBK3D advanced model parameters.

Subset size, overlap factor, and number of simulations are specific parameters of EBK and they refer to the configuration of the simulations: (i) the subset size specifies the maximum number of points (samples) in each subset; (ii) the overlap factor describes the degree of overlap between subsets, increasing the smoothness of the resulting surface; and (iii) the number of simulations stipulates the number of semivariograms that will be simulated for each subset. An increase in each of these parameters raises the processing time. In this sense, we tested different values and maintained the default ones since the improvements were negligible, considering the delay in processing time.

The transformation type refers to the transformation to which the input data is subjected (e.g., empirical, log empirical). Although the log empirical transformation could be useful because it requires all samples to be positive and it assures that all predictions will be positive (which is true for the selected OVs), this option is not available when using a measurement error value. For this reason, we selected no transformation.

The semivariogram type specifies the semivariogram modelling function. The K-Bessel is generally described as more flexible and accurate (Esri, 2022g). However, when compared with the Exponential function, K-Bessel produced more geometric artifacts in the resulting surface,

[16] The 25 m depth is the maximum depth in area B, for which all the stations (CTD and water) have samples.
[17] Except for DO and pH, for which we selected "None".

as exemplified in Annex C, Figure C-1. These artifacts are not noticeable in the geostatistical layer that is created after the interpolation, which hampers its evaluation. They only become visible after converting the geostatistical layer into a raster surface and changing the symbology from a discrete classification to a stretch scheme, as illustrated in Figure C-2. On the other hand, K-Bessel had better cross-validation results, as presented in Table C-1, although the magnitude of this improvement was not significant except for the RMSSE statistic.

In this sense, we selected the Exponential modelling function, based on several criteria. It created horizontal surfaces at 25 m depth with a superior visual aspect and more realistic behavior. Some authors argue that variography is a subjective process that seeks to adjust a realistic model and not a goal by itself (Goovaerts, 1997). Therefore, the choice of a specific modelling function should take into account not only the fit to the experimental semivariogram (Larrondo, et al., 2003), but also the accuracy of the predictions (Li, et al., 2018b) and the specific knowledge of the physical phenomenon (Costa, et al., 2008). Accuracy is usually assessed through cross-validation (Goovaerts, 1997). Although this method is used to evaluate the overall EBK3D performance, it only compares the predictions to the original samples, giving no feedback on unsampled locations, meaning that, for instance, it may be occurring model overfitting with K-Bessel. Nevertheless, the cross-validation results with K-Bessel only showed a substantial improvement regarding one statistic (RMSSE). Lastly, (Goovaerts, 1997) has highlighted several limitations to the use of cross-validation to select the semivariogram modelling function.

The order of trend removal is a specific parameter of EBK3D, and it is used to remove linear trends (first order) in the vertical direction. As indicated in the literature review and presented in ESDA, all OVs showed a close relationship with depth, as their values became larger or smaller as depth increased or decreased. For DO and pH, we noticed that selecting “none” instead of “first” in this parameter would improve the visual aspect of the horizontal surfaces at 25 m depth, as illustrated in Figure C-3. This could be explained by the fact that these specific OVs showed in ESDA a lower correlation with depth.

Using the configuration described in Table 7, we also conducted a manual adjustment of the search neighborhood parameters (minimum and maximum number of neighbors, sector type, and search radius) and the EIF.

The search radius was set to 100 000 m, a value that exceeded the study area so that the neighborhood parameters were only controlled by the number of neighbors and the sector type. The minimum and maximum number of neighbors concern each sector of the neighborhood, so an adjustment that decreases the number of sectors, e.g., from 12 (dodecahedron) to 4

(tetrahedron), can support an increase in the number of neighbors. The number of neighbors considered affects the processing time and, generally, 10 to 20 neighbors are enough for accurate and stable predictions (Esri, 2022e). The minimum number of neighbors was set to 1 and the maximum was mostly set to 4, with some exceptions where we needed to smooth the surface and it was set up to 15. We evaluated several sector types and, globally, 8 (octahedron) generated the best visual aspect, as illustrated in Figure C-4.

EIF accounts for different changing scales in the horizontal and vertical directions. Generally, OV values change faster vertically than horizontally. EIF stretches the locations of the points to statistically match the horizontal and vertical distances, which is crucial for an accurate estimation of the semivariogram model and an adequate three-dimensional neighbor selection. In this sense, EIF compensates fast changing values vertically, while the order of trend removal corrects consistent changes in OVs values as depth rises or drops, regardless of how swift that change is (Esri, 2022e). As performed for order of trend removal, this parameter was only manually adjusted for the OVs DO and pH.

4.4. Grid definition and prediction of ocean variables

3D interpolation with EBK3D in ArcGIS produces a geostatistical layer with a cubic extent limited by the minimum spatial bounding box of the input samples[18]. This extent is not adequate for the scope of this work, as the study areas are marine regions with specific boundary conditions like the bathymetry or the coastline. In this sense, we adopted the methodology illustrated in Figure 14 and detailed below:

- Create a generic 3D cubic grid (fishnet), enclosing all the input samples for each study area (2 areas), with a horizontal resolution of 250 m and a depth resolution of 5 m;
- Clip the fishnet prediction points that fall below the bathymetry or beyond (coastwards) the coastline. In this task we used the EMODnet bathymetric model, with, approximately, 115 m horizontal resolution (EMODnet, 2021);
- Compute the minimum bounding volume[19] (MBV). Depending on the distribution of the OVs samples, we calculated three MBV for each study area (2 areas) and for each campaign (4 campaigns): (i) for temperature and salinity; (ii) for DO, nutrients and pH; and (iii) for chlorophyll a;

[18] The minimum spatial bounding box is the smallest cubic box that can be generated to encompass all the input sample points.

[19] The minimum bounding volume is the smallest volume of space occupied by each dataset. In this computation we used the convex hull method, which provided an adequate level of detail.

- Extract the fishnet points inside each MBV (24 datasets);
- Predict the OV values at the extracted fishnet locations (64 datasets).

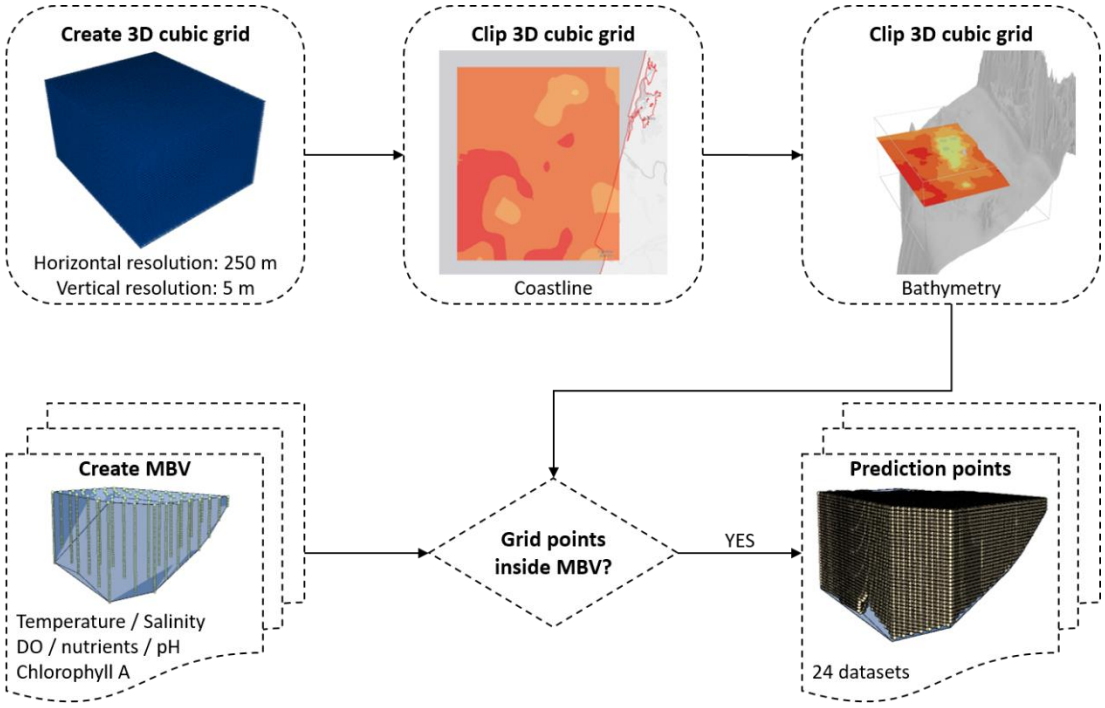


Figure 14. Methodology for extracting OVs prediction values.

For the selection of the fishnet resolution and the subsequent voxel size, we considered two factors: (i) the uncertainty of the sample values and (ii) the computational expense.

The uncertainty of the observed values can be divided into four components: (i) the uncertainty in the measurement of the OVs, resulting from the method and described in Table 3; (ii) the uncertainty in the horizontal positioning, resulting from the equipment (GNSS) measurement error and the horizontal oscillation of the rosette; (iii) the uncertainty in the vertical positioning, resulting from the equipment (pressure sensor) measurement error and the vertical oscillation of the rosette; and the temporal uncertainty, resulting from the fact that the samples were not collected simultaneously in every stations, across the same campaign (for instance, in C1B, data collection took approximately two days).

An exhaustive uncertainty assessment for estimating OVs values, as done, for instance, in bathymetry estimation (Calder & Wells, 2007), submarine landslides identification (Gamboa, et al., 2021) or marine geological classification (Kint, et al., 2020), was beyond the scope of this work. The uncertainty in OVs measurements was included in the interpolation method. The magnitude of the other three components was considered in the selection of the fishnet resolution, namely by not having a voxel size smaller than the 3D radius of the uncertainty of each OV (Hengl, 2006).

Nevertheless, fishnet resolution was mainly restricted by computational processing limitations[20], considering the available time for the project development and the requirements of the key users (Mulder, et al., 2019). Based on the average horizontal and vertical distances between sample points, detailed in Annex A, Table A-1, and Figure 12, we tested three horizontal resolutions (100 x 100 m, 250 x 250 m, and 500 x 500 m) and one vertical resolution (5 m). Considering the number of points of the initial fishnet, we opted for the middle solution: horizontal resolution of 250 m and depth resolution of 5 m. Table 8 describes the number of prediction points that emerged after each extraction with the chosen resolution.

Fishnet	Number of prediction points	
	Area B	Area E
Fishnet (250 x 250 x 5 m) – Initial	3 303 280	4 880 480
Fishnet (250 x 250 x 5 m) – Coastline clip	3 179 560	4 503 200
Fishnet (250 x 250 x 5 m) – Bathymetry clip	1 728 577	1 446 733

Table 8. Number of prediction points for different fishnets.

The descriptive statistics of the OVs predictions at the fishnet points were compared to the statistics of the input samples to assess the reliability of EBK3D. Globally, the predictions had similar statistics, except for NO_X, for which some of the predicted values in some of the campaigns (C2B, C1E, and C4E) had negative values. As mentioned before, this could be avoided with the use of the log empirical transformation, with the drawback of not allowing the input of a measurement error value. In this sense, the problem was solved by changing the negative values to 0 in post-processing.

OVs predictions at the specified locations were then integrated, depending on the succeeding tasks:

- For cluster analysis, they were joined (intersected) by campaign;
- For voxel generation, they were joined (united) by study area and by campaign.

4.5. Cluster analysis

Cluster analysis was performed in KNIME Analytics Platform 4.3.3[21], using the K-means method. We used seven OVs (temperature, salinity, DO, nutrients, and pH) and excluded chlorophyll a, since it has a restricted depth distribution (approximately, no data values below 67 m), which would excessively limit the extent of the analysis. However, this specific OV could be further included in more circumscribed studies.

[20] We used a desktop computer with an Intel Core i7-10700 processor and 32 GB RAM.
[21] https://www.knime.com/knime-analytics-platform [accessed 12 September 2022].

K-means (Lloyd, 1982) (MacQueen, 1967) (Hartigan & Wong, 1979) is one of the most popular clustering methods. Generally, the algorithm has four stages: (i) select a pre-determined number (k) of initial centroids (μ_k); (ii) for each data point, find the closest centroid (μ_k) and assign it with its class (cluster) (C_k); (iii) recompute the centroids of each cluster based on the mean of all the assigned data points; and (iv) repeat step (ii) until there are no changes in the seed (μ_k) location (convergence). In this sense, K-means requires three main inputs: distance metric, number of clusters, and cluster initialization (centroids).

As we used Euclidean distance, before any analysis we normalized all OVs values. We tested and compared two different methods: Min-Max (Equation 7) and Z-Score (Equation 8).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Equation 7

In Equation 7, x' is the normalized value, x is the original value, min(x) is the minimum distribution value and max(x) is the maximum distribution value, for a 0 to 1 scale.

$$x' = \frac{x - \mu}{\sigma}$$

Equation 8

In Equation 8, x' is the normalized value, x is the original value, μ is the distribution mean, and σ is the distribution standard deviation.

K-means was applied to each campaign, analyzing both study areas simultaneously. Table 9 describes the number of data points used in each cluster analysis.

Campaign	Number of data points		
	Area B	Area E	Total
C1	1 083 638	637 260	1 720 898
C2	914 630	564 755	1 479 385
C3	1 120 331	596 901	1 717 232
C4	977 684	604 556	1 582 240

Table 9. Cluster analysis data points.

We used the silhouette coefficient (SC) (Rousseeuw, 1987) to evaluate the optimal number of clusters (k value). This score measures how similar an object (data point) is to its own cluster compared to other clusters, and it can range from -1.0 to 1.0, where higher values indicate that the object is well matched to its own cluster. A value of 0 specifies that the object is on the decision border between two adjacent clusters, and a negative value points out that the object might have been allocated to the wrong cluster (Pedregosa, et al., 2011). In Equation 9, a(i) is the mean intra-cluster distance (average dissimilarity) from i to all other objects in the same cluster, and b(i) is the mean inter-cluster distance from i to the closest cluster.

$$SC = \frac{b(i) - a(i)}{\max\{a(i),\ b(i)\}}$$

Equation 9

This score was computed from k=2 to k=30. As it is computationally expensive, we used a stratified (by cluster) sampling of 1% of the original dataset, to assess the optimal number of clusters. After the selection of the optimal k for each campaign, we recomputed the SC for the entire dataset to assess the reliability of the preliminary estimation.

For optimal k value validation purpose, we also computed the pseudo-F statistic (Calinski & Harabasz, 1974) in ArcGIS. The pseudo-F statistic determines the ratio of between-cluster (inter-cluster) variance to within-cluster (intra-cluster) variance. Therefore, high values of this statistic indicate a greater cluster separation. In Equation 10, N is the number of data points, K is the number of clusters, GSS is the between-cluster sum of squares and WSS is the within-cluster sum of squares.

$$pseudoF = \frac{(GSS)\ /\ (K-1)}{(WSS)\ /\ (N-K)}$$

Equation 10

This score was also computed from k=2 to k=30.

Apart the optimal k values provided by these statistics, the choice of the adequate number of clusters also considered the opinion of key users on the purpose of the work and its value for further studies.

For cluster initialization, we used random seed location and ran the tool several times to assess the consistency of the results.

4.6. Voxel layers

The gridded points and the corresponding OVs predictions and cluster labels were merged into eight comma-separated values (CSV) files, for each campaign and area. Each file contains, for each data point, the following information: horizontal position, depth, cluster, OVs predictions, and OVs standard errors. For CSV file handling, we developed specific Python scripts using the Pandas[22] module.

In order to create a volumetric grid, data had to be exported to an adequate multidimensional format (Esri, 2022b), e.g., NetCDF. For the conversion from CSV to NetCDF, we adapted a Python script, using several modules (e.g., netCDF4[23], NumPy[24], Csv[25], etc.).

[22] https://pandas.pydata.org [accessed 12 September 2022].
[23] https://unidata.github.io/netcdf4-python [accessed 12 September 2022].
[24] https://numpy.org [accessed 12 September 2022].
[25] https://docs.python.org/3/library/csv.html [accessed 12 September 2022].

The NetCDF files were created according to the Climate and Forecast (CF) Metadata Convention (Unidata, 2022a) (Eaton, et al., 2021) and they were opened and visualized in ArcGIS Pro and QGIS 3.16.8[26].

4.7. Web application

The web application was built using the ArcGIS Online infrastructure and a StoryMap layout. For 3D voxel layer exploration, we created four web mapping applications, one for each campaign, with the ArcGIS Web AppBuilder[27]. We additionally published the NetCDF files on a ncWMS server. ncWMS is a web map service (WMS) (OGC, 2022) that allows environmental data visualization and exploration in a browser through a specific interface named Godiva3 (University of Reading, 2010).

[26] https://www.qgis.org [accessed 12 September 2022].
[27] https://developers.arcgis.com/web-appbuilder [accessed 12 September 2022].

5. RESULTS AND DISCUSSION

This chapter presents and discusses the results achieved in this study, divided into five major sections: ESDA, 3D surfaces, EMU identification and characterization, voxel layers, and web application.

5.1. Exploratory and spatial data analysis

The results of the ESDA are presented in Annex B and they are structured into descriptive statistics (location, dispersion, symmetry, and kurtosis), histograms, and scatter plots (OVs versus depth), for each OV, in both areas, across the four campaigns.

5.1.1. Temperature

Temperature is described in Table B-1, Figure B-1, and Figure B-9.

The temperature mean and the standard deviation were greater in the early autumn campaigns (C1 and C3). C1 and C3 also had an upper coefficient of variation, indicating higher variability and more heterogeneous temperature values. Temperature data was positively skewed, except for C4B, C2E and C4E, showing a longer tail on the right. Kurtosis was lower than 3, except for C4B and C3E, indicating that the distribution had lighter tails and fewer extreme values than the equivalent normal distribution (Esri, 2022c). Temperature decreased with depth, revealing relatively high coefficient of determination (R^2) values ($0.42 \leq R^2 \leq 0.73$).

5.1.2. Salinity

Salinity is described in Table B-2, Figure B-2, and Figure B-10.

The salinity mean and the standard variation were stable and roughly equal across the four campaigns, while the coefficient of variation was extremely low ($\leq$ 0.7%). Salinity data was negatively skewed, except for C1E and C3E, showing a longer tail on the left. In area B, skewness revealed a pattern depending on the season, with lower skewness in the early autumn campaigns (C1 and C3). Kurtosis was higher than 3, except for C1E, C2E, and C4E, indicating that distribution had heavier tails and more extreme values than the equivalent normal distribution. While in area E, salinity decreased with depth ($0.46 \leq R^2 \leq 0.67$), in area B, salinity increased with depth. In this area, the R^2 for the early autumn campaigns (C1 and C3) ($0.04 \leq R^2 \leq 0.11$) was significantly lower than the one for the late winter/early spring campaigns (C2 and C4) ($0.37 \leq R^2 \leq 0.43$).

5.1.3. Dissolved oxygen

DO is described in Table B-3, Figure B-3, and Figure B-11.

The DO mean was higher in the late winter/early spring campaigns (C2 and C4), and the standard deviation and the coefficient of variation were stable and approximately equal across the four campaigns. DO data had low skewness and was positively skewed, except for C2B, C3B and C4E. Kurtosis was lower than 3, except for C4B, C1E, C3E and C4E. DO decreased with depth ($0.11 \leq R^2 \leq 0.37$) and in area B the R^2 for the early autumn campaigns (C1 and C3) ($0.34 \leq R^2 \leq 0.37$) was higher than the one for the late winter/early spring campaigns (C2 and C4) ($0.25 \leq R^2 \leq 0.26$).

5.1.4. Nutrients

Nutrients are described in: Table B-4, Figure B-4, and Figure B-12 (PO_4); Table B-5, Figure B-5, and Figure B-13 (NO_X); and Table B-6, Figure B-6, and Figure B-14 (SiO_2).

The nutrients (PO_4, NO_X and SIO_2) mean and the standard deviation were relatively stable and roughly equal across the four campaigns. They revealed high coefficient of variation values ($\geq$ 46.1 %), indicating high variability and heterogeneous nutrient values. Nutrient data was positively skewed, accumulating high counts on the left tail of the histogram due to the significant number of observations below the LOQ that, accordingly, were changed to half the LOQ. Kurtosis varied for the three nutrients: PO_4 was lower than 3, except for C1E; NO_X was lower than 3 for all campaigns; and SiO_2 was lower than 3, except for C1B, C3B, and C1E. Nutrient concentration increased with depth, with different R^2 values: PO_4 ($0.37 \leq R^2 \leq 0.61$); NO_X ($0.44 \leq R^2 \leq 0.72$); and SiO_2 ($0.11 \leq R^2 \leq 0.48$).

Regarding PO_4, ESDA identified two outliers in the C2B dataset. As the two observations had been previously flagged with quality control code 1 (good value), these conclusions were re-evaluated and the values were re-flagged with quality control codes 3 (probably bad value) and 4 (bad value) (SeaDataNet, 2022a) and, consequently, excluded from the analysis.

5.1.5. Chlorophyll a

Chlorophyll a is described in Table B-7, Figure B-7, and Figure B-15.

The chlorophyll a mean, and the standard deviation were relatively stable across the four campaigns, and the coefficient of variation was considerably high ($\geq$ 44.9%). In area B the mean was higher in the late winter/early spring campaigns (C2 and C4). Chlorophyll a data was positively skewed, except for C4E, and kurtosis was bigger than 3, except for C4B and C4E. Generally, chlorophyll a decreased with depth, with low R^2 values ($0.01 \leq R^2 \leq 0.29$), until approximately 67 m (C1E), below which there were no data values.

5.1.6. pH

pH is described in Table B-8, Figure B-8, and Figure B-16.

The pH mean and the standard deviation were relatively stable across the four campaigns, while the coefficient of variation was considerably low ($\leq$ 1.1%). In area B the mean was higher in the late winter/early spring campaigns (C2 and C4). pH data was negatively skewed, and kurtosis was bigger than 3, except for C1B. Generally, pH decreased with depth, with low R^2 values ($0 \leq R^2 \leq 0.36$).

5.2. Three-dimensional surfaces

EBK3D properties and cross-validation results are specified by OV, area, and campaign in Annex C, Table C-2 to Table C-17.

EBK3D performed reliably as an automated interpolator (Krivoruchko, 2012), as we only made small modifications to the initial configuration in 35.9% (23 of 64) of the interpolated surfaces. Once the configuration of the advanced model parameters was completed, the manual changes of the search neighborhood parameters and the EIF for each OV, area, and campaign were based more on the visual aspect of the surface (slice at 25 m depth) than on the improvement of the cross-validation results.

These manual modifications revealed some facts that could guide further specific studies, for instance: (i) sector type selection had a greater impact on the visual aspect of the surfaces built from the OVs with more samples (temperature and salinity); (ii) for the same OV, the optimal adjustment was not duplicated for different areas and campaigns, probably due to the dissimilar sampling schemes, which made it impracticable to program EBK3D configuration by OV; (iii) EIF values, with the first order of trend removal selected, were considerably lower than without it, as expected (Esri, 2022e); and (iv) DO and pH produced more geometric artifacts than the other OVs, which were reduced by increasing the number of neighbors.

Considering the principles described in Table 5 and Table 6, globally, EBK3D had a good but varying performance for each surface at the locations of the observations (Table 10). The ME and the SME were both close to 0, showing that the predictions were unbiased, and, with a few exceptions, the RMSE was low, indicating that the predictions were precise. For a few OVs (PO_4 and SiO_2), the ASE was close to the RMSE and the RMSSE was close to 1, signifying that the variability of the predictions was correctly assessed. Generally, RVar was close to or less than 1, meaning that the predictions had similar or less variability than the observations, and the average CRPS was low, representing short differences between the CDF of the predictions and

the CDF of each observation. The percentage of points inside the 95% confidence interval of the cross-validation was close to 95%, suggesting a good model performance.

	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Temperature								
Average CRPS	0.0101	0.0075	0.0123	0.0045	0.0298	0.0060	0.0125	0.0052
Percent in 95% Interval	99.8321	99.0486	99.7162	98.9747	99.3180	98.9687	99.4389	99.3082
ME	2.4E-04	1.5E-04	3.4E-04	-2.3E-05	1.5E-04	4.4E-05	4.7E-04	6.6E-05
RMSE	0.0166	0.0146	0.0200	0.0086	0.0785	0.0109	0.0281	0.0092
RMSSE	0.3623	0.4877	0.3563	0.5128	0.5248	0.5065	0.4369	0.4384
Salinity								
Average CRPS	0.0030	0.0034	0.0042	0.0036	0.0074	0.0023	0.0043	0.0022
Percent in 95% Interval	97.9181	98.4695	97.6465	98.8438	97.9260	98.2811	98.1492	98.6836
ME	-5.5E-06	-7.2E-05	2.9E-05	-9.4E-05	1.8E-05	2.4E-05	3.8E-05	1.3E-05
RMSE	0.0059	0.0069	0.0102	0.0074	0.0187	0.0044	0.0108	0.0041
RMSSE	0.6530	0.6024	0.7134	0.5270	0.7077	0.6438	0.6503	0.5518
Dissolved Oxygen								
Average CRPS	0.2999	0.2709	0.2309	0.2256	0.3071	0.3023	0.2391	0.1937
Percent in 95% Interval	98.4848	97.4576	98.4906	100.0000	96.3470	96.7442	98.3051	99.1701
ME	-3.5E-02	-6.0E-02	-4.8E-02	-3.8E-02	-5.1E-02	-6.0E-02	-5.3E-02	-3.7E-02
RMSE	0.5412	0.4785	0.4047	0.3625	0.5495	0.5473	0.4241	0.3231
RMSSE	0.8992	0.7008	0.7261	0.5501	0.8438	0.7859	0.7416	0.6269
PO_4								
Average CRPS	0.0605	0.0423	0.0711	0.0426	0.0964	0.0687	0.0488	0.0480
Percent in 95% Interval	93.8931	94.7826	94.6970	95.0226	95.9459	91.2844	91.5966	93.7500
ME	-2.8E-03	-2.0E-03	-1.3E-03	-1.7E-03	-7.7E-03	4.7E-05	4.2E-04	-2.8E-03
RMSE	0.1143	0.0783	0.1363	0.0805	0.2068	0.1284	0.0916	0.0898
RMSSE	0.9766	0.9404	1.0345	0.9731	1.0445	1.0629	1.0037	1.0171
NO_X								
Average CRPS	0.9449	0.6409	1.0683	0.7090	0.9601	1.4413	0.7612	0.8322
Percent in 95% Interval	95.0570	94.3966	94.3182	95.9459	93.2432	94.0367	93.6975	94.5607
ME	-2.4E-02	-3.3E-02	9.0E-03	-3.7E-02	-7.7E-02	2.2E-02	2.6E-02	-2.6E-02
RMSE	1.8259	1.1709	2.0309	1.3622	1.9803	2.7771	1.4310	1.5691
RMSSE	0.9895	0.9144	1.0121	0.9629	1.1534	1.0087	0.9907	1.0174
SiO_2								
Average CRPS	0.6670	0.4743	0.7697	0.4864	0.8838	0.3824	0.5615	0.4674
Percent in 95% Interval	92.7757	93.1034	92.7481	94.1441	94.1176	95.4128	94.9580	93.7500
ME	-4.1E-02	-4.1E-02	-3.9E-02	-1.5E-02	-8.9E-02	2.4E-03	-2.5E-04	-2.1E-02
RMSE	1.3428	0.8909	1.4924	0.9010	1.7124	0.7074	1.0109	0.8490
RMSSE	1.0440	1.0700	1.0590	1.0259	1.0750	1.0486	1.0283	1.0654
Chlorophyll a								
Average CRPS	0.1853	0.4235	0.1482	0.1927	0.1741	0.6368	0.4318	0.1191
Percent in 95% Interval	92.5134	94.0120	95.0276	97.4684	88.9610	91.5584	91.0714	94.6429
ME	-5.7E-03	-1.8E-02	-6.6E-03	-1.5E-03	-4.7E-03	-5.0E-02	-3.8E-02	3.6E-03
RMSE	0.3547	0.8794	0.2748	0.3476	0.3708	1.2353	0.8405	0.2153
RMSSE	1.0502	0.9936	0.9887	0.9100	1.2526	1.0871	1.1015	1.0103
pH								
Average CRPS	0.0210	0.0320	0.0284	0.0189	0.0491	0.0321	0.0352	0.0252
Percent in 95% Interval	96.9349	94.9153	95.0943	99.5495	92.8251	95.3704	94.9580	97.5104
ME	-1.7E-04	1.6E-04	1.2E-03	2.7E-05	2.8E-03	5.5E-04	-3.6E-04	5.4E-04
RMSE	0.0375	0.0599	0.0525	0.0326	0.0904	0.0656	0.0677	0.0457
RMSSE	0.7749	0.9484	0.9107	0.6592	1.0838	0.9868	0.9465	0.7679

Table 10. EBK3D cross-validation statistics.

The visual assessment of the empirical semivariances to the spectrum of simulated semivariograms, regarding C1B, is shown, as an example, in Annex C, Figure C-5. An empirical

semivariogram that falls in the median of the simulated semivariogram distribution indicates good model performance.

EBK3D performance varied depending on the OV, campaign, and area due to data distribution, sampling scheme, and measurement uncertainty. The results for each OV, in both areas, across the four campaigns are illustrated in Annex C, Figure C-6 to Figure C-11.

Regarding the average CRPS (Figure C-6), ME (Figure C-8), and RMSE (Figure C-9), the best results were achieved for temperature, salinity, PO_4, and pH. These cross-validation statistics for these OVs were also stable across the different areas and campaigns.

Regarding the RMSSE (Figure C-10), the best results (between 0.90 and 1.10) were achieved for the nutrients (PO_4, and SiO_2), which were also relatively stable across the different areas and campaigns. As explained before, this parameter could be improved by changing the semivariogram modelling function from Exponential to K-Bessel, although this degrades the visual aspect of the surface.

Regarding the RVar (Figure C-11), the best results (close to 1) were achieved for temperature and salinity. All the other OVs had values lower than 1 (up to 0.2015, regarding DO), meaning that the predictions had significantly less variability than the observed values. This could be explained by the number of samples, where temperature and salinity had approximately 20 to 50 times more samples than the other OVs.

5.3. Ecosystems identification and characterization

EMUs were identified through cluster analysis with the K-means method.

Before deciding on the definitive K-means configuration, we evaluated different normalization methods (Min-Max and Z-Score) and OVs selections (with (7OV) and without (6OV) pH) for C1. These results are described in Figure 15.

The comparison of the mean SC values, from k = 2 to k = 30, for different configurations, showed that:

- Generally, the Min-Max normalization produced better results (higher mean SC values and clearer local peaks). The advantages of Z-Score over Min-Max are related to better outlier handling. As this eventual limitation was addressed during ESDA, we chose Min-Max based on the results and because of the characteristics of the method, which preserves the relationship between data values (scale) and the histogram format (Bação, 2020a);
- The selection of six OVs (disregarding pH, as the less important OV in the mix) produced better results than using all seven. Reducing the number of variables also reduces the

problem dimensionality (Bellman, 1961), which decreases complexity and improves processing time. However, we chose to maintain all seven OVs for two main reasons: (i) retaining pH in the analysis would be more useful for the users; and (ii) for the ideal number of clusters (k = 6, in this case), the mean SC values were almost equal.

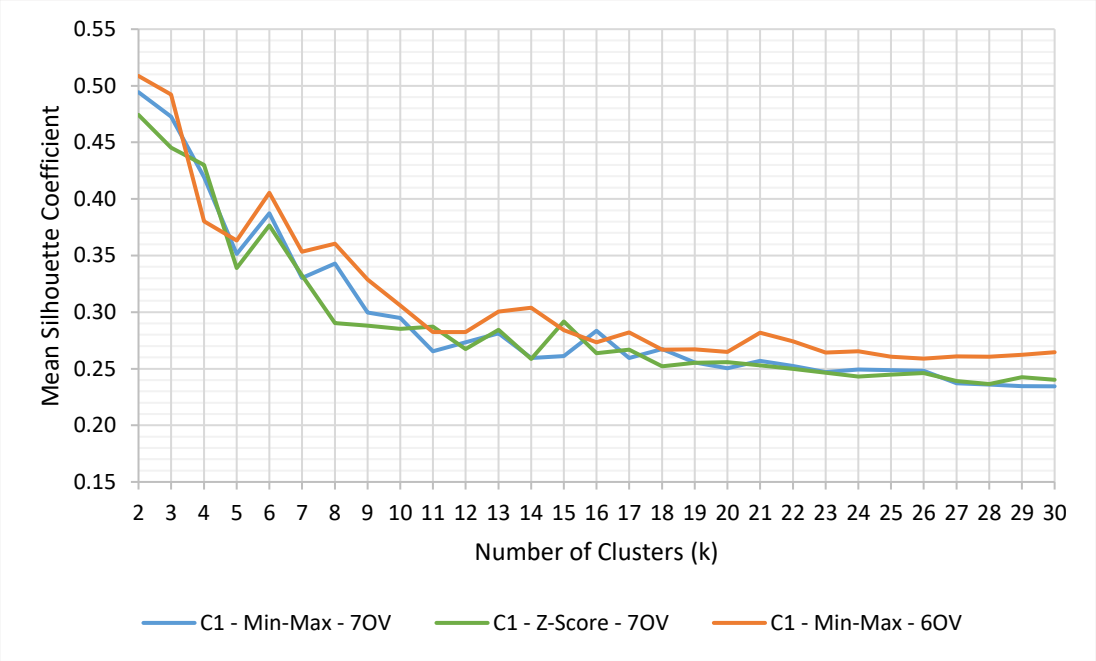


Figure 15. Mean SC of different configurations of K-means.

Figure 16 illustrates the mean SC values and Figure D-1, in Annex D, illustrates the pseudo-F statistic values, from k = 2 to k = 30, for the four campaigns.

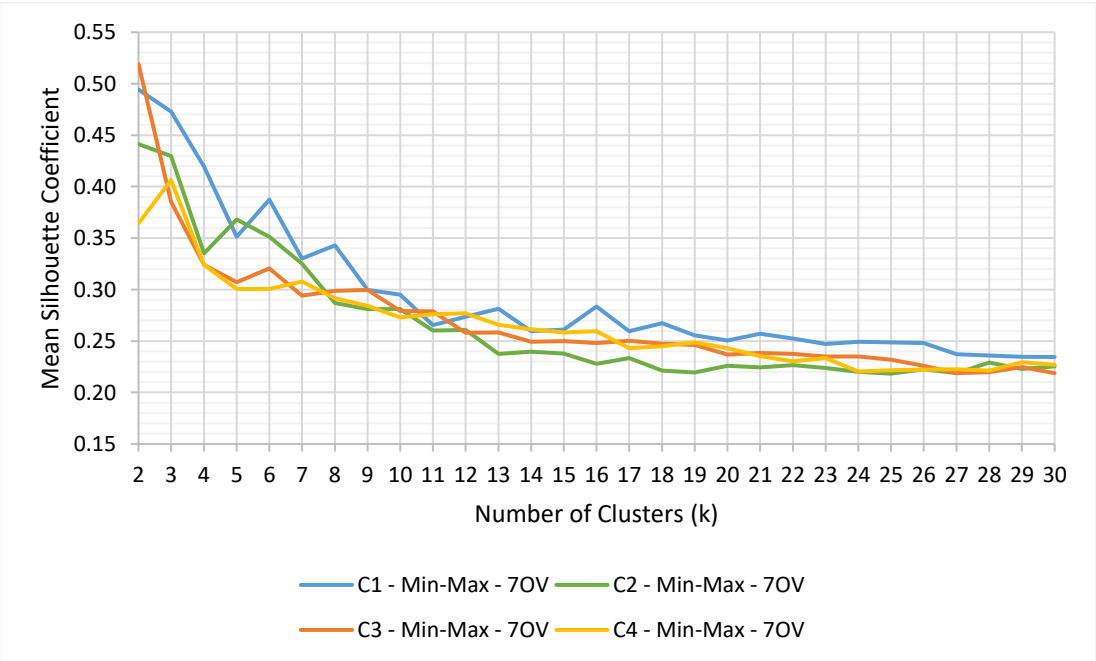


Figure 16. Mean SC of the four campaigns.

Both diagnoses (SC and pseudo-F) indicated that the best results were achieved with k = 2 for C1, C2 and C3, and k = 3 for C4. From a spatial perspective, this solution meant differentiating surface and deep-water masses, which is of minor interest to the users. In this sense, we chose the number of clusters based not only on these statistics (mean SC value), but also on the scope of the study and its importance for the users:

- C1: the appropriate k (besides the optimal 2) is k = 6 (SC = 0.387), where the graph (Figure 16) has a local maximum. After this k, the mean SC value drops significantly and the next local maximum occurs for k = 8 (SC = 0.343), which we considered an excessive number of clusters and, as further described, it would not match its seasonal pair (C3). As a result, we decided k = 6 was adequate;
- C2: the appropriate k (besides the optimal 2) is k = 5 (SC = 0.368), where the graph (Figure 16) has a local maximum. After this k, the mean SC value drops significantly. Therefore, we decided k = 5 was adequate;
- C3: the appropriate k (besides the optimal 2) is k = 6 (SC = 0.320), where the graph (Figure 16) has a local maximum. After this k, the mean SC value drops significantly, and the next local maximum occurs for k = 9 (SC = 0.300), which we considered an excessive number of clusters, and it would not match its seasonal pair (C1). As a result, we decided k = 6 was adequate;
- C4: the appropriate k (besides the optimal 3) is k = 7 (SC = 0.308), where the graph (Figure 16) has a local maximum. After this k, the next local maximum occurs for k = 12 (SC = 0.277), which we considered an excessive number of clusters. The difference for k = 5 (SC = 0.301), which would match its seasonal pair (C2) is not substantial. However, comparing the mean SC for each individual cluster, k = 5 and k = 7 (Annex D, Figure D-2), the minimum SC for k = 5 (0.186) is lower than the one for k = 7 (0.216). As a result, we decided k = 7 was adequate.

The selected k and the mean SC value for each campaign are illustrated in Figure 17. C1 has the best mean SC value, while C4 has the worst.

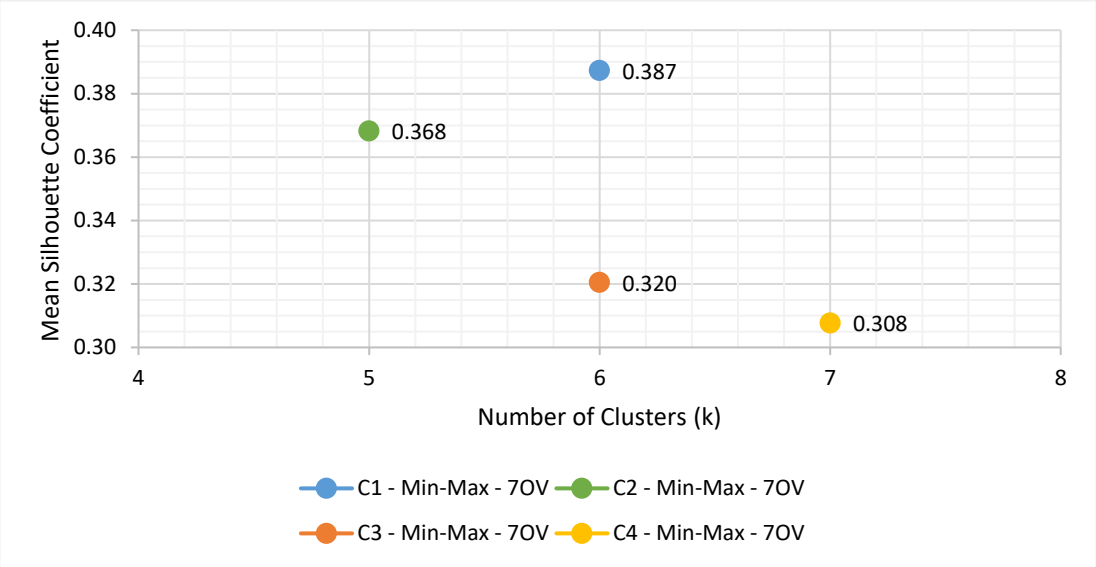


Figure 17. Optimal number of clusters of each campaign.

The SC of each cluster for the stratified sample and the entire dataset is presented in Annex D, Table D-1. The comparison of both values indicates that the sampling method provided a reliable estimation of the broad results, requiring considerably less time. The SC distribution across the different clusters for each campaign is illustrated in Figure D-3. The width of each distribution represents the cluster size, as also illustrated, for instance, by Figure 19. These SC plots can also be used to assess the adequate k. Generally, inappropriate k can be identified by clusters with all the SC scores below the mean SC, as well as great variations in cluster size (Pedregosa, et al., 2011). Accordingly, examining figure D-3, we can notice that C1 could have the most arguable choice of k as it has a considerably larger cluster (cluster 13) and a cluster with almost all the SC values below the mean SC (cluster 14).

The cluster centroids, for each campaign, are presented in Table 11. All the clusters have unique centroids because we performed the analysis independently for each campaign. In order to facilitate their distinction, we labelled them with two digits: the first one referring to the campaign and the second one referring to the sequence produced by the analysis (e.g., cluster 10 is the first cluster of C1, and cluster 21 is the second cluster of C2). Even if we can find some similarities between clusters of different campaigns, their numbering does not control their resemblance (for instance, cluster 11 is not equivalent to cluster 21).

	Temperature (°C)	Salinity (PSU)	DO (mgO_2/L)	PO_4 (µmol/L)	NO_x (µmol/L)	SiO_2 (µmol/L)	pH
C1							
Cluster 10	16.6132	35.8263	7.3848	0.1986	1.6249	2.0001	8.0738
Cluster 11	14.6385	35.9031	6.9169	0.4126	5.4358	2.9920	8.0357
Cluster 12	15.2831	36.1664	6.7590	0.5255	6.9887	5.4138	8.0693
Cluster 13	13.4938	35.8732	6.6814	0.6177	9.1058	4.1700	8.0051
Cluster 14	17.6557	36.3205	7.3249	0.2452	2.0013	3.0208	8.0949
Cluster 15	21.5553	36.6221	7.4851	0.1704	0.7732	1.7976	8.1197
C2							
Cluster 20	14.7870	35.5768	8.2405	0.1256	1.0164	1.3155	8.0941
Cluster 21	13.8696	35.8041	7.7607	0.3331	4.5901	2.5288	8.0646
Cluster 22	13.4388	35.8363	7.4223	0.5778	8.7194	3.9230	8.0402
Cluster 23	14.5885	36.0567	7.3407	0.4927	7.8917	2.9389	8.0494
Cluster 24	15.4786	36.1302	8.0060	0.2321	3.4453	1.8337	8.0736
C3							
Cluster 30	16.8832	36.1235	6.9569	0.1750	1.7550	2.0442	7.9844
Cluster 31	14.8644	35.9320	6.8004	0.3751	5.0528	3.2844	7.9966
Cluster 32	13.6923	35.8587	6.5827	0.5385	7.7037	4.4191	8.0144
Cluster 33	13.0512	35.8046	6.2489	0.7196	10.2335	5.6776	8.0097
Cluster 34	16.5455	35.7407	7.1390	0.1774	1.5831	2.2566	8.0117
Cluster 35	14.1477	35.9792	6.5516	0.6068	9.3095	4.9261	7.9390
C4							
Cluster 40	14.3005	35.5626	8.1843	0.1270	1.2769	1.5068	8.1105
Cluster 41	14.2215	35.8676	7.5800	0.3052	4.6716	2.3823	8.0795
Cluster 42	14.5386	35.7908	7.5993	0.1262	1.1676	1.6169	8.0585
Cluster 43	13.7508	35.9213	7.4270	0.4474	7.5193	3.2065	8.0974
Cluster 44	15.5667	36.3168	7.2016	0.1469	1.9164	1.9270	8.0621
Cluster 45	13.9182	36.0655	6.9437	0.4725	8.4493	3.4266	8.0163
Cluster 46	14.8935	36.2203	7.0967	0.3189	5.0899	2.7738	8.0543

Table 11. Cluster centroids of each campaign.

The clusters characterization for C1 is presented in Figure 18 to Figure 20. The characterization for the other campaigns is presented in Annex D, Figure D-4 to Figure D-12.

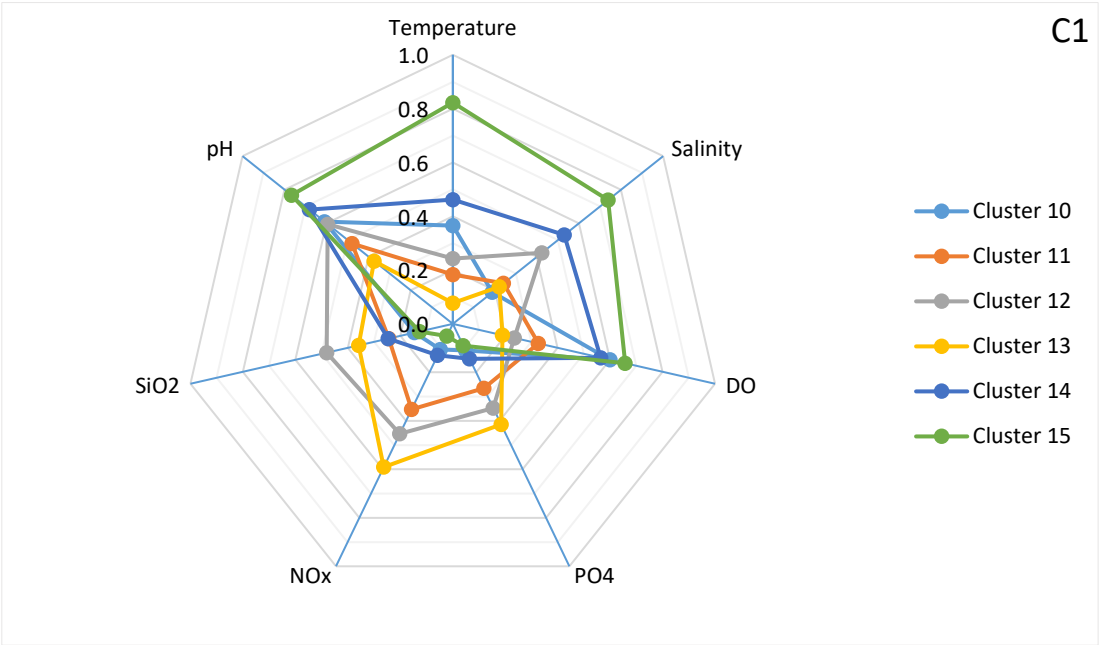


Figure 18. Clusters centroids of C1 (normalized values).

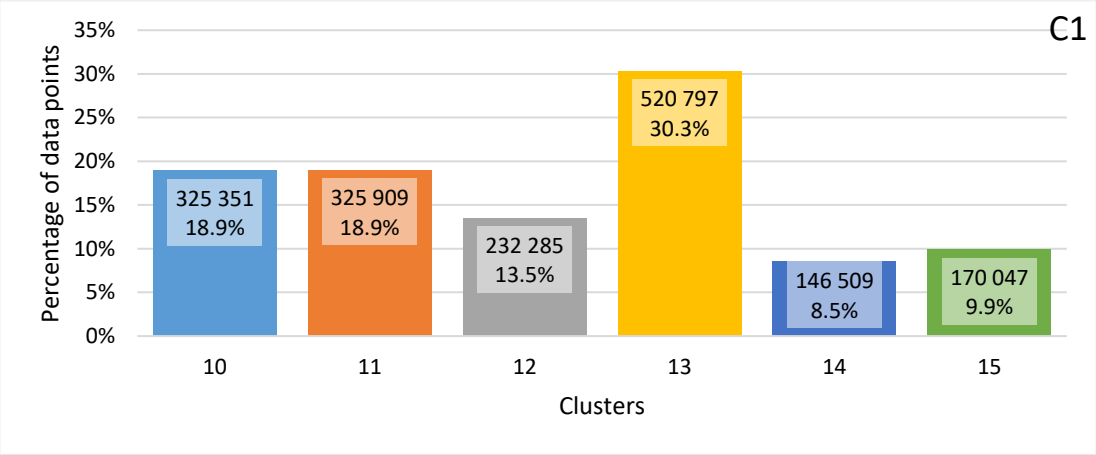


Figure 19. Size of the clusters of C1.

Figure 20. Area distribution of the clusters of C1.

When we look at the EMUs centroids values for each OV, we can see that high temperatures and salinity (shallower waters) are paired with low nutrients. This pattern is more perceptible in early autumn (C1-C3) than in late winter/early spring (C2-C4) and it indicates that EMUs distribution is mainly stratified by depth. Analyzing the graphs for the four campaigns (Figure 18, Figure D-4, Figure D-7 and Figure D-10), we can also assess that temperature is the main driver in EMUs distinction, as centroids values of this OV are less prone to overlap along its axis, particularly in C1 and C2.

Regarding EMUs size (Figure 19, Figure D-5, Figure D-8, and Figure D-11) we can notice a more even distribution in C3 and C4 than in C1 and C2. In both C1 and C2 there are two clusters (13 with 30.3% and 22 with 28.8%) that stand out from the rest. These two EMUs have relatively high values of nutrients, which means they are probably located at the higher depths of the study areas.

On the subject of EMUs distribution (Figure 20, Figure D-6, Figure D-9 and Figure D-12), it is noticeable that across the four campaigns, there are few clusters that have a substantial

distribution (> 5%) in both areas (clusters 11, 13, 22, 31, 32, and 45), which demonstrates physical and chemical seawater differences between the two areas. Data also shows that in early autumn (C1-C3) there is a greater probability of having similar EMUs in the different areas.

We compared the EMUs created in this study with the global ones generated by (Sayre, et al., 2017b), although the two analyses were made at different scales (global versus local) and with different data (in-situ collection versus archival means). In Annex D, Figure D-13, and Figure D-14, we can see that area B exhibits only one global EMU (EMU 11, from 0 to 175 m depth) and area E exhibits two (EMU 21, from 0 to 75 m depth, and EMU 11, from 75 m to 375 m depth). Table D-2 describes the characteristics of these two global EMUs, while Table D-3 reveals a probable correspondence between the global and local EMUs.

5.4. Voxel layers

The NetCDF files were opened and explored in ArcGIS Pro as multidimensional voxel layers. The visualization of the data as voxel layers provides an appropriate display of the 3D interpolation results and a prolific tool for data analysis, for instance, allowing the creation of horizontal and vertical cross-sections or isosurfaces (Esri, 2022f), as illustrated in Figure 21 and Figure 22. The voxel layers for C1B and C1E are illustrated in Annex E, Figure E-1 and Figure E-2, and the voxel layers for the OVs in areas B and E across the four campaigns, are illustrated in Figure E-3 to Figure E-20.

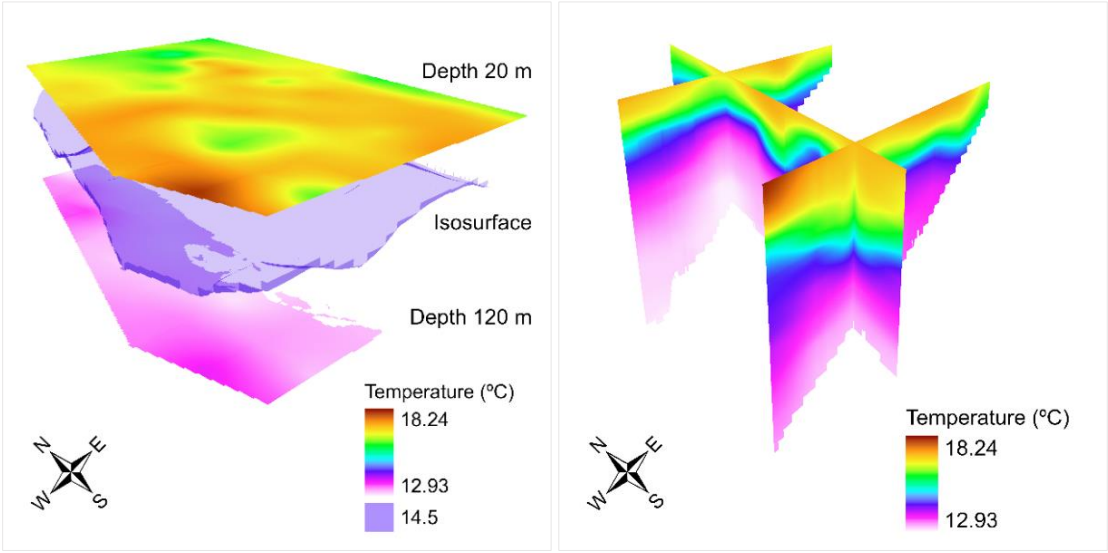


Figure 21. Voxel layer of temperature for C1B: isosurface and cross-sections.

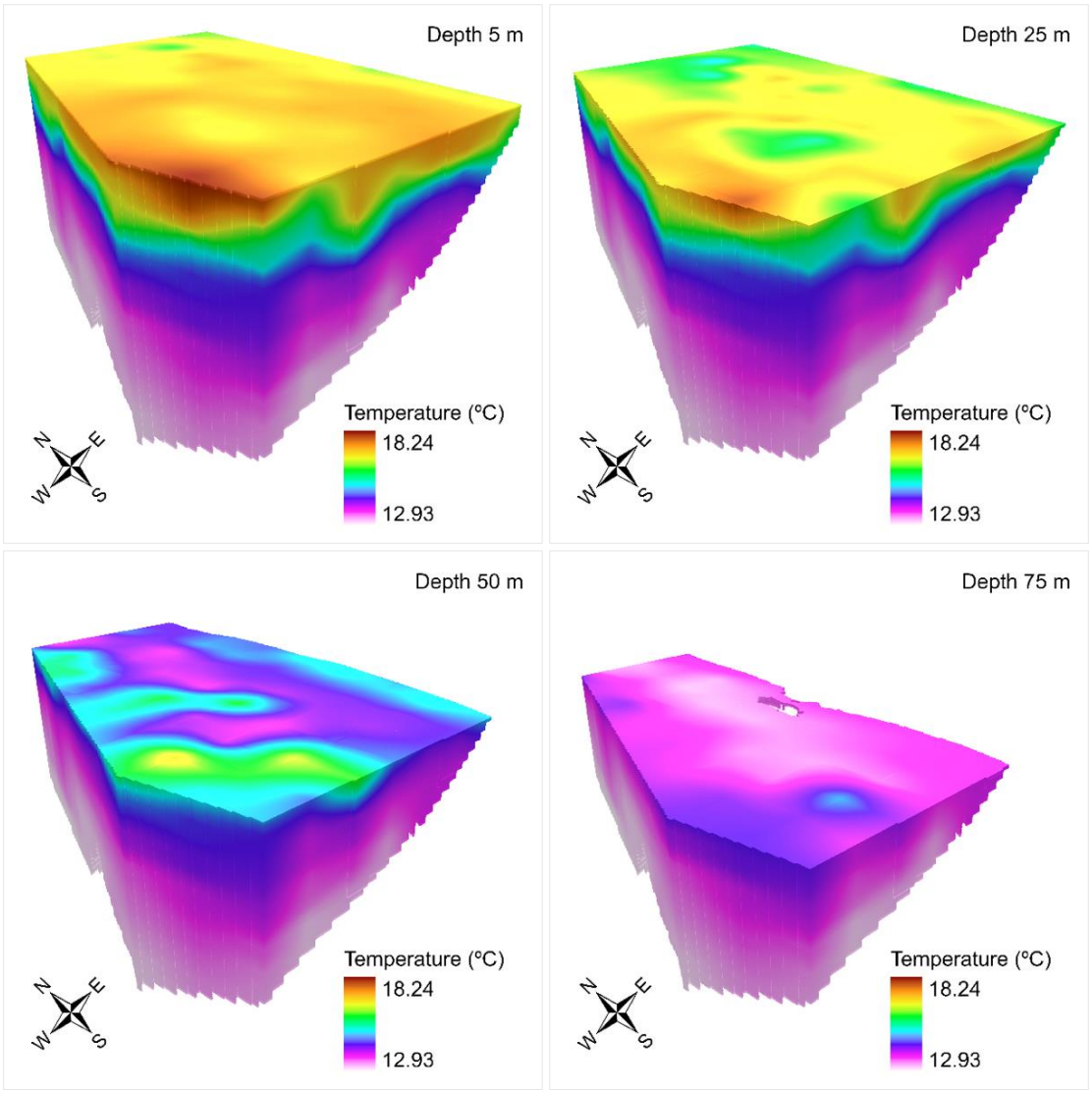


Figure 22. Voxel layer of temperature for C1B: different depths.

As a multidimensional format, the NetCDF files allow the comparison of OVs distribution at multiple depths and across different campaigns, as illustrated in Annex E, Figure E-21 to Figure E-38. Furthermore, the visualization as 3D voxel layers can enhance multiple and broader analysis.

The technical interpretation and application of the results is reserved for specific users and is beyond the scope of this work. However, some conclusions may be drawn, based on the visualization and the comparison between voxel layers of each OV, for different depths and campaigns.

5.4.1. Temperature

Temperature distribution is in accordance with that described in the literature review. In both areas (B and E) we can notice similar distributions among the seasonal pairs (C1-C3 (early

autumn) and C2-C4 (late winter/early spring)). For instance, the vertical stratification and the temperature range are higher in C1-C3 than in C2-C4. Generally, C3 has the greatest vertical stratification, while C4 has the lowest, displaying a higher mixture of temperature values in depth (Figure E-3 and Figure E-12). Regarding the horizontal distribution, we can also notice this similarity between seasonal pairs. In area B (Figure E-21), in early autumn (C1-C3) we can notice a greater horizontal variance, with higher temperature values in the southwest region and lower values near the coast, which can be explained by upwelling, which is a common process on the Portuguese west coast (Lemos, 2004). In area E (Figure E-30), there is also a superior horizontal difference in early autumn (C1-C3), but the minor temperatures near the coast are less evident, as the upwelling in the south coast is not so frequent (Fiúza, 1983).

5.4.2. Salinity

Salinity distribution is generally in accordance with that described in the literature review. In area B, we can notice that the higher values are not frequently at the surface (Figure E-4), while in area E they usually are (Figure E-13). The vertical stratification and the salinity range are also normally higher in C1-C3 than in C2-C4. Regarding the horizontal distribution, in area B there is a noticeably higher heterogeneity in C2-C4 than in C1-C3 (Figure E-22), while in area E the opposite occurs (higher heterogeneity in C1-C3) (Figure E-31). This could be explained by the freshwater discharge from the rivers, which is usually greater in late winter/early spring. While area B has two major rivers (Vouga and Mondego), area E has minor ones (Arade e Gilão) with significantly lower discharges.

5.4.3. Dissolved oxygen

DO distribution is in accordance with that described in the literature review. The minimum value in both areas and across the four campaigns (5.96 mgO_2/L) is above the critical level (5 mgO_2/L) (Cravo, et al., 2020) or the hypoxia level (below 2-3 mgO_2/L) (Vaquer-Sunyer & Duarte, 2008). Generally, the higher values are near the surface, where we can notice some moving hotspots across the four campaigns (Figure E-5 and Figure E-14). Regarding the horizontal distribution, in area B there is a clear pattern of higher values in C2-C4 than in C1-C3 (Figure E-23). In area E, this pattern is less evident, as there are still greater values in C2-C4 but also substantial high values in C1 (Figure E-32). DO higher values are usually associated with less biological productivity.

5.4.4. Nutrients

The distribution of nutrients (PO_4, NO_X, SiO_2) is in accordance with that described in the literature review. There are lower values near the surface and higher values in deeper waters. In area B, we can notice some hotspots along the coast, probably due to upwelling (Figure E-6 to Figure E-8). In area B, this pattern is also present along the coast, although in a less perceptible manner (Figure E-15 to Figure E-17). The horizontal cross-sections give a clear perspective of this pattern that seems to be heightened as depth increases (Figure E-24 to Figure E-26 and Figure E-33 to Figure E-35).

5.4.5. Chlorophyll a

Chlorophyll a distribution is in accordance with that described in the literature review. In both areas, there are higher values along the coast (Figure E-9 and Figure E-18), although in the vertical distribution, the DCM is not always easily distinguishable. Regarding the horizontal distribution, in area B there is a clear pattern of higher values in C2-C4 than in C1-C3 (Figure E-27). In area E, this pattern is also present, as there are greater values in C2-C4 but also substantial high values in C3 (Figure E-36).

5.4.6. pH

The distribution of pH is in accordance with that described in the literature review. There are many factors affecting pH distribution (Lauvset, et al., 2020) and, while in C1, in areas B and E, we can recognize a depth gradient, that pattern is less evident in the other campaigns (Figure E-10 and Figure E-19). There, pH values seem to maintain the same horizontal structure in depth, as better illustrated in Figure E-28 and Figure E-37. From C1 to C3, there is a slight decrease in pH minimum values in both areas. However, in C4, this minimum value is higher than the previous highest minimum.

5.4.7. Clusters

The distribution of clusters throughout the campaigns must be analyzed with care, as clusters are not equivalent for different campaigns (e.g., cluster 11 is not equal to cluster 21). To attain this, cluster analysis would have to be conducted with data merged from all four campaigns instead of data from each campaign as was done. Nevertheless, it is possible to pair similar clusters across campaigns, for instance, cluster 15 and 24. Generally, the clusters are stratified in depth (except for C4B), as this is one of the major drivers of OVs variability (Figure E-11 and Figure E-20). In both areas, as depth increases, the number of clusters usually rises, and its spatial distribution is approximately parallel to the coast (Figure E-29 and Figure E-38).

5.5. Web application

Publishing the results through the web in standardized formats (NetCDF, WMS) is an effective way of sharing data and improve its findability, accessibility, interoperability, and reusability by stakeholders (Wilkinson, et al., 2016).

The StoryMap produced for this study is published at https://arcg.is/1yeOHm. It includes a brief introduction, data and study areas, methods, examples of the results and tools for data visualization and exploration. As the analysis tools implemented in the ArcGIS web mapping application were limited, we also published the NetCDF files on a ncWMS server, which allows data analysis in a standardized format (WMS), for instance, with multiple depth or specific profile and transect views, as illustrated in Figure 23 and in Annex F, Figure F-1 and Figure F-2.

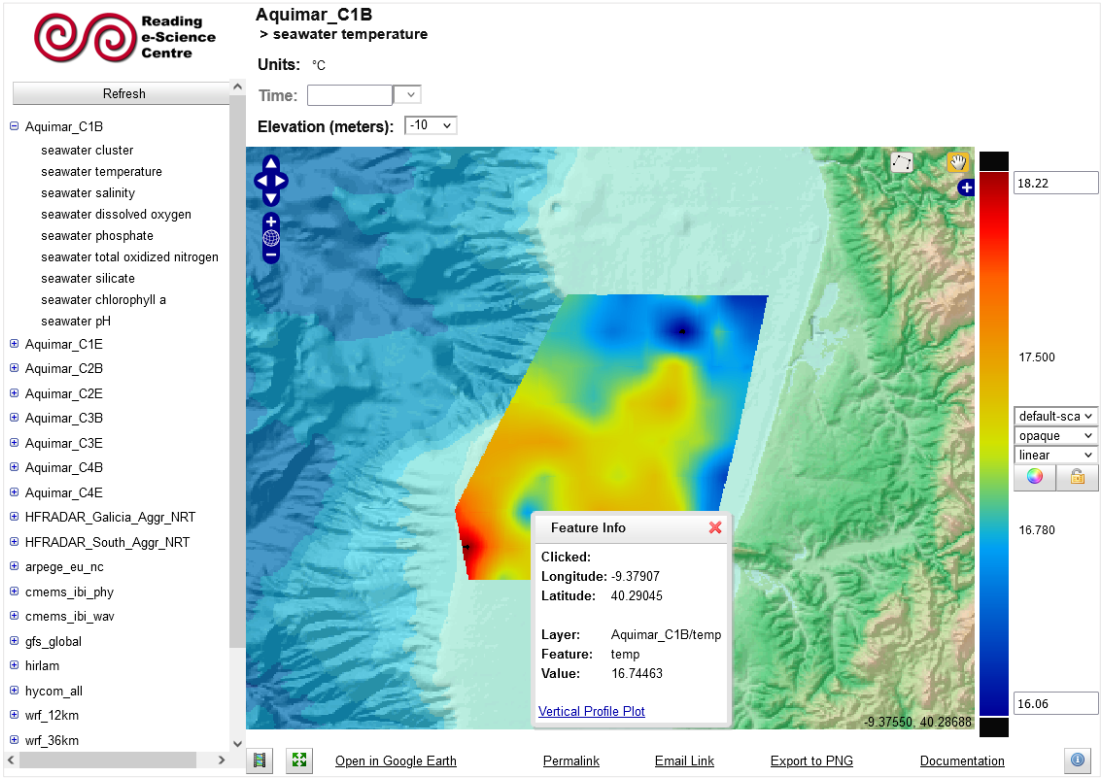


Figure 23. NetCDF file exploration in Godiva3 interface (ncWMS).

6. CONCLUSION

Marine ecosystems are one of the most productive ecosystems on Earth. They are key producers of oxygen (O_2) and absorbers of carbon dioxide (CO_2). Regardless of being valuable, their preservation and sustainability are at risk since coastal regions encompass approximately 40% of the world's population, accompanied by the highest rates of population growth and urbanization, and 90% of the fish catches. Therefore, there is an urgent need to produce better and broader scientific knowledge to support and drive effective policies and actions. One of the main challenges of this mission, globally incorporated by the Ocean Decade, is ocean digitalization.

The general objective of this work was to map, in 3D, the physical and chemical properties of seawater and characterize EMUs on the Portuguese coast. This aim can be translated as creating digital representations of OVs distribution in the water column, in limited areas and at specific times, and identifying distinct ecosystems based on OVs values. The study used data acquired for the project AQUIMAR and focused on two areas of the Portuguese coast: one on the northwestern region, approximately between Aveiro and Figueira da Foz, and another on the south region, approximately between Sagres and Tavira.

We also established three SO. SO 1 was to create 3D surfaces of the physical and chemical properties of seawater through spatial interpolation. We achieved this by selecting eight OVs (temperature, salinity, DO, nutrients (phosphate (PO_4), TON (NO_X), and silicate (SiO_2)), chlorophyll a, and pH) whose data had been previously prepared and characterized. The 3D surfaces of each OV were built using the EBK3D method. This method had reliable performance as an automated interpolator since, after the configuration of its advanced model parameters, we only had to make specific modifications in some OVs and, in some cases, adjustments in the search neighborhood parameters. EBK3D estimations were assessed using the cross-validation method. The disparity, in some circumstances, between the two assessments used (cross-validation and visual), underlined their complementarity and, in future work, to ideally include a third method, for instance, a validation dataset, not subject to the typical cross-validation drawbacks like model overfitting. Regarding the visual assessment of the interpolation results, we discerned some improvements that could be made in the ArcGIS interface, namely presenting the results in a raster with a specific resolution and colored in a stretch scheme, rather than using geostatistical layers, colored by editable contours.

SO 2 was to identify and describe EMUs through cluster analysis. We accomplished this by using the K-means method and selecting seven OVs (chlorophyll a was excluded, due to its limited

extent in depth). The optimal number of clusters was assessed based on the mean SC and the pseudo-F statistic. Although both diagnoses indicated that the optimal k values were 2 (for C1, C2, and C3) and 3 (for C4), we considered that this partition would have an inconsequential interest for the users as it only distinguished surface and deep-water masses. This division would be more relevant, for instance, in a macroscale analysis and could resemble the ecosystem identification of other works. Consequently, we chose the number of EMUs based not only on the mean SC value but also on the scope of the study and the further data usage by stakeholders. We then validated this decision through the SC distribution plots. Cluster analysis was performed independently for each campaign since the goal was to obtain insights from the data and infer its natural structure from each dataset. Thus, EMUs are not correspondents for different campaigns, although we can find some resemblances. In a specific evolution study, in which we would like to visualize the spatial distribution of EMUs through time, cluster analysis should be run with data from all the campaigns, to produce EMUs transversal to all of them. From another analysis perspective, if the characteristics of the EMUs (OVs values) were predetermined, we could also run a supervised method for EMUs classification. We leave these multiple research paths for further work.

SO3 was to publish the results on a web application. We attained this by building a StoryMap hosted by the ArcGIS Online infrastructure. The StoryMap contains a brief description of the work that was done and links to two exploration tools: a web mapping application, where the voxel layers can be visualized in 3D; and the Godiva3 interface of the ncWMS server, where the NetCDF files can be explored. The specific development of additional widgets and the enhancement of the visualization tools were left for further work.

The contributions from the earlier marine ecosystem mapping studies were essential for the fulfillment of the research aim. Like in previous studies, we selected physical and chemical properties of the seawater to derive and map ecosystems through an unsupervised data-driven method (cluster analysis). However, unlike those studies, we used in-situ observed data instead of averaged records, and we mapped OVs and EMUs in 3D for the entire water column of the study areas instead of a specific and limited horizontal layer.

Low-cost ocean observing technologies are proliferating (Butler & Pagniello, 2021), creating marine monitoring networks with higher spatial resolutions and sampling rates, producing more accurate datasets, and increasing data volume and availability. Consequently, the use of in-situ collected data, preferably in real time, will become a common trend and input in ocean modelling. This remark to the data's temporal reference should remind the users of the main

limitation of this study, which is related to the uncertainty of the 3D maps involving OVs values. Although we know that the sample points were not measured simultaneously, 3D interpolation assumes they were, or that the OVs variation between measurements can be ignored.

The workflow developed in this work can be seen as a breakthrough in the use of observed data to build 3D maps of OVs and EMUs. These ocean state base maps are essential for ocean preservation, sustainability, and governance, and for effectively creating DTOs.

Concerning future research topics, we prioritized the following actions:

- Extend this work to other marine areas of the AQUIMAR Project with in-situ data collection;
- Improve the web interface and user experience for data visualization and exploration;
- Develop the characterization of EMUs from other perspectives (e.g., time-series, supervised methods);
- Study the use of EBK3D with other datasets, detailing how to adjust its parameters and testing other methods to assess its predictions.

REFERENCES

AbdelRahman, M. A. E., Zakarya, Y. M., Metwaly, M. M. & Koubouris, G., 2021. Deciphering Soil Spatial Variability through Geostatistics and Interpolation Techniques. *Sustainability,* Volume 13(1), p. 194.

Amiri, M., Tarkesh, M., Jafari, R. & Jetschke, G., 2020. Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can perform better in species distribution modeling?. *Ecological Informatics,* Volume 57, p. 101060.

Anselin, L., 2005. Interactive techniques and exploratory spatial data analysis. In: *Geographical Information Systems.* 2.º ed. New York: Wiley, pp. 253-266.

Arbic, B. K., 2015. *Physical Properties of Seawater.* [Online] Available at: https://coessing.files.wordpress.com/2016/08/arbic_25august2015.pdf [Accessed 12 September 2022].

Ashburner, J. & Friston, K. J., 2009. Voxel-Based Morphometry. *Encyclopedia of Neuroscience,* pp. 471-477.

Attorre, F. et al., 2007. Comparison of interpolation methods for mapping climate and bioclimate variables at regional scale. *International Journal of Climatology,* Volume 27(13), pp. 1825-1843.

Bação, F. L., 2020a. Módulo 3 - A metodologia do Data Mining. In: *Data Mining Geo-Espacial.* Lisboa: Universidade Nova de Lisboa - Instituto Superior de Estatística e Gestão da Informação.

Bação, F. L., 2020b. Módulo 5 - Ferramentas de Exploração. In: *Data Mining Geo-Espacial.* Lisboa: Universidade Nova de Lisboa - Instituto Superior de Estatística e Gestão de Informação.

Baird, R. & Bridgewater, L., 2017. *Standard Methods for the Examination of Water and Wastewater.* 23 ed. Washington D.C.: American Public Health Association.

Barbier, E. B., 2017. Marine ecosystem services. *Current Biology,* Volume 27, pp. R507-R510.

Beal, S. L., 2001. Ways to Fit a PK Model with Some Data Below the Quantification Limit. *Journal of Pharmacokinetics and Pharmacodynamics,* Volume 28, pp. 481-504.

Bellman, R. E., 1961. *Adaptive Control Processes.* Princeton: Princeton University Press.

Betrand, A., Ballón, M. & Chaigneau, A., 2010. Acoustic Observation of Living Organisms Reveals the Upper Limit of the Oxygen Minimum Zone. *PLoS ONE,* Volume 5(4), p. e10330.

Bettencourt, A. M. et al., 2004. *Typology and Reference Conditions for Portuguese Transitional and Coastal Waters - Development of Guidelines for the Application of the European Union Water Framework Directive.* Lisbon: Instituto da Água (INAG).

Bristow, L. A., 2017. Nutrients that limit growth in the ocean. *Current Biology,* Volume 27(11), pp. R474-R478.

Britannica, 2022a. *Ecosystem.* [Online] Available at: https://www.britannica.com/science/ecosystem [Accessed 12 September 2022].

Britannica, 2022b. *Pelagic Zone.* [Online] Available at: https://www.britannica.com/science/pelagic-zone [Accessed 12 September 2022].

Buonocore, E. et al., 2021. Trends and Evolution in the Concept of Marine Ecosystem Services: An Overview. *Water,* Volume 13:2060, p. 2060.

Butler, J. & Pagniello, C. M. L. S., 2021. Emerging, Low-Cost Ocean Observing Technologies to Democratize Access to the Ocean. *Oceanography,* Volume 34(4), pp. 94-95.

Calder, B. & Wells, D., 2007. *CUBE User's Manual.* 1.13 ed. New Hampshire: University of New Hampshire.

Calinski, T. & Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics,* Volume 3:1, pp. 1-27.

Calvert, J. et al., 2014. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. *ICES Journal of Marine Science,* Volume 72(5), p. 1498–1513.

Capone, D., Bronk, D., Mulholland, M. & Carpenter, E., 2008. *Nitrogen in the Marine Environment.* 2nd ed. s.l.:Academic Press.

Celebi, M. E., Kingravi, H. A. & Velab, P. A., 2013. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications,* Volume 40(1), pp. 200-210.

Chmielewski, S. & Tompalski, P., 2017. Estimating outdoor advertising media visibility with voxel-based approach. *Applied Geography,* Volume 87, pp. 1-13.

CICES, 2022. *Common International Classification of Ecosystem Services.* [Online] Available at: https://cices.eu [Accessed 12 September 2022].

CIESIN, 2012. *National Aggregates of Geospatial Data Collection: Population, Landscape, And Climate Estimates, Version 3 (PLACE III).* [Online] Available at: https://bit.ly/3vIOH3H [Accessed 12 September 2022].

Constable, A. J. et al., 2016. Developing priority variables ("ecosystem Essential Ocean Variables" - eEOVs) for observing dynamics and change in Southern Ocean ecosystems. *Journal of Marine Systems,* Volume 161, pp. 26-41.

Costa, A. C., Durão, R., Soares, A. & Pereira, M. J., 2008. A Geostatistical Exploratory Analysis of Precipitation Extremes in Southern Portugal. *Revstat - Statistical Journal,* Volume 6(1), pp. 21-32.

Costello, M. J. & Breyer, S., 2017a. Ocean Depths: The Mesopelagic and Implications for Global Warming. *Current Biology,* Volume 27, pp. R36-R38.

Costello, M. J. & Chaudhary, C., 2017b. Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current Biology,* Volume 27(11), pp. R511-R527.

Couto, A. B. et al., 2016. Inter-comparison of OC-CCI chlorophyll-a. *International Journal of Remote Sensing,* Volume 37:18, pp. 4337-4355.

Cravo, A., Rosa, A., Jacob, J. & Correia, C., 2020. Dissolved oxygen dynamics in Ria Formosa Lagoon (South Portugal) - A real time monitoring station observatory. *Marine Chemistry,* Volume 223, p. 103806.

Cuesta, A., 2020. *Filling in the gaps: 3D mapping Arizona's basin and range aquifer in the prescott active management area.* Los Angeles: University of Southern California.

Danovaro, R. et al., 2020. Ecological variables for developing a global deep-ocean monitoring and conservation strategy. *Nature Ecology & Evolution,* Volume 4, pp. 181-192.

De La Cruz, P. M. C., 2021. The Knowledge Status of Coastal and Marine Ecosystem Services - Challenges, Limitations and Lessons Learned From the Application of the Ecosystem Services Approach in Management. *Frontiers in Marine Science,* Volume 8:684770.

Delua, J., 2021. *Supervised vs. Unsupervised Learning: What's the Difference?.* [Online] Available at: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning [Accessed 12 September 2022].

DGRM, 2019. *Plano de Situação do Ordenamento do Espaço Marítimo Nacional.* [Online] Available at: https://www.psoem.pt [Accessed 12 September 2022].

DITTO, 2022. *Welcome to DITTO.* [Online] Available at: https://ditto-oceandecade.org [Accessed 12 September 2022].

Doney, S. C., Busch, D. S., Cooley, S. R. & Kroeker, K. J., 2020. The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities. *Annual Review of Environment and Resources,* Volume 45, pp. 83-112.

Durant, J. M. et al., 2019. Contrasting effects of rising temperatures on trophic interactions in marine ecosystem. *Scientific Reports,* Volume 9, p. 15213.

Duxbury, A. C. & Cenedese, C., 2022. *Ocean - Encyclopedia Britannica.* [Online] Available at: https://www.britannica.com/science/ocean [Accessed 12 September 2022].

Eaton, B. et al., 2021. *NetCDF Climate and Forecast (CF).* 1.9 ed. s.l.:s.n.

EC, 2020. *A transparent & accessible ocean: Towards a Digital Twin of the Ocean.* [Online] Available at: https://ec.europa.eu/info/sites/default/files/research_and_innovation/green_deal/gdc_stakeholder_engagement_topic_09-3_digital_ocean.pdf [Accessed 12 September 2022].

EC, 2022a. *Destination Earth.* [Online] Available at: https://digital-strategy.ec.europa.eu/en/policies/destination-earth [Accessed 12 September 2022].

EC, 2022b. *EU Mission: Restore our Ocean and Waters.* [Online] Available at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters_en#what-are-eu-missions [Accessed 12 September 2022].

EC, 2022c. *European Digital Twin of the Ocean (European DTO).* [Online] Available at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters/european-digital-twin-ocean-european-dto_en [Accessed 12 September 2022].

EEA, 2021. *Oxygen concentrations in European coastal and marine waters.* [Online] Available at: https://www.eea.europa.eu/data-and-maps/indicators/oxygen-concentrations-in-coastal-and/assessment [Accessed 12 September 2022].

Emery, W. J. & Meincke, J., 1986. Global water masses: summary and review. *Oceanologica Acta,* Volume 9(4), pp. 383-391.

EMODnet, 2021. *European Marine Observation and Data Network (EMODnet).* [Online] Available at: https://emodnet.ec.europa.eu [Accessed 12 September 2022].

Encyclopedia, 2016. *Voxel.* [Online] Available at: https://www.encyclopedia.com/science-and-technology/computers-and-electrical-engineering/computers-and-computing/voxel [Accessed 12 September 2022].

EPA, 2021. *Climate Change Indicators: Sea Surface Temperature.* [Online] Available at: https://www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature [Accessed 12 September 2022].

ESA, 2021. *Surface ocean pH.* [Online] Available at: https://www.esa.int/ESA_Multimedia/Images/2015/01/Surface_ocean_pH [Accessed 12 September 2022].

Esri, 2022a. *Cross Validation (Geostatistical Analyst).* [Online] Available at: https://pro.arcgis.com/en/pro-app/2.8/tool-reference/geostatistical-analyst/cross-validation.htm [Accessed 12 September 2022].

Esri, 2022b. *GA Layer 3D To NetCDF (Geostatistical Analyst).* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/geostatistical-analyst/ga-layer-3d-to-netcdf.htm [Accessed 12 September 2022].

Esri, 2022c. *Interact with statistics.* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geoprocessing/data-engineering/view-statistics.htm [Accessed 12 September 2022].

Esri, 2022d. *Performing cross-validation and validation.* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/performing-cross-validation-and-validation.htm [Accessed 12 September 2022].

Esri, 2022e. *What is Empirical Bayesian Kriging 3D?.* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-3d-.htm [Accessed 12 September 2022].

Esri, 2022f. *What is a voxel layer?.* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/what-is-a-voxel-layer-.htm [Accessed 12 September 2022].

Esri, 2022g. *What is empirical Bayesian kriging?.* [Online] Available at: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm [Accessed 12 September 2022].

Fabijańczyk, P., Zawadzki, J. & Magiera, T., 2017. Magnetometric assessment of soil contamination in problematic area using empirical Bayesian and indicator kriging: A case study in Upper Silesia, Poland. *Geoderma,* Volume 308, pp. 69-77.

Fernandes, M. L. et al., 2020. Mapping the future: Pressures and impacts in the Portuguese maritime spatial planning. *Science of the Total Environment,* Volume 715, p. 136863.

Fiúza, A. F. G., 1983. Upwelling Patterns off Portugal. In: E. Suess & J. Thiede, eds. *Coastal Upwelling Its Sediment Record.* Boston: Springer, pp. 85-98.

Gamboa, D., Omira, R. & Terrinha, P., 2021. A database of submarine landslides offshore West and Southwest Iberia. *Scientific Data,* Volume 8, p. 185.

Gneiting, T. & Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association,* Volume 102(477), p. 359–378.

Göl, C., Bulut, S. & Bolat, F., 2017. Comparison of different interpolation methods for spatial distribution of soil organic carbon and some soil properties in the Black Sea backward region of Turkey. *Journal of African Earth Sciences,* Volume 134, pp. 85-91.

Goodchild, M. F., 2009. First Law of Geography. In: *International Encyclopedia of Human Geography.* s.l.:Elsevier, pp. 179-182.

GOOS, 2021. *Essential ocean variables.* [Online] Available at: https://www.goosocean.org/index.php?option=com_content&view=article&layout=edit&id=283&Itemid=441 [Accessed 12 September 2022].

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

Gregor, L. & Gruber, N., 2021. OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data,* Volume 13, pp. 777-808.

Gribov, A. & Krivoruchko, K., 2020. Empirical Bayesian kriging implementation and usage. *Science of the Total Environment,* Volume 722, p. 137290.

Grieves, M., 2019. Virtually Intelligent Product Systems: Digital and Physical Twins. In: S. Flumerfelt, K. G. Schwartz, D. Mavris & S. Briceno, eds. *Complex Systems Engineering: Theory and Practice.* s.l.:American Institute of Aeronautics and Astronautics, pp. 175-200.

Groom, S. et al., 2019. Satellite Ocean Colour: Current Status and Future Perspective. *Frontiers in Marine Science,* Volume 6, p. 485.

Gupta, A., Kamble, T. & Machiwal, D., 2017. Comparison of ordinary and Bayesian kriging techniques in depicting rainfall variability in arid and semi-arid regions of north-west India. *Environmental Earth Sciences,* Volume 76, p. 512.

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology,* Volume 332, pp. 144-157.

Harmouch, M., 2021. *17 Clustering Algorithms Used in Data Science and Mining.* [Online] Available at: https://towardsdatascience.com/17-clustering-algorithms-used-in-data-science-mining-49dbfa5bf69a [Accessed 12 September 2022].

Harris, P. T. & Whiteway, T., 2009. High seas marine protected areas: Benthic environmental conservation priorities from a GIS analysis of global ocean biophysical data. *Ocean & Coastal Management,* Volume 52(1), pp. 22-38.

Hartigan, J. A. & Wong, M. A., 1979. A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* Volume 28 (1), pp. 100-108.

Hayes, K. R. et al., 2015. Identifying indicators and essential variables for marine ecosystems. *Ecological Indicators,* Volume 57, pp. 409-419.

Hecht, M. et al., 2018. Utilization of data below the analytical limit of quantitation in pharmacokinetic analysis and modeling: promoting interdisciplinary debate. *Bioanalysis,* Volume 10:15, pp. 1229-1248.

Hengl, T., 2006. Finding the right pixel size. *Computer & Geosciences,* Volume 32, pp. 1283-1298.

Hennes, L., 2020. *Increase in Surface Temperature and Deep Layer Nitrate in the California Current: A Spatiotemporal Analysis of Four-Dimensional Hydrographic Data.* Los Angeles: University of Southern California.

Hughes, J. F. et al., 2014. *Computer Graphics: Principles and Practice.* 3.º ed. Willard, Ohio: Addison-Wesley.

IH, 2020. *Relatório de Missão: Cruzeiro IHPT-AQUIMAR2018-2.* Lisbon: Instituto Hidrográfico.

Iliad, 2022. *What is the Iliad Digital Twin of the Ocean?.* [Online] Available at: https://www.ocean-twin.eu [Accessed 12 September 2022].

Intel, 2011. *Voxel-Based Graphics on Intel Architectures.* [Online] Available at: https://software.intel.com/content/www/us/en/develop/articles/voxel-based-graphics-on-intel-architectures.html [Accessed 13 October 2021].

IOC; SCOR; IAPSO, 2010. *The international thermodynamic equation of seawater - 2010: Calculation and use of thermodynamic properties.* Paris: Intergovernamental Oceanographic Comission.

IPBES, 2022. *Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.* [Online] Available at: https://ipbes.net [Accessed 12 September 2022].

IPCC, 2021. 2021: Summary for Policymakers. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press, pp. 3-32.

Isaaks, E. H. & Srivastava, R. M., 1989. *Applied Geostatistics.* New York: Oxford University Press.

ISO, 2017. *ISO/IEC 17025:2017 - General requirements for the competence of testing and calibration laboratories.* [Online] Available at: https://www.iso.org/standard/66912.html [Accessed 12 September 2022].

Jain, A. K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters,* Volume 31(8), pp. 651-666.

Jayne, S. R. & Marotzke, J., 1999. A Destabilizing Thermohaline Circulation–Atmosphere–Sea Ice Feedback. *Journal of Climate,* Volume 12(2), pp. 642-651.

Jeffrey, S. W. & Humphrey, G. F., 1975. New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. *Biochemie und Physiologie der Pflanzen,* Volume 167:2, pp. 191-194.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K. & Lucas, N., 2001. *Using ArcGIS Geostatistical Analyst.* Redlands: Esri.

Jones, D. et al., 2020. Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology,* Volume 29, Part A, pp. 36-52.

Jürgens, K. & Taylor, G. T., 2018. Microbial Ecology and Biogeochemistry of Oxygen-Deficient Water Columns. In: J. W. &. Sons, ed. *Microbial Ecology of the Oceans.* Hoboken: Wiley Blackwell, pp. 231-288.

Kim, Y. H. et al., 2020. Application of satellite remote sensing in monitoring dissolved oxygen variabilities: A case study for coastal waters in Korea. *Environmental International,* Volume 134, p. 105301.

Kint, L. et al., 2020. Uncertainty assessment applied to marine subsurface datasets. *Quarterly Journal of Engineering Geology and Hydrogeology,* Volume 54, pp. 1-13.

Kravchenko, A. N., 2003. Influence of spatial structure on accuracy of interpolation methods. *Soil Science Society of America Journal,* Volume 67(5), pp. 1564-1571.

Krivoruchko, K., 2011. *Spatial Statistical Data Analysis for GIS Users.* Redlands: Esri.

Krivoruchko, K., 2012. *Empirical Bayesian Kriging - Implemented in ArcGIS Geostatistical Analyst.* [Online] Available at: https://www.esri.com/news/arcuser/1012/files/ebk.pdf [Accessed 12 September 2022].

Krivoruchko, K. & Gribov, A., 2019. Evaluation of empirical Bayesian kriging. *Spatial Statistics,* Volume 32, p. 100368.

Kutuzov, S. et al., 2019. Volume changes of Elbrus glaciers from 1997 to 2017. *Frontiers in Earth Science,* Volume 7, p. 153.

Larrondo, P. F., Neufeld, C. T. & Deutsch, C. V., 2003. VARFIT: A program for semi-automatic variogram modelling. *Centre for Computational Geostatistics Report 5,* p. 122.

Lauvset, S. K. et al., 2020. Processes Driving Global Interior Ocean pH Distribution. *Global Biogeochemical Cycles,* Volume 34(1), p. e2019GB006229.

Lauvset, S. K. et al., 2016. A new global interior ocean mapped climatology: the 1º x 1º GLODAP version 2. *Earth System Science Data,* Volume 8(2), pp. 325-340.

Leitão, F. et al., 2019. A 60-Year Time Series Analyses of the Upwelling along the Portuguese Coast. *Water,* Volume 11(6), p. 1285.

Lemos, R. T., 2004. The upwelling regime off the West Portuguese Coast, 1941–2000. *International Journal of Climatology,* Volume 24(4), pp. 511-524.

Levitus, S. et al., 2015. *World Ocean Atlas 2013 (NCEI Accession 0114815).* [Online] Available at: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0114815 [Accessed 12 September 2022].

Li, J. & Heap, A. D., 2008. *A Review of Spatial Interpolation Methods for Environmental Scientists.* Canberra: Australian Government - Geoscience Australia.

Li, J. & Heap, A. D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics,* Volume 6, pp. 228-241.

Li, J. & Heap, A. D., 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software,* Volume 53, pp. 173-189.

Linstone, H. A. & Turoff, M., 2002. *The Delphi Method: Techniques and Applications.* Newark: New Jersey Institute of Technology.

Li, T. et al., 2018a. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. *Science of The Total Environment,* Volume 628–629, pp. 1446-1459.

Li, Y. et al., 2020. Empirical Bayesian Kriging method to evaluate inter-annual water-table evolution in the Cuenca Alta del Río Laja aquifer, Guanajuato, México. *Journal of Hydrology,* Volume 582, p. 124517.

Li, Z. et al., 2018b. An automatic variogram modeling method with high reliability fitness and. *Computers and Geosciences,* Volume 120, pp. 48-59.

Lloyd, S. P., 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory,* Volume 28 (2), pp. 129-137.

Longhurst, A. R., 2007. *Ecological Geography of the Sea.* 2º ed. s.l.:Academic Press.

Lowe, A. T., Bos, J. & Ruesink, J., 2019. Ecosystem metabolism drives pH variability and modulates long-term ocean acidification in the Northeast Pacific coastal ocean. *Scientific Reports,* Volume 9, p. 963.

Lu, G. Y. & Wong, D. W., 2008. An adaptative inverse-distance weighting spatial interpolation technique. *Computer & Geosciences,* Volume 34(9), pp. 1044-1055.

MacKenzie, L. & Adamson, J., 2004. Water column stratification and the spatial and temporal distribution of phytoplankton biomass in Tasman Bay, New Zealand: Implications for aquaculture. *New Zealand Journal of Marine and Freshwater Research,* Volume 38:4, pp. 705-728.

Mackie, D., McGraw, C. & Hunter, K., 2011. *Ocean Adification.* [Online] Available at: https://skepticalscience.com/print.php?n=918 [Accessed 12 September 2022].

MacQueen, J., 1967. *Some methods for classification and analysis of multivariate observations.* Berkeley, University of California Press, pp. 281-297.

Mandal, S., Islam, M. S., Biswas, M. H. A. & Akter, S., 2022. A mathematical model applied to investigate the potential impact of global warming on marine ecosystems. *Applied Mathematical Modelling,* Volume 101, pp. 19-37.

Marrari, M., Piola, A. R. & Valla, D., 2017. Variability and 20-Year Trends in Satellite-Derived Surface Chlorophyll Concentrations in Large Marine Ecosystems around South and Western Central America. *Frontiers in Marine Science,* Volume 4:372.

Martínez, M. L. et al., 2007. The coasts of our world: Ecological, economic and social importance. *Ecological Economics,* pp. 254-272.

Martínez-Santos, P., Aristizábal, H. F., Díaz-Alcaide, S. & Gómez-Escalonilla, V., 2021. Predictive mapping of aquatic ecosystems by means of support vector machines and random forests. *Journal of Hydrology,* Volume 595, p. 126026.

MEA, 2005. *Millennium Ecosystem Assessment.* [Online] Available at: https://www.millenniumassessment.org [Accessed 12 September 2022].

Merriam-Webster, 2022a. *Ecosystem.* [Online] Available at: https://www.merriam-webster.com/dictionary/ecosystem [Accessed 12 November 2022].

Merriam-Webster, 2022b. *Voxel.* [Online] Available at: https://www.merriam-webster.com/dictionary/voxel [Accessed 12 September 2022].

Milbrandt, E. C. et al., 2021. Geospatial distribution of hypoxia associated with a Karenia brevis bloom. *Estuarine, Coastal and Shelf Science,* Volume 259, p. 107446.

Miloslavich, P., Bax, N. J. & Simmons, S. E., 2018. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology,* Volume 24(6), pp. 2416-2433.

Minnett, P. J. et al., 2019. Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment,* Volume 233, p. 111366.

Mirzaei, R. & Sakizadeh, M., 2016. Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. *Environmental Science and Pollution Research,* Volume 23(3), pp. 2758-2769.

Mulder, M. J. et al., 2019. Size and shape matter: The impact of voxel geometry on the identification of small nuclei. *PLoS ONE,* Volume 14(4), p. e0215382.

Murphy, J. & Riley, J. P., 1962. A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta,* Volume 27, pp. 31-36.

NASA, 2022. *Chlorophyll.* [Online] Available at: https://earthobservatory.nasa.gov/global-maps/MY1DMM_CHLORA [Accessed 12 September 2022].

Neumann, B., Vafeidis, A. T., Zimmermann, J. & Nicholls, R. J., 2015. Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment. *PLoS One,* Volume 10(3), p. e0118571.

NOAA, 2020. *Sea Surface Temperature - Pathfinder.* [Online] Available at: https://www.ncdc.noaa.gov/cdr/oceanic/sea-surface-temperature-pathfinder [Accessed 23 November 2020].

NOAA, 2022. *Global Drifter Program.* [Online] Available at: https://www.aoml.noaa.gov/global-drifter-program/ [Accessed 12 September 2022].

Nobre, A., Godinho, S., Nunes, P. & Mosqueira, M., 2020. *Interpolação de variáveis oceanográficas com distribuição tridimensional através do método Empirical Bayesian Kriging 3D.* Lisboa, Instituto Hidrográfico, pp. 454-457.

Nocco, M. A., Smail, R. A. & Kucharik, C. J., 2019. Observation of irrigation-induced climate change in the midwest United States. *Global Change Biology,* Volume 25(10), pp. 3472-3484.

Nogueira, A. A. et al., 2019. Spatial variation in phylogenetic diversity of communities of Atlantic Forest harvestmen (Opiliones, Arachnida). *Insect Conservation and Diversity,* Volume 12(5), pp. 414-426.

Noi, E. & Murray, A. T., 2022. Interpolation biases in assessing spatial heterogeneity of outdoor air quality in Moscow, Russia. *Land Use Policy,* Volume 112, p. 105783.

O'Keefe, J., 2014. *Spatial Cells in the Hippocampal Formation.* London, University College London.

Oceana, 2022. *Marine Life.* [Online] Available at: https://oceana.org/marine-life/marine-science-and-ecosystems/open-ocean [Accessed 12 September 2022].

OGC, 2022. *Web Map Service.* [Online] Available at: https://www.ogc.org/standards/wms [Accessed 12 September 2022].

O'Higgins, T. G., Lago, M. & DeWitt, T. H., 2021. *Ecosystem-Based Management, Ecosystem Services and Aquatic Biodiversity.* Cham: Springer Open.

Oliver, M. J. et al., 2004. Bioinformatic approaches for objective detection of water masses on continental shelves. *Journal of Geophysical Research - Oceans,* Volume 109, p. C07S04.

Oliver, M. J. & Irwin, A. J., 2008. Objective global ocean biogeographic provinces. *Geophysical Research Letters,* Volume 35(15), p. L15601.

Pastuszak, M. et al., 1999. Impact of water temperature on nutrient concentrations in the Oder estuary in 1996–1998. *Deutsche Hydrografische Zeitschrift,* Volume 51, pp. 423-439.

Pawlowicz, R., 2013. Key Physical Variables in the Ocean: Temperature, Salinity, and Density. *Nature Education Knowledge,* Volume 4(4):13.

PCM, 2021a. *Resolução do Conselho de Ministros n.º 68/2021: Aprova a Estratégia Nacional para o Mar 2021-2030.* [Online] Available at: https://dre.pt/application/file/a/164651425 [Accessed 12 September 2022].

PCM, 2021b. *Resolução do Conselho de Ministros n.º 120/2021: Aprova o plano de ação da Estratégia Nacional para o Mar 2021/2030.* [Online] Available at: https://dre.pt/application/file/a/170591877 [Accessed 12 September 2022].

Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* Volume 12, pp. 2825-2830.

Pickard, G. L. & Emery, W. J., 1990. *Descriptive Physical Oceanography: An Introduction.* 5th ed. Burlington: Butterworth-Heinemann.

Pilz, J. & Spöck, G., 2008. Why do we need and how should we implement Bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment,* Volume 22, pp. 621-632.

Pinet, P. R., 2009. *Invitation to Oceanography.* 5th ed. Sudbury: Jones and Bartlett Publishers.

Pottier, A. et al., 2021. Mapping coastal marine ecosystems of the National Park of Banc d'Arguin (PNBA) in Mauritania using Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation,* Volume 102, p. 102419.

Reijden, K. J. et al., 2021. Beyond connecting the dots: A multi-scale, multi-resolution approach to marine habitat mapping. *Ecological Indicators,* Volume 128, p. 107849.

Reul, N. et al., 2020. Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010–2019). *Remote Sensing of Environment,* Volume 242, p. 111769.

Reygondeau, G. et al., 2017a. Global biogeochemical provinces of the mesopelagic zone. *Journal of Biogeography,* Volume 45(2), pp. 500-514.

Reygondeau, G. et al., 2017b. Biogeochemical regions of the Mediterranean Sea: An objective multidimensional and multivariate environmental approach. *Progress in Oceanography,* Volume 151, pp. 138-148.

Reygondeau, G. et al., 2013. Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles,* Volume 27(4), pp. 1046-1058.

Roelfsema, C. M. et al., 2020. Habitat maps to enhance monitoring and management of the Great Barrier Reef. *Coral Reefs,* Volume 39, pp. 1039-1054.

Rousseeuw, P. J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* Volume 20, pp. 53-65.

Ryabinin, V. et al., 2019. The UN Decade of Ocean Science for Sustainable Development. *Frontiers in Marine Science,* Volume 6, p. 470.

Sahlin, J., Mostafavi, M. A., Forest, A. & Babin, M., 2014. Assessment of 3D Spatial Interpolation Methods for Study of the Marine Pelagic Environment. *Marine Geodesy,* Volume 37:2, pp. 238-266.

Samsonova, V. P., Blagoveshchenskii, Y. N. & Meshalkina, Y. L., 2017. Use of empirical Bayesian kriging for revealing heterogeneities in the distribution of organic carbon on agricultural lands. *Eurasian Soil Science,* Volume 50, pp. 305-311.

Sayre, R. et al., 2017a. *A New Map of Global Ecologic Marine Units - An Environmental Stratification Approach.* Washington, DC: American Association of Geographers.

Sayre, R. et al., 2017b. A Three-Dimensional Mapping of the Ocean Based on Environmental Data. *Oceanography,* Volume 30 (1), pp. 90-103.

Seabed 2030, 2022. *Seabed 2030 Project - Mapping progress.* [Online] Available at: https://seabed2030.org/mapping-progress [Accessed 12 September 2022].

SeaDataNet, 2022a. *L20 (SeaDataNet Measurand Qualifier Flags).* [Online] Available at: http://seadatanet.maris2.nl/v_bodc_vocab_v2/browse.asp?order=conceptid&formname=search&screen=0&lib=l20 [Accessed 12 September 2022].

SeaDataNet, 2022b. *Pan-European Infrastructure for Ocean & Marine Data Management.* [Online] Available at: https://www.seadatanet.org [Accessed 12 September 2022].

Sherman, K. & Alexander, L. M., 1986. *Variability and Management of Large Marine Ecosystems.* Boulder: Westview Press.

Sigman, D. M. & Hain, M. P., 2012. The Biological Productivity of the Ocean. *Nature Education Knowledge,* Volume 3(10):21.

Smithsonian, 2018. *Ocean Acidification.* [Online] Available at: https://ocean.si.edu/ocean-life/invertebrates/ocean-acidification [Accessed 12 September 2022].

Smyth, K. & Elliott, M., 2016. Effects of changing salinity on the ecology of the marine environment. In: *Stressors in the Marine Environment: Physiological and ecological responses; societal implications.* s.l.:Oxford Scholarship Online, pp. 161-174.

Soni, D., 2018. *Supervised vs. Unsupervised Learning.* [Online] Available at: https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d [Accessed 12 September 2022].

SOOS, 2022. *Developing ecosystem essential ocean variables for the Southern Ocean ecosystem (eEOVs) and routine delivery of products.* [Online] Available at: https://www.soos.aq/activities/task-teams/eeovs [Accessed 12 September 2022].

Spalding, M. D., Agostini, V. N., Rice, J. & Grant, S. M., 2012. Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management,* Volume 60, pp. 19-30.

Spalding, M. D. et al., 2007. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience,* Volume 57(7), pp. 573-583.

Spalding, M. D. et al., 2014. The role of ecosystems in coastal protection: Adapting to climate change and coastal hazards. *Ocean & Coastal Management,* Volume 90, pp. 50-57.

Stewart, R. H., 2008. *Introduction To Physical Oceanography.* 1.ª ed. College Station: Texas A&M University.

Stramma, L. et al., 2012. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. *Nature Climate Change,* Volume 2, pp. 33-37.

Strickland, J. D. H. & Parsons, T. R., 1972. *A practical handbook of seawater analysis.* Ottawa: Fisheries Research Board of Canada.

Sutton, T. T., 2017. A global biogeographic classification of the mesopelagic zone. *Deep-Sea Research Part I,* Volume 126, pp. 85-102.

Sverdrup, H. U., Johnson, M. W. & Fleming, R. H., 1942. *The Oceans - Their Physics, Chemistry and General Biology.* New York: Prentice-Hall.

TEEB, 2022. *The Economics of Ecosystems & Biodiversity.* [Online] Available at: http://teebweb.org [Accessed 12 September 2022].

Thushara, V. et al., 2019. Vertical distribution of chlorophyll in dynamically distinct regions of the southern Bay of Bengal. *Biogeosciences,* Volume 16, p. 1447–1468.

Tibshirani, R., Walther, G. & Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society - Series B - Statistical Methodology,* Volume 63(2), pp. 411-423.

Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography,* Volume 46(2), pp. 234-240.

Tomczak, M., 1999. Some historical, theoretical and applied aspects of quantitative water mass analysis. *Journal of Marine Research,* Volume 57, pp. 275-303.

Troupin, C. et al., 2012. Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA). *Ocean Modelling,* Volume 52-53, pp. 90-101.

UN, 2007. *Percentage of total population living in coastal areas.* [Online] Available at: https://www.un.org/esa/sustdev/natlinfo/indicators/methodology_sheets/oceans_seas_coasts/pop_coastal_areas.pdf [Accessed 12 September 2022].

UN, 2015. *Transforming our world: the 2030 Agenda for Sustainable Development.* [Online] Available at: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf [Accessed 12 September 2022].

UN, 2020. *The Science We Need for the Ocean We Want.* [Online] Available at: https://www.oceandecade.org/wp-content/uploads//2021/10/346780-The%20Science%20We%20Need%20For%20The%20Ocean%20We%20Want [Accessed 12 September 2022].

UN, 2021. *Ocean Decade - Implementation Plan.* [Online] Available at: https://www.oceandecade.org/wp-content/uploads//2021/09/337567-Ocean%20Decade%20Implementation%20Plan%20-%20Full%20Document [Accessed 12 September 2022].

UN, 2022. *10 Ocean Decade Challenges for collective impact.* [Online] Available at: https://www.oceandecade.org/challenges [Accessed 12 September 2022].

UNEP, 2006. *Marine and coastal ecosystems and human well-being: A synthesis report based on the findings of the Millennium Ecosystem Assessment.* Nairobi, Kenya: United Nations Environment Programme.

UNEP, 2011. *Taking Steps toward Marine and Coastal Ecosystem-Based Management - An Introductory Guide.* Nairobi, Kenya: United Nations Environment Programme.

UNESCO, 2009. *Global Open Oceans and Deep Seabed (GOODS) - Biogeographic Classification.* IOC Technical Series, 84 ed. Paris: UNESCO-IOC.

Unidata, 2022a. *NetCDF Conventions.* [Online] Available at: https://www.unidata.ucar.edu/software/netcdf/conventions.html [Accessed 12 September 2022].

Unidata, 2022b. *Network Common Data Form (NetCDF).* [Online] Available at: https://www.unidata.ucar.edu/software/netcdf [Accessed 12 September 2022].

University of Reading, 2010. *ncWMS User Guide.* [Online] Available at: https://reading-escience-centre.gitbooks.io/ncwms-user-guide/content [Accessed 12 September 2022].

Vaquer-Sunyer, R. & Duarte, C. M., 2008. Thresholds of hypoxia for marine biodiversity. *PNAS,* Volume 105 (40), pp. 15452-15457.

Vinogradova, N. et al., 2019. Satellite Salinity Observing System: Recent Discoveries and the Way Forward. *Frontiers in Marine Science,* Volume 6:243.

Visbeck, M., 2018. Ocean science research is key for a sustainable future. *Nature Communications,* Volume 9, p. 690.

Voxelmaps, 2021. *Bulding a 4D Volumetric "Digital Twin" of the Planet.* [Online] Available at: https://www.voxelmaps.com [Accessed 12 September 2022].

Wang, D. et al., 2018. Satellite Retrieval of Surface Water Nutrients in the Coastal Regions of the East China Sea. *Remote Sensing,* Volume 10(12), p. 1896.

Wang, M. et al., 2017. Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China. *Remote Sensing,* Volume 9(12), p. 1278.

Watling, L., Guinotte, J., Clark, M. R. & Smith, C. R., 2013. A proposed biogeography of the deep ocean floor. *Progress in Oceanography,* Volume 111, pp. 91-112.

Webb, P., 2021. *Introduction to Oceanography.* s.l.:Pressbooks.

Weber, S. C. et al., 2019. Habitat Delineation in Highly Variable Marine Environments. *Frontiers in Marine Science,* Volume 6, p. 112.

WEF, 2021. *The Global Risks Report 2021.* 16 ed. Geneva: World Economic Forum.

WikiGIS, 2011. *Voxel.* [Online] Available at: http://wiki.gis.com/wiki/index.php/Voxel [Accessed 12 September 2022].

Wilkinson, M. D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data,* Volume 3, p. 160018.

Wilson, J., Ucharm, G. & Beman, J. M., 2019. Climatic, physical, and biogeochemical changes drive rapid oxygen loss and recovery in a marine ecosystem. *Scientific Reports,* Volume 9, p. 16114.

Wong, D., Yuan, L. & Perlin, S., 2004. Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science & Environmental Epidemiology,* Volume 14, pp. 404-415.

WOR, 2010. *Oxygen in the ocean.* [Online] Available at: https://worldoceanreview.com/en/wor-1/ocean-chemistry/oxygen [Accessed 12 September 2022].

Wunsch, C., 2002. What Is the Thermohaline Circulation?. *Science,* Volume 298(5596), pp. 1179-1181.

Wu, Y.-H. & Hung, M.-C., 2016. Comparison of Spatial Interpolation Techniques Using Visualization and Quantitative Assessment. In: *Applications of Spatial Statistics.* s.l.:IntechOpen.

Yan, P., Peng, H., Luobin, Y. & Lin, K., 2019. Spatial Variability of Soil Physical Properties Based on GIS and Geo-Statistical Methods in the Red Beds of the Nanxiong Basin, China. *Polish Journal of Environment Studies,* Volume 28(4), pp. 2961-2972.

Yuval, Levy, I. & Broday, D. M., 2017. Improving modeled air pollution concentration maps by residual interpolation. *Science of the Total Environment,* Volume 598, pp. 780-788.

Zamboni, N. S., Filho, E. M. N. & Carvalho, A. R., 2021. Unfolding differences in the distribution of coastal marine ecosystem services values among developed and developing countries. *Ecological Economics,* Volume 189, p. 107151.

Zhao, Q., Basher, Z. & Costello, M. J., 2020. Mapping near surface global marine ecosystems through cluster analysis of environmental data. *Ecological Research,* Volume 35(2), pp. 327-342.

Zhao, Q. & Costello, M. J., 2019. Summer and winter ecosystems of the world ocean photic zone. *Ecological Research,* Volume 34, pp. 457-471.

ANNEX A – DATA AND STUDY AREAS

Stations	Statistics (m)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
CTD Stations	Mean	7 093	7 356	7 122	7 250	4 541	4 550	4 526	4 521
	Median	7 226	7 315	7 315	7 293	4 729	4 596	4 546	4 482
	Standard Deviation	855	578	856	324	983	694	674	666
	Minimum	4 079	6 676	4 118	6 537	77	2 486	2 467	2 476
	Maximum	9 970	10 909	10 012	7 834	6 838	5 936	5 798	5 834
Water Stations	Mean	7 303	7 357	7 331	6 997	5 796	5 336	5 144	5 169
	Median	7 247	7 314	7 356	7 293	5 832	5 610	5 553	5 558
	Standard Deviation	622	577	557	1 385	1 626	866	950	957
	Minimum	6 407	6 686	6 600	6 535	3 233	3 386	3 300	3 207
	Maximum	11 355	10 907	10 964	7 833	8 954	6 284	6 253	6 287

Table A-1. Horizontal distance statistics between sampling stations.

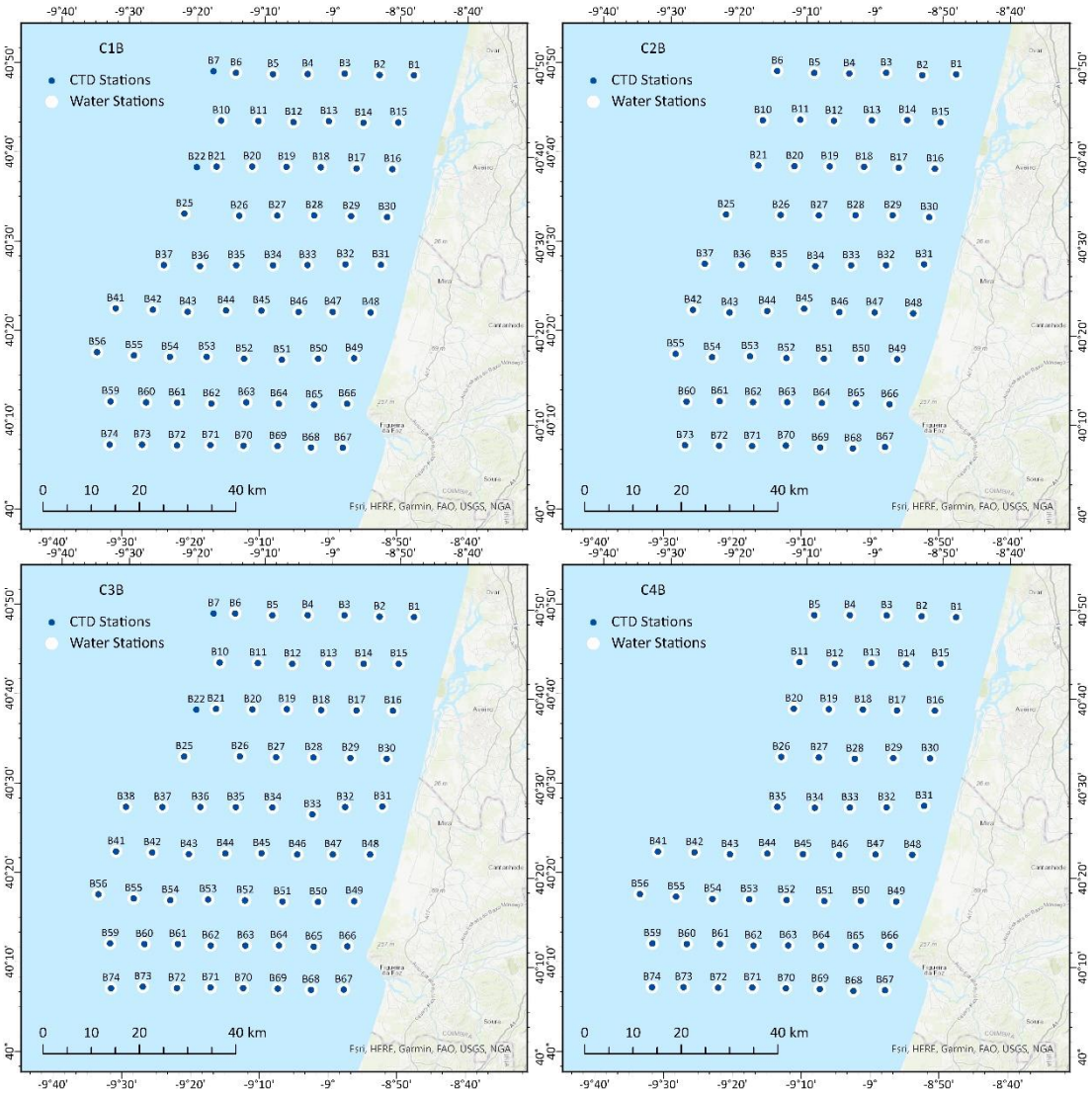


Figure A-1. Sampling stations in area B.

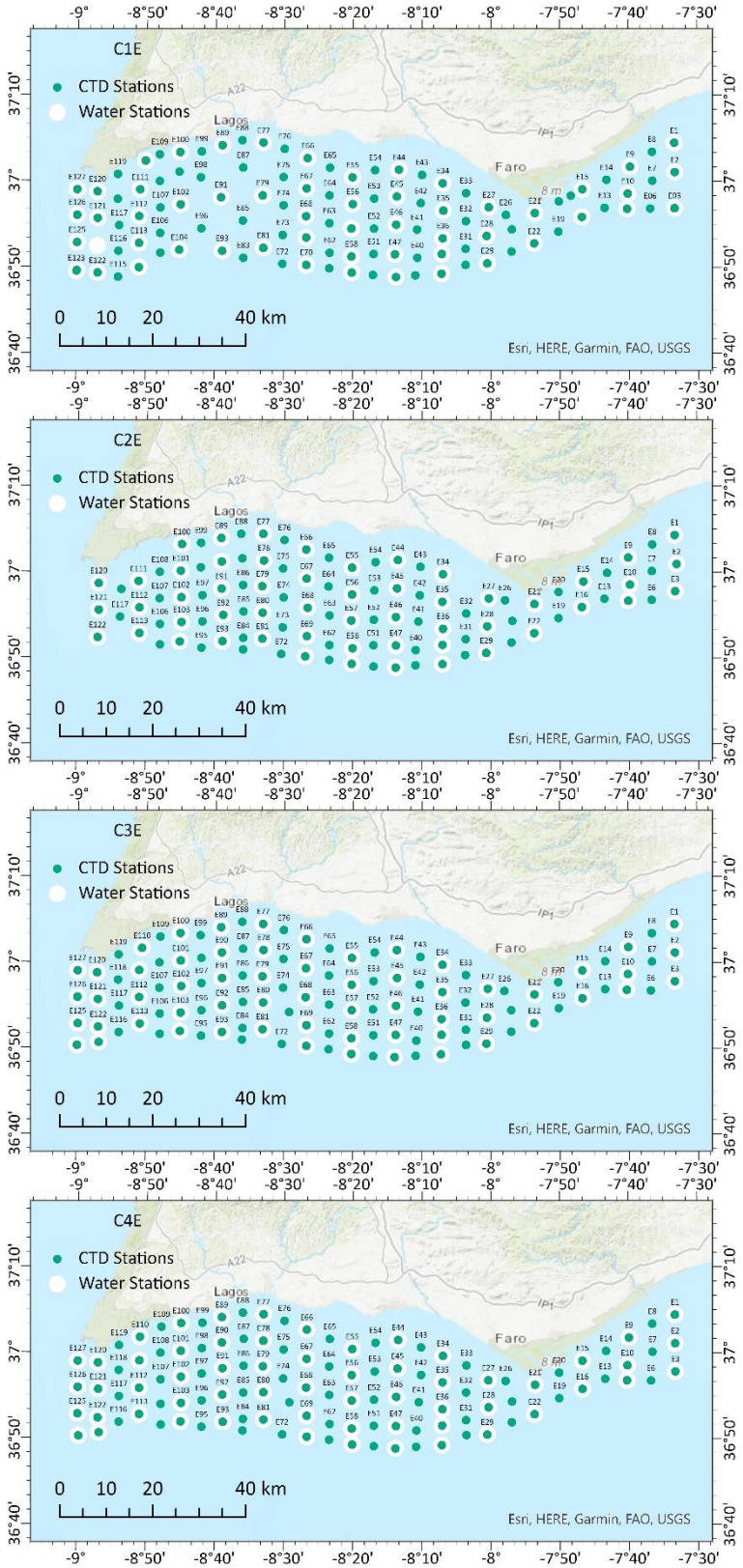


Figure A-2. Sampling stations in area E.

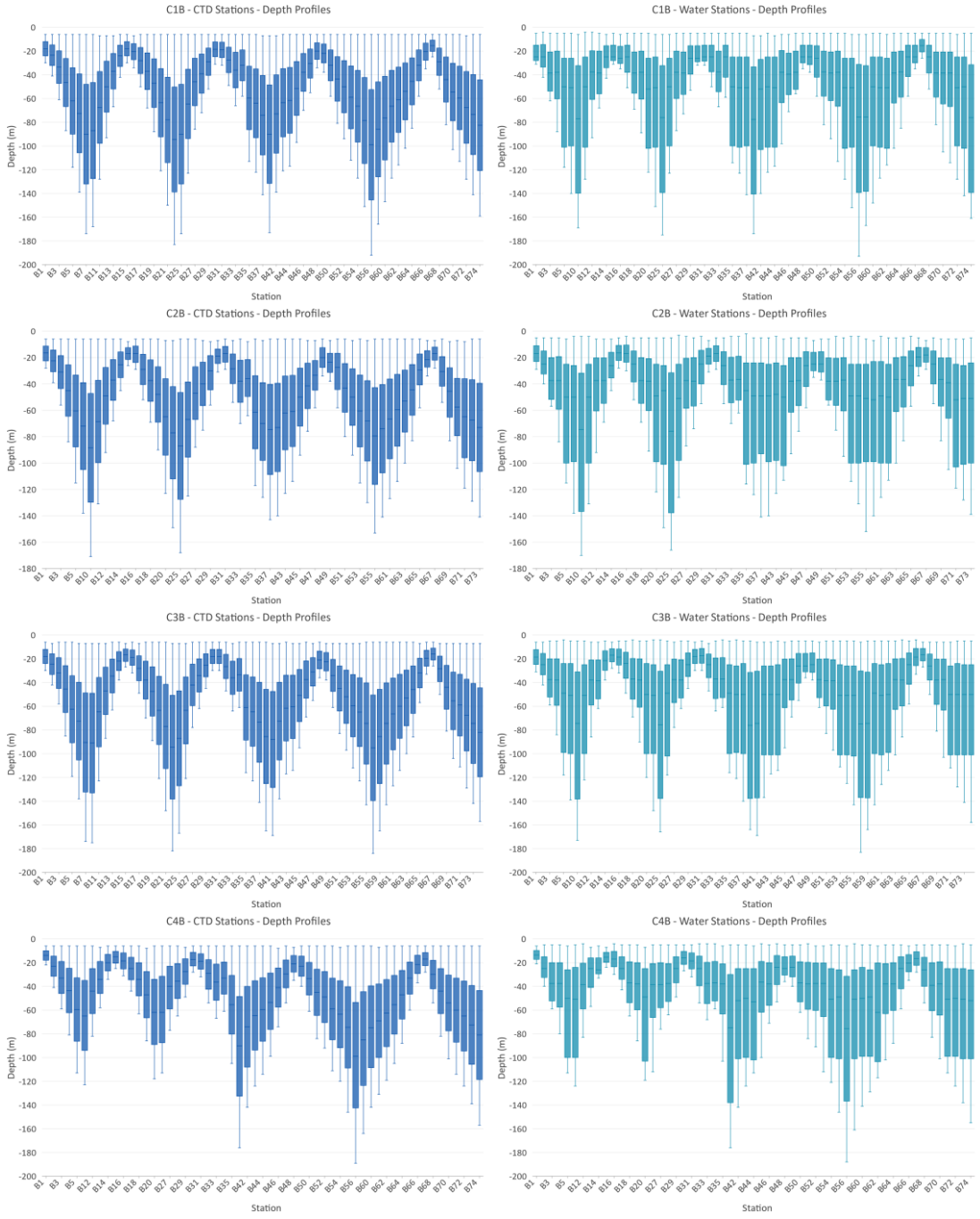


Figure A-3. Depth profiles in area B.

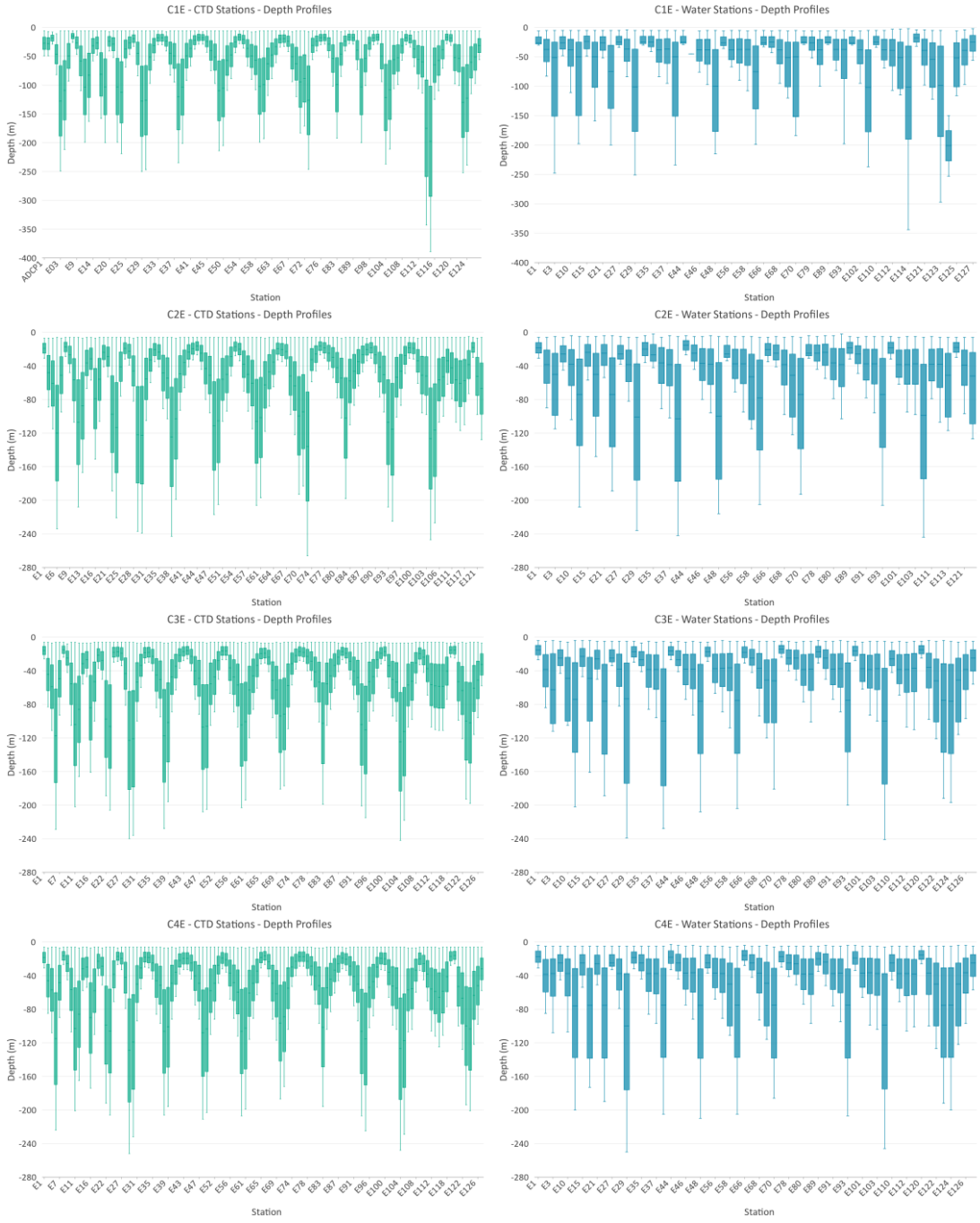


Figure A-4. Depth profiles in area E.

ANNEX B – EXPLORATORY AND SPATIAL DATA ANALYSIS

Temperature (°C)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	14.78	13.99	14.55	14.21	16.99	14.75	15.14	14.94
Median	14.30	13.85	13.99	14.27	15.73	14.79	14.68	15.14
Standard Deviation	1.50	0.73	1.58	0.36	3.16	0.79	1.44	0.80
Coefficient of Variation	10.1%	5.2%	10.8%	2.5%	18.6%	5.4%	9.5%	5.4%
Samples	5 956	4 835	5 991	4 584	10 704	9 890	10 158	10 407
Minimum	12.95	12.31	12.56	12.80	12.65	12.71	13.42	12.91
Maximum	18.28	16.09	18.72	14.87	23.67	16.54	20.53	16.25
Skewness	0.47	0.56	0.59	-0.98	0.73	-0.09	1.11	-0.73
Kurtosis	1.73	2.75	1.97	4.00	2.15	2.06	3.62	2.63

Table B-1. Descriptive statistics of temperature.

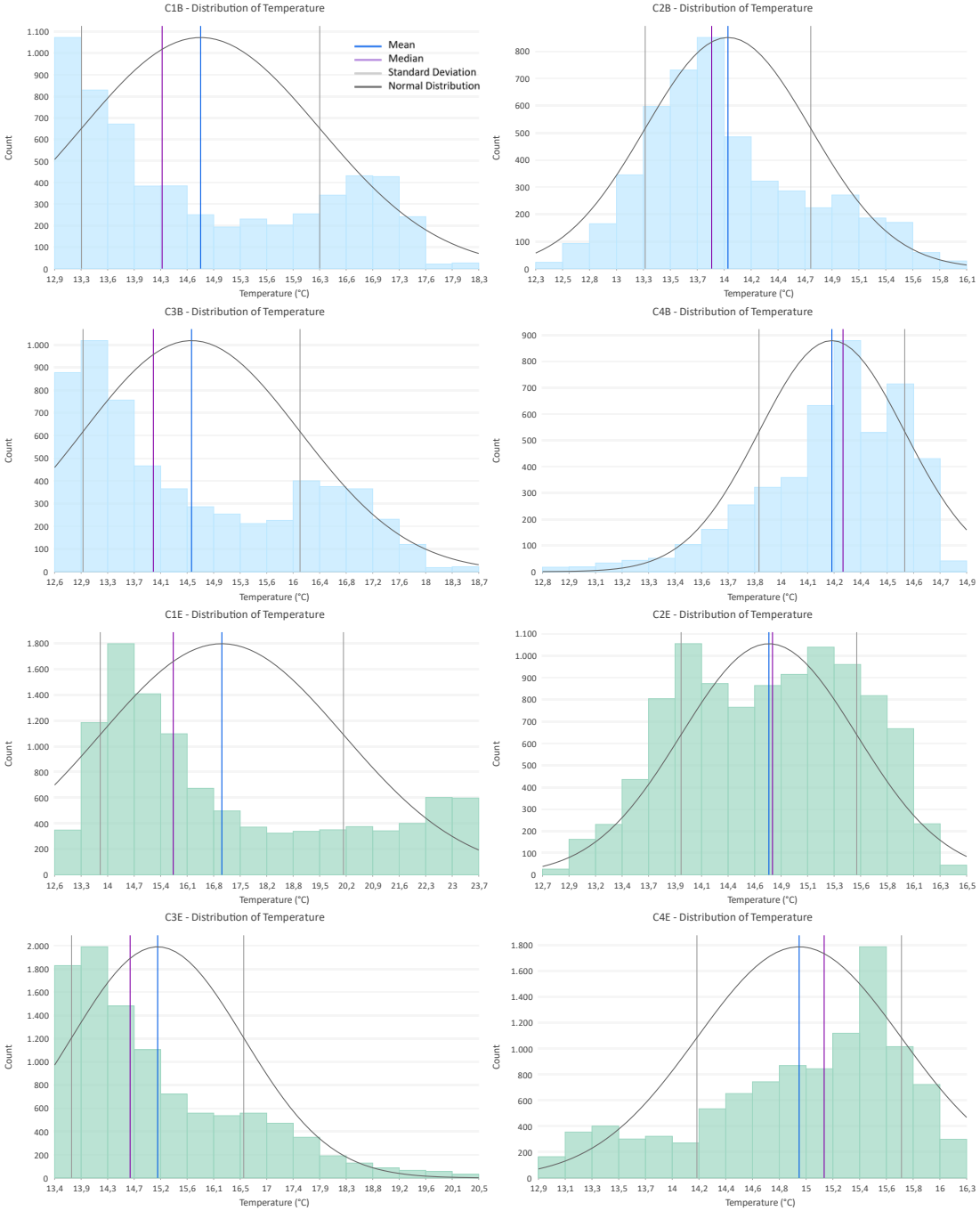


Figure B-1. Histograms of temperature.

Salinity (PSU)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	35.86	35.73	35.80	35.80	36.26	36.06	36.05	36.23
Median	35.86	35.80	35.81	35.89	36.21	36.07	36.04	36.25
Standard Deviation	0.05	0.15	0.09	0.19	0.26	0.10	0.10	0.11
Coefficient of Variation	0.1%	0.4%	0.2%	0.5%	0.7%	0.3%	0.3%	0.3%
Samples	5 956	4 835	5 991	4 584	10 704	9 890	10 158	10 407
Minimum	35.52	35.17	35.36	34.50	35.81	35.77	35.90	35.93
Maximum	36.01	35.88	36.06	36.03	37.06	36.32	36.66	36.48
Skewness	-0.72	-1.46	-0.70	-1.74	0.66	-0.32	1.67	-0.37
Kurtosis	5.41	4.09	5.10	6.47	2.82	2.42	7.62	2.69

Table B-2. Descriptive statistics of salinity.

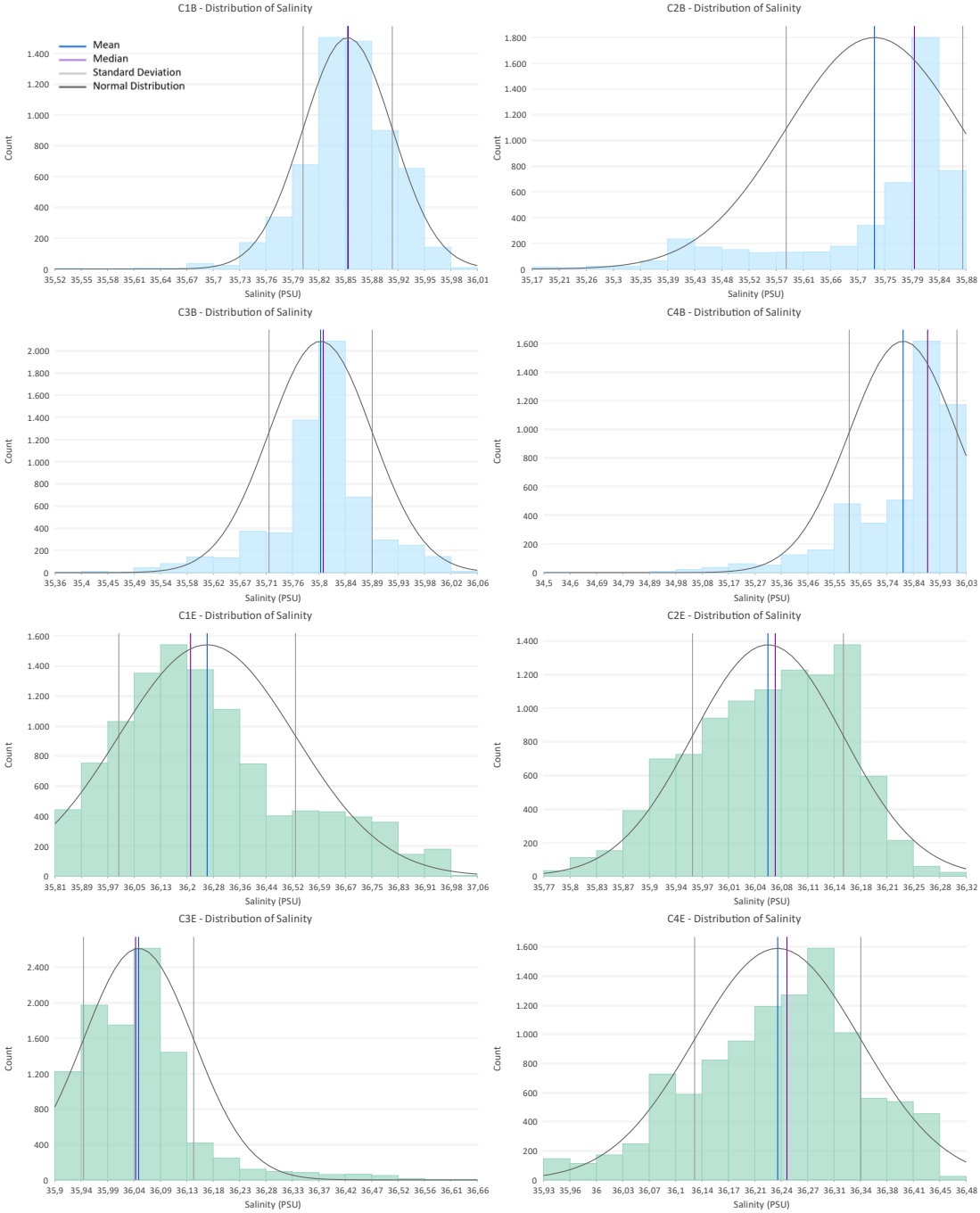


Figure B-2. Histograms of salinity.

Dissolved Oxygen (mgO_2/L)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	7.15	7.96	6.85	7.67	7.19	7.82	6.84	7.17
Median	7.22	8.07	6.93	7.57	7.17	7.87	6.88	7.21
Standard Deviation	0.72	0.67	0.61	0.61	0.70	0.78	0.50	0.43
Coefficient of Variation	10.1%	8.4%	8.8%	8.0%	9.8%	9.9%	7.2%	6.1%
Samples	264	236	265	185	219	215	236	241
Minimum	5.47	6.02	5.00	6.36	5.59	6.19	5.61	5.14
Maximum	9.45	9.95	8.46	9.70	9.33	9.36	8.75	8.15
Skewness	0.21	-0.21	-0.18	0.73	0.37	0.13	0.08	-0.83
Kurtosis	2.68	2.92	2.71	3.77	3.04	1.91	3.65	5.64

Table B-3. Descriptive statistics of DO.

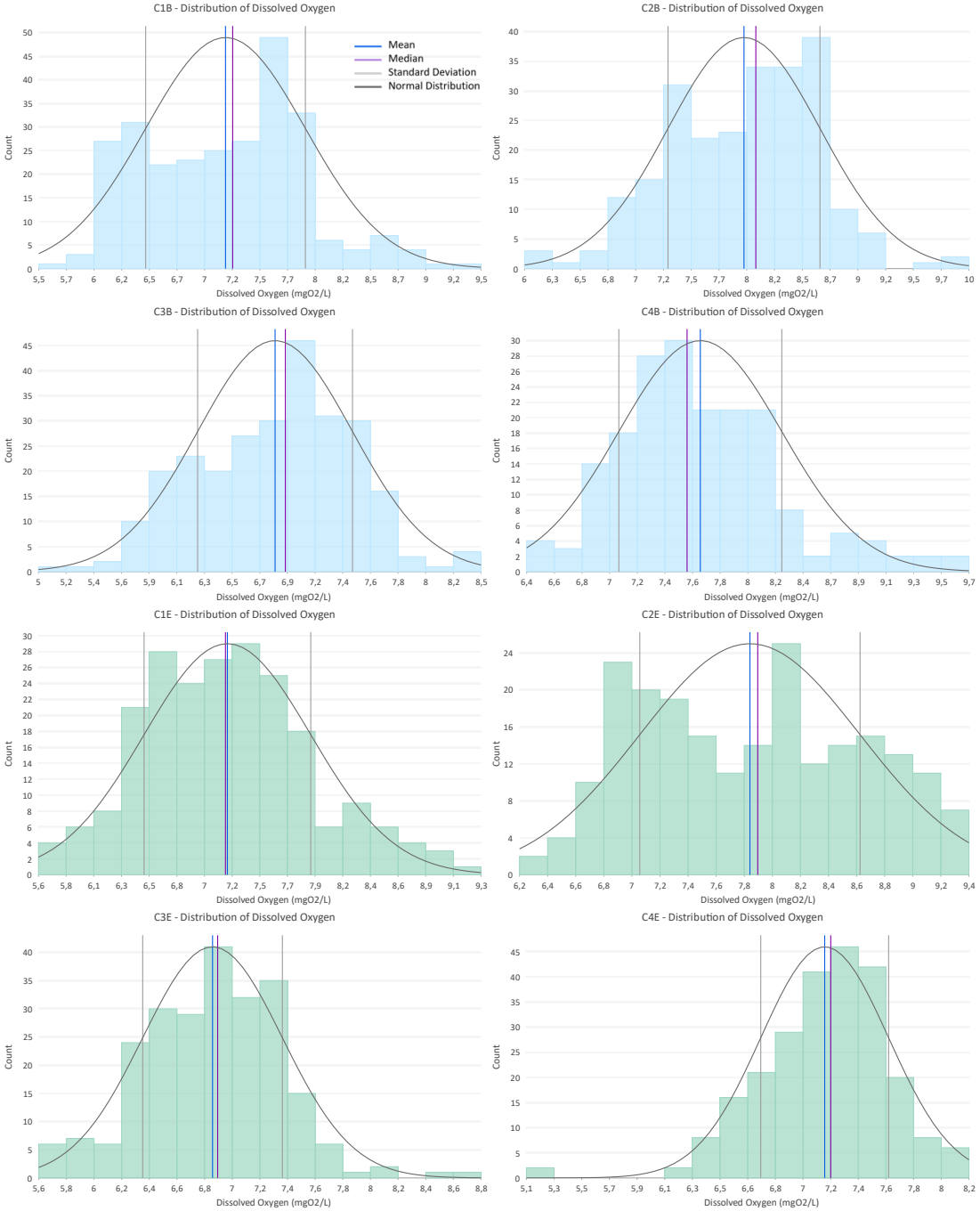


Figure B-3. Histograms of DO.

PO_4 (µmol/L)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	0.37	0.32	0.40	0.27	0.35	0.37	0.35	0.25
Median	0.36	0.25	0.36	0.21	0.25	0.32	0.34	0.21
Standard Deviation	0.23	0.22	0.27	0.18	0.28	0.24	0.23	0.17
Coefficient of Variation	61.4%	69.9%	68.4%	66.0%	79.4%	65.2%	65.9%	68.9%
Samples	262	230	264	221	222	218	238	240
Minimum	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Maximum	1.06	0.77	0.95	0.64	1.51	1.02	0.77	0.73
Skewness	0.28	0.34	0.28	0.41	1.26	0.40	0.16	0.83
Kurtosis	1.92	1.44	1.62	1.60	4.85	1.87	1.41	2.60

Table B-4. Descriptive statistics of PO_4.

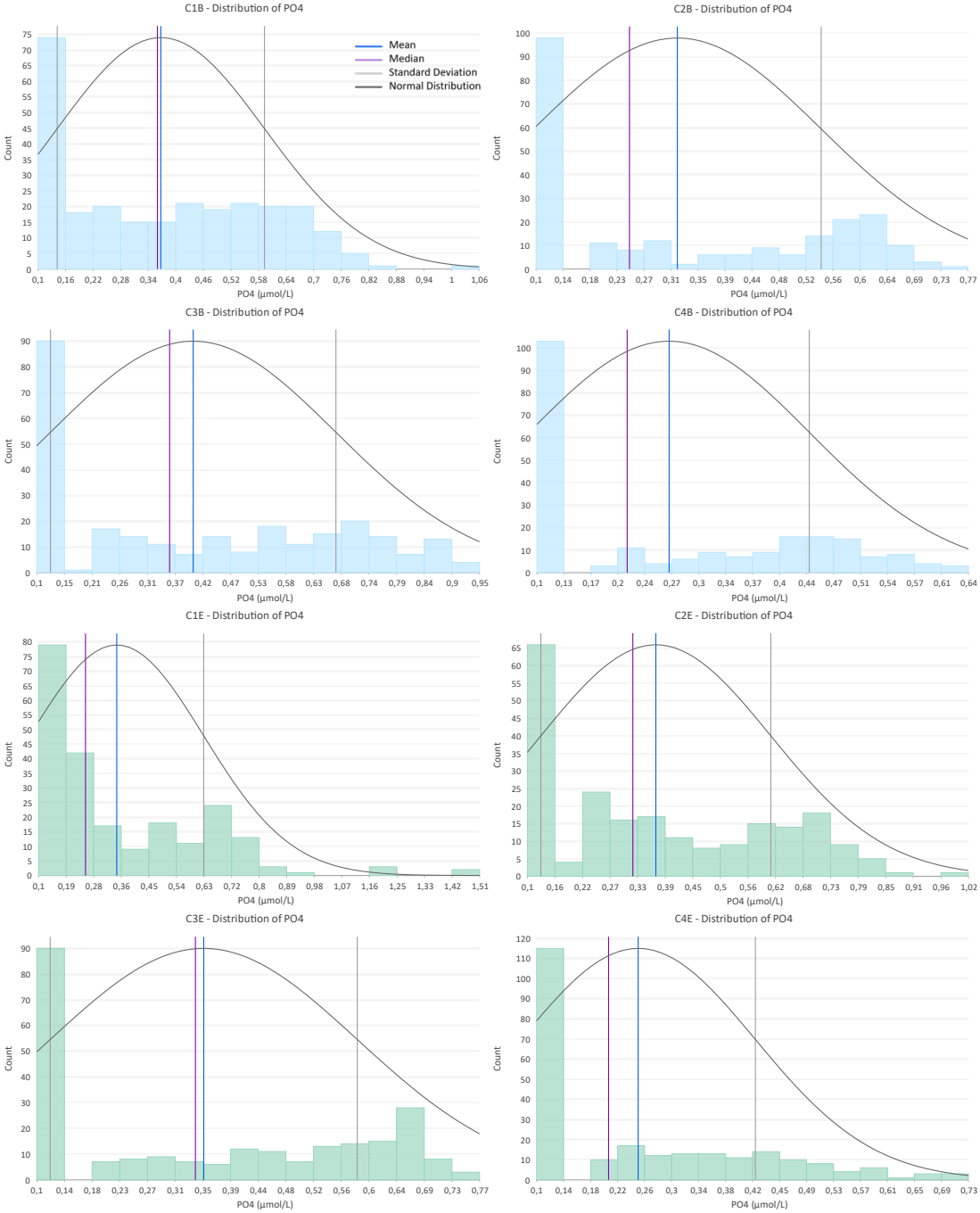


Figure B-4. Histograms of PO_4.

NO_X (µmol/L)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	4.63	4.22	5.09	3.76	3.70	5.83	4.91	3.85
Median	3.90	3.00	4.50	3.02	1.50	5.80	4.50	2.31
Standard Deviation	3.89	3.71	4.31	3.22	4.11	4.30	4.04	3.44
Coefficient of Variation	84.0%	88.0%	84.7%	85.7%	111.2%	73.9%	82.4%	89.5%
Samples	263	232	264	222	222	218	238	239
Minimum	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Maximum	13.10	11.10	14.00	14.40	15.10	19.00	12.10	13.59
Skewness	0.38	0.30	0.27	0.47	1.06	0.49	0.18	0.85
Kurtosis	1.69	1.43	1.56	2.07	2.80	2.55	1.40	2.56

Table B-5. Descriptive statistics of NO_X.

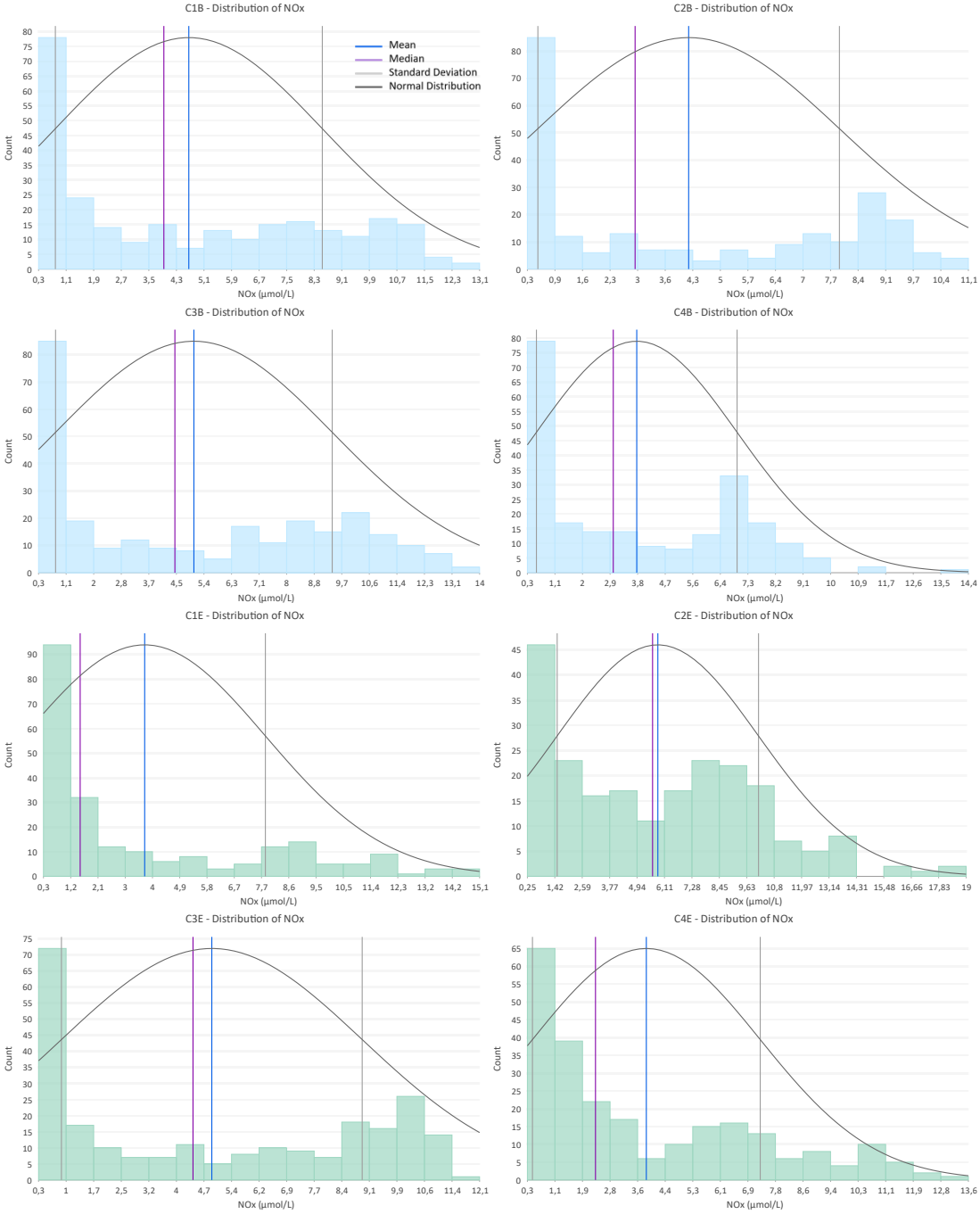


Figure B-5. Histograms of NO_X.

SiO_2 (µmol/L)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	3.17	2.53	3.75	2.43	3.24	2.43	3.09	2.45
Median	2.70	2.00	3.30	2.39	2.60	2.30	3.10	2.38
Standard Deviation	1.80	1.64	2.06	1.29	2.20	1.20	1.69	1.13
Coefficient of Variation	56.9%	64.8%	54.8%	53.1%	68.0%	49.4%	54.5%	46.1%
Samples	263	232	262	222	221	218	238	240
Minimum	0.90	0.36	0.33	0.42	0.60	0.60	0.60	0.80
Maximum	12.00	8.20	12.00	6.56	11.00	6.10	7.60	5.61
Skewness	1.24	0.70	0.67	0.53	1.37	0.43	0.41	0.49
Kurtosis	5.23	2.79	3.06	2.68	4.75	2.48	2.27	2.43

Table B-6. Descriptive statistics of SiO_2.

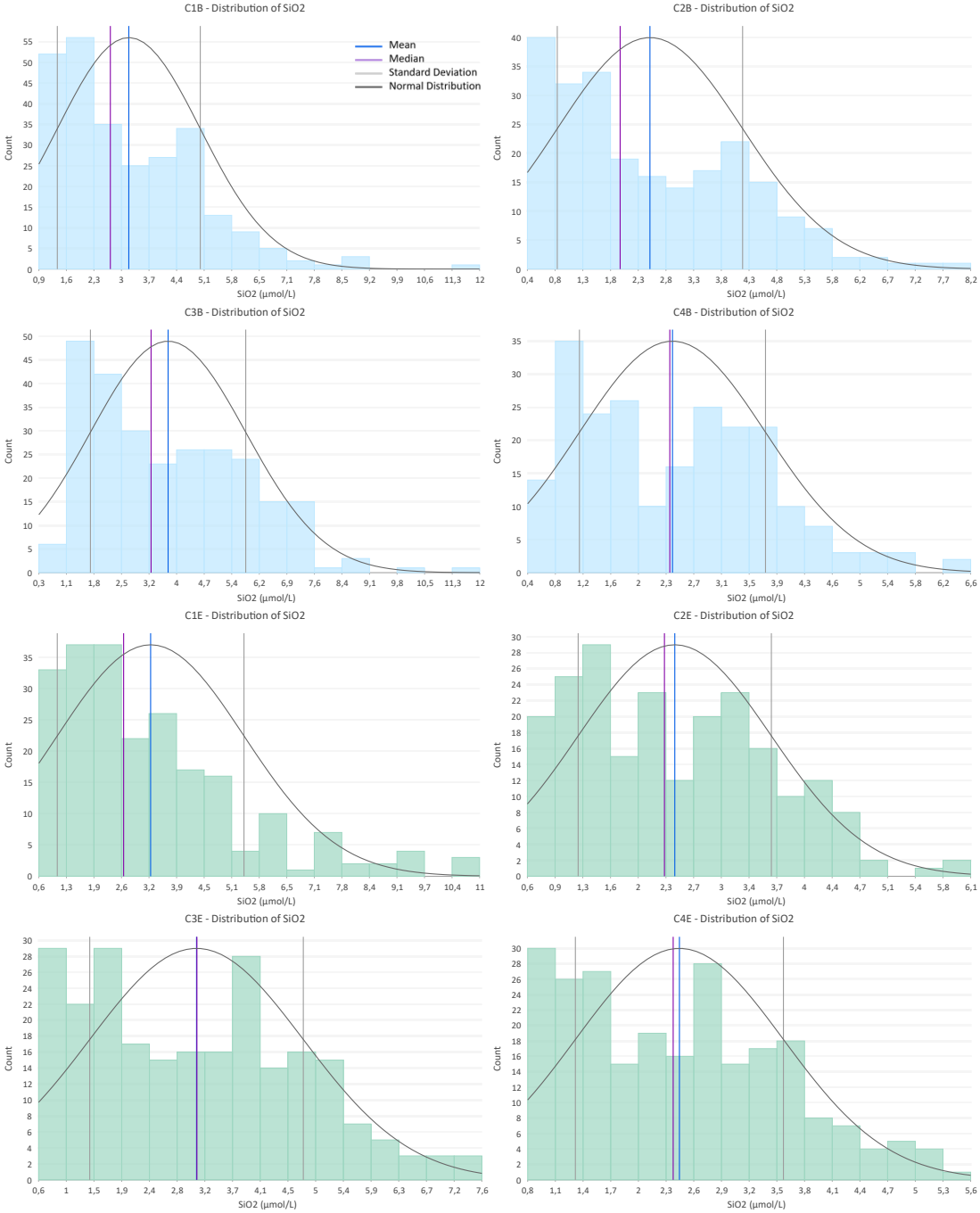


Figure B-6. Histograms of SiO_2.

Chlorophyll a (mg/m³)	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	0.75	1.30	0.83	1.09	0.51	1.84	1.13	0.68
Median	0.62	0.78	0.68	1.06	0.41	1.00	0.75	0.68
Standard Deviation	0.52	1.40	0.53	0.54	0.42	1.83	1.08	0.31
Coefficient of Variation	68.6%	107.7%	63.9%	49.6%	81.8%	99.7%	95.6%	44.9%
Samples	187	167	181	158	154	154	168	168
Minimum	0.05	0.05	0.14	0.17	0.05	0.15	0.05	0.05
Maximum	2.70	8.40	2.70	2.95	3.50	8.80	4.90	1.43
Skewness	1.56	2.51	1.22	0.52	3.30	1.57	1.80	-0.08
Kurtosis	5.58	10.66	4.19	2.83	21.22	4.76	5.71	2.68

Table B-7. Descriptive statistics of chlorophyll a.

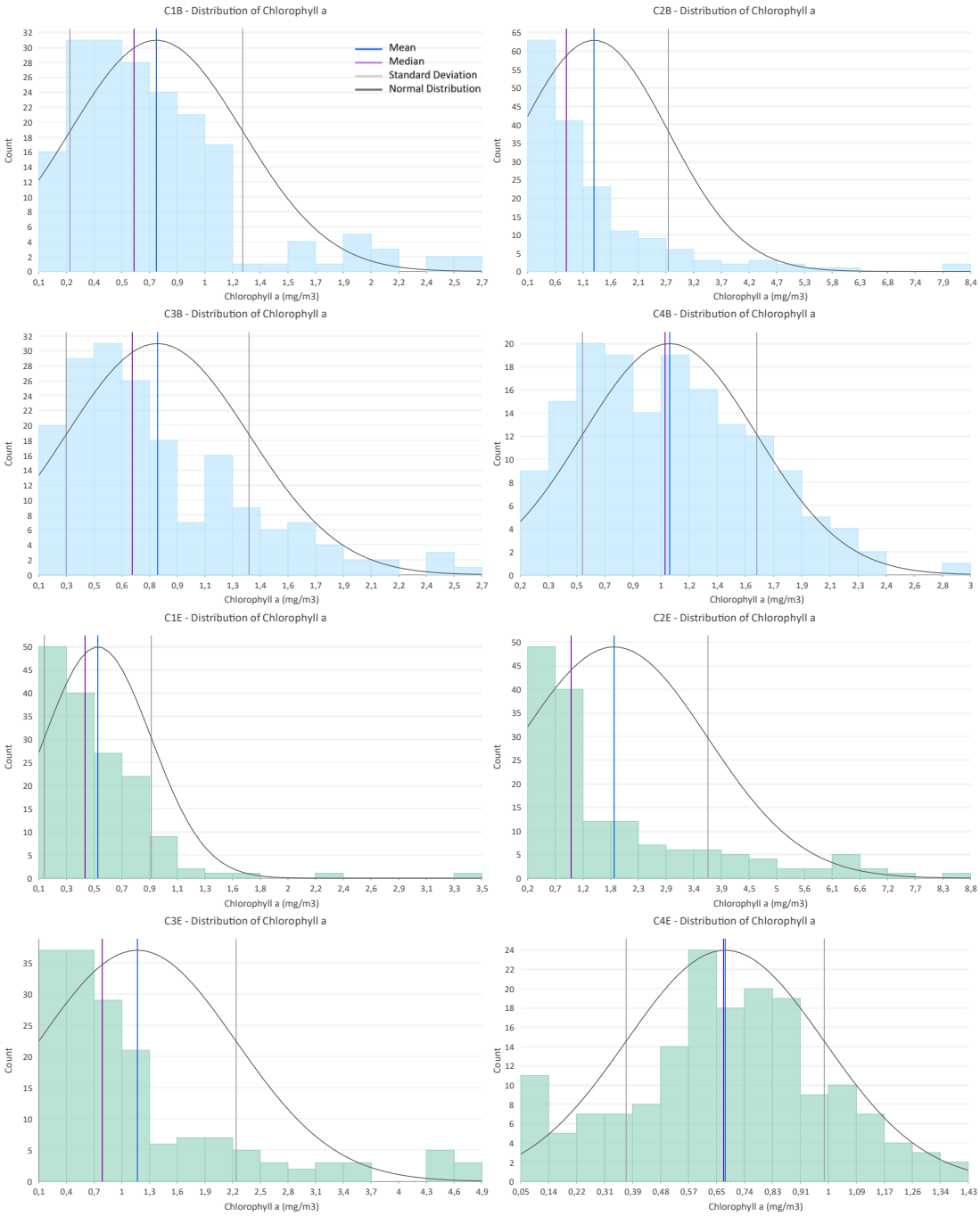


Figure B-7. Histograms of chlorophyll a.

pH	C1B	C2B	C3B	C4B	C1E	C2E	C3E	C4E
Mean	8.04	8.07	8.01	8.08	8.09	8.05	7.98	8.05
Median	8.05	8.08	8.02	8.09	8.12	8.06	8.00	8.07
Standard Deviation	0.06	0.08	0.07	0.05	0.09	0.07	0.09	0.07
Coefficient of Variation	0.7%	1.0%	0.9%	0.7%	1.1%	0.9%	1.1%	0.8%
Samples	261	236	265	222	223	216	238	241
Minimum	7.90	7.77	7.75	7.87	7.82	7.69	7.58	7.81
Maximum	8.14	8.22	8.17	8.19	8.21	8.20	8.11	8.15
Skewness	-0.53	-1.25	-1.02	-0.64	-1.26	-1.44	-1.46	-0.82
Kurtosis	2.73	5.07	4.50	3.36	4.03	7.17	5.31	3.09

Table B-8. Descriptive statistics of pH.

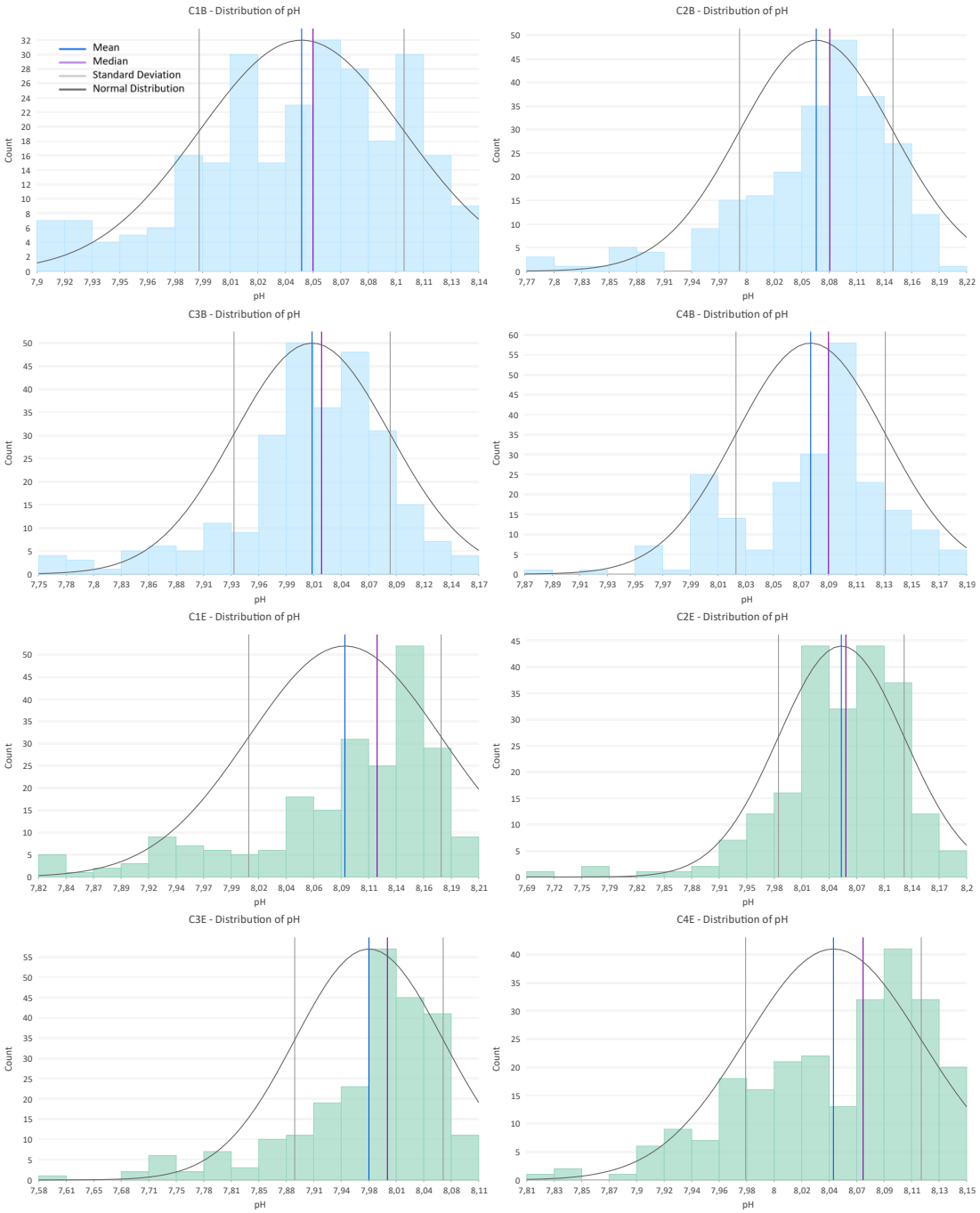


Figure B-8. Histograms of pH.

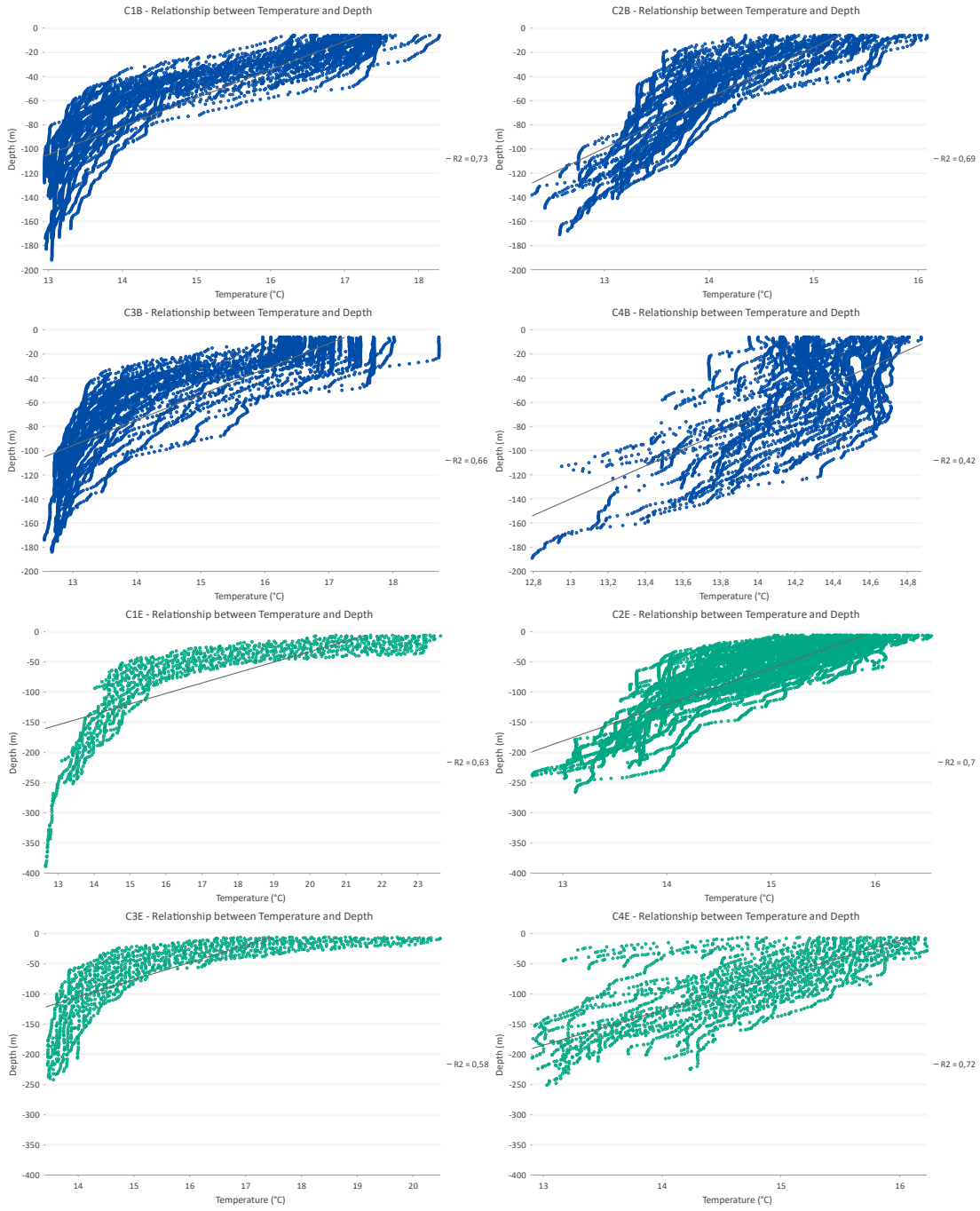


Figure B-9. Scatter plots of temperature.

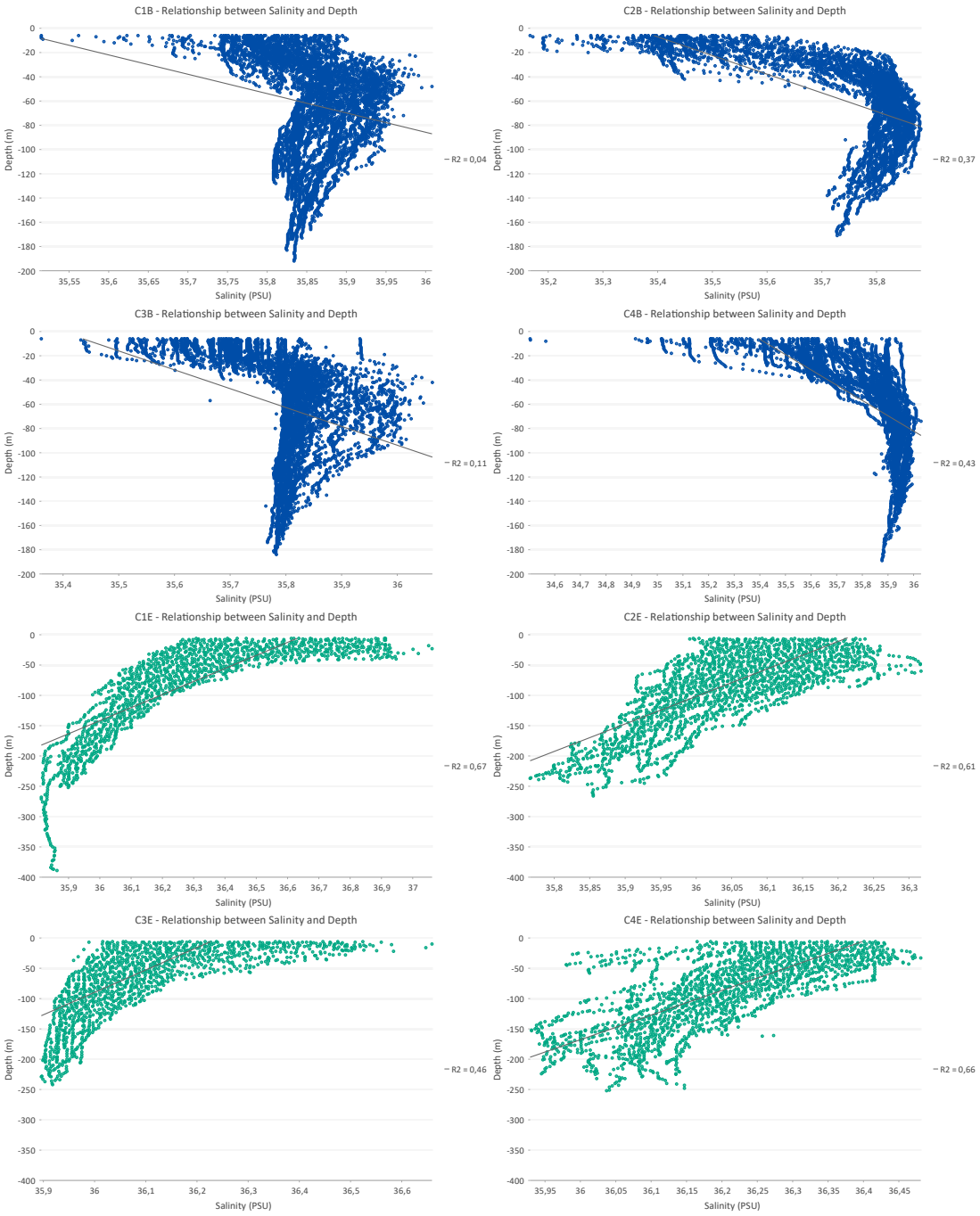


Figure B-10. Scatter plots of salinity.

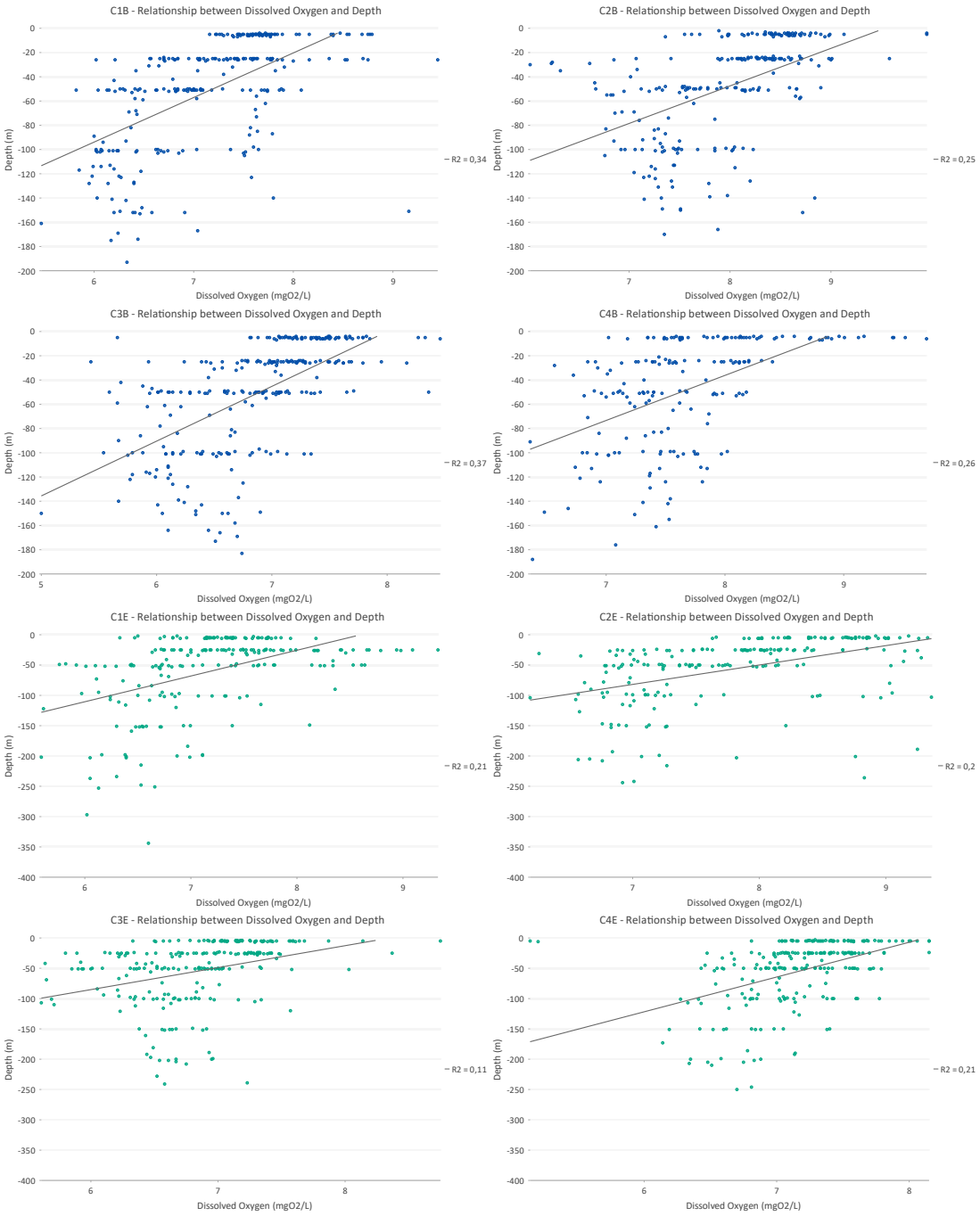


Figure B-11. Scatter plots of DO.

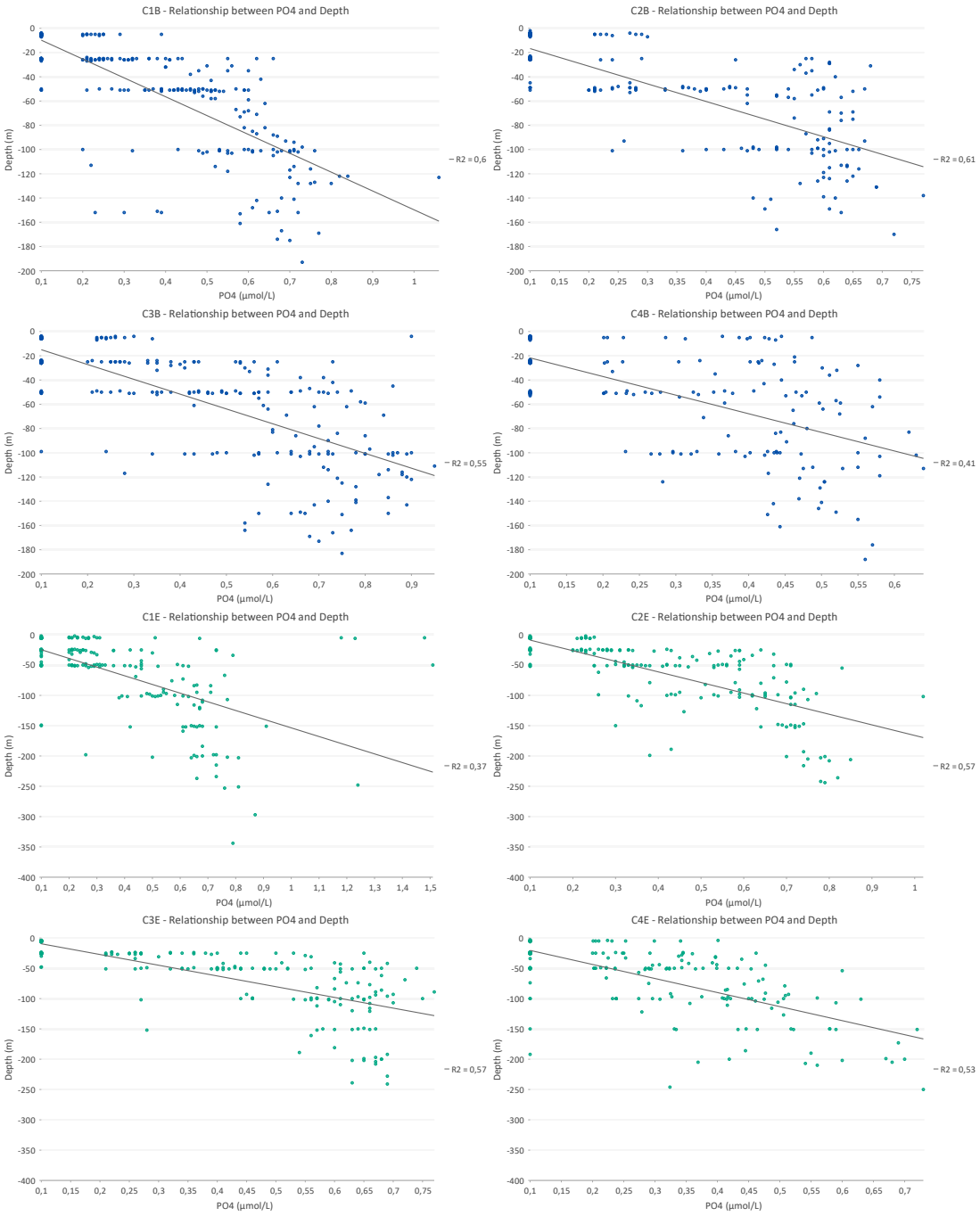


Figure B-12. Scatter plots of PO_4.

Figure B-13. Scatter plots of NO_X.

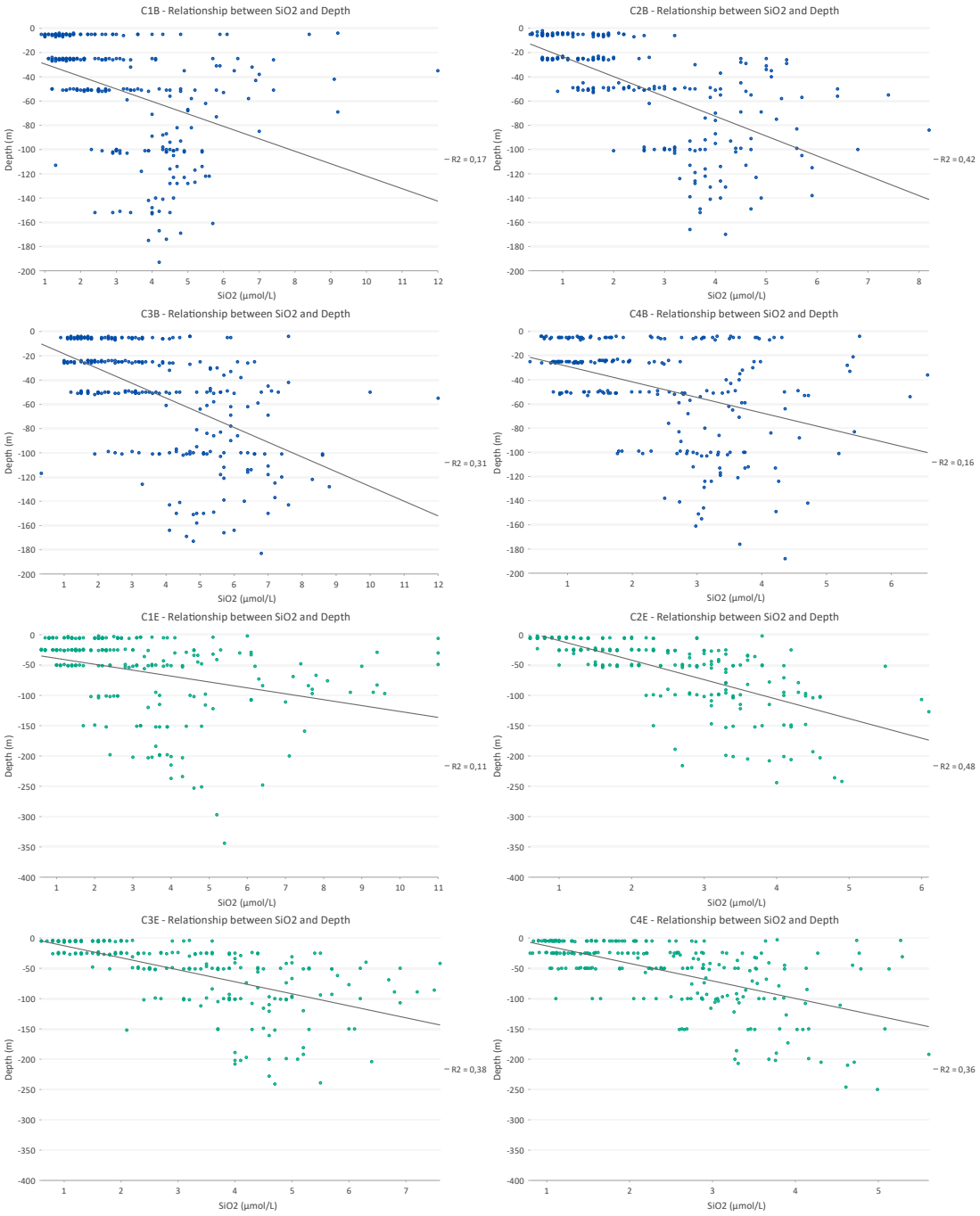


Figure B-14. Scatter plots of SiO_2.

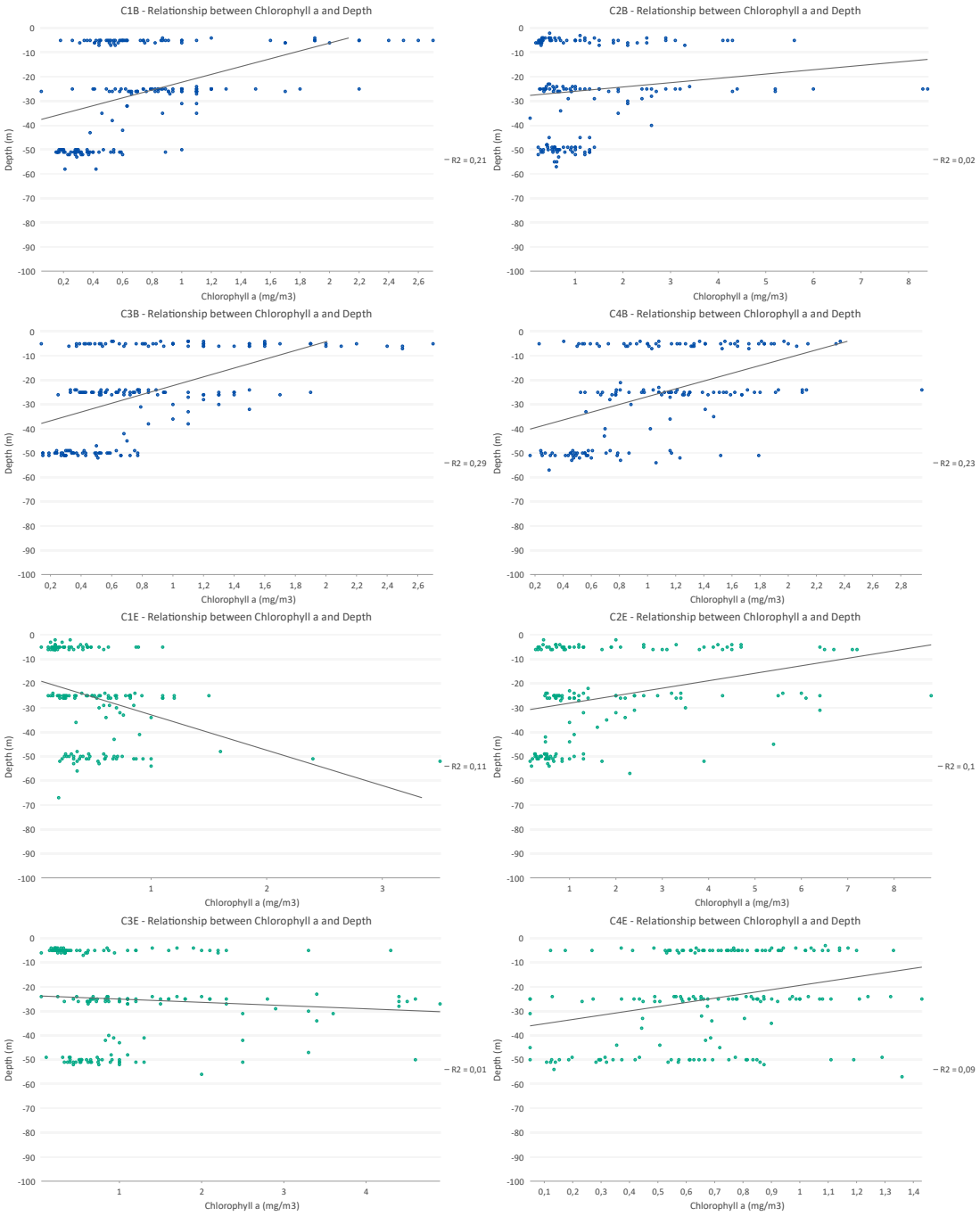


Figure B-15. Scatter plots of chlorophyll a.

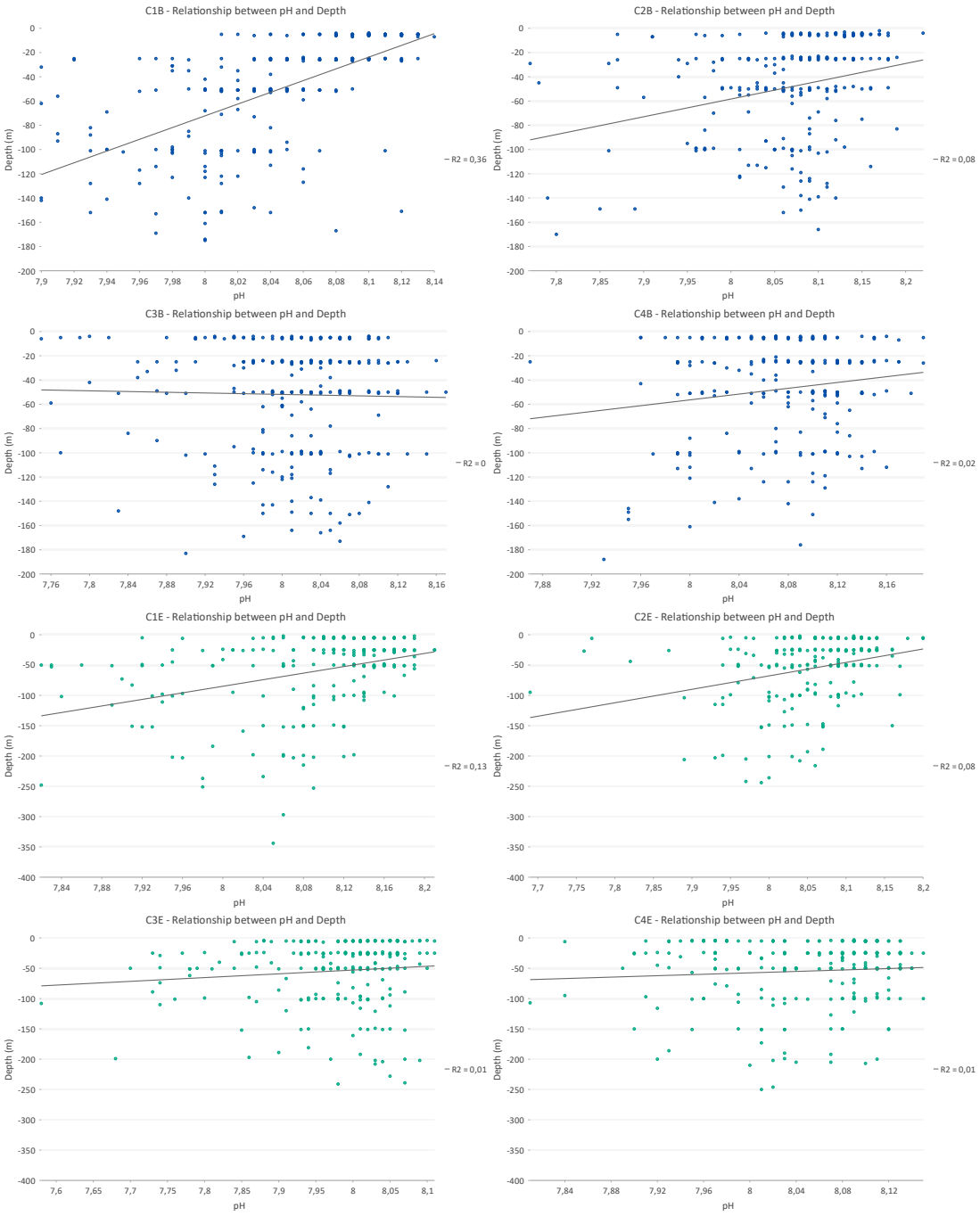


Figure B-16. Scatter plots of pH.

ANNEX C – THREE-DIMENSIONAL INTERPOLATION

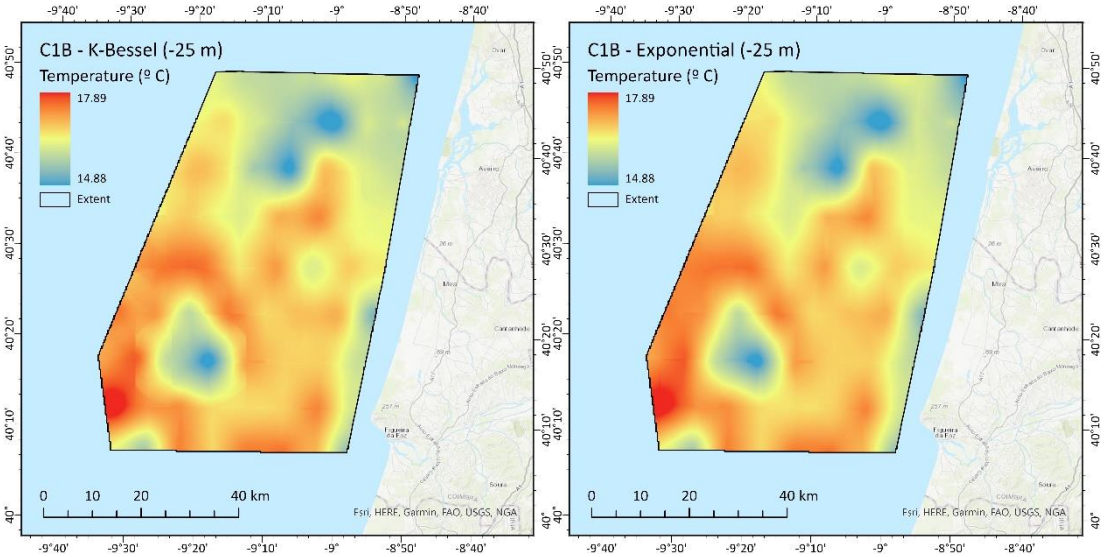


Figure C-1. EBK3D using different semivariogram types: K-Bessel and Exponential.

This comparison was performed for both areas and for all the OVs, obtaining identical results. This example illustrates the analysis for C1B and temperature at 25 m depth and with a 250 m horizontal resolution. All the EBK3D parameters were the same except for the semivariogram type. It is perceptible, in the extracted horizontal surface, that K-Bessel generated more geometric artifacts than Exponential.

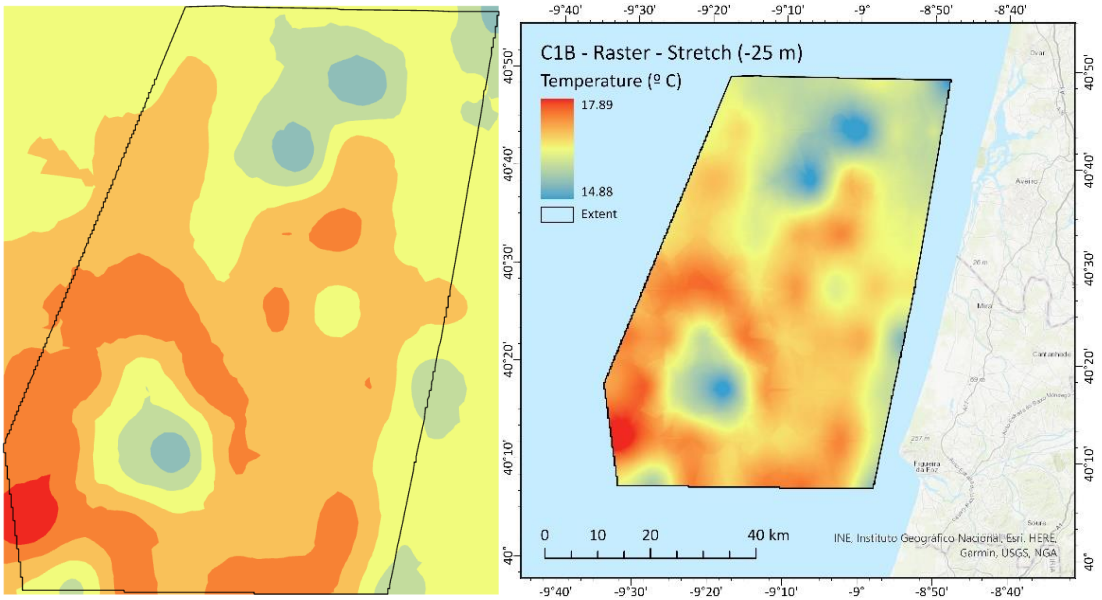


Figure C-2. EBK3D visual assessment: geostatistical layer and raster surface.

During the interpolation process, the result is a geostatistical layer, which is like a continuous function without a specific resolution. For visualization purposes, ArcGIS uses a classification scheme based on the range of the input samples, colored by editable contours. This does not allow a true visual assessment of the interpolation. Only when the geostatistical layer is extracted to a raster with a specific resolution (250 m, in this example) and its symbology is changed to a stretch scheme, are we able to rigorously evaluate the result and, in this case, perceive several geometric artifacts.

Temperature	C1B	
	K-Bessel	Exponential
EBK3D properties		
Elevation Inflation Factor	489	489
Maximum Neighbors	4	4
Minimum Neighbors	1	1
Sector Type	8	8
Radius	100 000	100 000
Cross-validation results		
Count	5 956	5 956
Average Continuous Ranked Probability Score (CRPS)	0.0051	0.0100
Percent in 95% Interval	96.4741	99.8321
Mean Error (ME)	-5.2E-05	2.2E-04
Root Mean Square Error (RMSE)	0.0117	0.0166
Mean Standardized Error (SME)	-0.0025	0.0058
Root Mean Square Standardized Error (RMSSE)	0.7755	0.3610
Average Standard Error (ASE)	0.0136	0.0397

Table C-1. Cross-validation results using K-Bessel and Exponential.
In this example, all the EBK3D parameters were the same except for the semivariogram type.

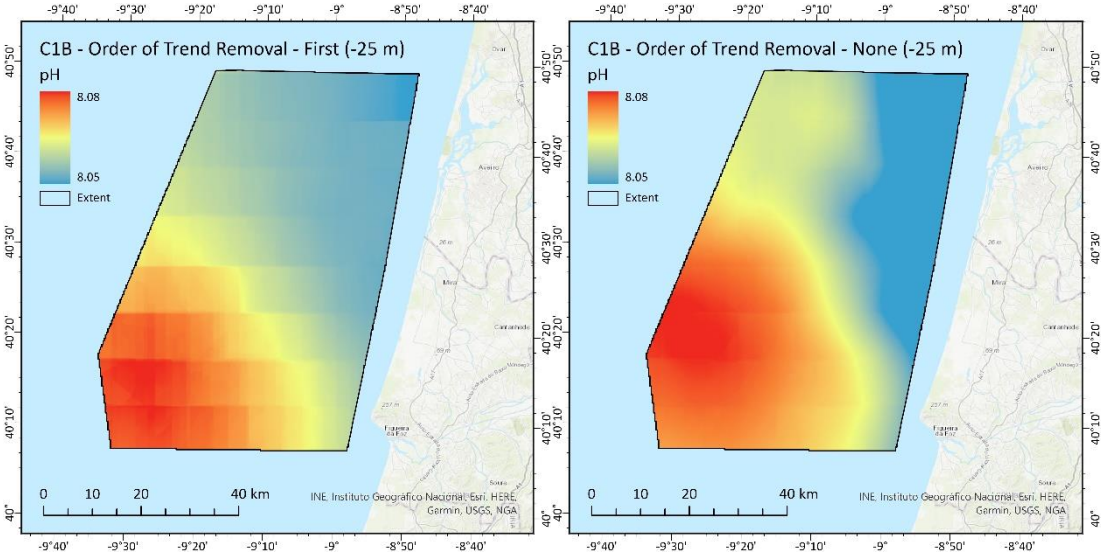


Figure C-3. EBK3D using different order of trend removal configurations: first and none.
This comparison was performed for both areas, and the results were improved for DO and pH. This example illustrates the analysis for C1B and pH at 25 m depth and with a 250 m horizontal resolution. All the EBK3D parameters were the same except for the order of trend removal. It is perceptible, in the extracted horizontal surface, that the configuration “first” generated more geometric artifacts than “none”.

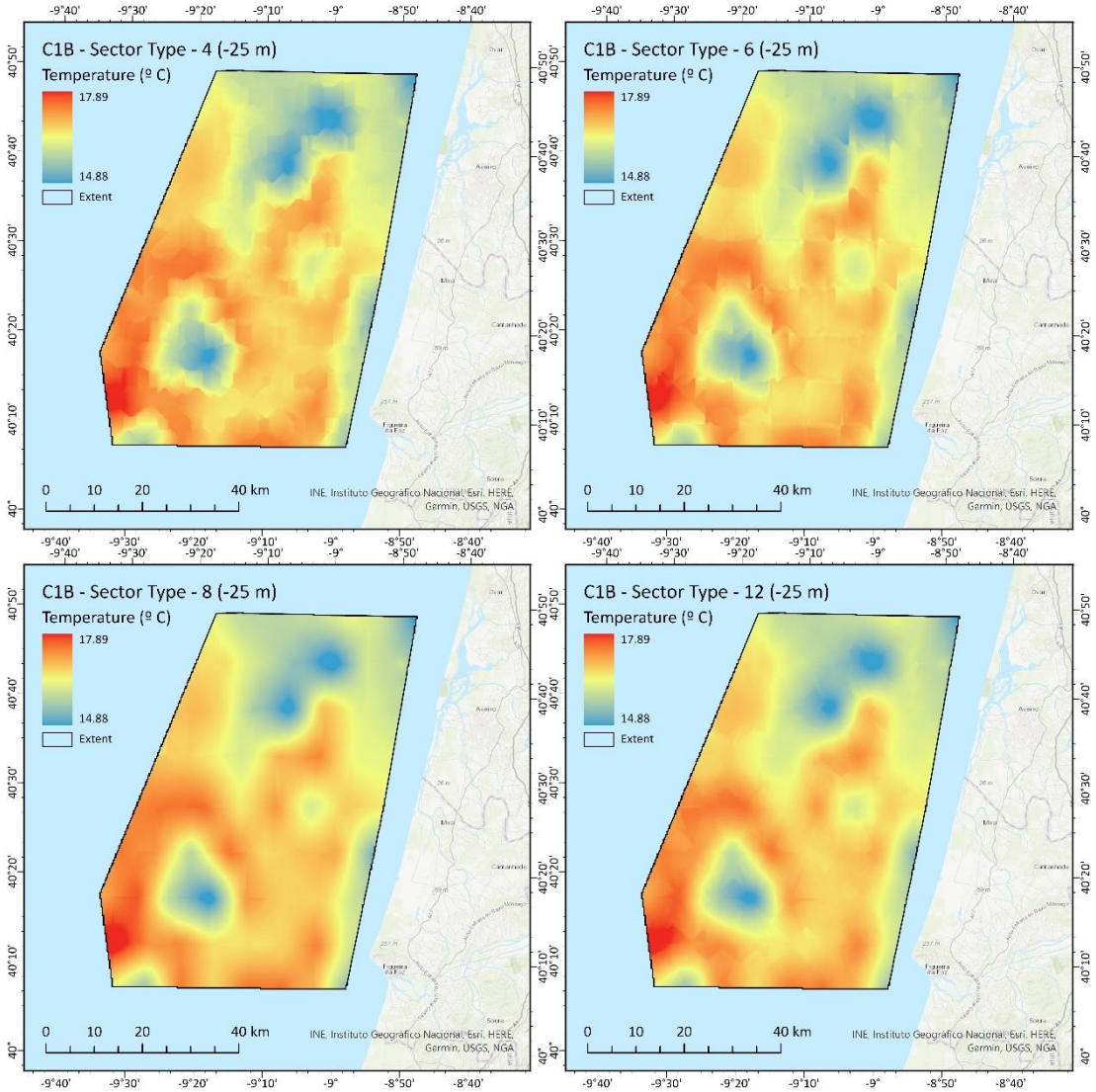


Figure C-4. EBK3D using different sector type configurations: 4, 6, 8 and 12.

This comparison was performed for both areas and for all the OVs, obtaining identical results. This example illustrates the analysis for C1B and temperature at 25 m depth and with a 250 m horizontal resolution. All the EBK3D parameters were the same except for the sector type. It is perceptible, in the extracted horizontal surface, that sector type 8 (octahedron) generated fewer geometric artifacts than the others.

Temperature (°C)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	489	489	621	489
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	5 956	4 835	5 991	4 584
Average Continuous Ranked Probability Score (CRPS)	0.0101	0.0075	0.0123	0.0045
Percent in 95% Interval	99.8321	99.0486	99.7162	98.9747
Mean Error (ME)	2.4E-04	1.5E-04	3.4E-04	-2.3E-05
Root Mean Square Error (RMSE)	0.0166	0.0146	0.0200	0.0086
Mean Standardized Error (SME)	0.0067	0.0058	0.0071	-0.0020
Root Mean Square Standardized Error (RMSSE)	0.3623	0.4877	0.3563	0.5128
Average Standard Error (ASE)	0.0400	0.0260	0.0509	0.0152
Ratio of the Variance	0.9987	0.9982	0.9983	0.9967
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-2. EBK3D properties and cross-validation results of temperature in area B.

Temperature (°C)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	329	127	281	174
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	10 704	9 890	10 158	10 407
Average Continuous Ranked Probability Score (CRPS)	0.0307	0.0059	0.0125	0.0052
Percent in 95% Interval	99.1031	98.9687	99.4389	99.3082
Mean Error (ME)	-5.2E-05	3.7E-05	4.7E-04	6.6E-05
Root Mean Square Error (RMSE)	0.0832	0.0107	0.0281	0.0092
Mean Standardized Error (SME)	-0.0009	0.0025	0.0101	0.0011
Root Mean Square Standardized Error (RMSSE)	0.5559	0.5046	0.4369	0.4384
Average Standard Error (ASE)	0.1164	0.0196	0.0508	0.0198
Ratio of the Variance	-	-	0.9985	0.9988
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	329	127	-	-
Maximum Neighbors	12	12	-	-
Minimum Neighbors	1	1	-	-
Sector Type	8	8	-	-
Radius	100 000	100 000	-	-
Cross-validation results				
Count	10 704	9 890	-	-
Average Continuous Ranked Probability Score (CRPS)	0.0298	0.0060	-	-
Percent in 95% Interval	99.3180	98.9687	-	-
Mean Error (ME)	1.5E-04	4.4E-05	-	-
Root Mean Square Error (RMSE)	0.0785	0.0109	-	-
Mean Standardized Error (SME)	-0.0010	0.0027	-	-
Root Mean Square Standardized Error (RMSSE)	0.5248	0.5065	-	-
Average Standard Error (ASE)	0.1173	0.0200	-	-
Ratio of the Variance	0.9987	0.9990	-	-

Table C-3. EBK3D properties and cross-validation results of temperature in area E.

Salinity (PSU)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	489	621	574	489
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	5 956	4 835	5 991	4 584
Average Continuous Ranked Probability Score (CRPS)	0.0030	0.0034	0.0042	0.0036
Percent in 95% Interval	97.9181	98.4695	97.6465	98.8438
Mean Error (ME)	-5.5E-06	-7.2E-05	2.9E-05	-9.4E-05
Root Mean Square Error (RMSE)	0.0059	0.0069	0.0102	0.0074
Mean Standardized Error (SME)	0.0006	-0.0076	0.0091	-0.0077
Root Mean Square Standardized Error (RMSSE)	0.6530	0.6024	0.7134	0.5270
Average Standard Error (ASE)	0.0080	0.0111	0.0122	0.0138
Ratio of the Variance	0.9693	0.9918	0.9752	0.9924
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-4. EBK3D properties and cross-validation results of salinity in area B.

Salinity (PSU)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	356	127	281	174
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	10 704	9 890	10 158	10 407
Average Continuous Ranked Probability Score (CRPS)	0.0075	0.0023	0.0043	0.0022
Percent in 95% Interval	97.6084	98.2204	97.9228	98.6259
Mean Error (ME)	1.4E-06	3.1E-05	3.8E-05	1.4E-05
Root Mean Square Error (RMSE)	0.0189	0.0044	0.0108	0.0041
Mean Standardized Error (SME)	0.0001	0.0044	0.0043	-0.0002
Root Mean Square Standardized Error (RMSSE)	0.7300	0.6448	0.6668	0.5550
Average Standard Error (ASE)	0.0220	0.0065	0.0131	0.0066
Ratio of the Variance	-	-	-	-
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	356	127	281	174
Maximum Neighbors	12	12	12	12
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	10 704	9 890	10 158	10 407
Average Continuous Ranked Probability Score (CRPS)	0.0074	0.0023	0.0043	0.0022
Percent in 95% Interval	97.9260	98.2811	98.1492	98.6836
Mean Error (ME)	1.8E-05	2.4E-05	3.8E-05	1.3E-05
Root Mean Square Error (RMSE)	0.0187	0.0044	0.0108	0.0041
Mean Standardized Error (SME)	0.0010	0.0031	0.0060	0.0003
Root Mean Square Standardized Error (RMSSE)	0.7077	0.6438	0.6503	0.5518
Average Standard Error (ASE)	0.0221	0.0065	0.0131	0.0067
Ratio of the Variance	0.9930	0.9951	0.9772	0.9950

Table C-5. EBK3D properties and cross-validation results of salinity in area E.

Dissolved Oxygen (mgO_2/L)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Order of Trend Removal	First	First	First	First
Elevation Inflation Factor	1 000	788	189	137
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	264	236	265	185
Average Continuous Ranked Probability Score (CRPS)	0.2967	0.2687	0.2258	0.2183
Percent in 95% Interval	96.9697	97.8814	98.8679	100.0000
Mean Error (ME)	-4.1E-02	-5.7E-02	-2.2E-02	-2.7E-02
Root Mean Square Error (RMSE)	0.5449	0.4768	0.3908	0.3479
Mean Standardized Error (SME)	-0.0027	-0.0462	0.0016	-0.0119
Root Mean Square Standardized Error (RMSSE)	0.9146	0.7113	0.7109	0.5360
Average Standard Error (ASE)	0.5749	0.6665	0.5688	0.6411
Ratio of the Variance	-	-	-	-
		Modified Model		
EBK3D properties				
Order of Trend Removal	None	None	None	None
Elevation Inflation Factor	1 000	788	1 000	329
Maximum Neighbors	10	15	15	15
Minimum Neighbors	1	1	1	1
Sector Type	12	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	264	236	265	185
Average Continuous Ranked Probability Score (CRPS)	0.2999	0.2709	0.2309	0.2256
Percent in 95% Interval	98.4848	97.4576	98.4906	100.0000
Mean Error (ME)	-3.5E-02	-6.0E-02	-4.8E-02	-3.8E-02
Root Mean Square Error (RMSE)	0.5412	0.4785	0.4047	0.3625
Mean Standardized Error (SME)	0.0135	-0.0507	-0.0431	-0.0251
Root Mean Square Standardized Error (RMSSE)	0.8992	0.7008	0.7261	0.5501
Average Standard Error (ASE)	0.5943	0.6748	0.5725	0.6526
Ratio of the Variance	0.2962	0.3657	0.4571	0.3319

Table C-6. EBK3D properties and cross-validation results of DO in area B.

Dissolved Oxygen (mgO_2/L)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Order of Trend Removal	First	First	First	First
Elevation Inflation Factor	1 000	1 000	1	1
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	219	215	236	241
Average Continuous Ranked Probability Score (CRPS)	0.3038	0.3110	0.2263	0.1931
Percent in 95% Interval	95.8904	96.7442	98.3051	99.1701
Mean Error (ME)	-5.8E-02	-8.1E-02	-4.4E-02	-3.3E-02
Root Mean Square Error (RMSE)	0.5447	0.5674	0.3963	0.3205
Mean Standardized Error (SME)	-0.0435	-0.0668	-0.0403	-0.0283
Root Mean Square Standardized Error (RMSSE)	0.8369	0.8212	0.6796	0.6204
Average Standard Error (ASE)	0.6411	0.6708	0.5693	0.5808
Ratio of the Variance	-	-	-	-
		Modified Model		
EBK3D properties				
Order of Trend Removal	None	None	None	First
Elevation Inflation Factor	1000	1000	1000	1
Maximum Neighbors	12	12	12	15
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	219	215	236	241
Average Continuous Ranked Probability Score (CRPS)	0.3071	0.3023	0.2391	0.1937
Percent in 95% Interval	96.3470	96.7442	98.3051	99.1701
Mean Error (ME)	-5.1E-02	-6.0E-02	-5.3E-02	-3.7E-02
Root Mean Square Error (RMSE)	0.5495	0.5473	0.4241	0.3231
Mean Standardized Error (SME)	-0.0310	-0.0433	-0.0531	-0.0345
Root Mean Square Standardized Error (RMSSE)	0.8438	0.7859	0.7416	0.6269
Average Standard Error (ASE)	0.6430	0.6738	0.5665	0.5796
Ratio of the Variance	0.3151	0.4177	0.2015	0.2685

Table C-7. EBK3D properties and cross-validation results of DO in area E.

PO_4 (µmol/L)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	1 000	788	1 000	853
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	262	230	264	221
Average Continuous Ranked Probability Score (CRPS)	0.0605	0.0423	0.0711	0.0426
Percent in 95% Interval	93.8931	94.7826	94.6970	95.0226
Mean Error (ME)	-2.8E-03	-2.0E-03	-1.3E-03	-1.7E-03
Root Mean Square Error (RMSE)	0.1143	0.0783	0.1363	0.0805
Mean Standardized Error (SME)	-0.0151	-0.0107	-0.0024	-0.0086
Root Mean Square Standardized Error (RMSSE)	0.9766	0.9404	1.0345	0.9731
Average Standard Error (ASE)	0.1180	0.0822	0.1332	0.0822
Ratio of the Variance	0.8046	0.8665	0.7955	0.8226
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-8. EBK3D properties and cross-validation results of PO_4 in area B.

PO_4 (µmol/L)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	853	304	788	240
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	222	218	238	240
Average Continuous Ranked Probability Score (CRPS)	0.0964	0.0687	0.0488	0.0480
Percent in 95% Interval	95.9459	91.2844	91.5966	93.7500
Mean Error (ME)	-7.7E-03	4.7E-05	4.2E-04	-2.8E-03
Root Mean Square Error (RMSE)	0.2068	0.1284	0.0916	0.0898
Mean Standardized Error (SME)	-0.0116	0.0011	0.0061	-0.0155
Root Mean Square Standardized Error (RMSSE)	1.0445	1.0629	1.0037	1.0171
Average Standard Error (ASE)	0.1864	0.1222	0.0955	0.0843
Ratio of the Variance	0.5952	0.7689	0.8786	0.7625
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-9. EBK3D properties and cross-validation results of PO_4 in area E.

NO_X (µmol/L)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	1 000	853	1 000	1 000
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	263	232	264	222
Average Continuous Ranked Probability Score (CRPS)	0.9449	0.6409	1.0683	0.7090
Percent in 95% Interval	95.0570	94.3966	94.3182	95.9459
Mean Error (ME)	-2.4E-02	-3.3E-02	9.0E-03	-3.7E-02
Root Mean Square Error (RMSE)	1.8259	1.1709	2.0309	1.3622
Mean Standardized Error (SME)	-0.0075	-0.0153	0.0047	-0.0121
Root Mean Square Standardized Error (RMSSE)	0.9895	0.9144	1.0121	0.9629
Average Standard Error (ASE)	1.8481	1.2892	2.0229	1.3650
Ratio of the Variance	0.8537	0.9062	0.8264	0.8378
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-10. EBK3D properties and cross-validation results of NO_X in area B.

NO_X (µmol/L)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	1 000	189	788	189
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	222	218	238	239
Average Continuous Ranked Probability Score (CRPS)	0.9601	1.4413	0.7612	0.8322
Percent in 95% Interval	93.2432	94.0367	93.6975	94.5607
Mean Error (ME)	-7.7E-02	2.2E-02	2.6E-02	-2.6E-02
Root Mean Square Error (RMSE)	1.9803	2.7771	1.4310	1.5691
Mean Standardized Error (SME)	-0.0297	0.0067	0.0119	-0.0050
Root Mean Square Standardized Error (RMSSE)	1.1534	1.0087	0.9907	1.0174
Average Standard Error (ASE)	1.6795	2.7421	1.5065	1.5261
Ratio of the Variance	0.8493	0.6805	0.9127	0.8076
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-11. EBK3D properties and cross-validation results of NO_X in area E.

SiO_2 (µmol/L)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	672	788	1 000	489
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	263	232	262	222
Average Continuous Ranked Probability Score (CRPS)	0.6670	0.4743	0.7697	0.4864
Percent in 95% Interval	92.7757	93.1034	92.7481	94.1441
Mean Error (ME)	-4.1E-02	-4.1E-02	-3.9E-02	-1.5E-02
Root Mean Square Error (RMSE)	1.3428	0.8909	1.4924	0.9010
Mean Standardized Error (SME)	-0.0169	-0.0284	-0.0160	-0.0061
Root Mean Square Standardized Error (RMSSE)	1.0440	1.0700	1.0590	1.0259
Average Standard Error (ASE)	1.2348	0.8123	1.3850	0.8793
Ratio of the Variance	0.5426	0.7304	0.5974	0.5732
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-12. EBK3D properties and cross-validation results of SiO_2 in area B.

SiO_2 (µmol/L)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	1 000	240	788	161
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	221	218	238	240
Average Continuous Ranked Probability Score (CRPS)	0.8838	0.3824	0.5615	0.4674
Percent in 95% Interval	94.1176	95.4128	94.9580	93.7500
Mean Error (ME)	-8.9E-02	2.4E-03	-2.5E-04	-2.1E-02
Root Mean Square Error (RMSE)	1.7124	0.7074	1.0109	0.8490
Mean Standardized Error (SME)	-0.0296	0.0084	0.0075	-0.0122
Root Mean Square Standardized Error (RMSSE)	1.0750	1.0486	1.0283	1.0654
Average Standard Error (ASE)	1.5992	0.6743	1.0074	0.8060
Ratio of the Variance	0.4656	0.7046	0.7331	0.5882
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

Table C-13. EBK3D properties and cross-validation results of SiO_2 in area E.

Chlorophyll a (mg/m³)	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	1 000	1 000	621	386
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	187	167	181	158
Average Continuous Ranked Probability Score (CRPS)	0.1853	0.4206	0.1482	0.1927
Percent in 95% Interval	92.5134	95.2096	95.0276	97.4684
Mean Error (ME)	-5.7E-03	-2.1E-02	-6.6E-03	-1.5E-03
Root Mean Square Error (RMSE)	0.3547	0.8769	0.2748	0.3476
Mean Standardized Error (SME)	-0.0051	-0.0140	-0.0124	0.0017
Root Mean Square Standardized Error (RMSSE)	1.0502	0.9939	0.9887	0.9100
Average Standard Error (ASE)	0.3229	0.7834	0.2747	0.3775
Ratio of the Variance	0.5948	-	0.7238	0.5735
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	1 000	-	-
Maximum Neighbors	-	3	-	-
Minimum Neighbors	-	1	-	-
Sector Type	-	6	-	-
Radius	-	100 000	-	-
Cross-validation results				
Count	-	167	-	-
Average Continuous Ranked Probability Score (CRPS)	-	0.4235	-	-
Percent in 95% Interval	-	94.0120	-	-
Mean Error (ME)	-	-1.8E-02	-	-
Root Mean Square Error (RMSE)	-	0.8794	-	-
Mean Standardized Error (SME)	-	-0.0063	-	-
Root Mean Square Standardized Error (RMSSE)	-	0.9936	-	-
Average Standard Error (ASE)	-	0.7943	-	-
Ratio of the Variance	-	0.5789	-	-

Table C-14. EBK3D properties and cross-validation results of chlorophyll a in area B.

Chlorophyll a (mg/m³)	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Elevation Inflation Factor	304	386	418	137
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	154	154	168	168
Average Continuous Ranked Probability Score (CRPS)	0.1741	0.6368	0.4318	0.1191
Percent in 95% Interval	88.9610	91.5584	91.0714	94.6429
Mean Error (ME)	-4.7E-03	-5.0E-02	-3.8E-02	3.6E-03
Root Mean Square Error (RMSE)	0.3708	1.2353	0.8405	0.2153
Mean Standardized Error (SME)	-0.0093	-0.0313	-0.0352	0.0106
Root Mean Square Standardized Error (RMSSE)	1.2526	1.0871	1.1015	1.0103
Average Standard Error (ASE)	0.2662	1.1304	0.7197	0.2124
Ratio of the Variance	0.3737	0.4569	0.4084	0.5002
		Modified Model		
EBK3D properties				
Elevation Inflation Factor	-	-	-	-
Maximum Neighbors	-	-	-	-
Minimum Neighbors	-	-	-	-
Sector Type	-	-	-	-
Radius	-	-	-	-
Cross-validation results				
Count	-	-	-	-
Average Continuous Ranked Probability Score (CRPS)	-	-	-	-
Percent in 95% Interval	-	-	-	-
Mean Error (ME)	-	-	-	-
Root Mean Square Error (RMSE)	-	-	-	-
Mean Standardized Error (SME)	-	-	-	-
Root Mean Square Standardized Error (RMSSE)	-	-	-	-
Average Standard Error (ASE)	-	-	-	-
Ratio of the Variance	-	-	-	-

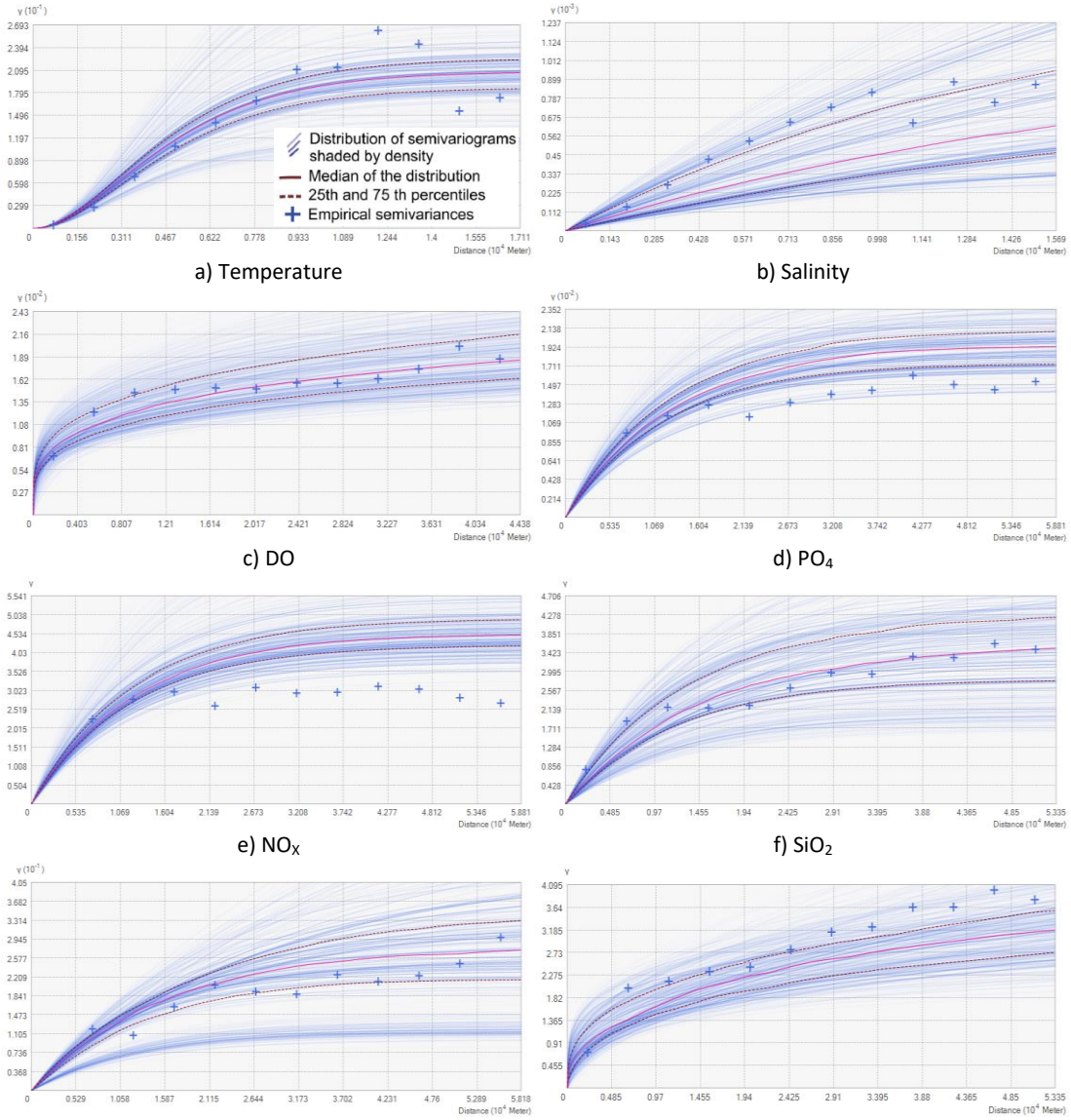
Table C-15. EBK3D properties and cross-validation results of chlorophyll a in area E.

pH	C1B	C2B	C3B	C4B
		Adjusted Model		
EBK3D properties				
Order of Trend Removal	First	First	First	First
Elevation Inflation Factor	924	45	20	1
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	261	236	265	222
Average Continuous Ranked Probability Score (CRPS)	0.0212	0.0315	0.0283	0.0185
Percent in 95% Interval	96.9349	94.9153	94.3396	99.5495
Mean Error (ME)	-1.9E-04	1.2E-03	6.7E-04	1.8E-04
Root Mean Square Error (RMSE)	0.0384	0.0587	0.0523	0.0317
Mean Standardized Error (SME)	-0.0015	0.0226	0.0031	0.0017
Root Mean Square Standardized Error (RMSSE)	0.8036	0.9987	0.9261	0.6394
Average Standard Error (ASE)	0.0475	0.0578	0.0556	0.0495
Ratio of the Variance	-	-	-	-
		Modified Model		
EBK3D properties				
Order of Trend Removal	None	None	None	None
Elevation Inflation Factor	1000	108	22	1
Maximum Neighbors	15	8	12	12
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	261	236	265	222
Average Continuous Ranked Probability Score (CRPS)	0.0210	0.0320	0.0284	0.0189
Percent in 95% Interval	96.9349	94.9153	95.0943	99.5495
Mean Error (ME)	-1.7E-04	1.6E-04	1.2E-03	2.7E-05
Root Mean Square Error (RMSE)	0.0375	0.0599	0.0525	0.0326
Mean Standardized Error (SME)	-0.0057	0.0027	0.0099	-0.0010
Root Mean Square Standardized Error (RMSSE)	0.7749	0.9484	0.9107	0.6592
Average Standard Error (ASE)	0.0482	0.0620	0.0567	0.0493
Ratio of the Variance	0.4531	0.3544	0.4041	0.4806

Table C-16. EBK3D properties and cross-validation results of pH in area B.

pH	C1E	C2E	C3E	C4E
		Adjusted Model		
EBK3D properties				
Order of Trend Removal	First	First	First	First
Elevation Inflation Factor	30	57	11	9
Maximum Neighbors	4	4	4	4
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	223	216	238	241
Average Continuous Ranked Probability Score (CRPS)	0.0352	0.0315	0.0343	0.0220
Percent in 95% Interval	93.7220	94.9074	93.2773	97.9253
Mean Error (ME)	2.5E-03	5.8E-05	-6.4E-04	1.0E-03
Root Mean Square Error (RMSE)	0.0664	0.0648	0.0664	0.0403
Mean Standardized Error (SME)	0.0295	-0.0035	-0.0075	0.0132
Root Mean Square Standardized Error (RMSSE)	0.9699	0.9965	0.9875	0.7382
Average Standard Error (ASE)	0.0686	0.0621	0.0659	0.0536
Ratio of the Variance	-	-	-	-
		Modified Model		
EBK3D properties				
Order of Trend Removal	None	None	None	None
Elevation Inflation Factor	924	108	108	108
Maximum Neighbors	4	12	12	12
Minimum Neighbors	1	1	1	1
Sector Type	8	8	8	8
Radius	100 000	100 000	100 000	100 000
Cross-validation results				
Count	223	216	238	241
Average Continuous Ranked Probability Score (CRPS)	0.0491	0.0321	0.0352	0.0252
Percent in 95% Interval	92.8251	95.3704	94.9580	97.5104
Mean Error (ME)	2.8E-03	5.5E-04	-3.6E-04	5.4E-04
Root Mean Square Error (RMSE)	0.0904	0.0656	0.0677	0.0457
Mean Standardized Error (SME)	0.0231	0.0040	-0.0084	0.0056
Root Mean Square Standardized Error (RMSSE)	1.0838	0.9868	0.9465	0.7679
Average Standard Error (ASE)	0.0850	0.0635	0.0701	0.0582
Ratio of the Variance	0.2517	0.2431	0.2642	0.3266

Table C-17. EBK3D properties and cross-validation results of pH in area E.



a) Temperature b) Salinity

c) DO d) PO_4

e) NO_X f) SiO_2

g) Chlorophyll a h) pH

Figure C-5. Examples of semivariograms of each OV, in C1B.

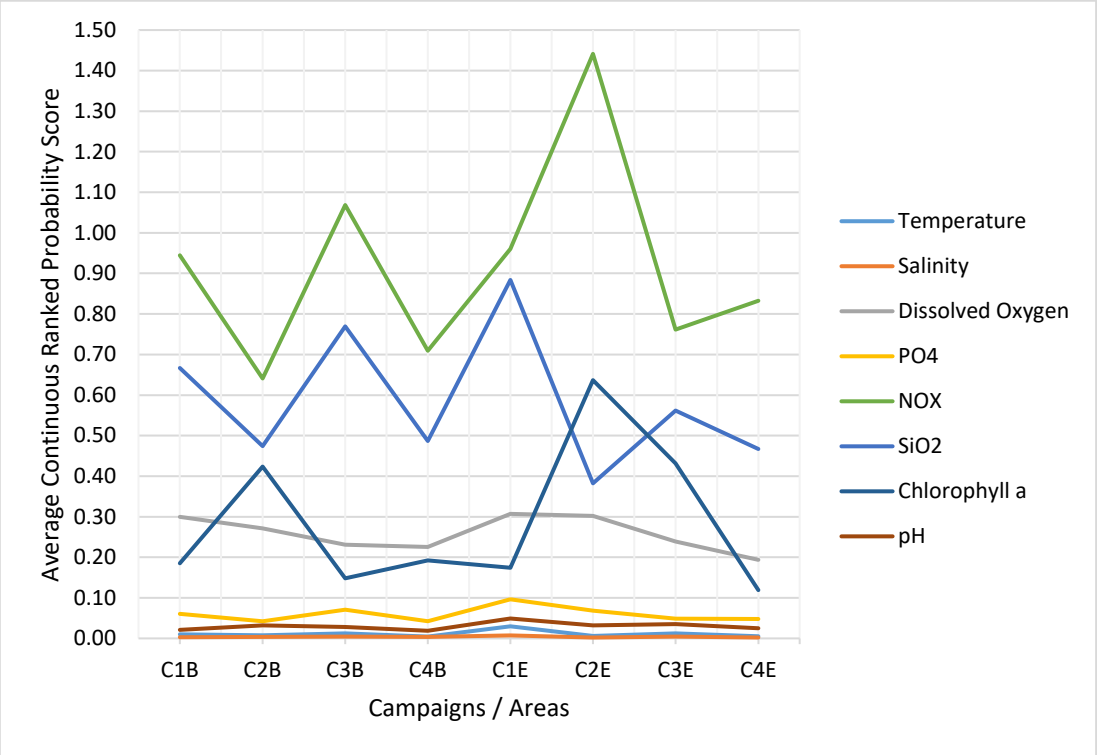


Figure C-6. Average CRPS of each OV.

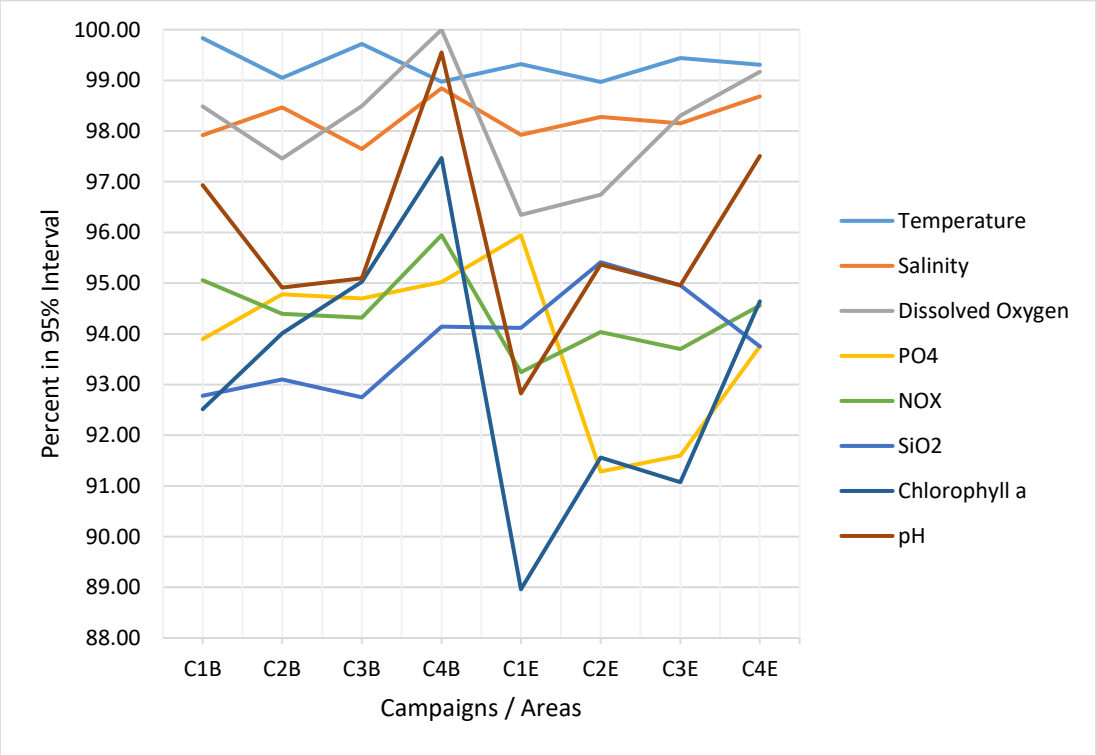


Figure C-7. Percent in 95% Interval of each OV.

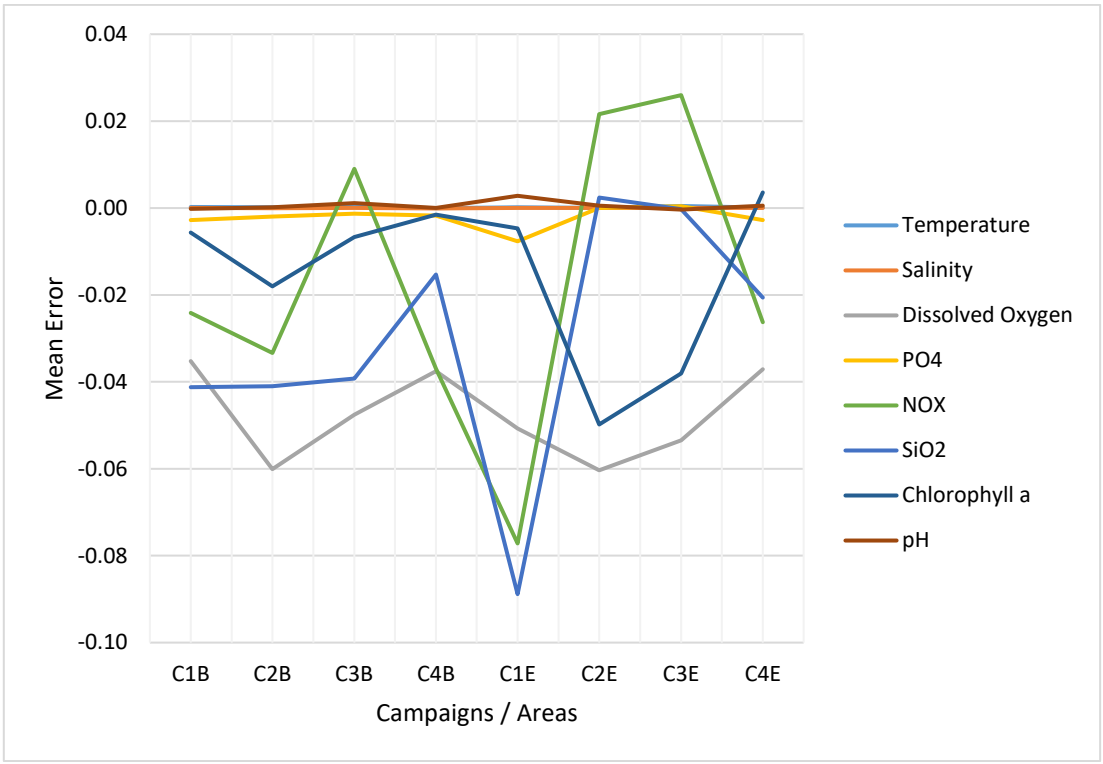


Figure C-8. ME of each OV.

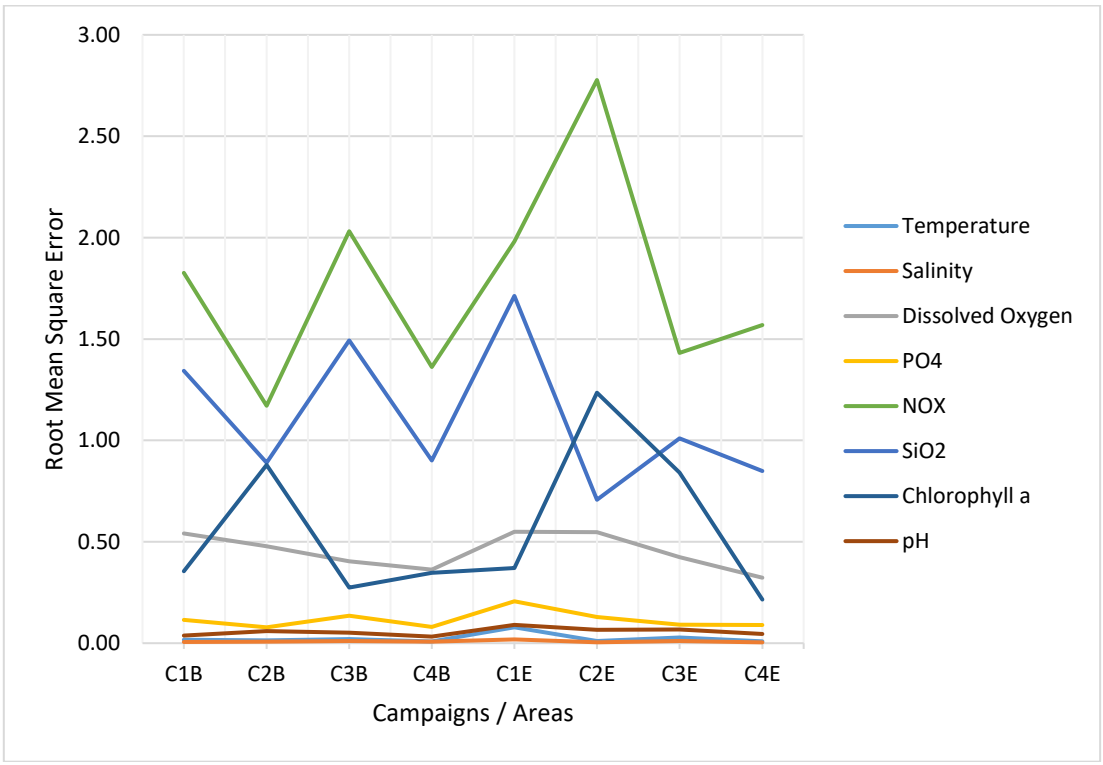


Figure C-9. RMSE of each OV.

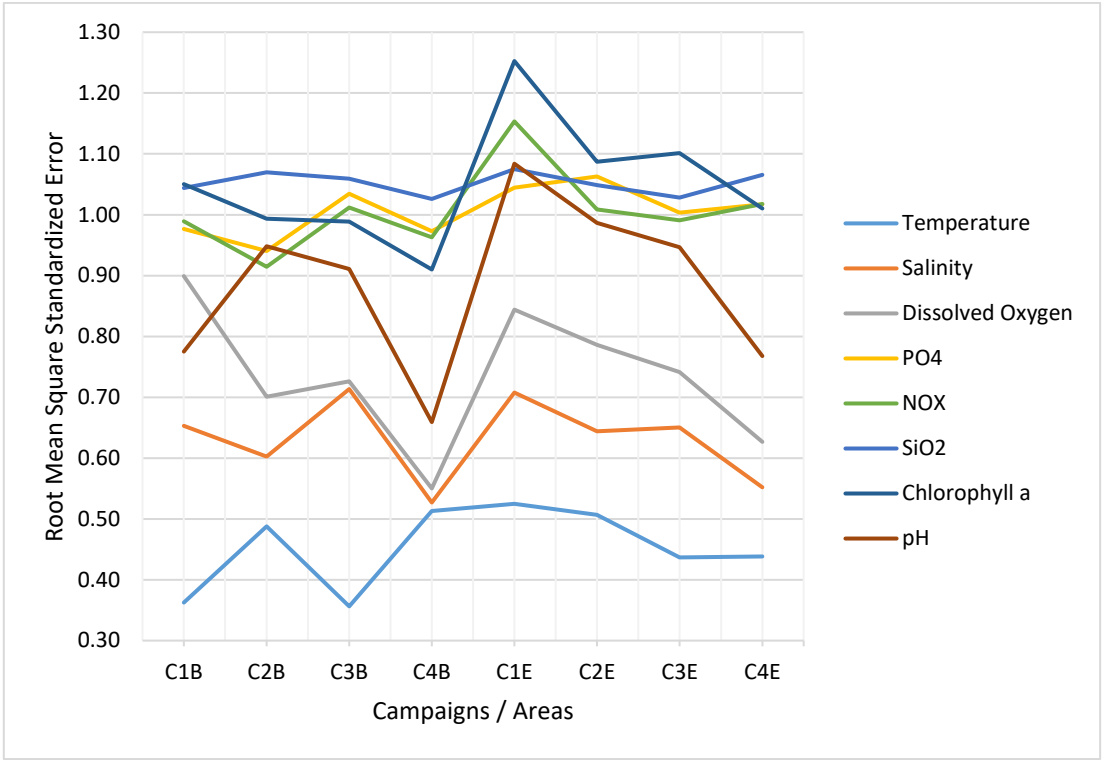


Figure C-10. RMSSE of each OV.

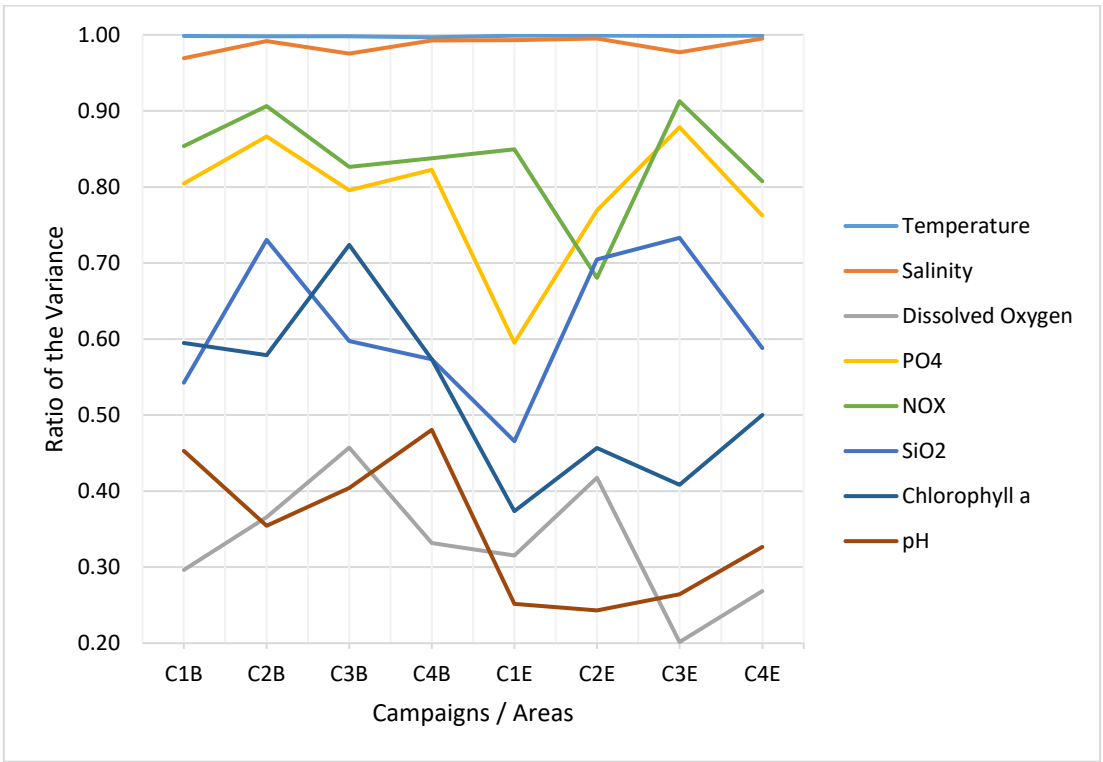


Figure C-11. RVar of each OV.

ANNEX D – CLUSTER ANALYSIS

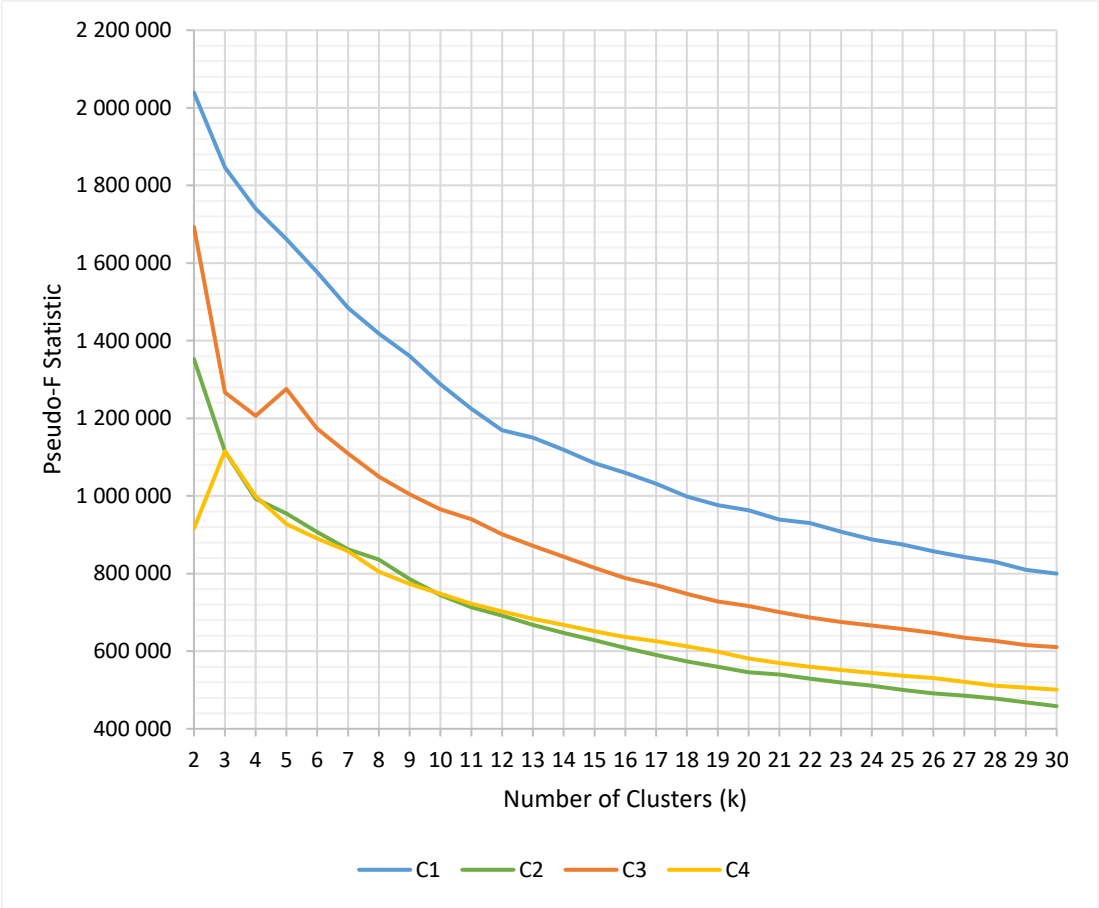


Figure D-1. Pseudo-F statistic of the four campaigns.

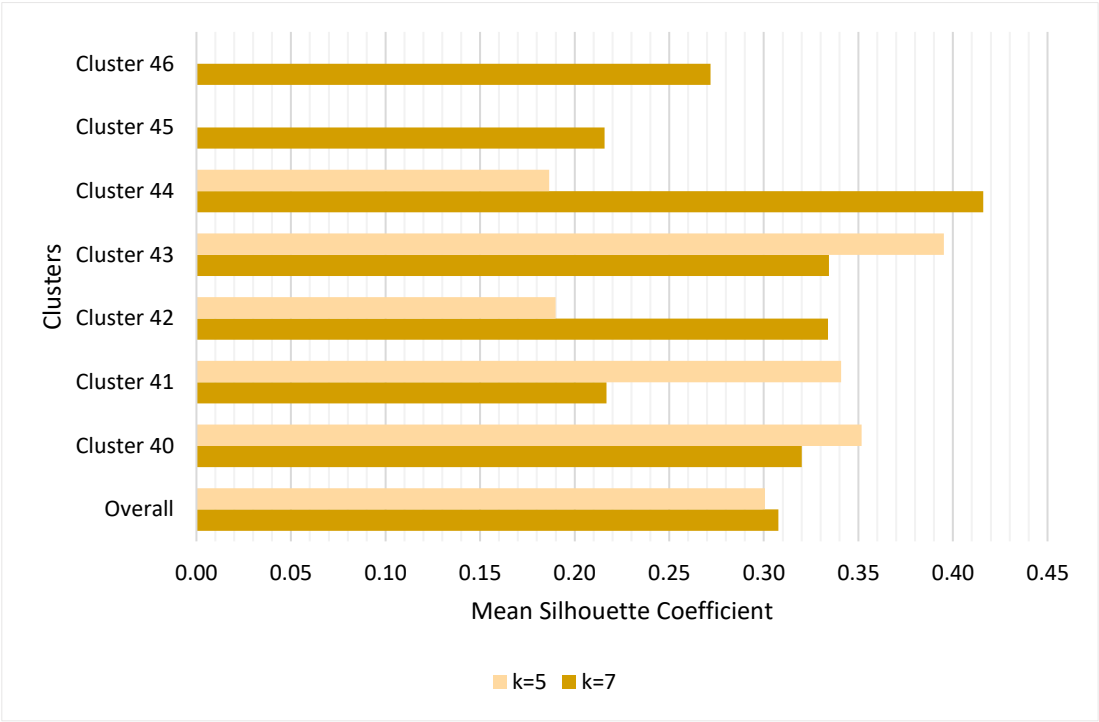


Figure D-2. Mean SC of each cluster (k = 5 and k = 7) in C4.

		Silhouette Coefficient						
Clusters	**Overall**	**10**	**11**	**12**	**13**	**14**	**15**	
Stratified sample	0.387	0.508	0.296	0.251	0.475	0.177	0.430	
Entire dataset	0.386	0.507	0.304	0.240	0.475	0.178	0.417	
Clusters	**Overall**	**20**	**21**	**22**	**23**	**24**		
Stratified sample	0.368	0.491	0.395	0.335	0.250	0.382		
Entire dataset	0.366	0.485	0.391	0.334	0.249	0.383		
Clusters	**Overall**	**30**	**31**	**32**	**33**	**34**	**35**	
Stratified sample	0.320	0.338	0.235	0.250	0.431	0.401	0.288	
Entire dataset	0.320	0.336	0.229	0.253	0.437	0.399	0.287	
Clusters	**Overall**	**40**	**41**	**42**	**43**	**44**	**45**	**46**
Stratified sample	0.308	0.320	0.217	0.334	0.334	0.416	0.216	0.272
Entire dataset	0.307	0.310	0.221	0.333	0.333	0.416	0.222	0.267

Table D-1. Mean SC of each cluster, for the stratified sample and the entire dataset.

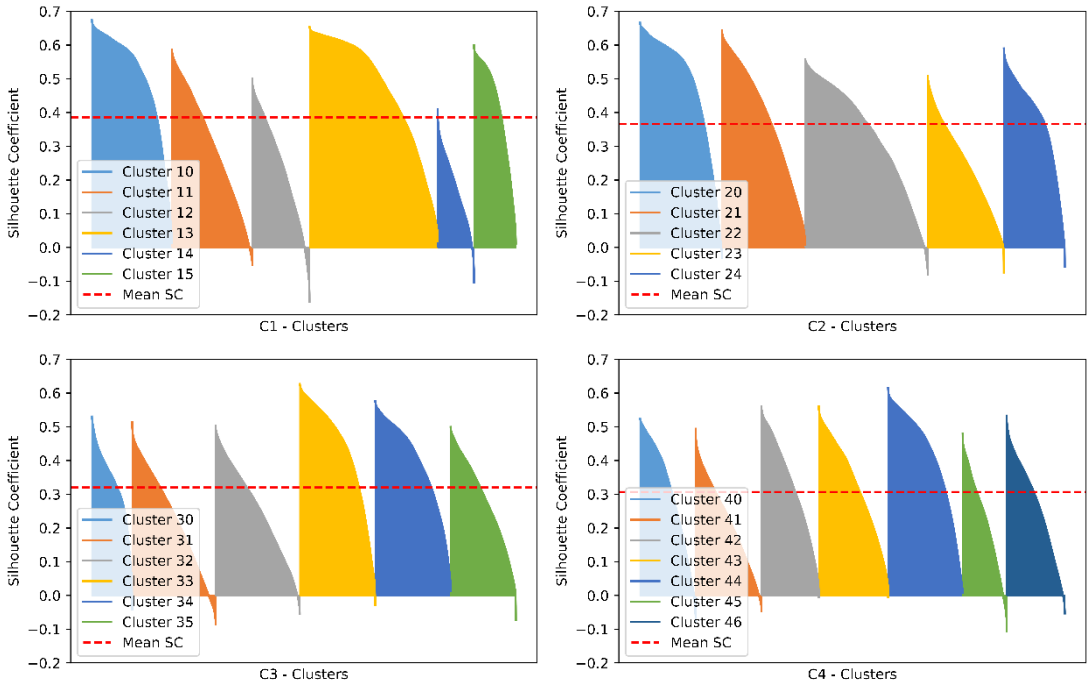


Figure D-3. SC distribution of each cluster in each campaign.

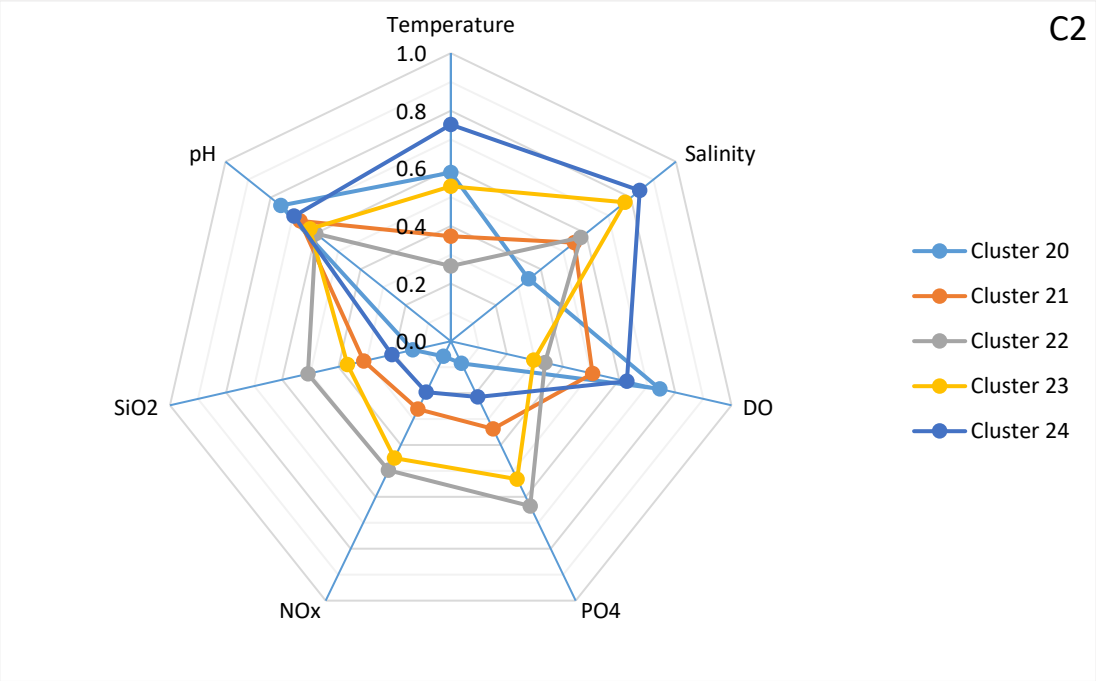


Figure D-4. Clusters centroids of C2 (normalized values).

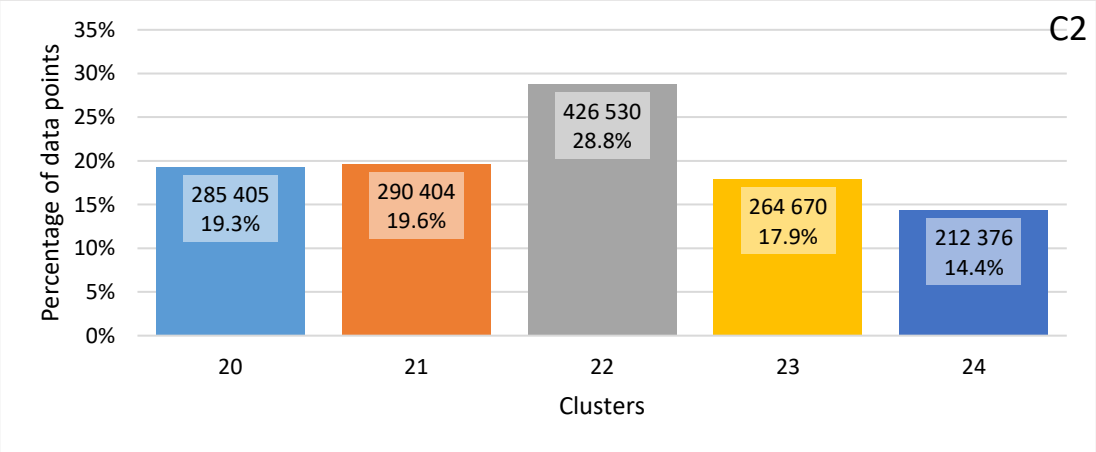


Figure D-5. Size of the clusters of C2.

Figure D-6. Area distribution of the clusters of C2.

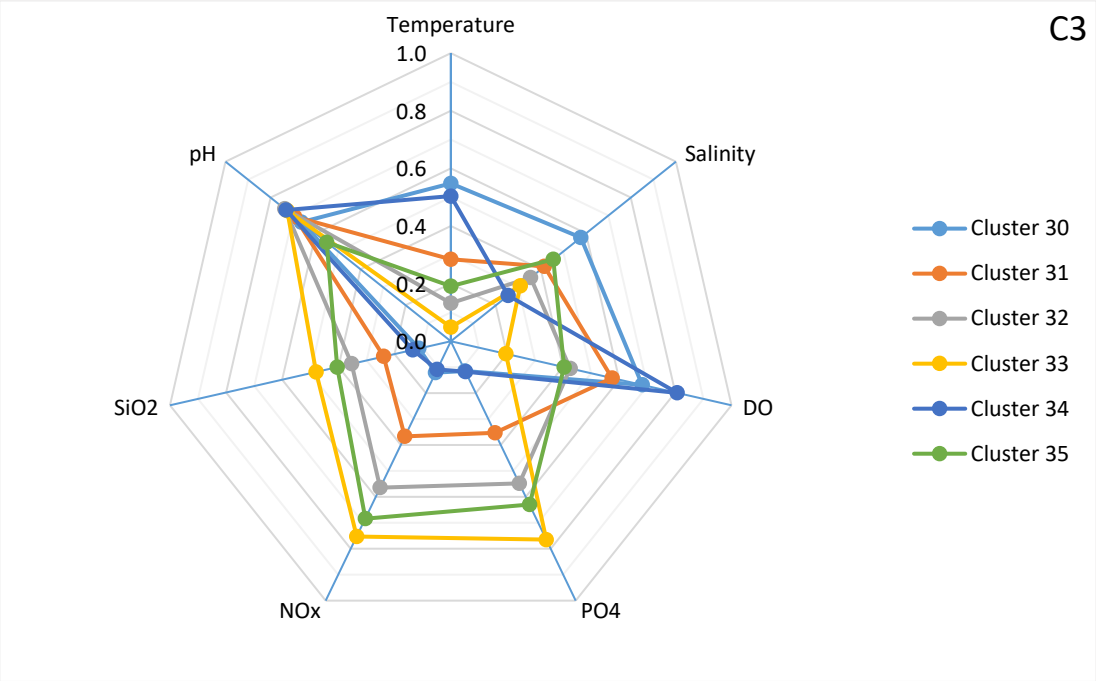


Figure D-7. Cluster centroids of C3 (normalized values).

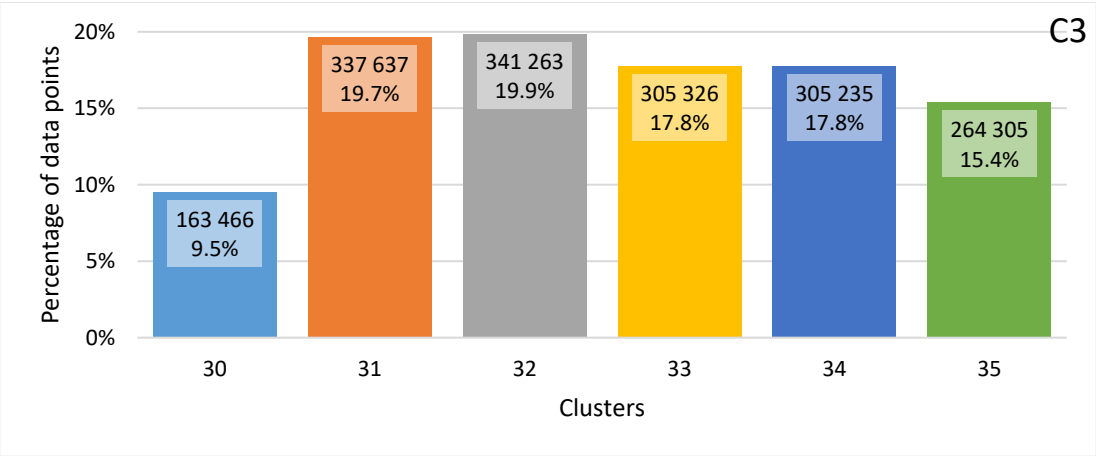


Figure D-8. Size of the clusters of C3.

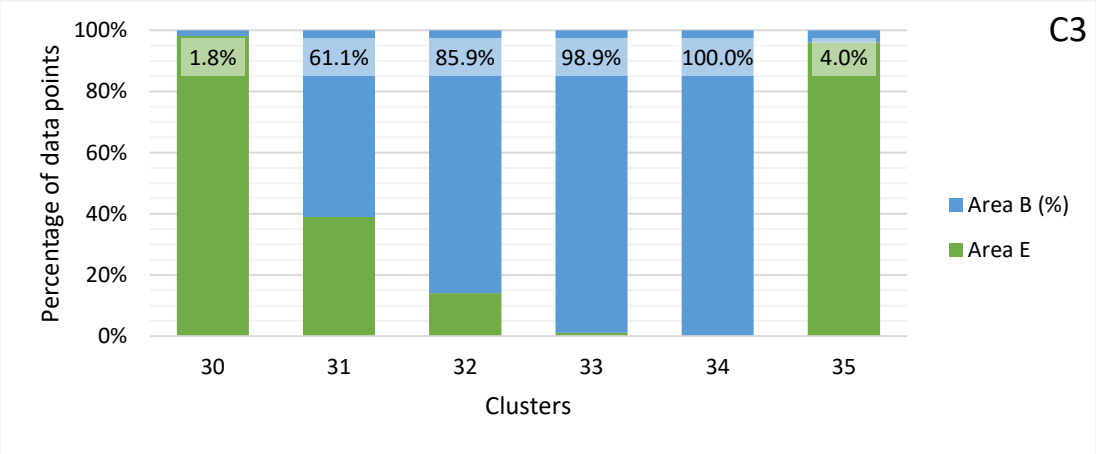


Figure D-9. Area distribution of the clusters of C3.

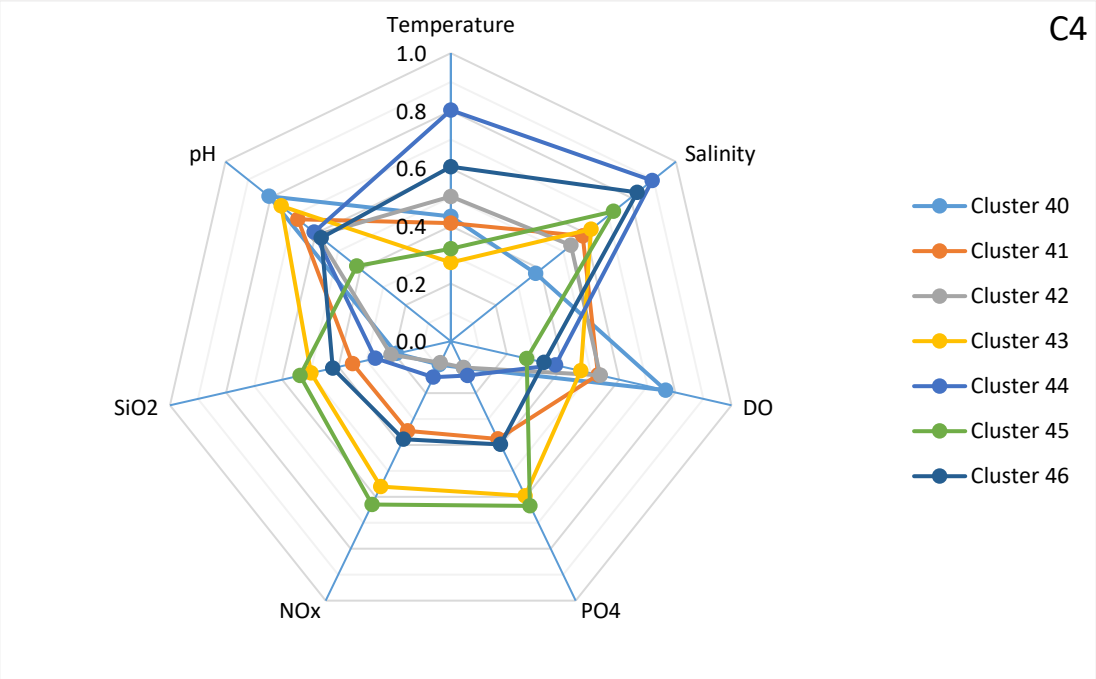


Figure D-10. Cluster centroids of C4 (normalized values).

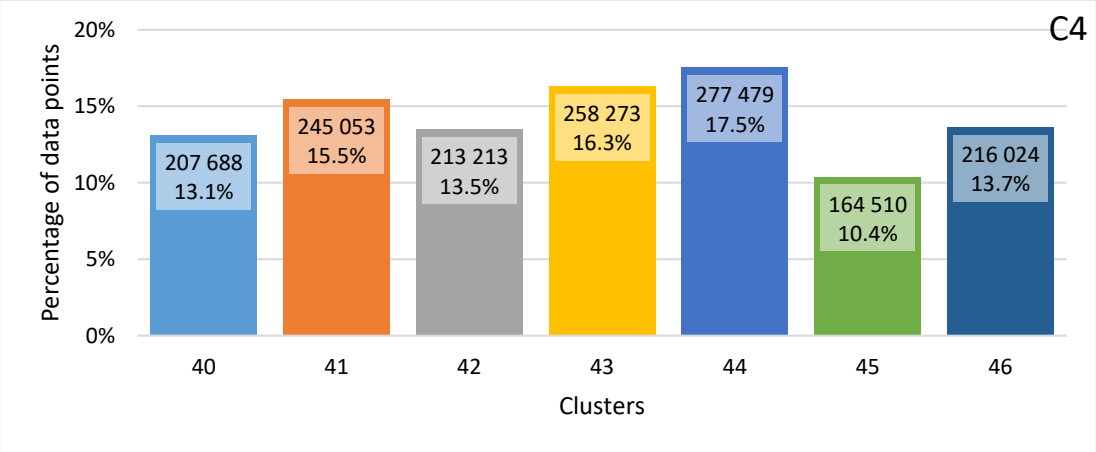


Figure D-11. Size of the clusters of C4.

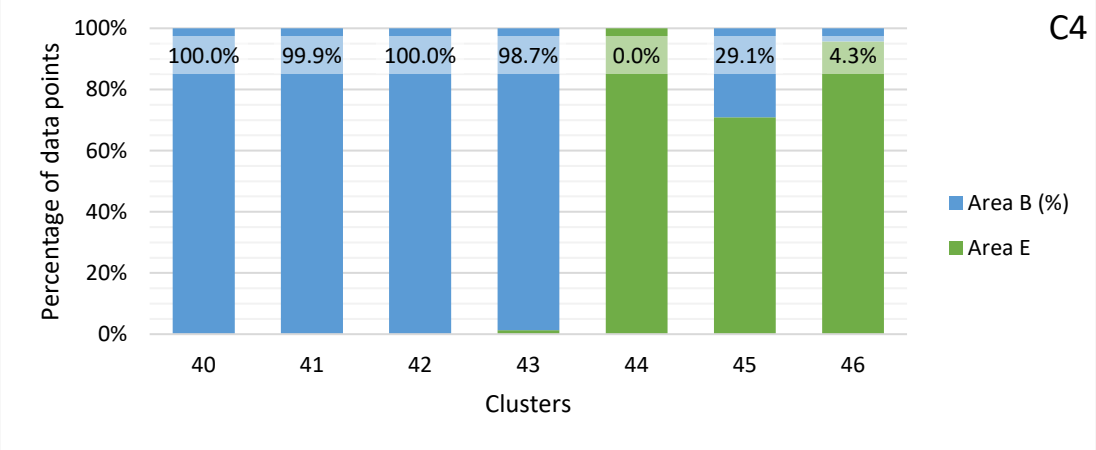


Figure D-12. Area distribution of the clusters of C4.

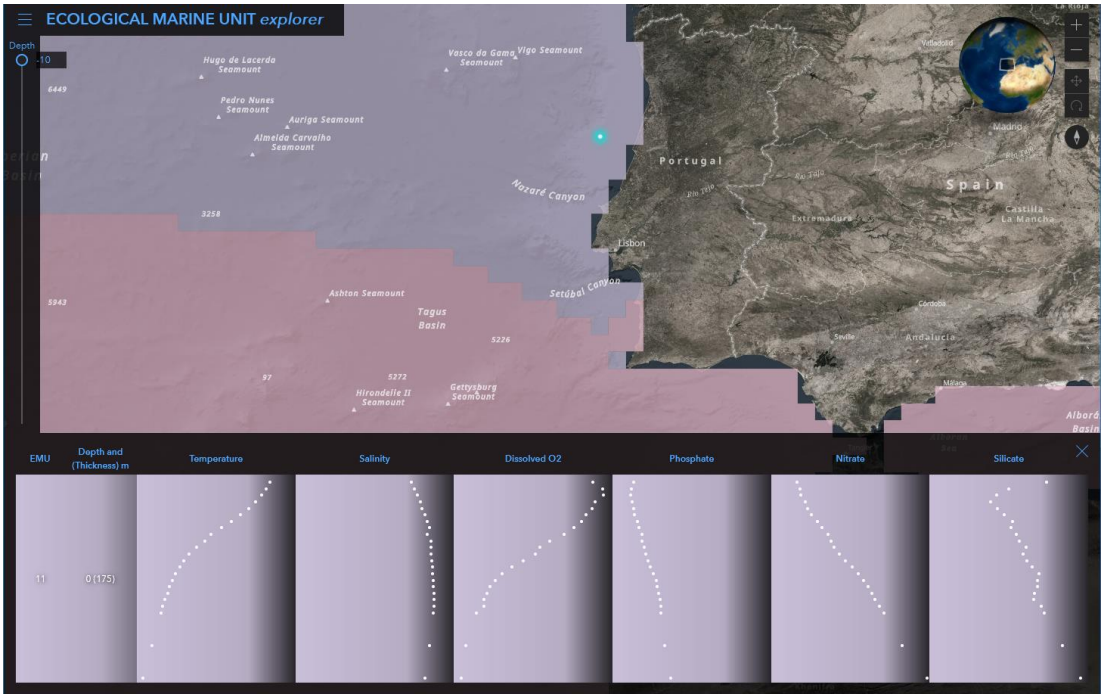


Figure D-13. EMUs in area B.
Source: https://livingatlas.arcgis.com/emu.

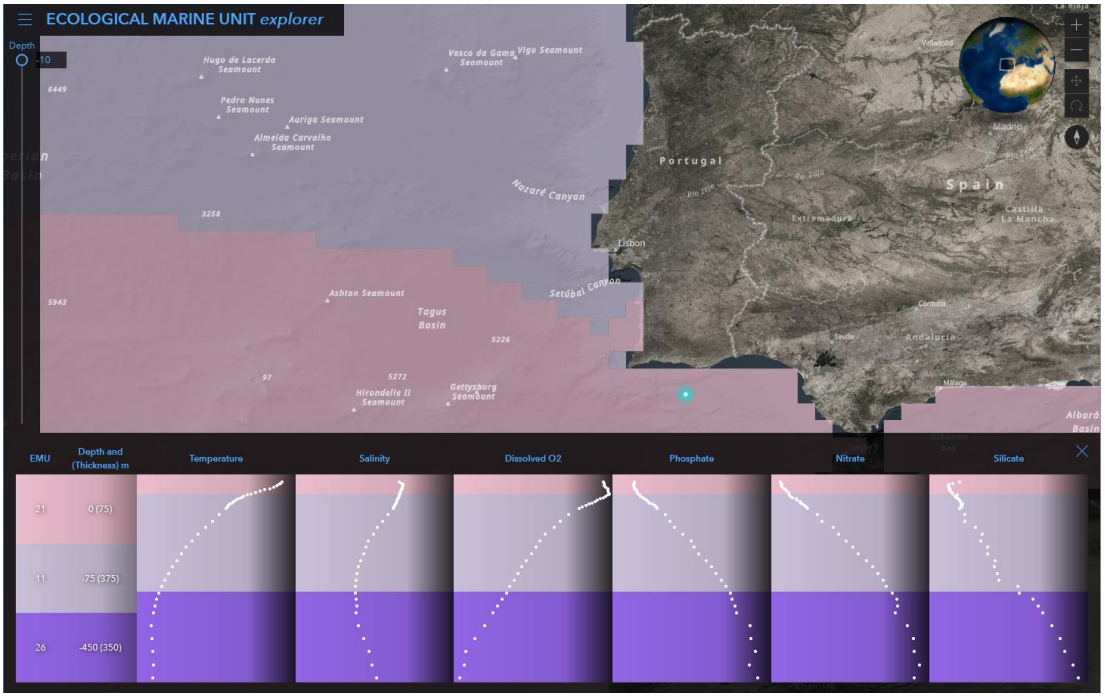


Figure D-14. EMUs in area E.
Source: https://livingatlas.arcgis.com/emu.

	Temperature (°C)	Salinity (PSU)	DO (µmol/L)	Phosphate (µmol/L)	Nitrate (µmol/L)	Silicate (µmol/L)
EMU 11	Epipelagic, Moderate to Cool, Euhaline, Oxic, Low Nitrate, Low Phosphate, Low Silicate					
Minimum	10.07	33.93	3.25	0.05	0.01	0.27
Mean	16.46	35.33	5.23	0.39	3.66	3.33
Maximum	21.67	36.46	6.46	1.40	19.98	20.34
Std. Dev.	2.26	0.41	0.46	0.19	3.27	1.86
EMU 21	Epipelagic, Warm to Very Warm, Euhaline, Oxic, Low Nitrate, Low Phosphate, Low Silicate					
Minimum	12.67	35.77	2.43	0.01	0.01	0.39
Mean	22.53	36.48	4.80	0.20	1.56	1.72
Maximum	29.73	38.13	5.88	1.40	16.50	18.82
Std. Dev.	3.00	0.36	0.40	0.15	2.34	1.14

Table D-2. Characteristics of global EMUs 11 and 21.
Source: (Sayre, et al., 2017b).

	Temperature (°C)	Salinity (PSU)	DO (mgO_2/L)	PO_4 (µmol/L)	NO_X (µmol/L)	SiO_2 (µmol/L)	Global EMU
C1							
Cluster 10	16.6132	35.8263	7.3848	0.1986	1.6249	2.0001	EMU 11
Cluster 11	14.6385	35.9031	6.9169	0.4126	5.4358	2.9920	
Cluster 12	15.2831	36.1664	6.7590	0.5255	6.9887	5.4138	
Cluster 13	13.4938	35.8732	6.6814	0.6177	9.1058	4.1700	
Cluster 14	17.6557	36.3205	7.3249	0.2452	2.0013	3.0208	EMU 21
Cluster 15	21.5553	36.6221	7.4851	0.1704	0.7732	1.7976	

Table D-3. Equivalence between global and local EMUs of C1.

ANNEX E – VOXEL LAYERS

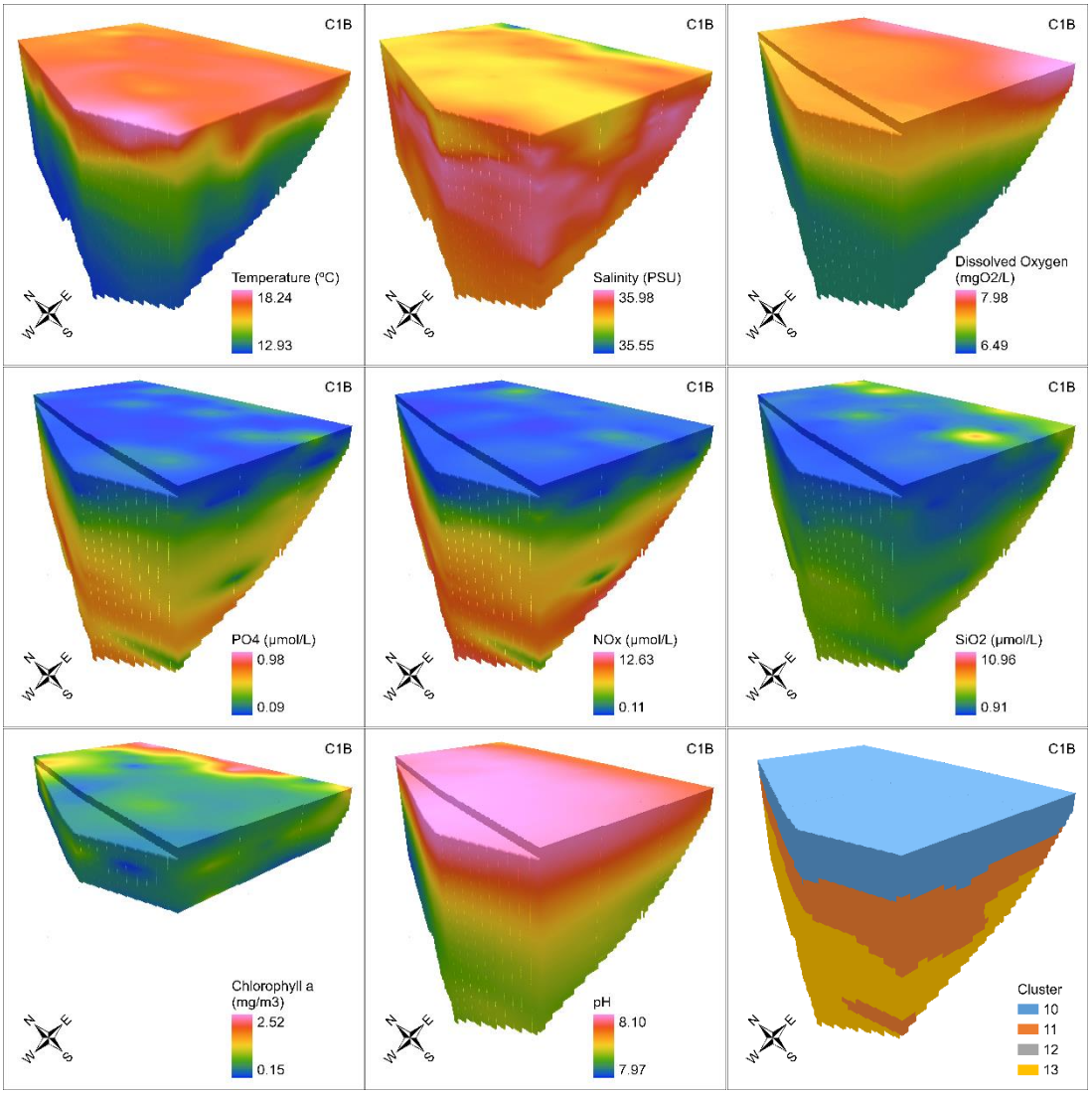


Figure E-1. Voxel layers of the different OVs of C1B.

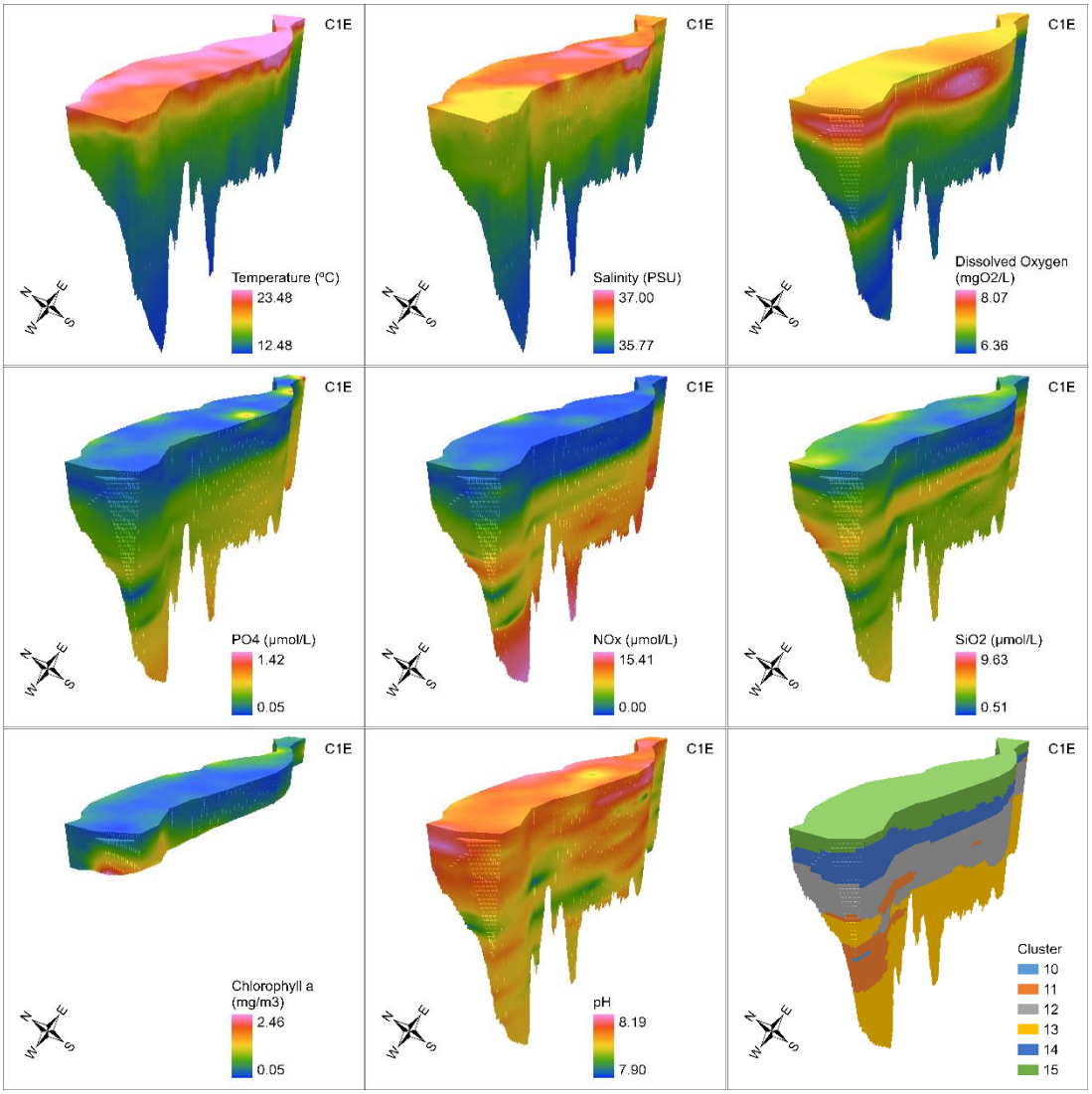


Figure E-2. Voxel layers of the different OVs of C1E.

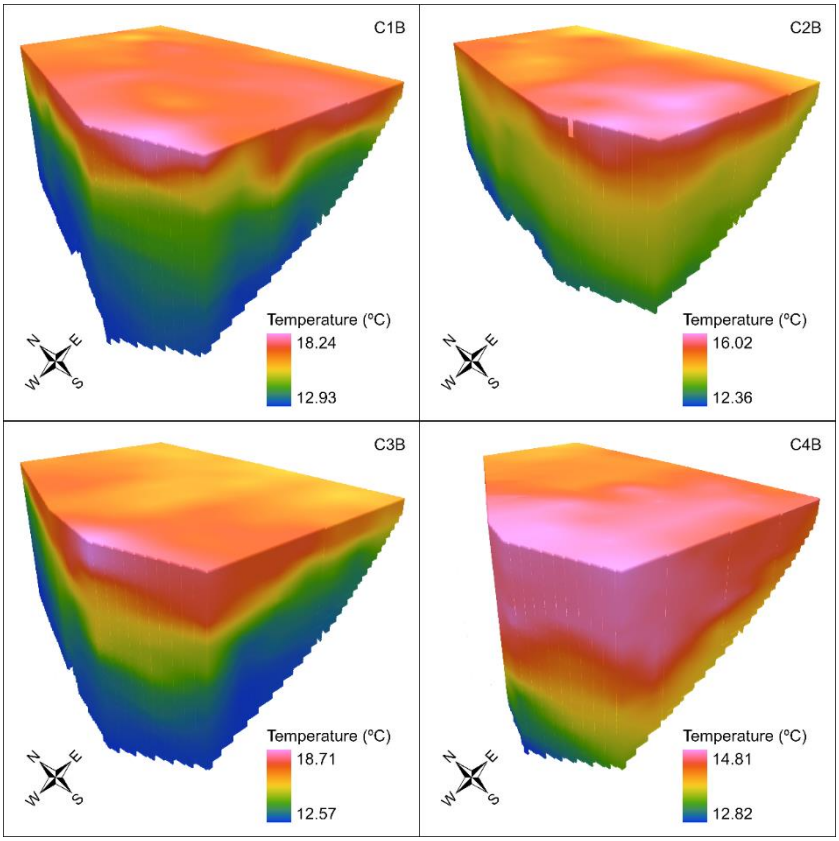


Figure E-3. Voxel layers of temperature in area B.

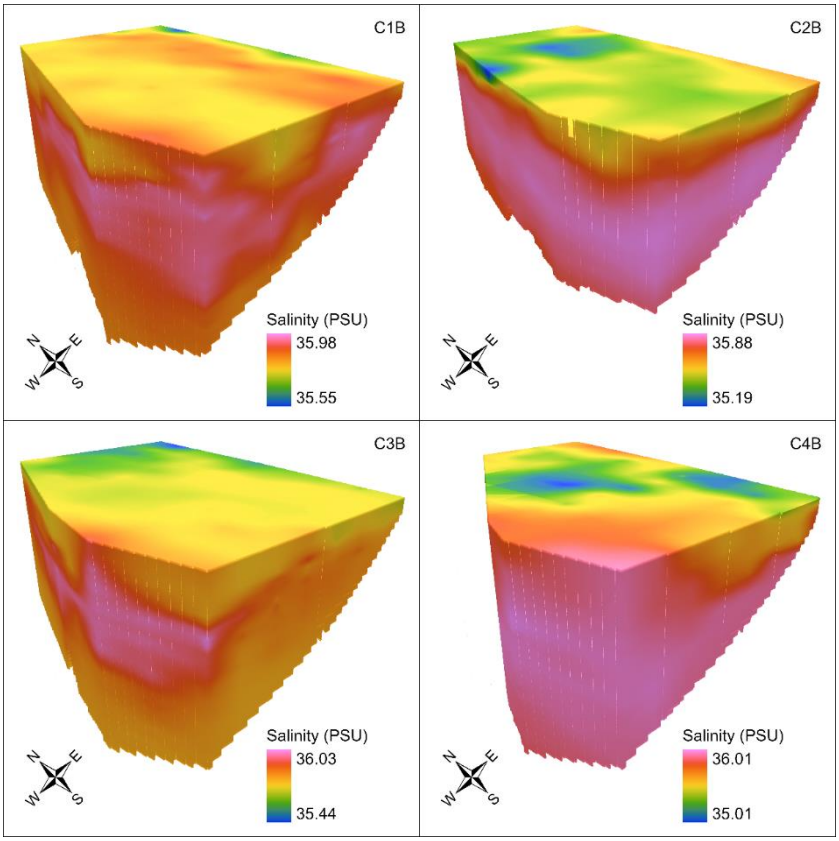


Figure E-4. Voxel layers of salinity in area B.

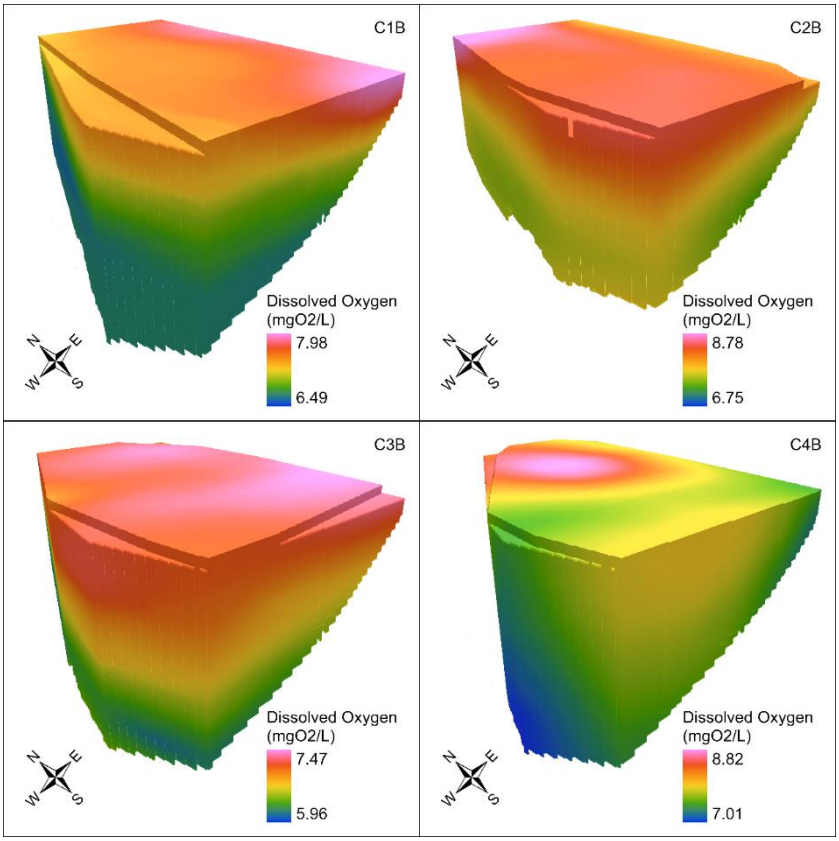


Figure E-5. Voxel layers of DO in area B.

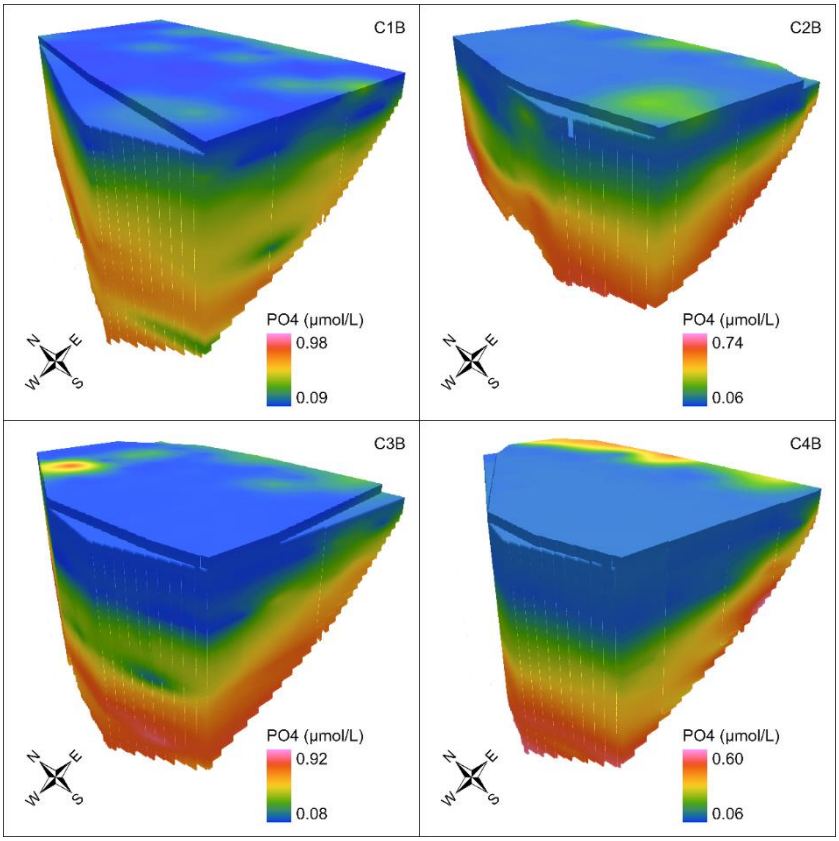


Figure E-6. Voxel layers of PO_4 in area B.

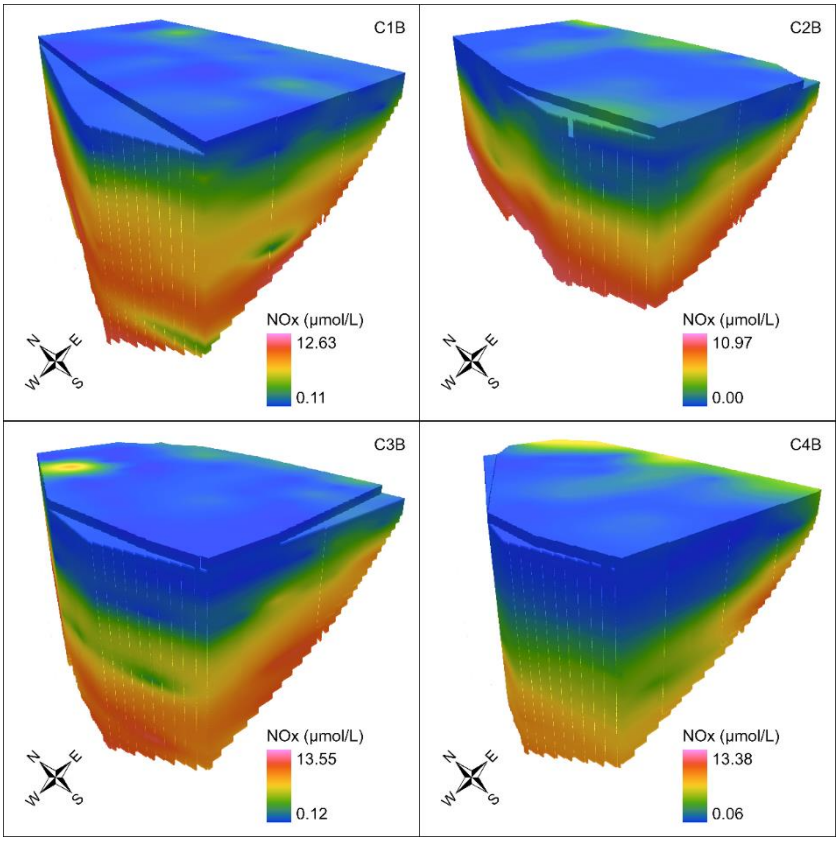


Figure E-7. Voxel layers of NO_X in area B.

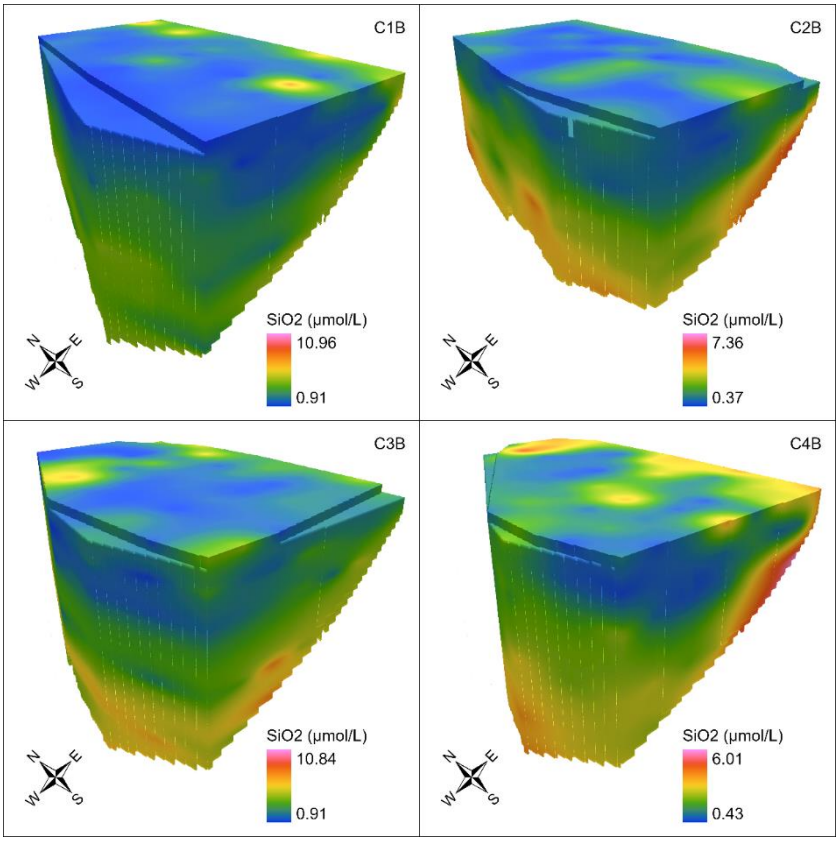


Figure E-8. Voxel layers of SiO_2 in area B.

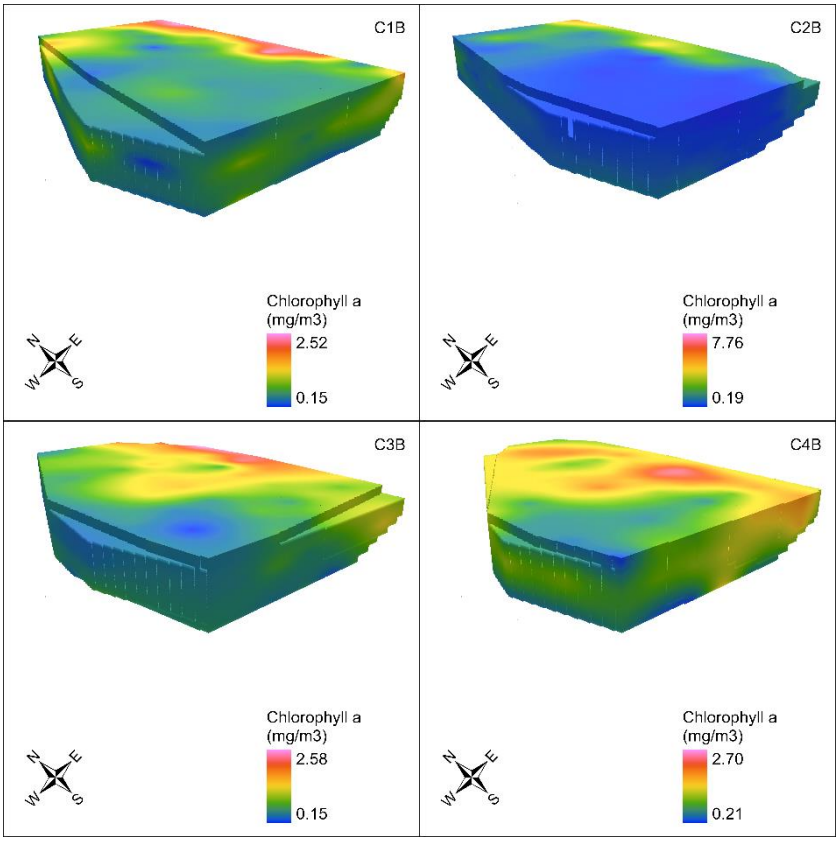


Figure E-9. Voxel layers of chlorophyll a in area B.

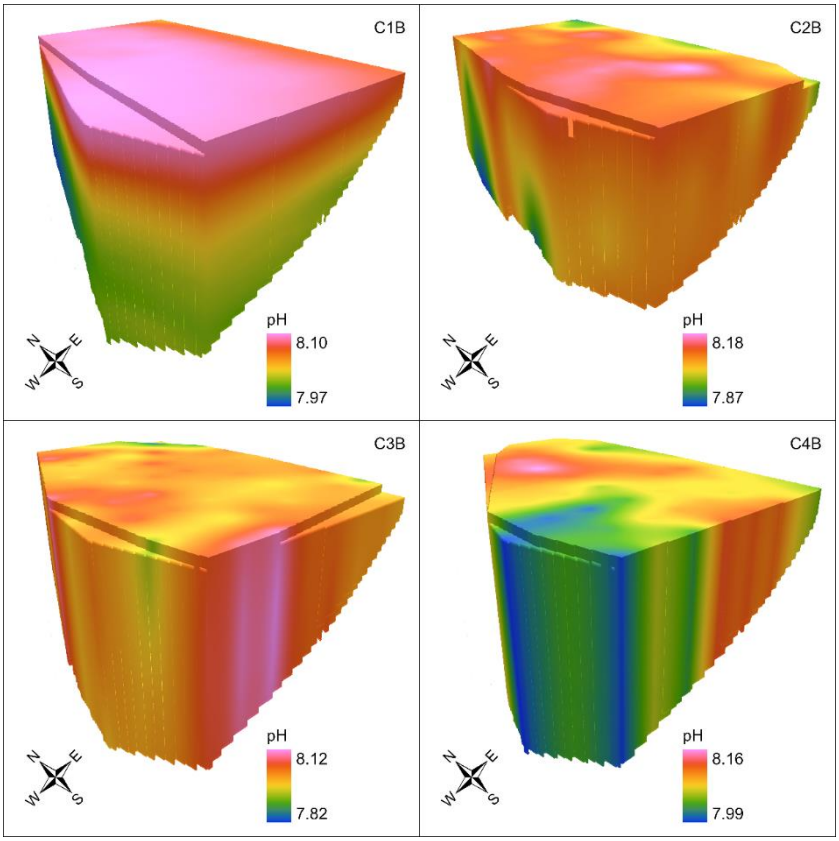


Figure E-10. Voxel layers of pH in area B.

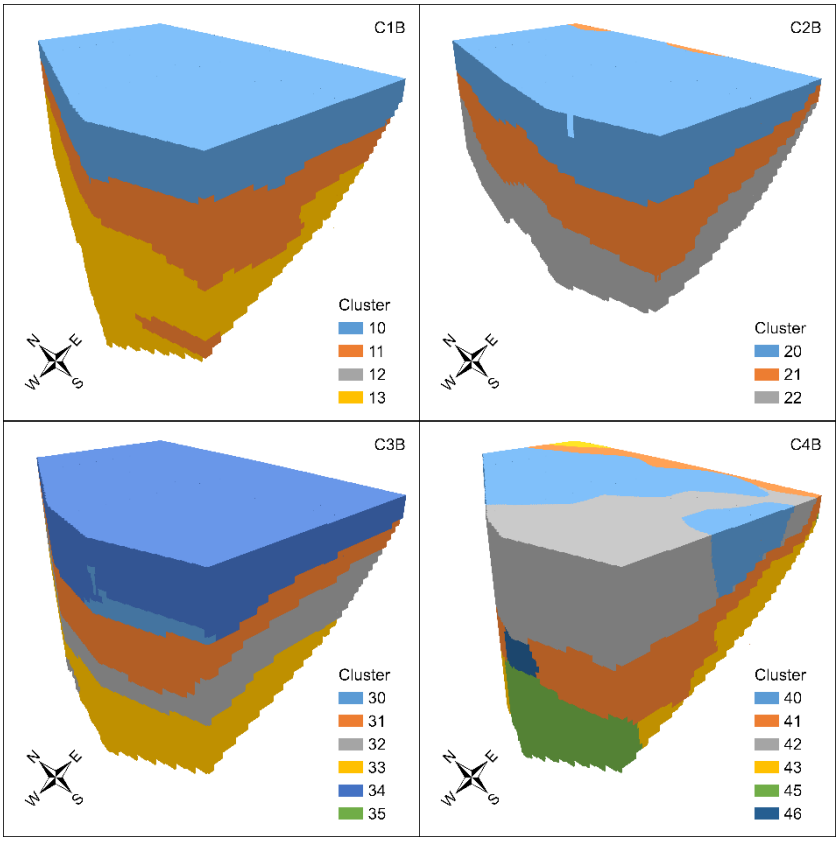


Figure E-11. Voxel layers of clusters in area B.

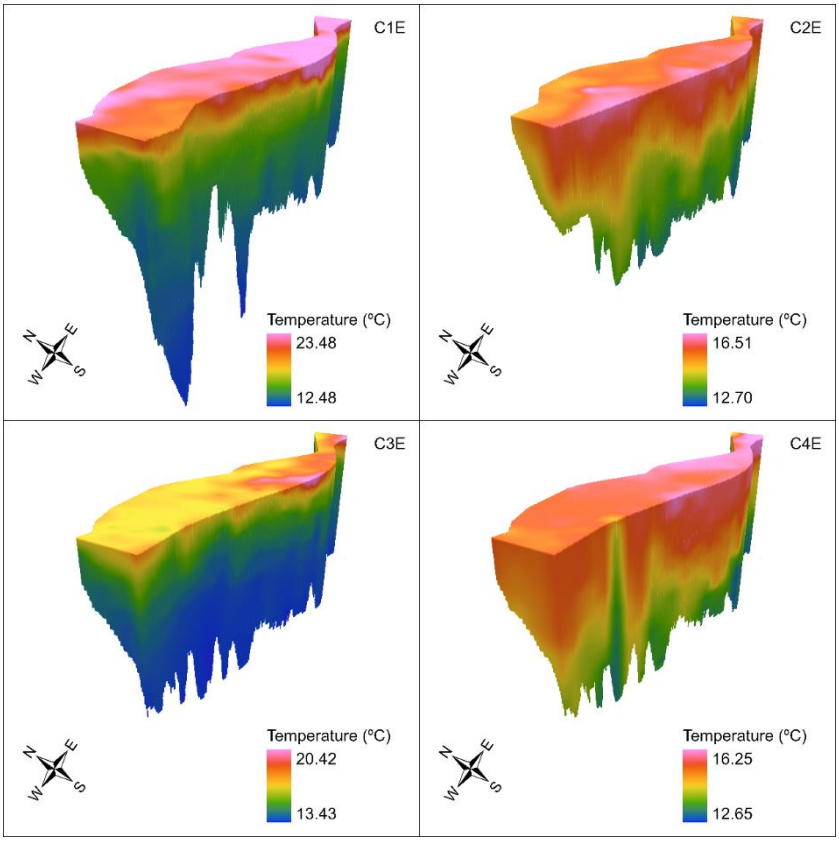


Figure E-12. Voxel layers of temperature in area E.

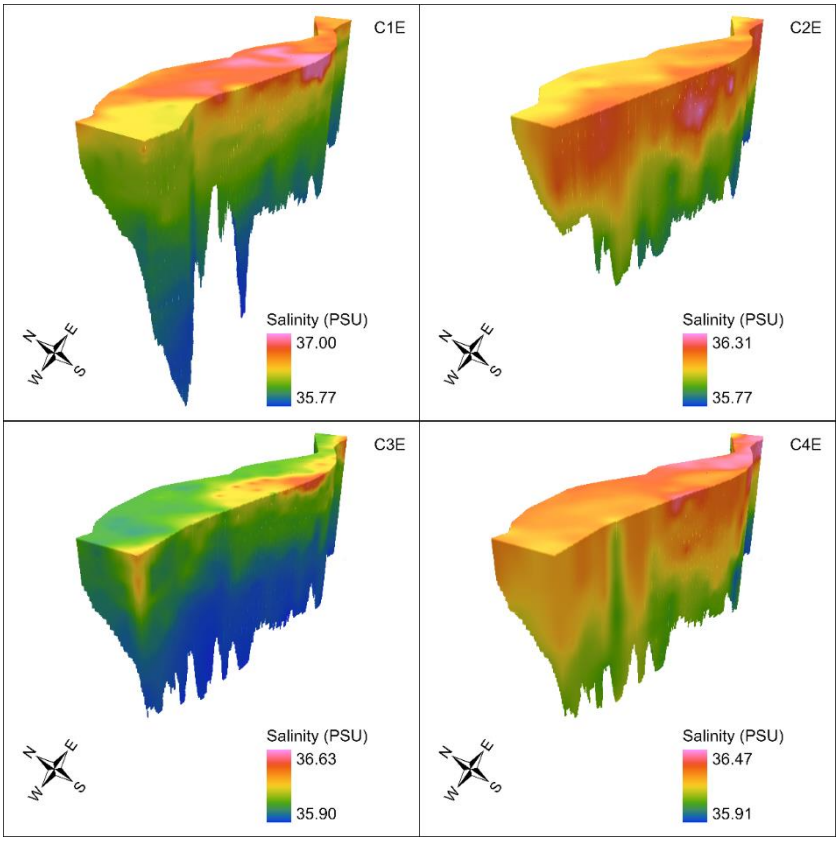


Figure E-13. Voxel layers of salinity in area E.

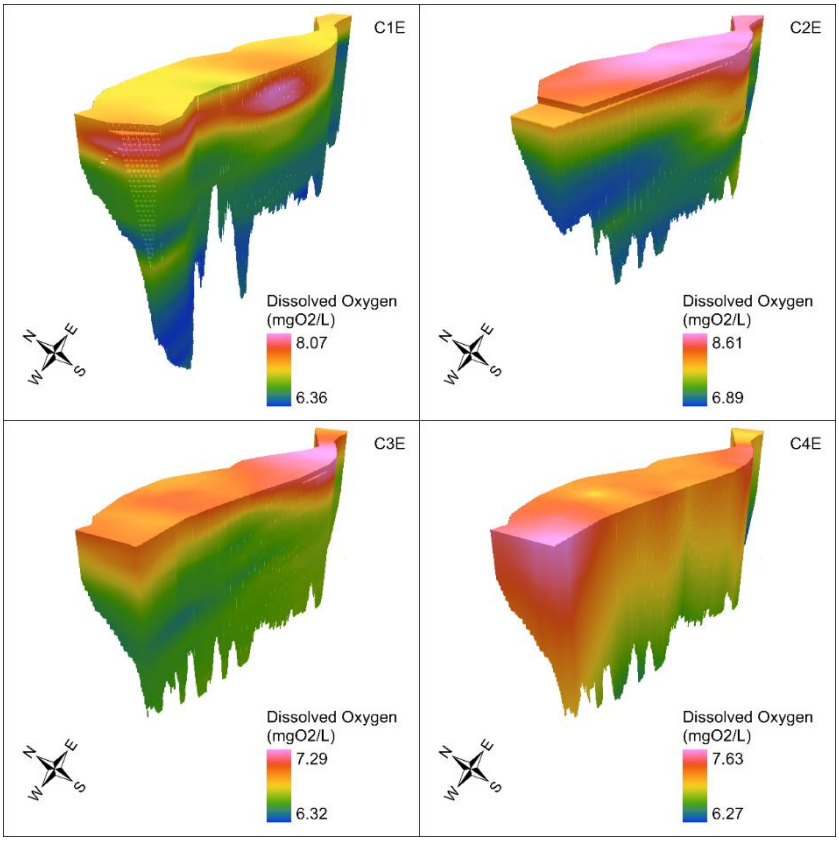


Figure E-14. Voxel layers of DO in area E.

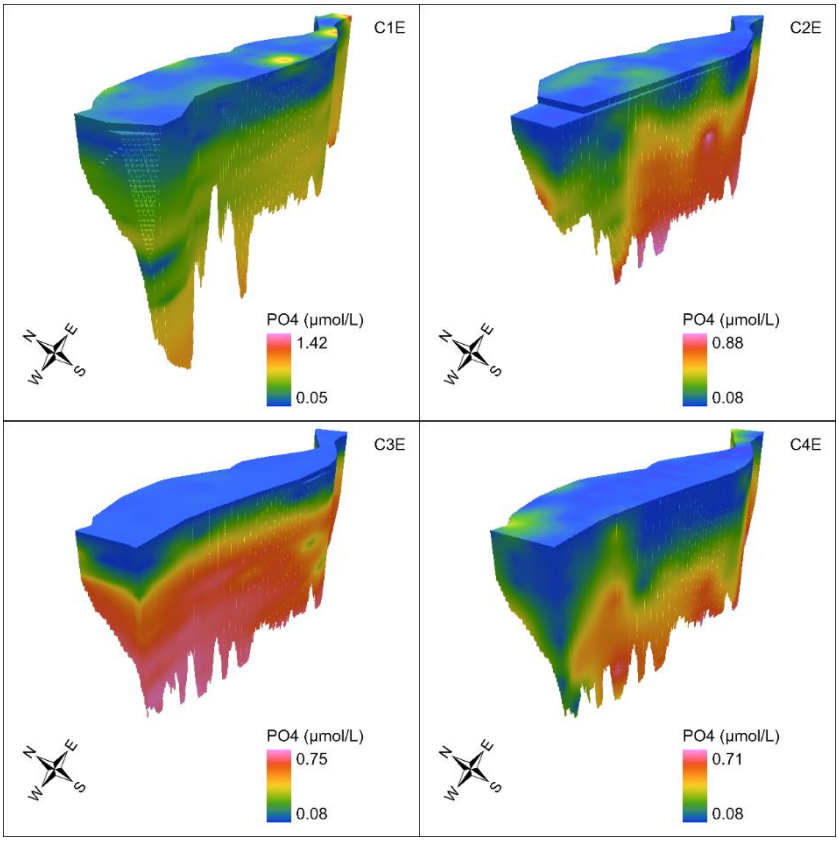


Figure E-15. Voxel layers of PO_4 in area E.

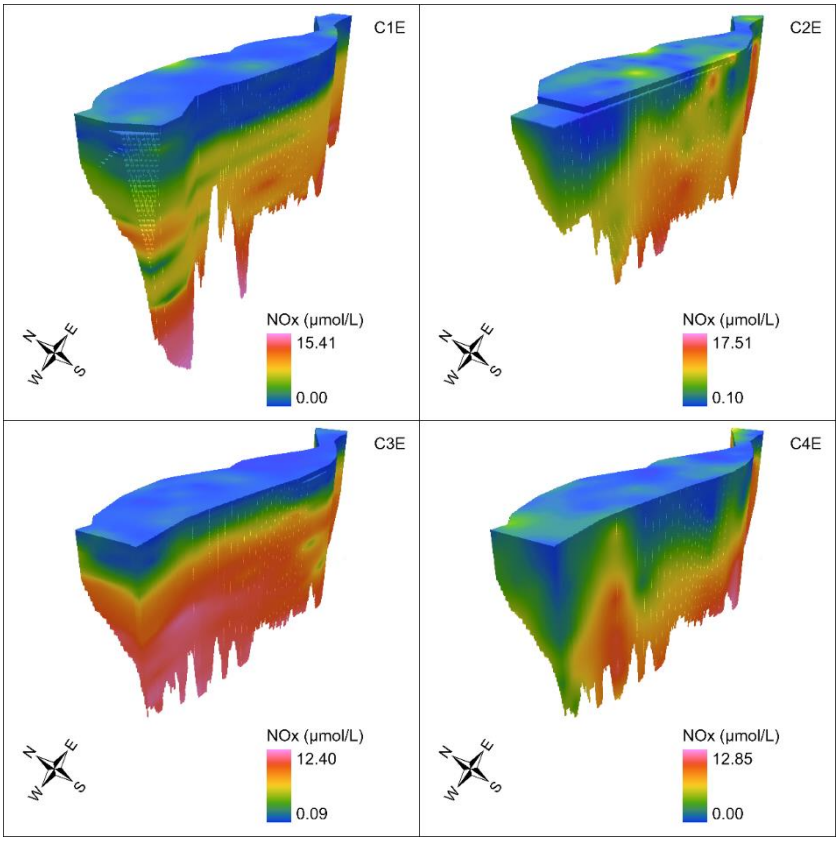


Figure E-16. Voxel layers of NO_X in area E.

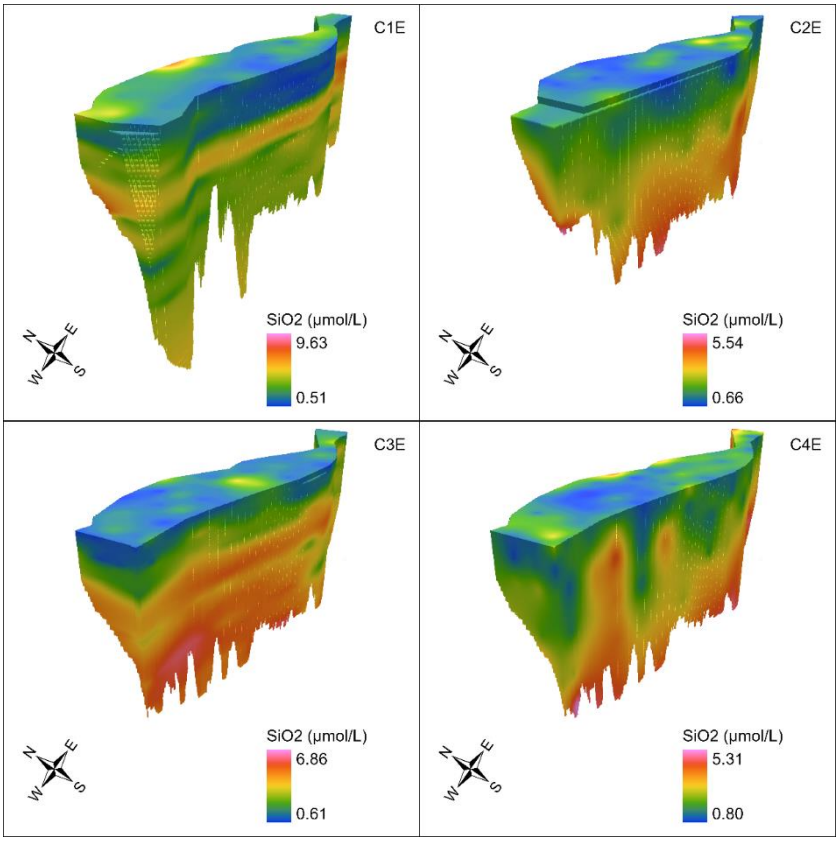


Figure E-17. Voxel layers of SiO_2 in area E.

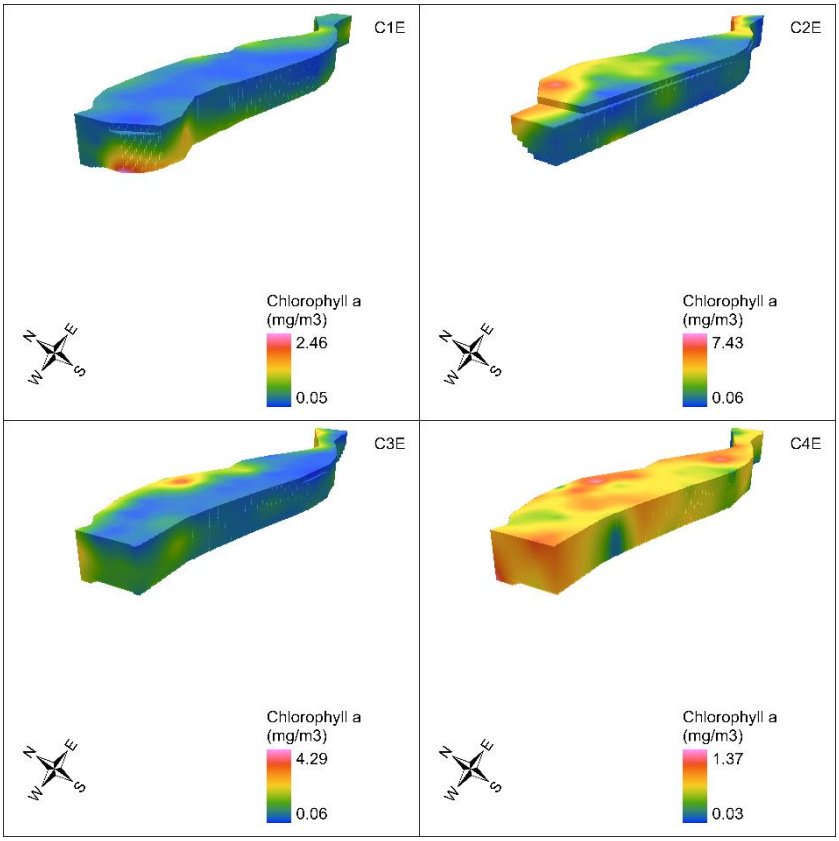


Figure E-18. Voxel layers of chlorophyll a in area E.

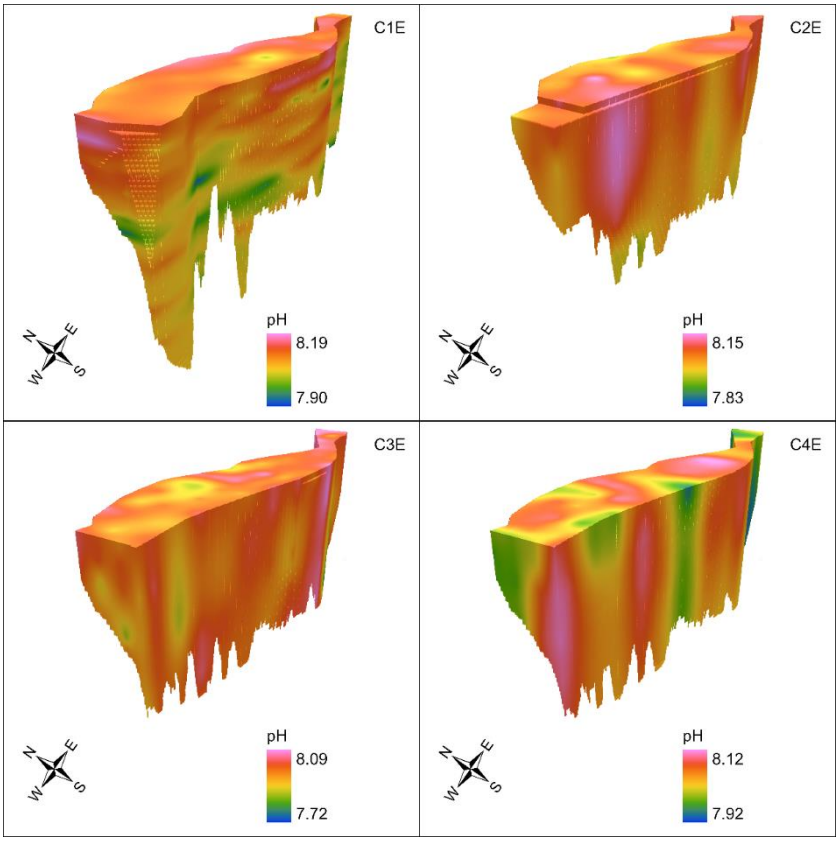


Figure E-19. Voxel layers of pH in area E.

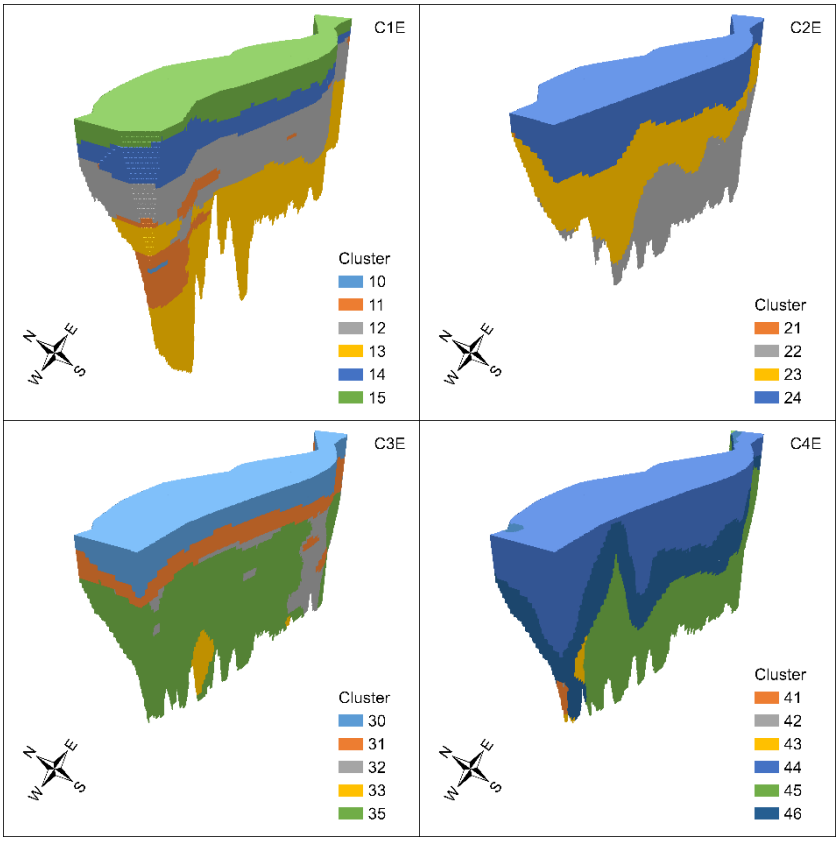


Figure E-20. Voxel layers of clusters in area E.

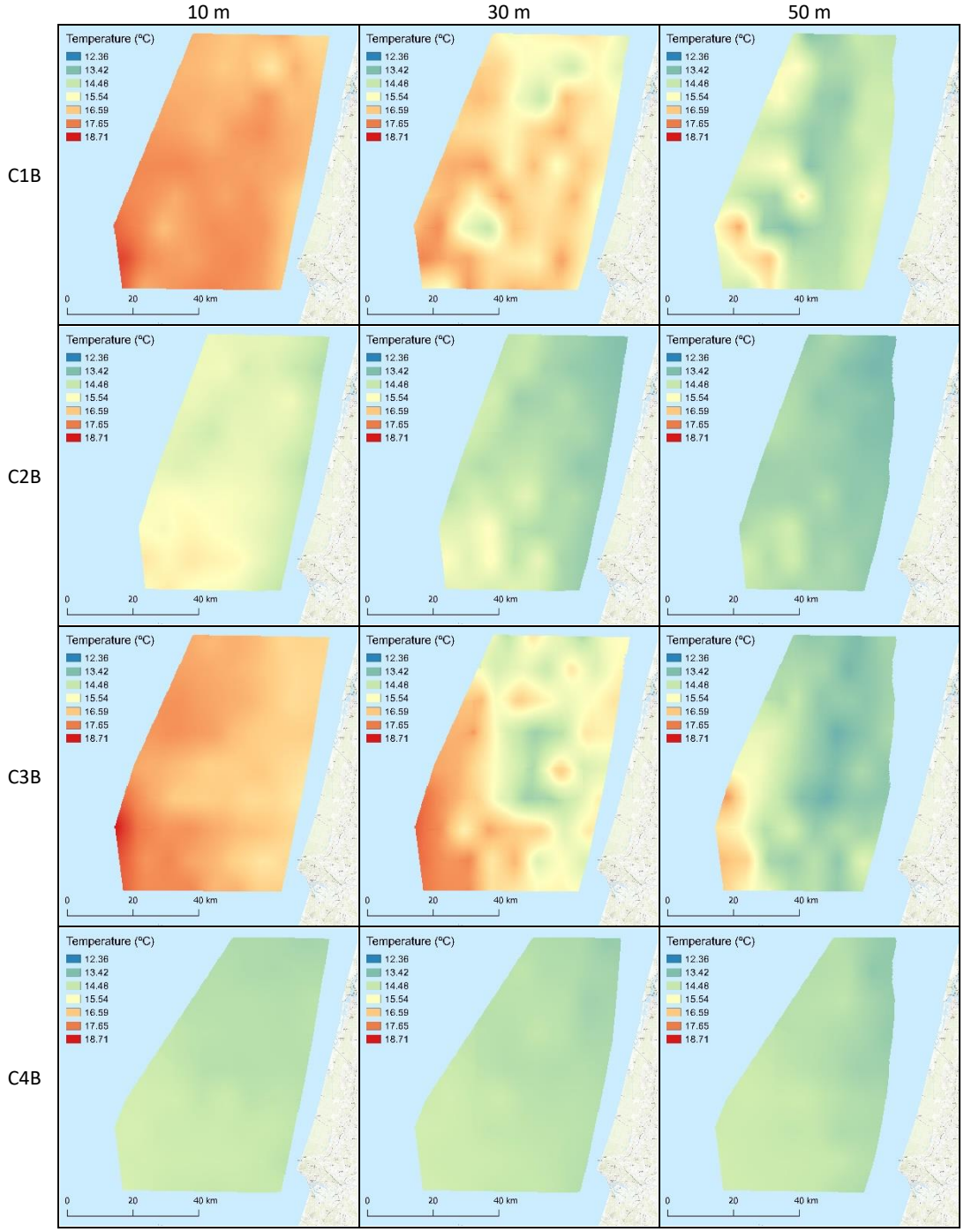


Figure E-21. Horizontal cross-sections of temperature in area B.

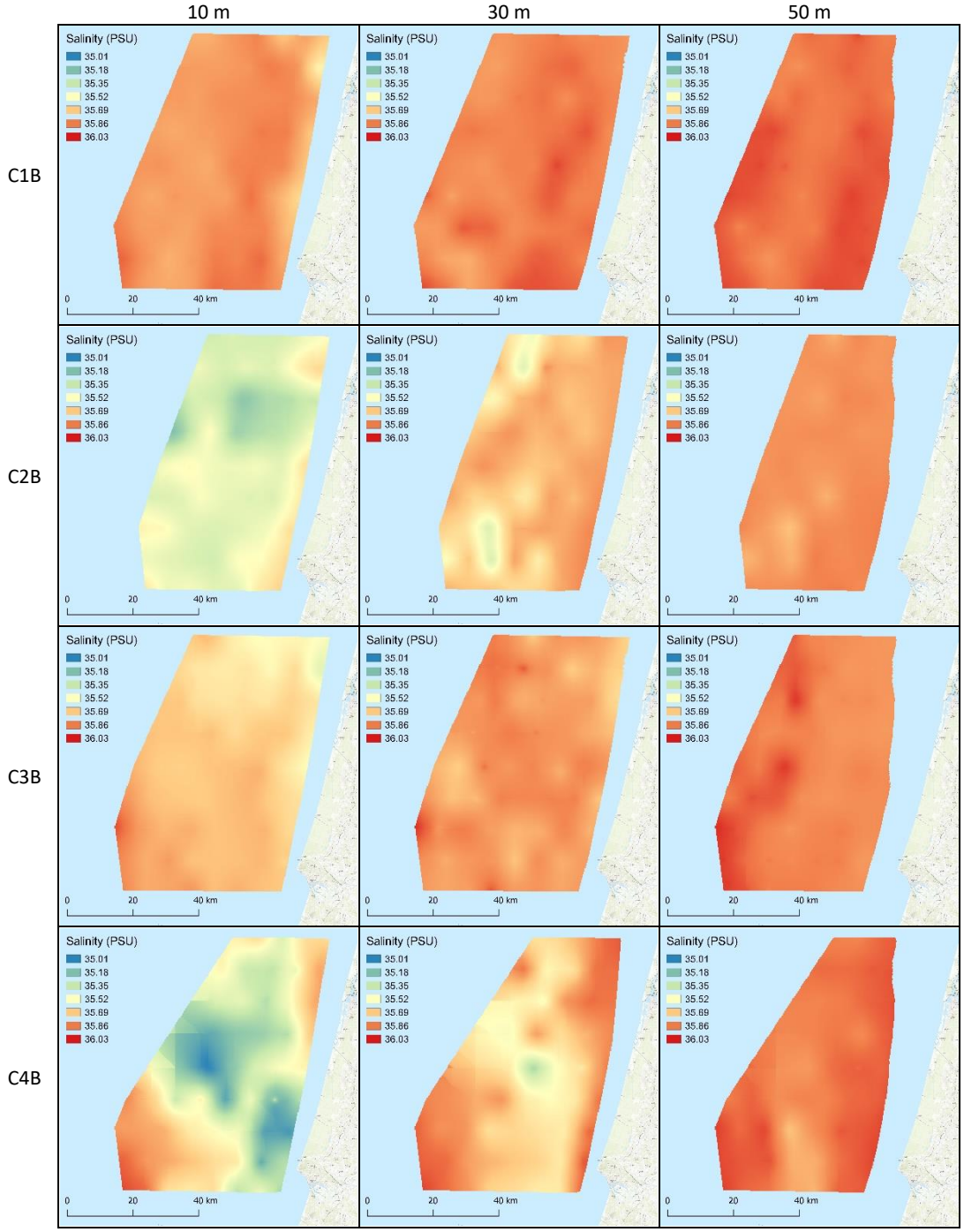


Figure E-22. Horizontal cross-sections of salinity in area B.

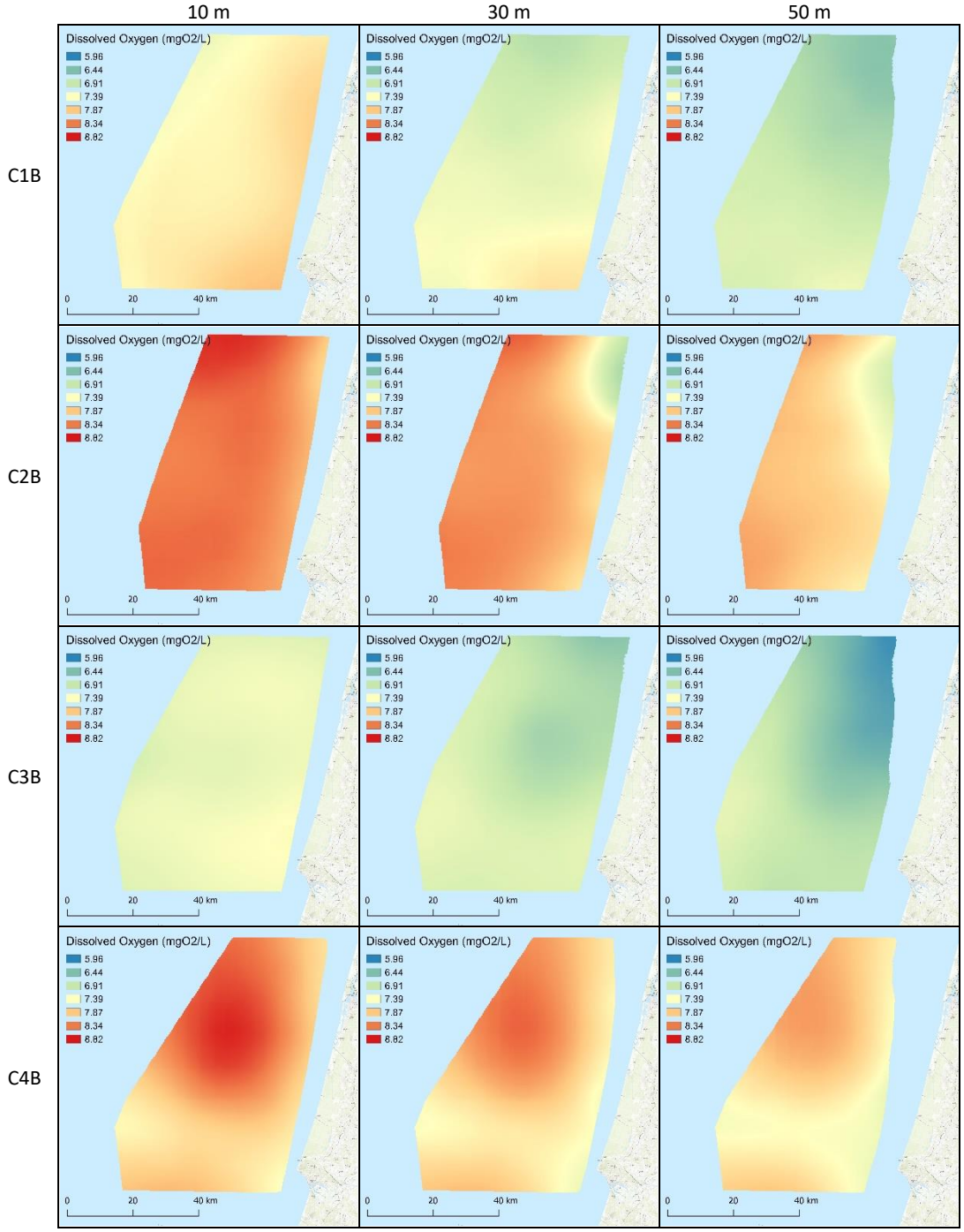


Figure E-23. Horizontal cross-sections of DO in area B.

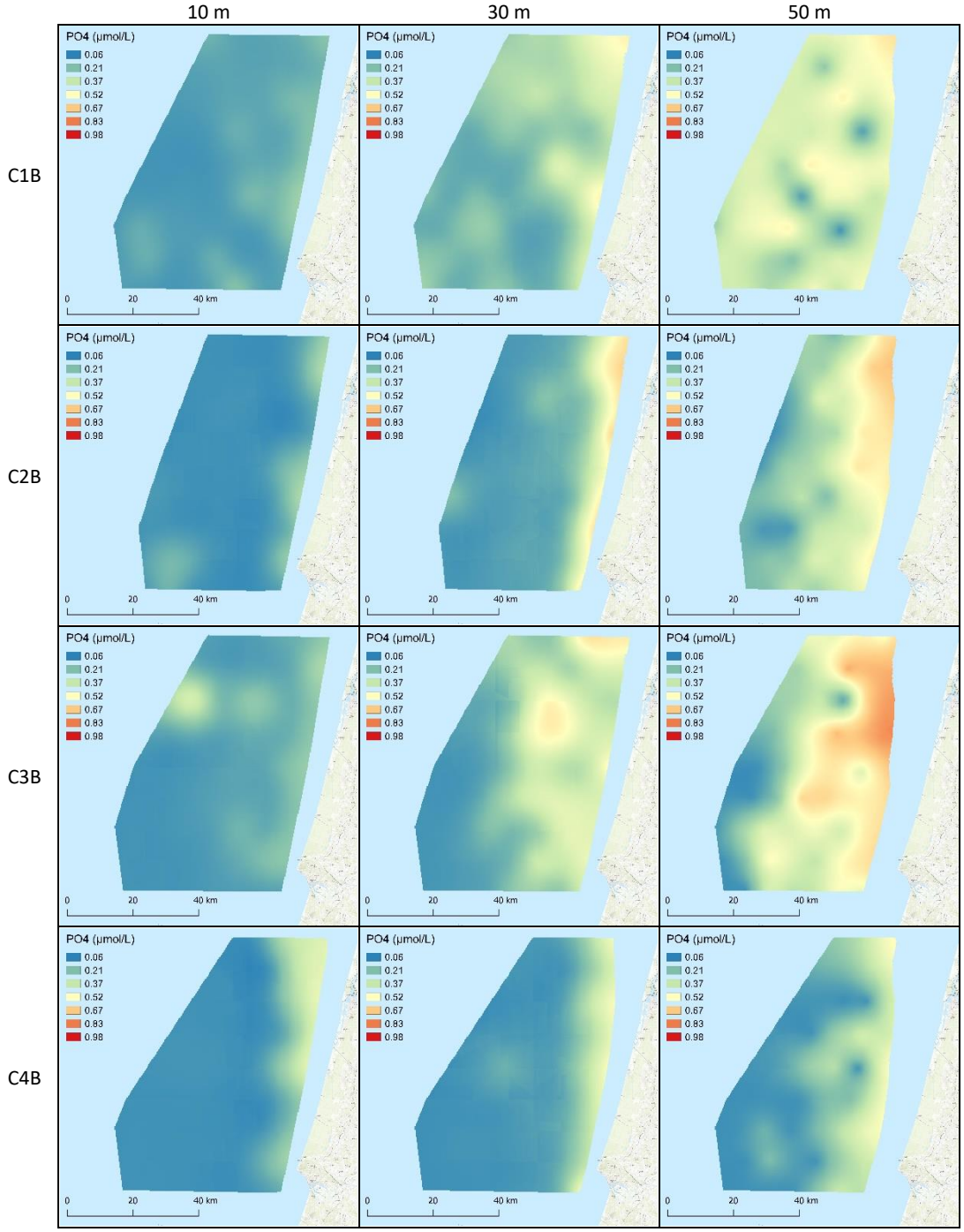


Figure E-24. Horizontal cross-sections of PO_4 in area B.

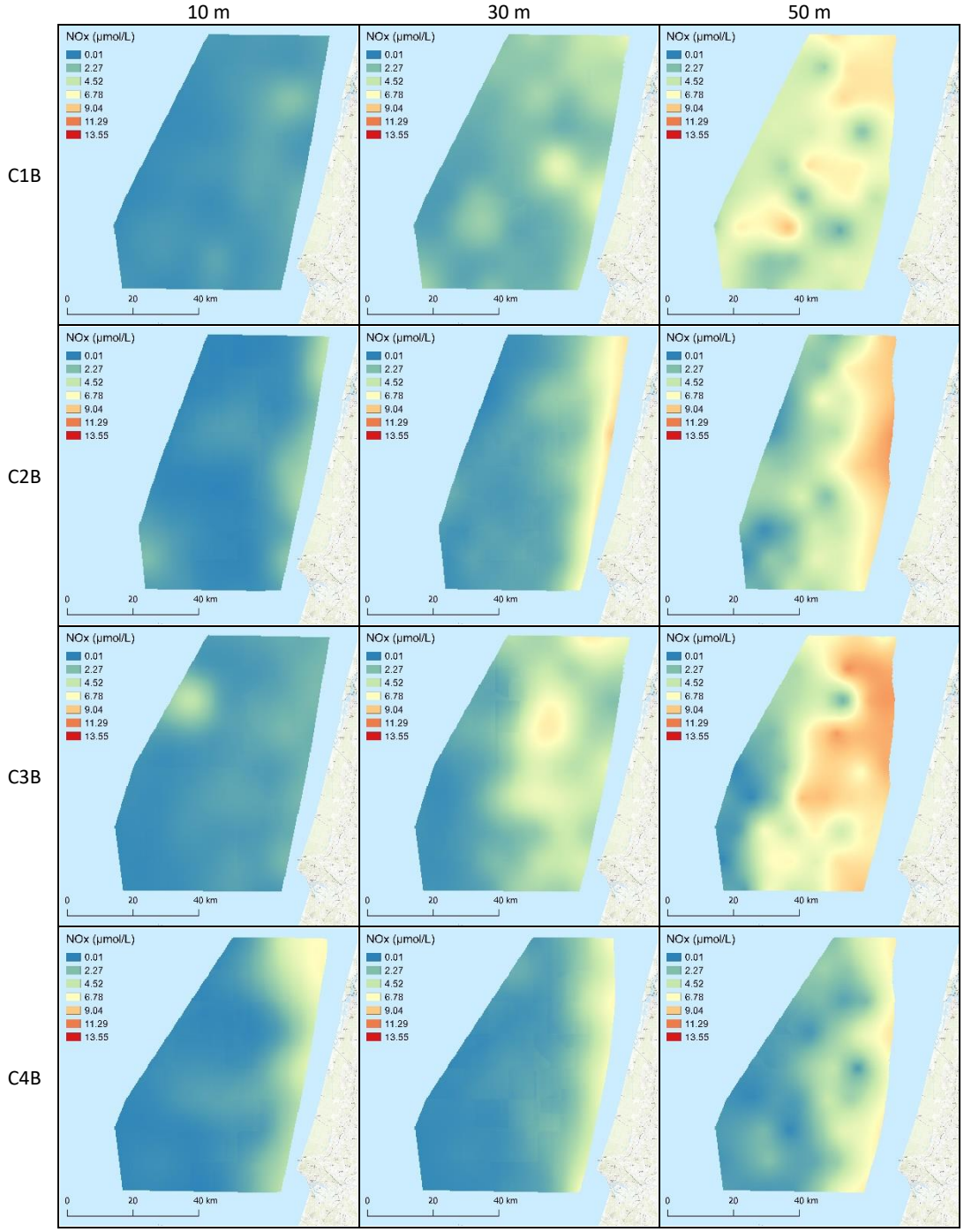


Figure E-25. Horizontal cross-sections of NO_X in area B.

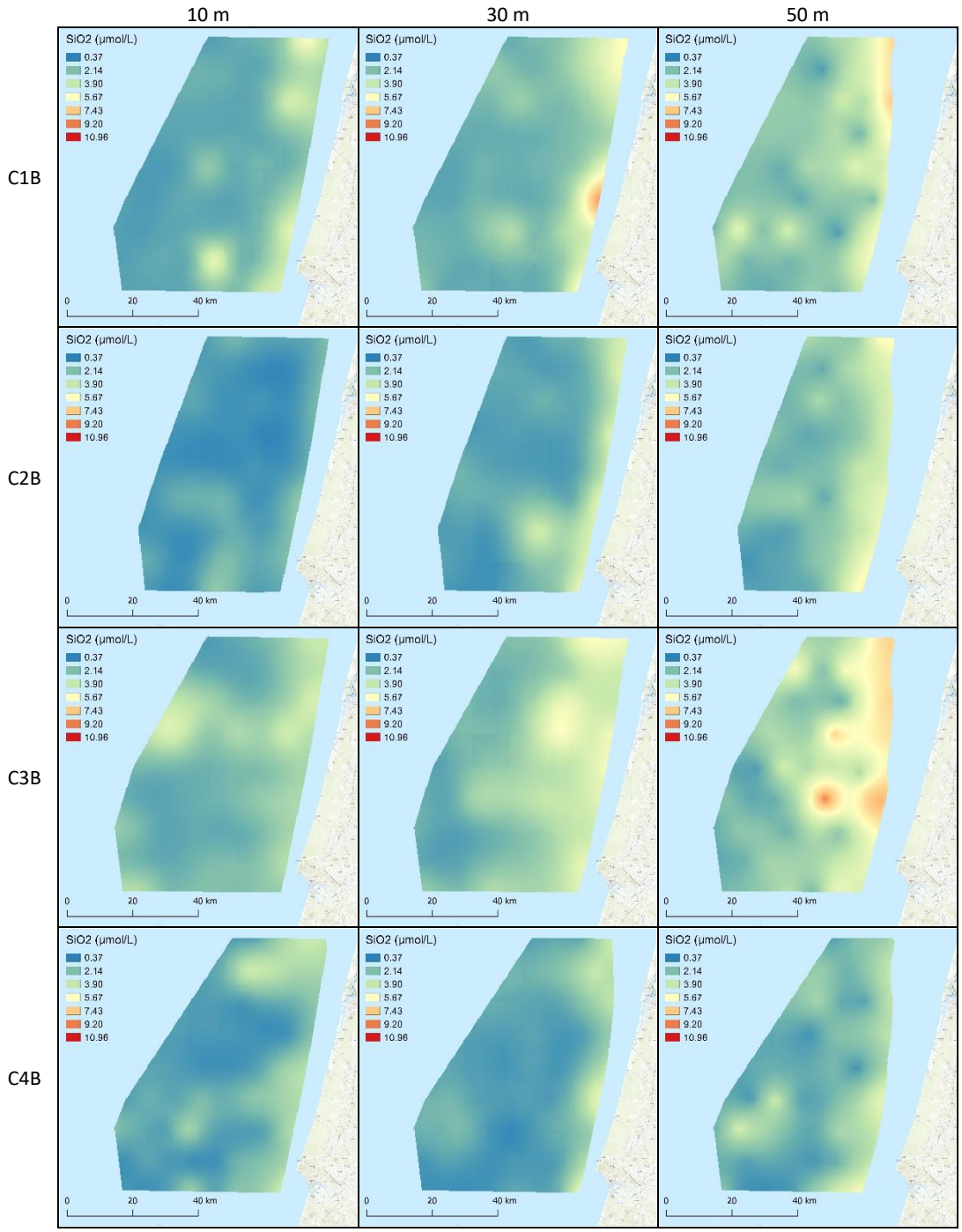


Figure E-26. Horizontal cross-sections of SiO_2 in area B.

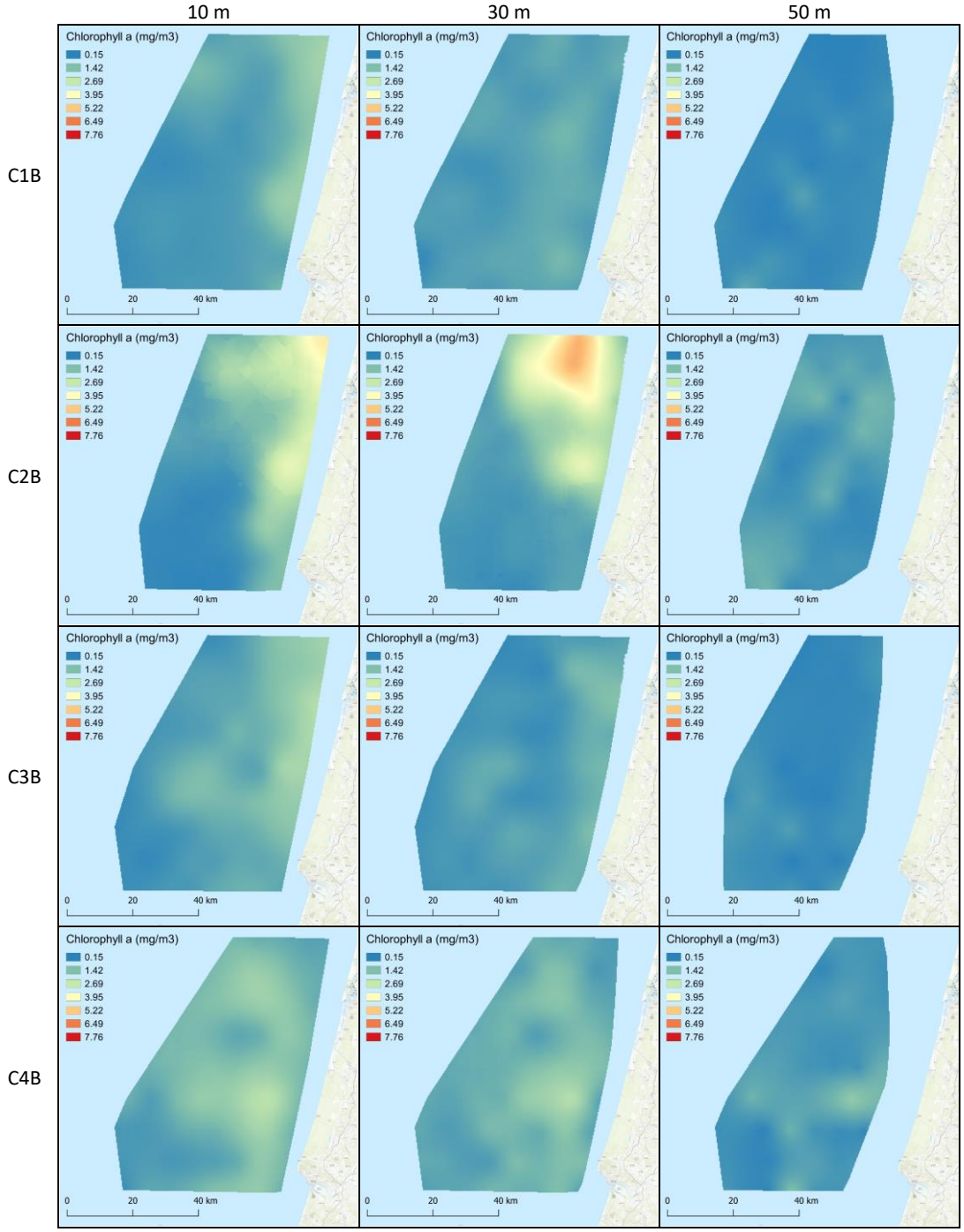


Figure E-27. Horizontal cross-sections of chlorophyll a in area B.

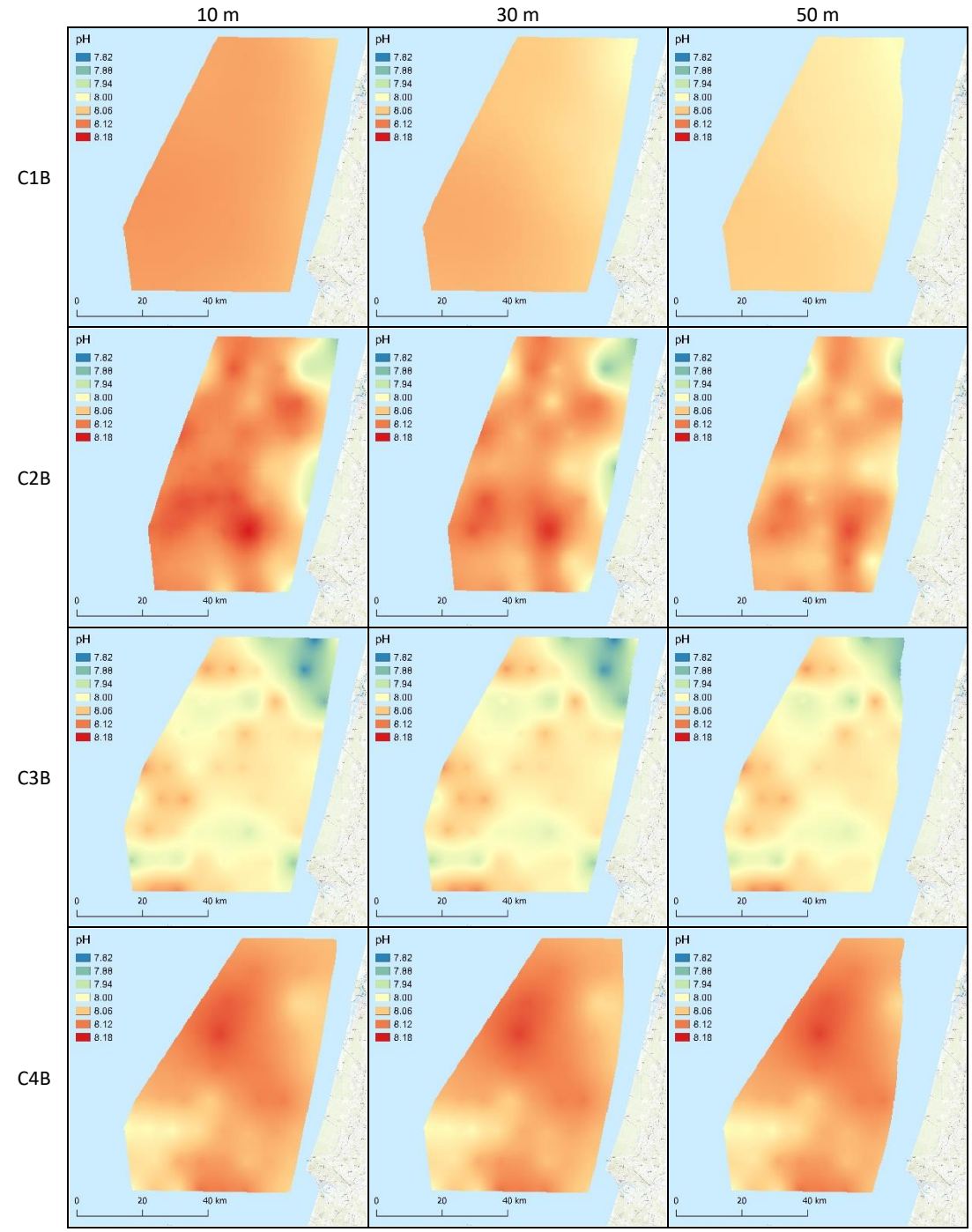


Figure E-28. Horizontal cross-sections of pH in area B.

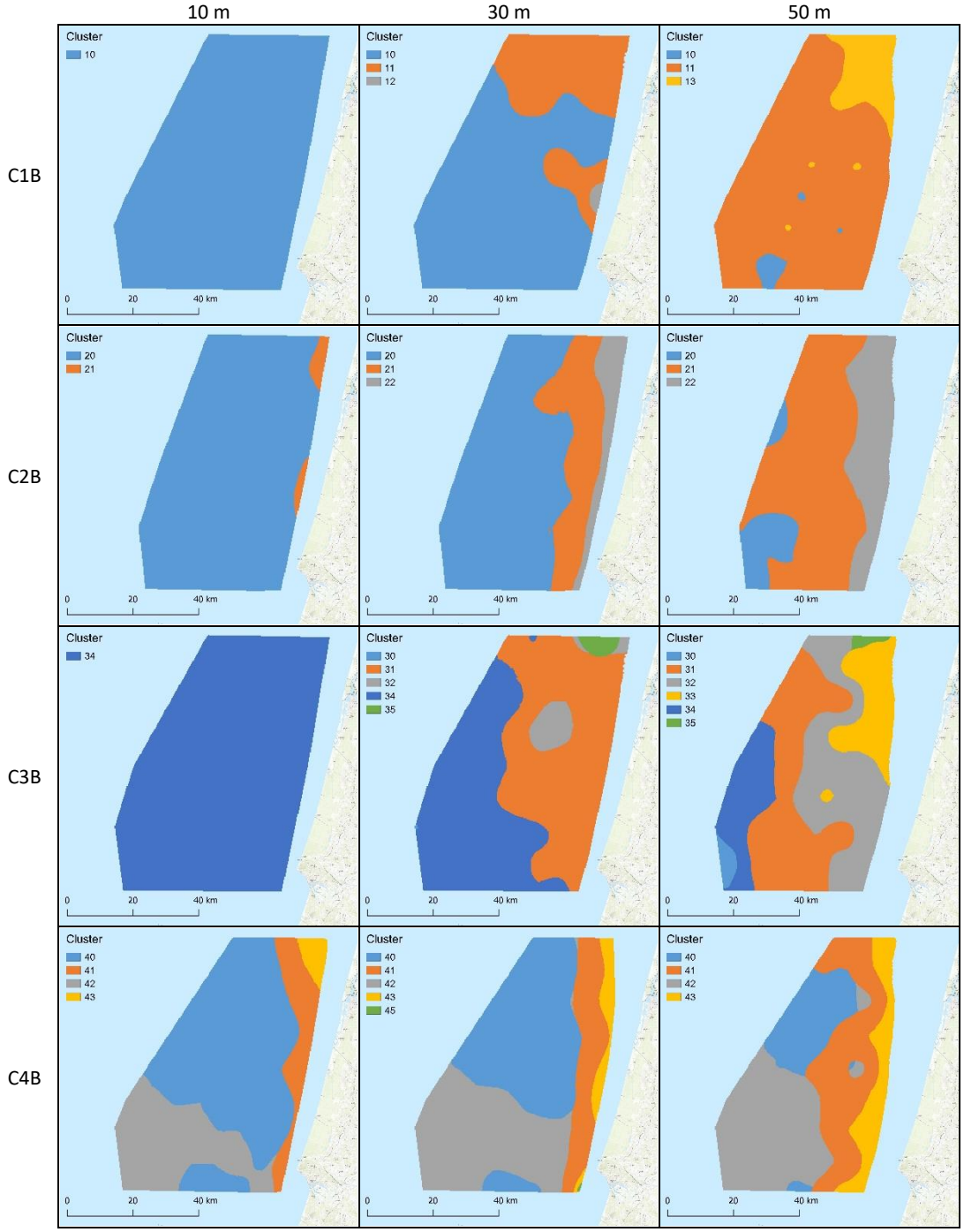


Figure E-29. Horizontal cross-sections of clusters in area B.

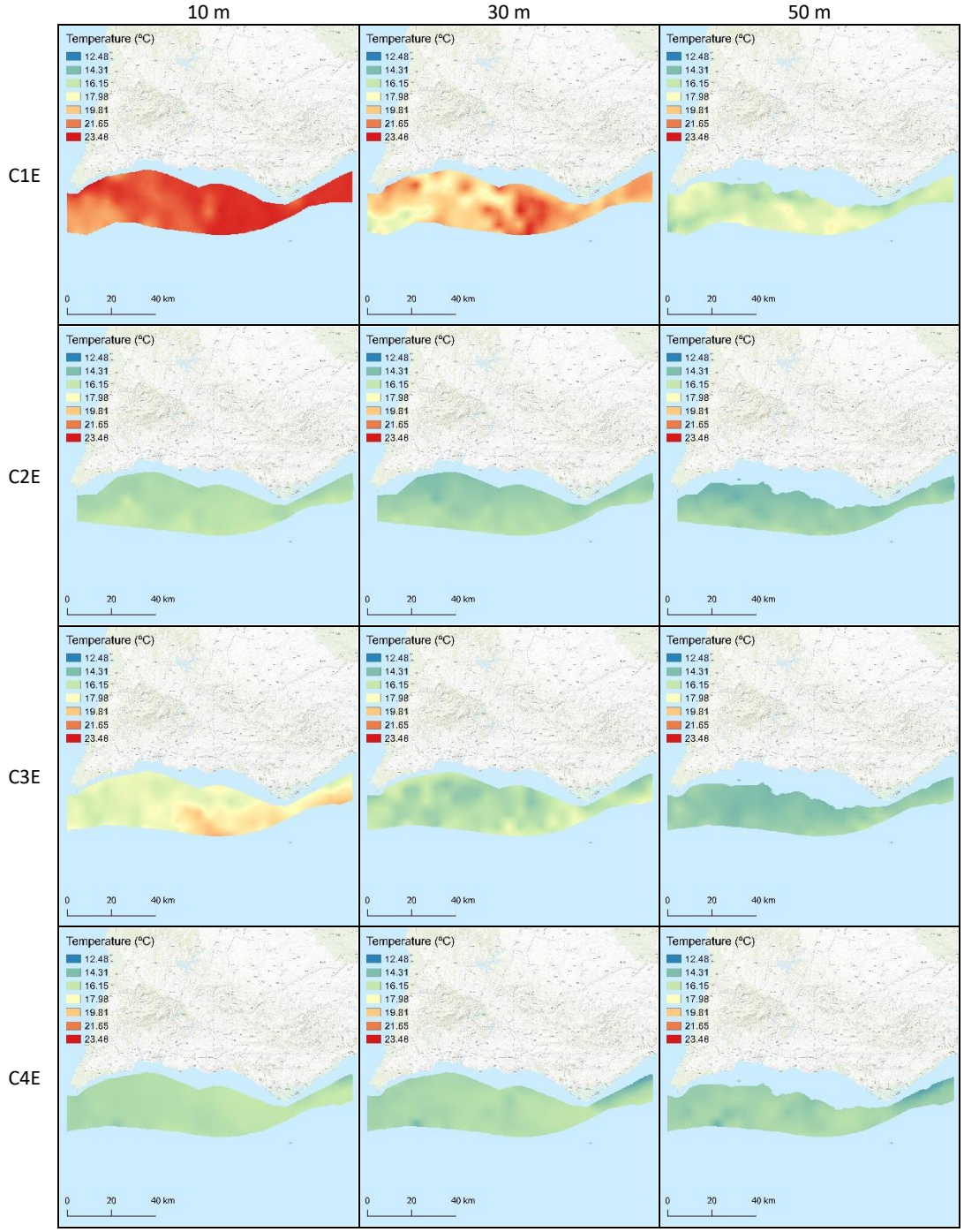


Figure E-30. Horizontal cross-sections of temperature in area E.

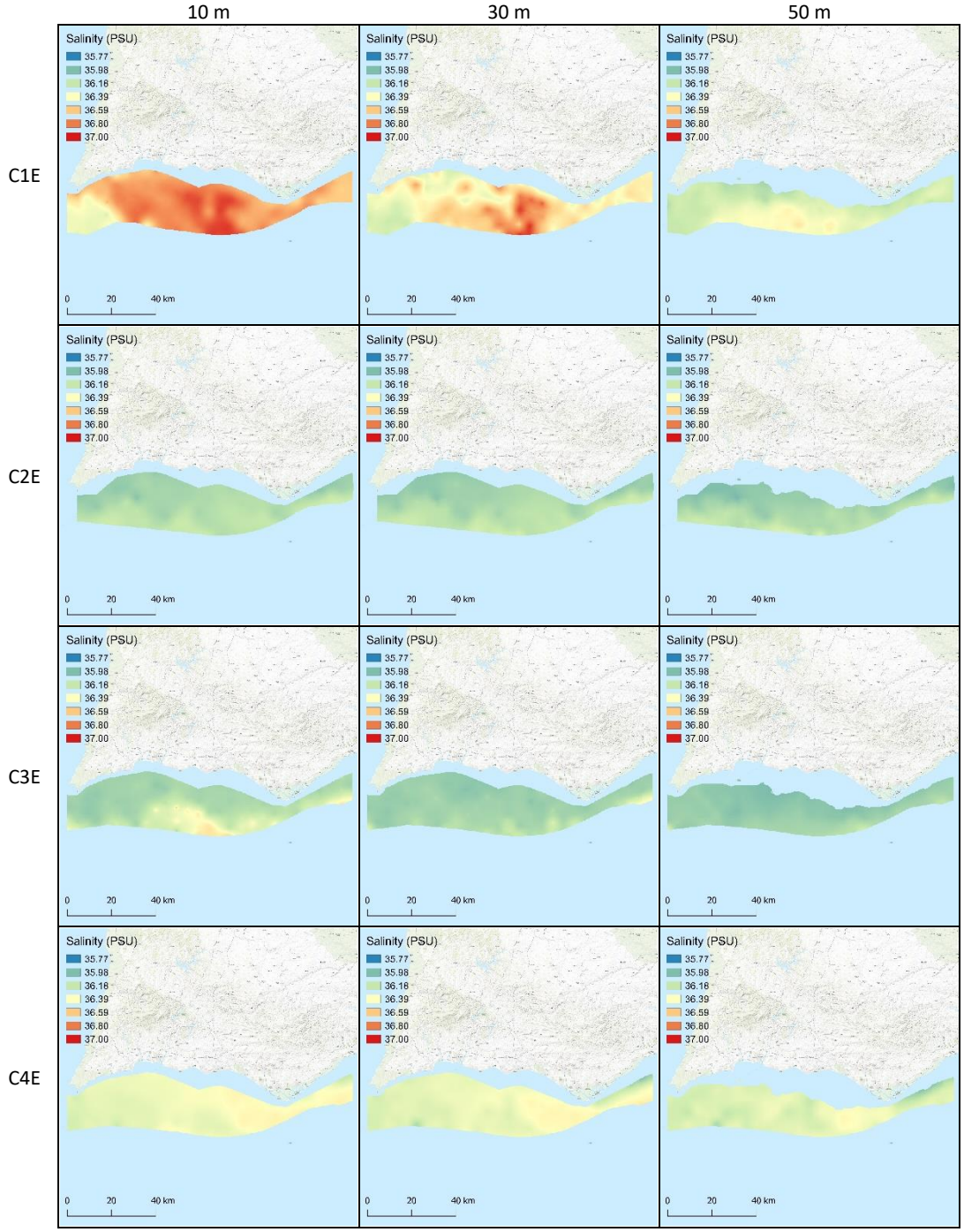


Figure E-31. Horizontal cross-sections of salinity in area E.

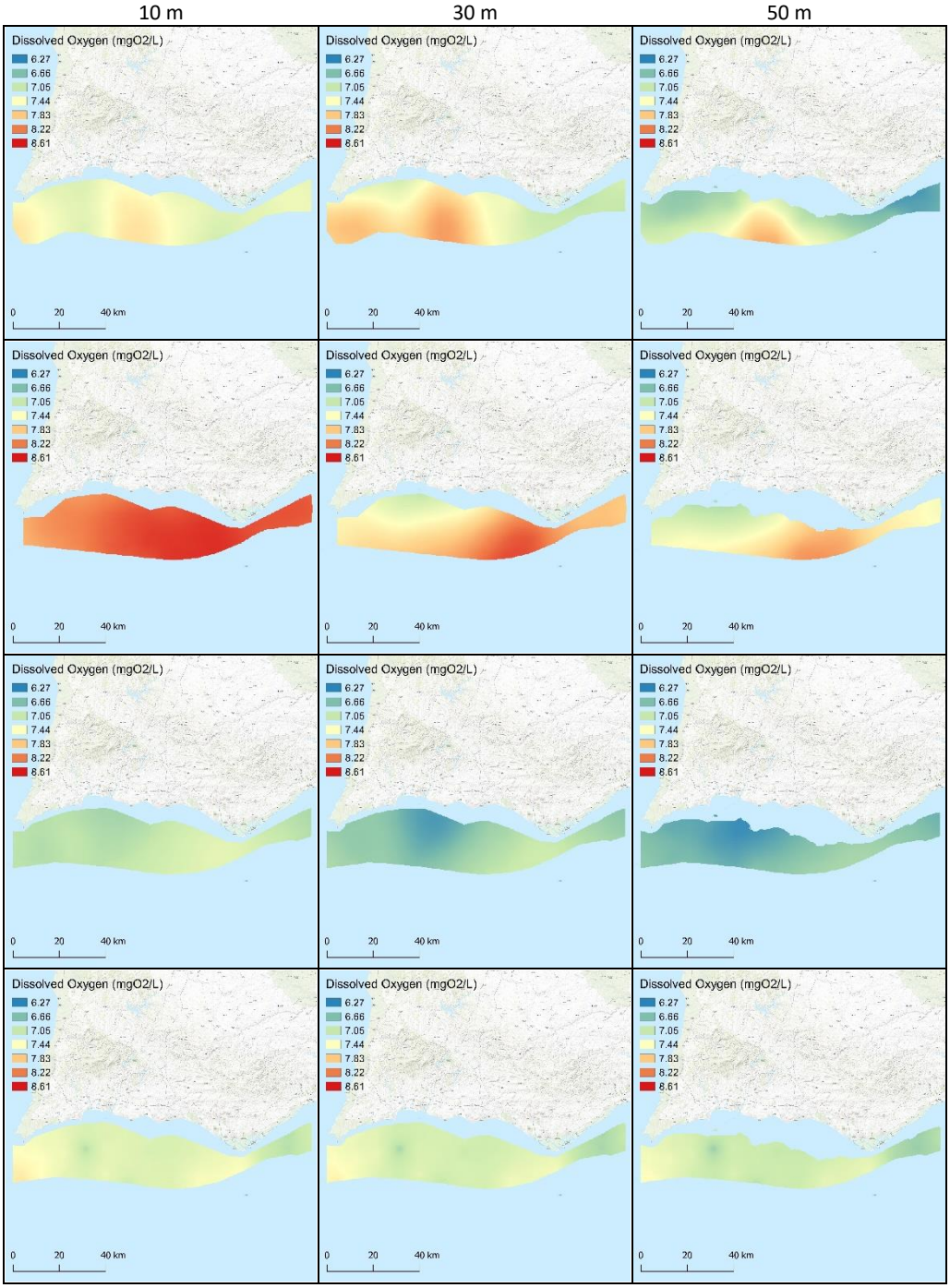


Figure E-32. Horizontal cross-sections of DO in area E.

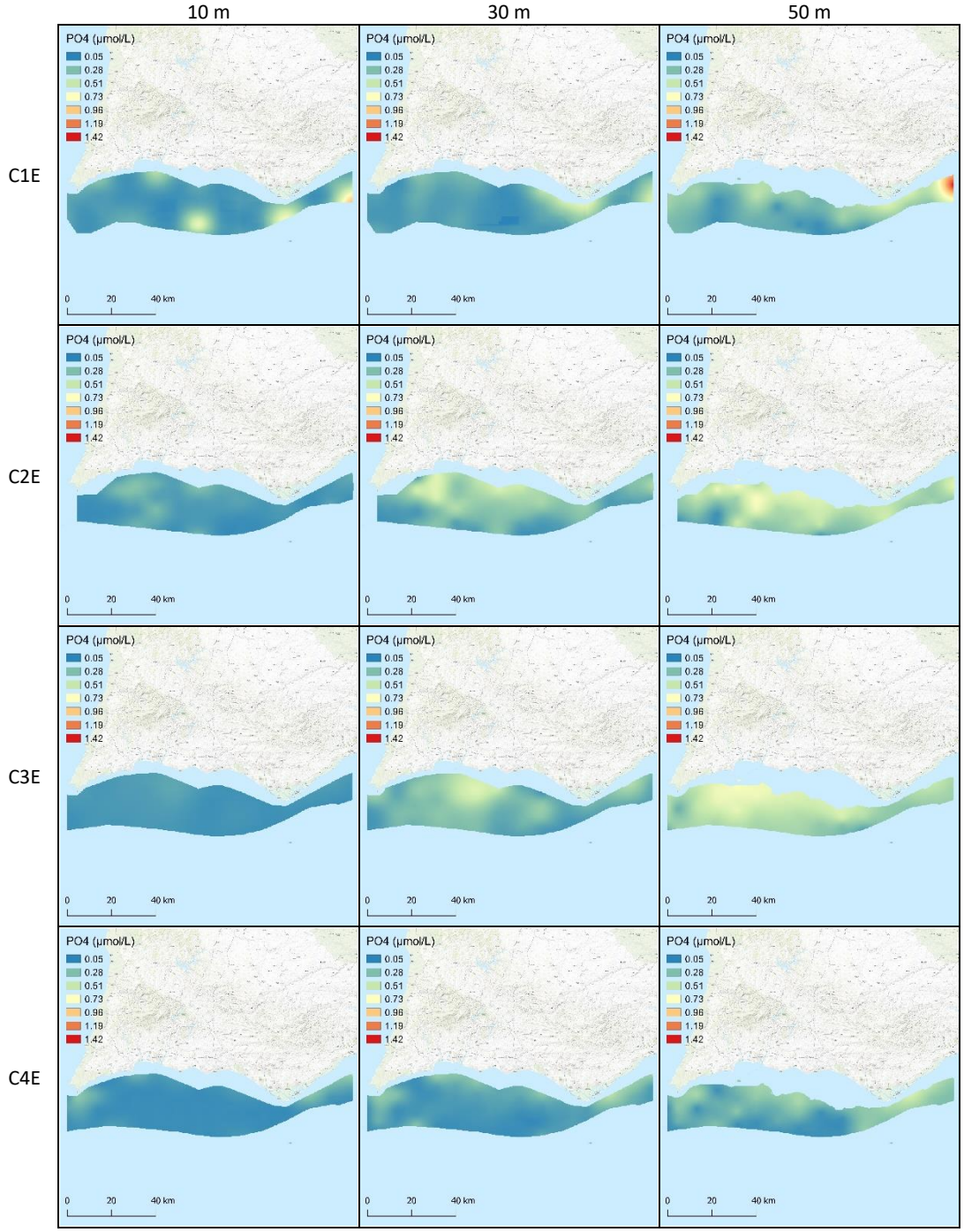


Figure E-33. Horizontal cross-sections of PO_4 in area E.

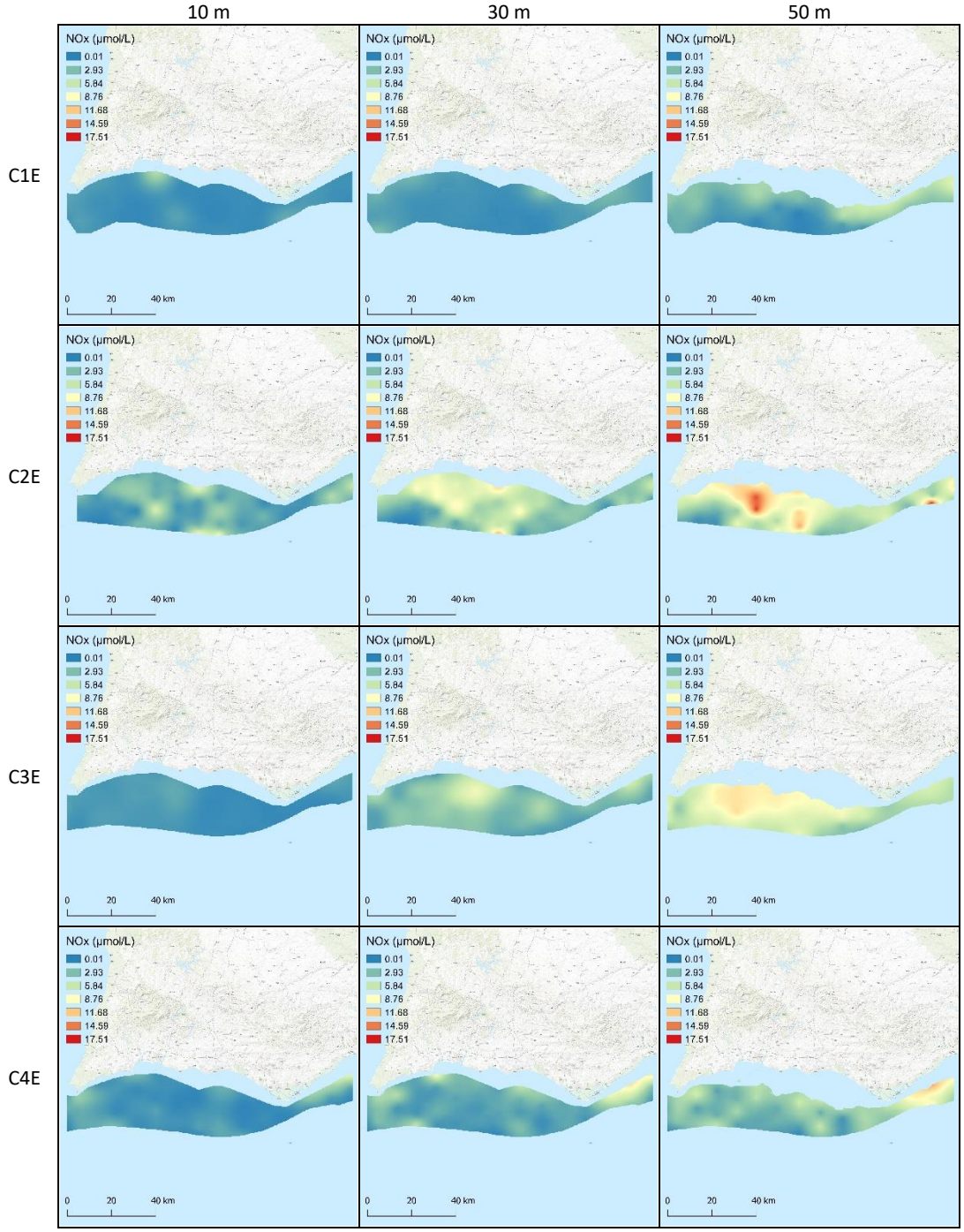


Figure E-34. Horizontal cross-sections of NO_X in area E.

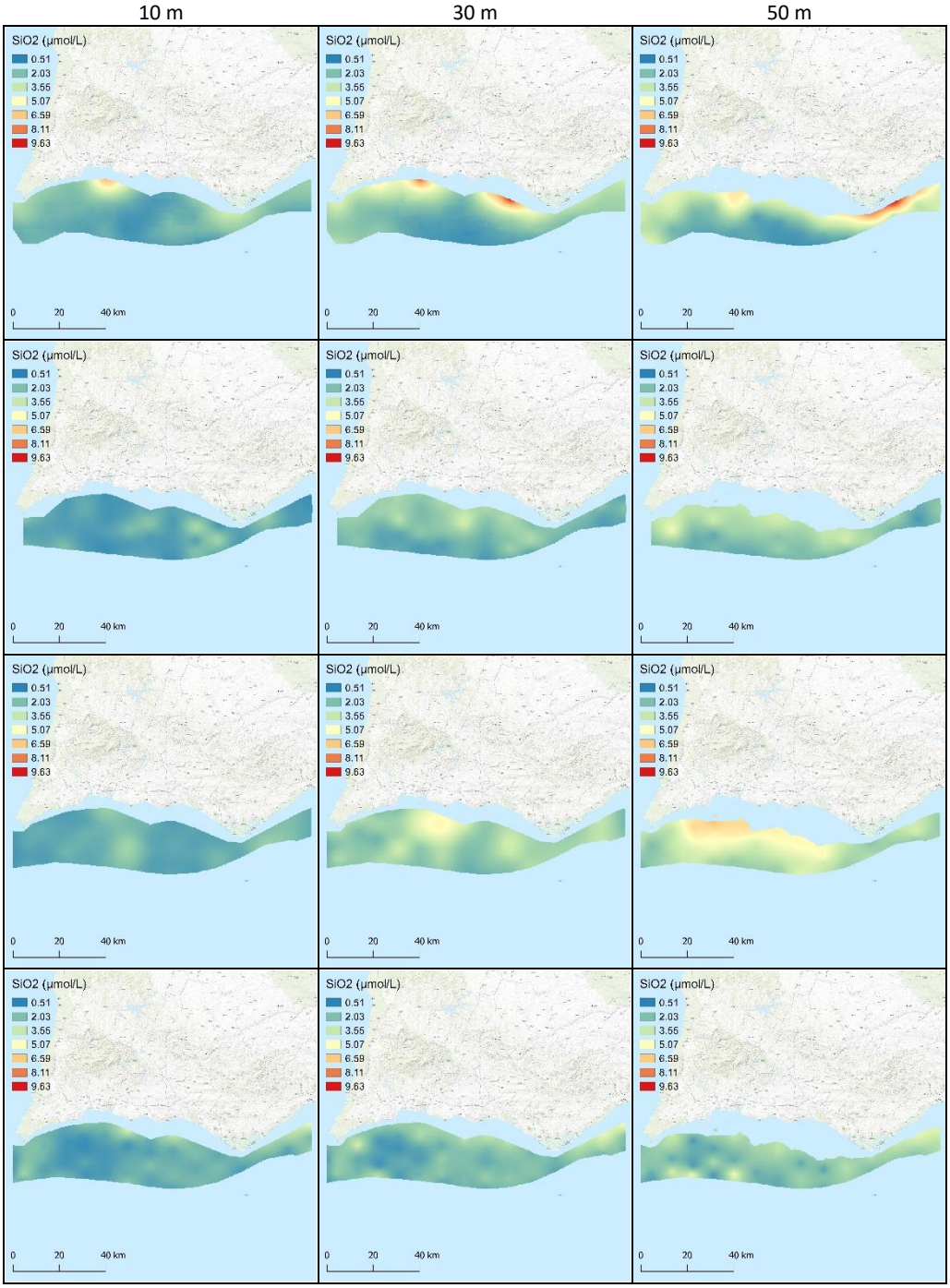


Figure E-35. Horizontal cross-sections of SiO_2 in area E.

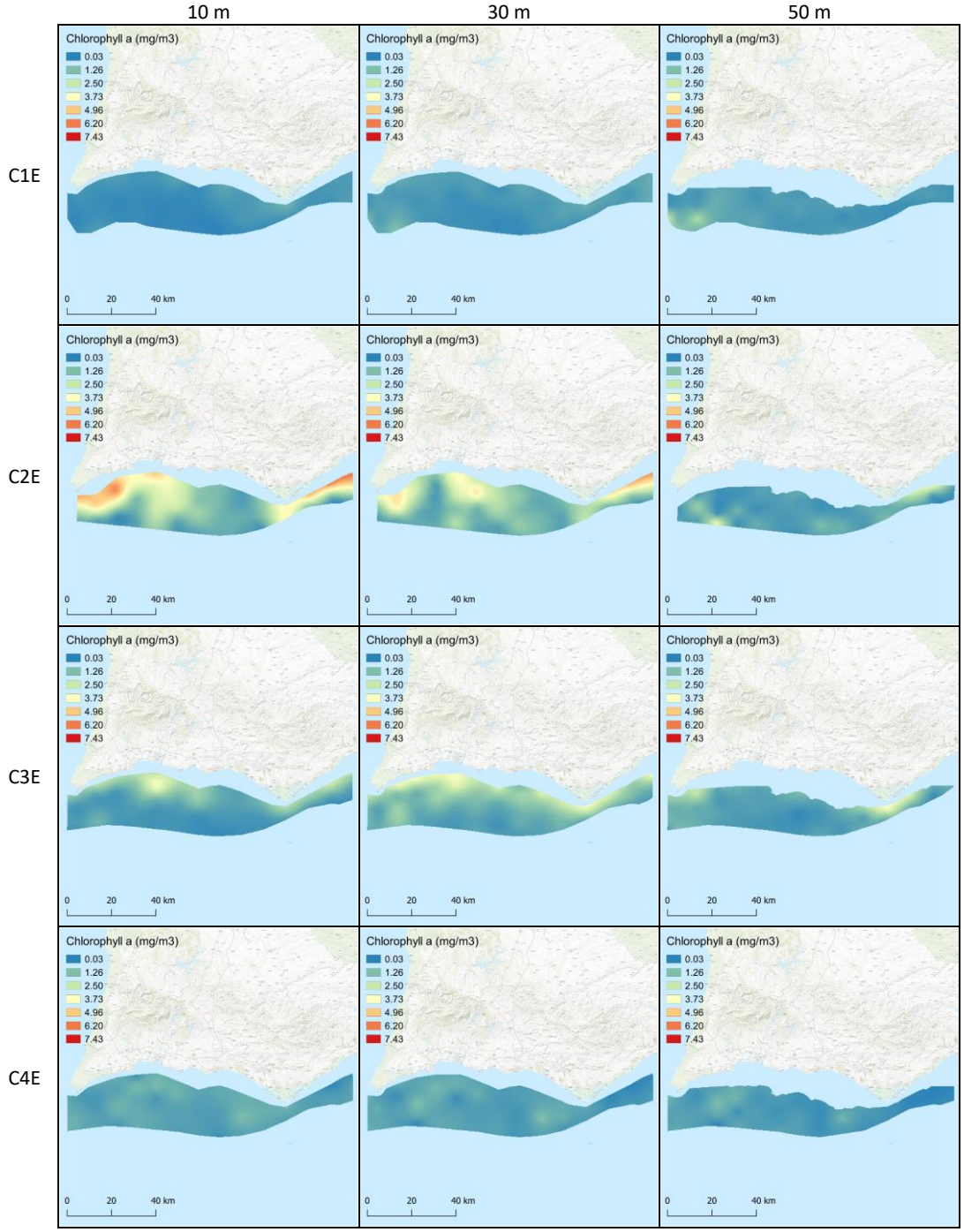


Figure E-36. Horizontal cross-sections of chlorophyll a in area E.

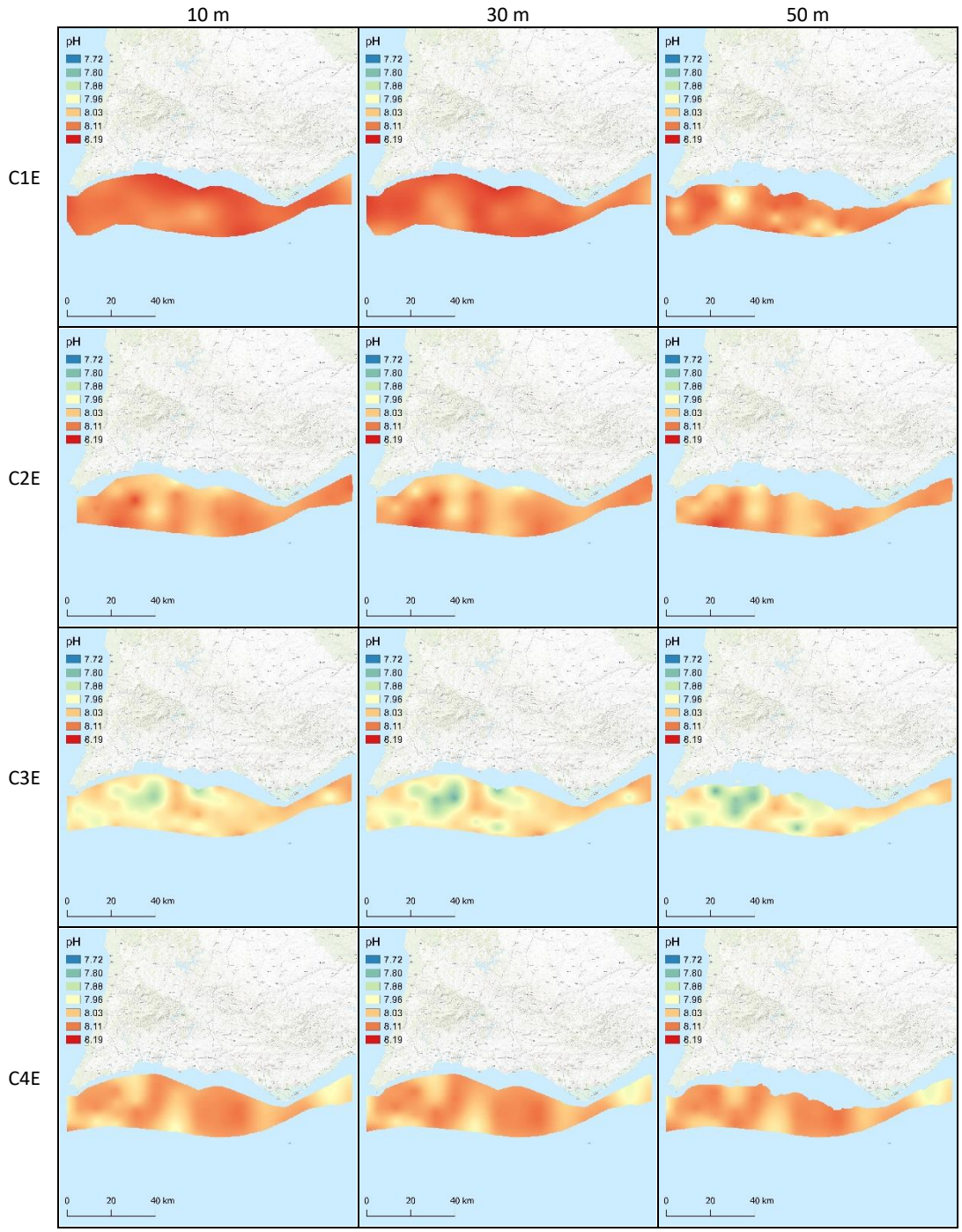


Figure E-37. Horizontal cross-sections of pH in area E.

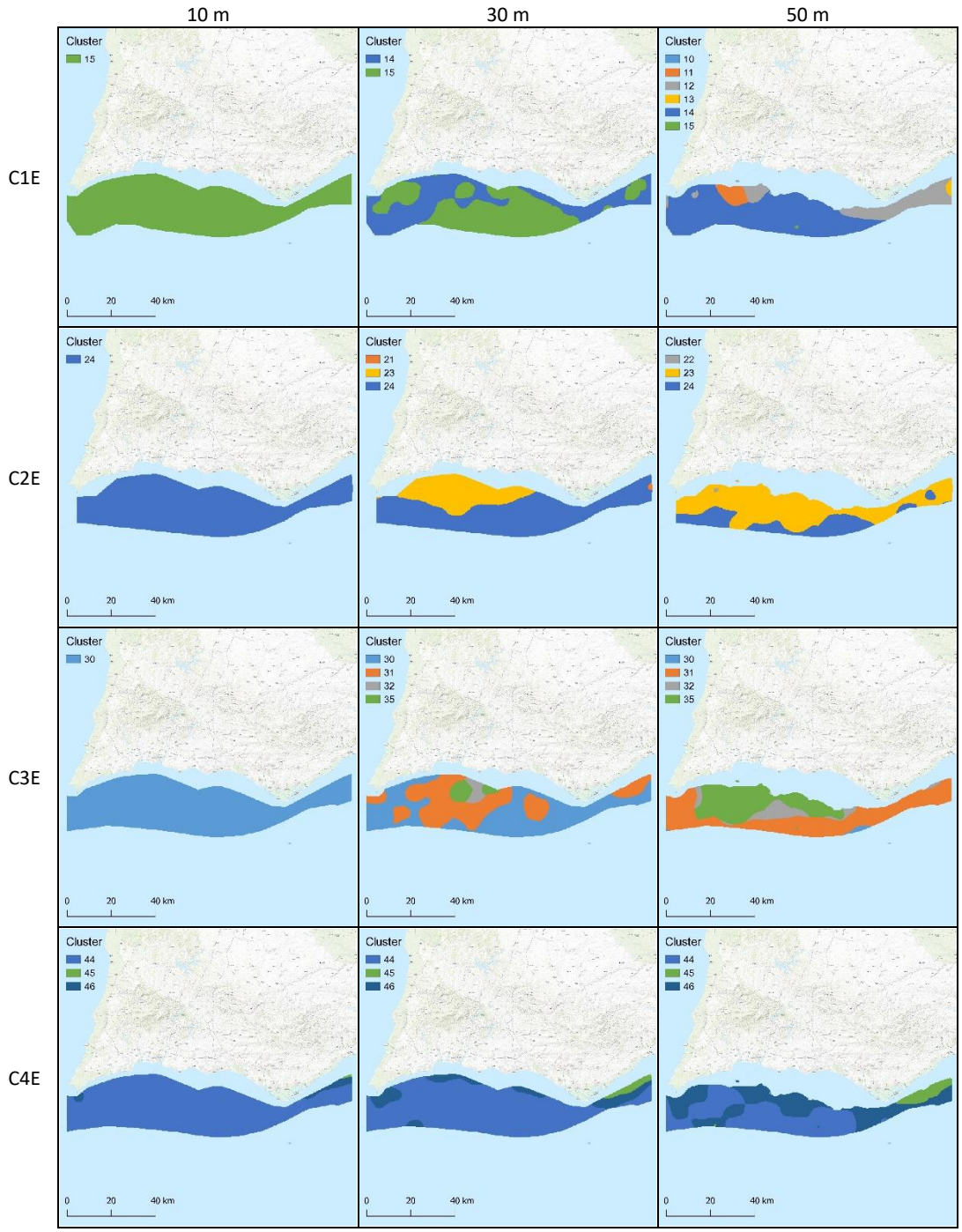


Figure E-38. Horizontal cross-sections of clusters in area E.

ANNEX F – WEB APPLICATION

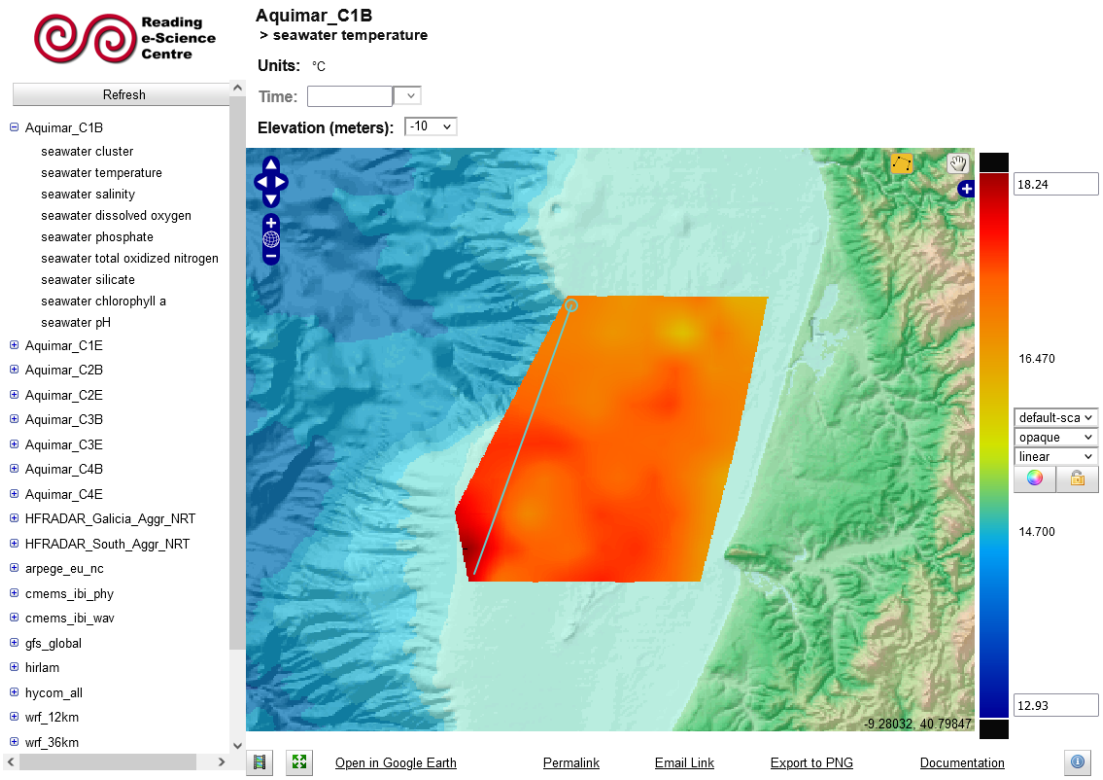


Figure F-1. Transect drawing in Godiva3 interface (ncWMS).

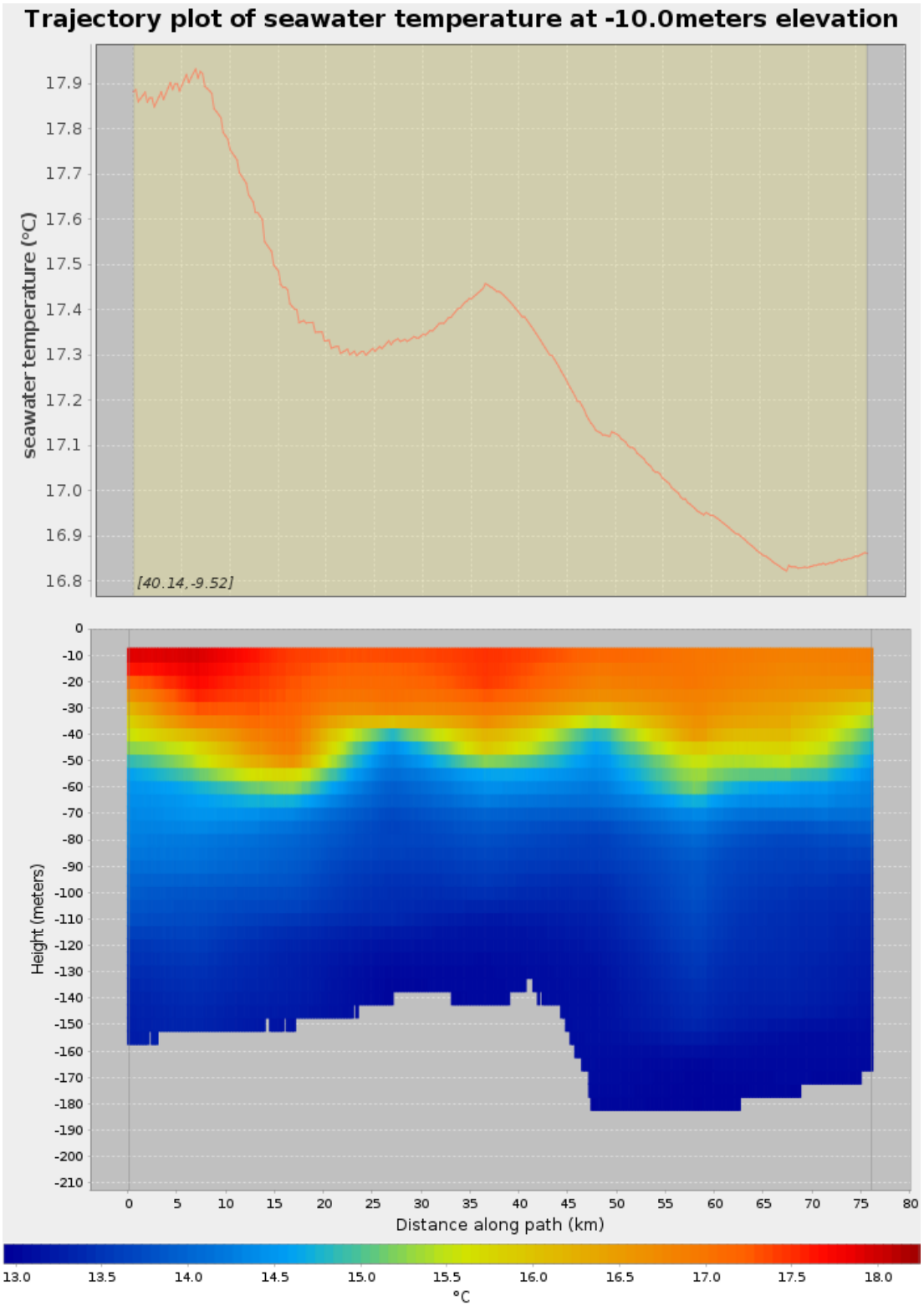


Figure F-2. Transect view in Godiva3 interface (ncWMS).

C&SIG

UNIGIS PT

NOVA IMS Information Management School