

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Corso di Laurea Magistrale in Fisica del Sistema Terra

**Estimating the export of biomass across  
the shelf break via AI applied to  
satellite images**

**Relatore:**  
Prof. Nadia Pinardi

**Correlatori:**  
Prof. Silvia Bianconcini  
Prof. Astrid Bracher

**Presentata da:**  
Federica Benassi

**Sessione III**  
**Anno Accademico 2022-2023**

The whole world at your fingertips,  
the ocean at your door

(Bo Burnham, *That Funny Feeling*)

A mia nonna Maria

## Abstract

Ocean carbon uptake has an important role in the global carbon cycle, absorbing one third of anthropogenic atmospheric carbon. The biological carbon pump, powered by phytoplankton primary production, transports organic carbon from surface waters to deeper ocean zones, effectively sequestering it. Shelf regions, characterized by shallow waters and nutrient abundance, exhibit elevated productivity compared to the open ocean.

Submesoscale filaments, or *streamers*, form through the instability of coastal currents, concentrating phytoplankton and chlorophyll into elongated structures that enhance long-distance transport. Streamers can transport cold, chlorophyll-rich shelf waters into the open ocean, potentially enhancing carbon fluxes to deeper layers.

This thesis presents an automated image analysis tool utilizing K-means clustering on an 18-year dataset of chlorophyll-*a* and sea surface temperature anomalies from Level 3 satellite data. Three coastal regions (California, Mauritania, and South China Sea) are studied. The algorithm effectively identifies streamers within a single cluster, enabling estimation of chlorophyll content, lateral export, and carbon fluxes.

The analysis of the chlorophyll content time series reveals distinct patterns. While the biomass export intensity in upwelling regions follows the primary production seasonality, the South China Sea shows high-frequency variabilities due to purely turbulent processes.

The annual lateral biomass export through streamers totals  $3.5 \pm 0.7 \text{ Tg yr}^{-1}$ . Carbon fluxes vary between regions, with upwelling areas having larger fluxes (approximately  $150 \text{ g m}^{-2} \text{ yr}^{-1}$ ) compared to the South China Sea ( $65 \text{ g m}^{-2} \text{ yr}^{-1}$ ), indicating that streamers have different areas and carbon transport capacities.

This study offers valuable insights into lateral carbon export from coastal to open ocean regions. Extending this research to a global estimate is essential to understand the role of this transport in the processes of carbon burial to depth.

## Sommario

L'assorbimento di carbonio negli oceani ha un ruolo importante nel ciclo globale del carbonio, assorbendo un terzo del carbonio atmosferico di origine antropica. La pompa biologica del carbonio, alimentata dalla produzione primaria del fitoplancton, trasporta il carbonio organico dalle acque superficiali alle zone oceaniche più profonde, sequestrandolo efficacemente. Le regioni della piattaforma continentale, caratterizzate da acque poco profonde e abbondanza di nutrienti, presentano una produttività elevata rispetto all'oceano aperto.

I filamenti di sottomesoscala, o *streamers*, si formano grazie all'instabilità delle correnti costiere, concentrando il fitoplancton e la clorofilla in strutture allungate che favoriscono il trasporto a lunga distanza. Gli *streamers* possono trasportare le acque fredde della piattaforma, ricche di clorofilla, nell'oceano aperto, potenzialmente aumentando i flussi verticali di carbonio verso gli strati più profondi.

Questa tesi presenta uno strumento di analisi automatica delle immagini che utilizza il clustering K-means su un set di dati di 18 anni di anomalie di clorofilla-*a* e temperatura superficiale del mare derivate da dati satellitari di livello 3. Questo sistema è stato applicato a tre regioni costiere specifiche (California, Mauritania e Mar Cinese Meridionale). L'algoritmo identifica efficacemente gli *streamers* all'interno di un singolo cluster, consentendo la stima del contenuto di clorofilla, del trasporto laterale e dei flussi di carbonio.

L'analisi delle serie temporali del contenuto di clorofilla rileva pattern distinti. Mentre l'intensità dell'esportazione di biomassa nelle regioni di upwelling segue la stagionalità della produzione primaria, il Mar Cinese Meridionale mostra variabilità ad alta frequenza dovute a processi puramente turbolenti.

L'esportazione laterale annuale di biomassa attraverso gli *streamers* è pari a  $3,5 \pm 0,7$  Tg anno<sup>-1</sup>. I flussi di carbonio variano tra le regioni, con le aree di upwelling che presentano flussi maggiori (circa  $150 \text{ g m}^{-2} \text{ yr}^{-1}$ ) rispetto al Mar Cinese Meridionale ( $65 \text{ g m}^{-2} \text{ yr}^{-1}$ ), indicando che gli *streamers* hanno aree e capacità di trasporto del carbonio diverse.

Questo studio offre importanti informazioni sul trasporto laterale di carbonio dalle regioni costiere all'oceano aperto. L'estensione di questa ricerca a una stima globale è essenziale per comprendere il ruolo di questo tipo di trasporto nei processi di seppellimento di carbonio in profondità.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Global carbon cycle	1
1.1.1	Components of global carbon cycle: sources and sinks	2
1.1.2	Air-sea fluxes and ocean carbon pumps	5
1.1.3	Biological carbon pump and carbon export	8
1.2	Mesoscale and submesoscale dynamics	12
1.2.1	Eddies and filaments	14
1.3	Lateral export or recycling?	20
1.4	Thesis objectives	21
<b>2</b>	<b>Satellite data</b>	<b>23</b>
2.1	Introduction to remote sensing	23
2.2	Sea surface temperature	25
2.3	Ocean color	26
2.3.1	Satellite chlorophyll- <i>a</i>	28
2.3.2	Satellite euphotic depth	30
2.4	Sea bottom topography	30
<b>3</b>	<b>Methods</b>	<b>32</b>
3.1	Introduction to Machine Learning	32
3.1.1	Unsupervised learning and K-means clustering	33
3.2	Coastal ocean image analysis	36
3.2.1	State of the art	36
3.2.2	Proposed clustering-based algorithm	37
3.3	Biomass content calculation	43
<b>4</b>	<b>Results and discussion</b>	<b>45</b>
4.1	K-means fitting	45
4.1.1	Results on 2003 subset	45
4.1.2	Application on 2004-2021 time series	47
4.2	Lateral transport estimation	50

4.2.1	Chlorophyll content time series . . . . .	50
4.2.2	Estimation of carbon export and fluxes . . . . .	54
<b>5</b>	<b>Conclusions and future work</b>	<b>57</b>

# Chapter 1

## Introduction

This chapter is a brief introduction to the topic of the global carbon cycle, the effects of human activities on carbon budget, and the role of the oceans in carbon export and burial. The role of submesoscale dynamics in carbon export and burial to depth is investigated.

Section 1.1 provides an overview of the global carbon cycle, its main reservoirs and fluxes, with a particular focus on air-sea fluxes and ocean carbon uptake. The biological carbon pump, which represents the main mechanism of carbon export and burial in the ocean, is presented. Section 1.2 presents the dynamical regimes of the ocean, with a specific focus on mesoscale and submesoscale. An overview of lateral export of carbon from the shelf to the open ocean is given in Section 1.3. Finally, in Section 1.4, the structure and objectives of the thesis are presented, providing an overview of what will be covered in the subsequent chapters.

### 1.1 Global carbon cycle

Carbon plays a central role in Earth's life and climate systems. It serves as the fundamental building block of a wide range of organic compounds, making up about 50% of the dry weight of living organisms, and it acts as an energy driver participating in both the process of photosynthesis and cellular respiration. Its role in climate regulation is fundamental, as it forms two of the most important greenhouse gases, carbon dioxide ( $\text{CO}_2$ ) and methane ( $\text{CH}_4$ ). Carbon is one of the most plentiful elements on Earth, and it is constantly exchanged between the different components of the Earth System in what is known as the global carbon cycle.

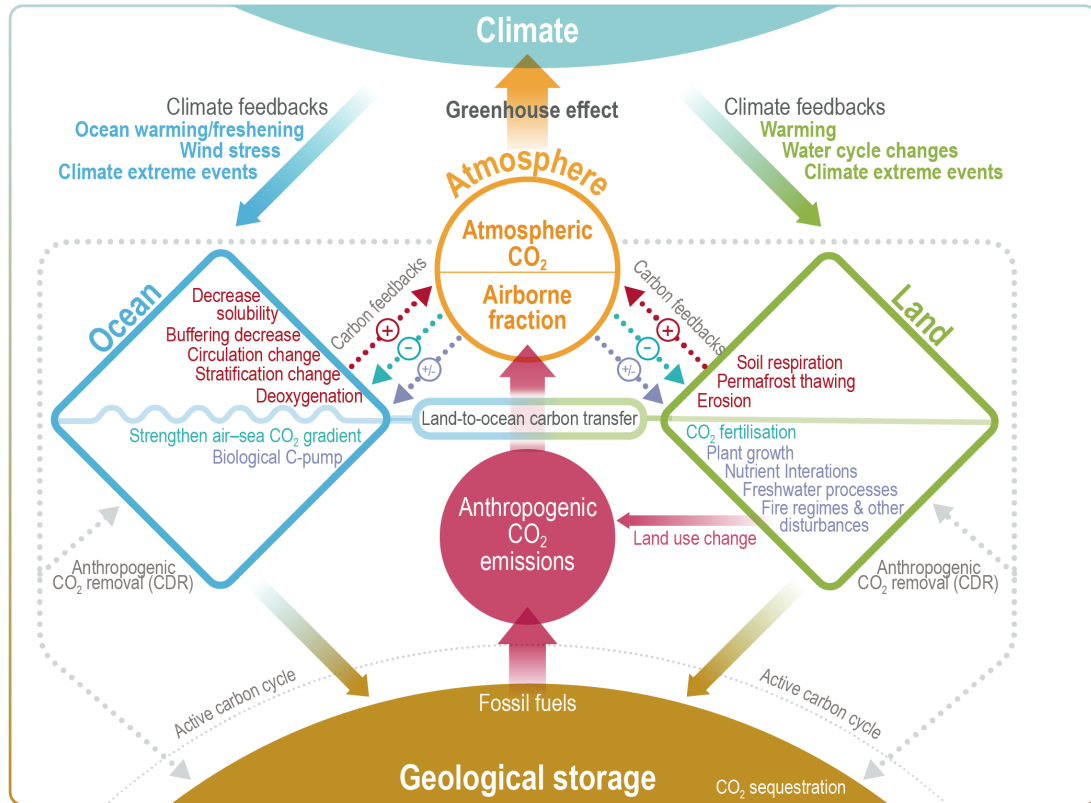


Figure 1.1: Diagram of components and interactions in the contemporary global cycle. Figure from sixth IPCC assessment report [1].

### 1.1.1 Components of global carbon cycle: sources and sinks

The global carbon cycle is a complex system characterized by interconnected reservoirs of CO<sub>2</sub> within the Earth system that interact exchanging fluxes of materials, heat, momentum and water. The four primary reservoirs in the cycle are solid Earth, ocean, terrestrial biosphere, and atmosphere, whose interactions are illustrated in Figure 1.1. Over time, human activities, like fossil fuel combustion, deforestation, and land use changes have emerged as an additional influential component, directly impacting the balances and feedbacks between the natural reservoirs.

The fluxes operating within the global carbon cycle exhibit a wide range of time scales, spanning from short-term to extremely long-term processes. The fifth Intergovernmental Panel on Climate Change (IPCC) assessment report [2] identifies two distinct modes:

1. a *fast* regime, characterized by significant exchange fluxes (tens of PgC yr<sup>-1</sup>,



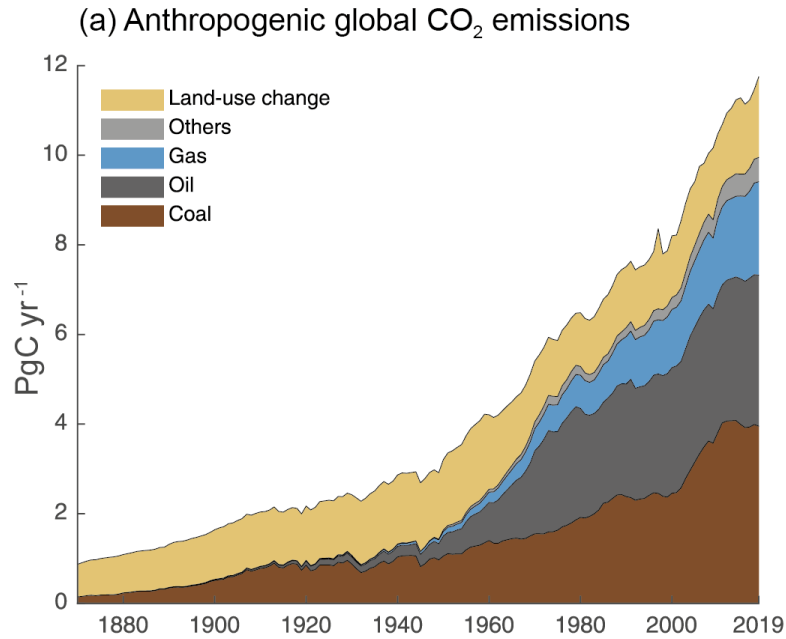


Figure 1.2: Time series of anthropogenic global CO<sub>2</sub> emissions by fossil fuels combustion (gas, oil, coal) and other sources (flaring, emissions from carbonates during cement manufacture). Figure from sixth IPCC assessment report [1].

[2]) and rapid turnover times, ranging from a few years to decades and millennia. This regime encompasses processes involving atmospheric carbon, oceans, surface ocean sediments, land vegetation, and soil;

2. a *slow* regime, with weaker fluxes ( $< 0.3 \text{ PgC yr}^{-1}$ , [2]) that operate on much longer time scales of 10,000 years or more. This mode primarily revolves around the interaction between the solid Earth, the largest carbon reservoir, and the other faster components. It involves processes such as volcanic activity, chemical weathering of rocks, erosion, and sediment formation on the sea floor.

The slow domain is characterized by processes that occur on such lengthy time scales that they can be considered steady and constant. Similarly, prior to the Industrial Era, the fast domain was also relatively stable and close to a steady state. However, with the advent of industrial activities, particularly the extraction and combustion of fossil fuels from geological reservoirs, a significant amount of fossil carbon was transferred from the slow domain to the fast domain, causing an unprecedented, major human-induced perturbation in the carbon cycle (Figure 1.2).

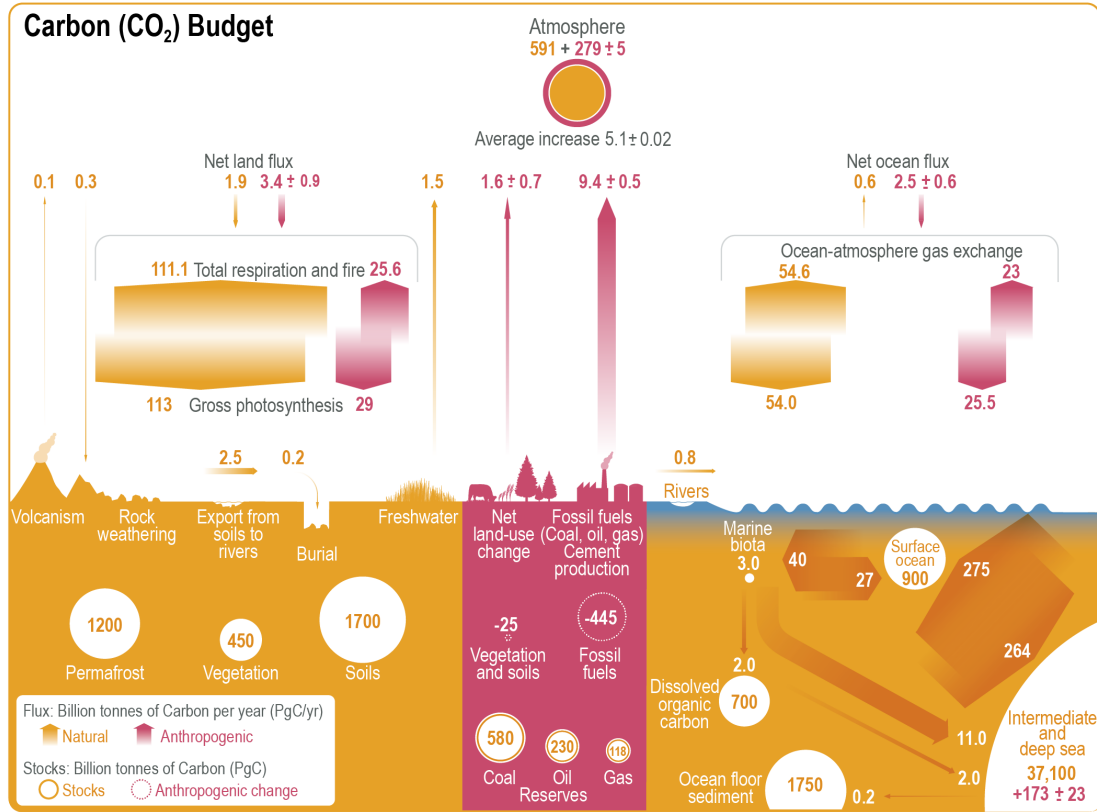


Figure 1.3: Diagram of components and interactions in the contemporary global CO<sub>2</sub> budget. Yellow arrows represent annual carbon fluxes (in PgC yr<sup>-1</sup>) associated with the natural carbon cycle, estimated for the time prior to the industrial era (around 1750). Pink arrows represent anthropogenic fluxes, averaged between 2010 and 2019. Figure reproduced from sixth IPCC assessment report [1].

The combustion of fossil fuels, as well as extensive deforestation and changes in land use, impacted on the global carbon cycle resulting in the additional release of carbon, mainly in the form of CO<sub>2</sub>, and in the perturbation of the long-established equilibrium of the natural carbon cycle.

Contemporary global CO<sub>2</sub> budget, whose sources, sinks and fluxes are represented in Figure 1.3, can be expressed by the following equation [1]:

$$E_{FOS} + E_{LULUCF} = G_{atm} + S_{land} + S_{ocean} + B_{imb} \quad (1.1)$$

where:

- the left-end side represents the human-induced carbon emissions, through fossil fuels and carbonate emissions (FOS) and land use, land use change and

forestry (LULUCF). It is estimated that the total average human-related carbon emissions in the decade of 2010–2019 were  $10.9 \pm 0.9 \text{ PgC yr}^{-1}$  (Figure 1.2, Figure 1.3) [1].

- the right-end side represents the natural response to the additional anthropogenic flux: growth rate in the atmosphere ( $G_{atm}$ ), land and ocean sinks ( $S_{land}$ ,  $S_{ocean}$ ) and an additional imbalance term accounting for uncertainties in the estimations ( $B_{imb}$ ). Anthropogenic  $\text{CO}_2$  is differently redistributed in these three Earth system components: in the 2010–2019 decade, the greater fraction, 46%, accumulated in the atmosphere ( $5.4 \pm 0.02 \text{ PgC yr}^{-1}$ ), while 31% was stored by vegetation in terrestrial ecosystems ( $3.4 \pm 0.9 \text{ PgC yr}^{-1}$ ) and 23% was taken up by the ocean ( $2.5 \pm 0.6 \text{ PgC yr}^{-1}$ ) [1].

This additional carbon uptake has significantly disrupted the delicate balance within each environmental reservoir.

Due to human activities, the concentration of carbon dioxide ( $\text{CO}_2$ ) in the atmosphere has risen by 47% compared to pre-industrial levels in 1750. As of April 2023, the atmospheric  $\text{CO}_2$  concentration stands at 420.54 parts per million [3]. This increase, along with the introduction of other greenhouse gases, is causing the global average temperature to rise at an unprecedented rate. This rapid warming is leading to extreme effects on local climates, including more frequent heatwaves, storms, and desertification.

Starting from the 1980s, the biosphere’s absorption of this additional carbon has been primarily driven by a fertilization effect resulting from rising atmospheric  $\text{CO}_2$  concentrations, and a decrease in global burned areas [1].

Additionally, there is strong confidence in the scientific community that the ocean’s capacity to absorb and store carbon has increased in response to global anthropogenic  $\text{CO}_2$  emissions in the last 50 years (Figure 1.4). This absorption of  $\text{CO}_2$  by the ocean is altering seawater chemistry, resulting in a decrease in pH, a phenomenon known as ocean acidification.

Given the purpose of this work, a more comprehensive overview of the processes involved in carbon uptake, export, and burial within the ocean will be provided. These processes are essential for understanding the role of the ocean in mitigating the effects of rising atmospheric  $\text{CO}_2$  levels and their impact on the environment.

### 1.1.2 Air-sea fluxes and ocean carbon pumps

The ocean uptake of carbon involves a two-step process. Initially, carbon dioxide is exchanged at the air-sea boundary and integrated into the surface mixed layer. Subsequently, it undergoes vertical transportation into the ocean interior, where it can be stored for extended periods, ranging from decades to millennia.

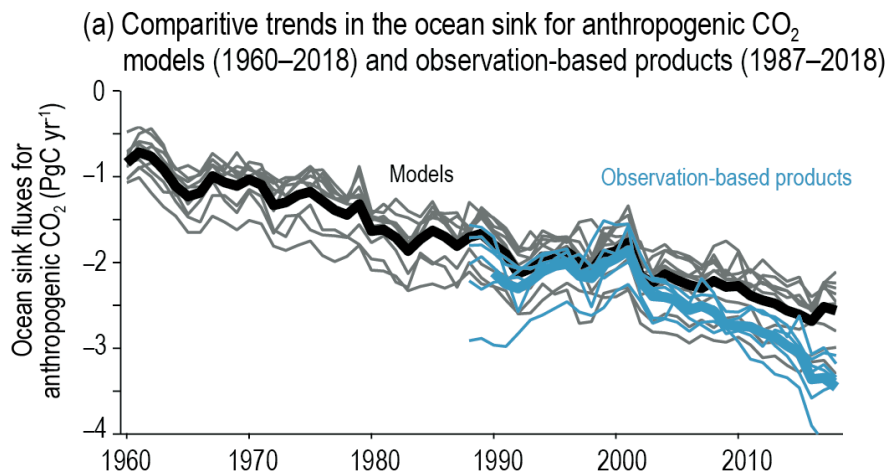


Figure 1.4: Multi-decadal (1960–2019) trends for the ocean sink of CO<sub>2</sub>, reconstructed from nine Global Ocean Biogeochemical Models forced with atmospheric re-analysis products, and observationally based gap-filling products from sparse observations of surface ocean partial pressure of CO<sub>2</sub> (1987–2018). Figure reproduced from sixth IPCC assessment report [1].

Carbon dioxide is primarily present in water as dissolved inorganic carbon (DIC), existing in the form of carbonate ( $\text{CO}_3^{2-}$ ) and bicarbonate ( $\text{HCO}_3^-$ ) ions, making up approximately 37,000 PgC. This represents the largest carbon reservoir within the ocean. Following DIC, dissolved organic carbon (DOC) and particulate organic carbon (POC)<sup>1</sup> contribute about 700 PgC, supporting marine food webs and ecosystems, while biomass, which includes marine biota such as phytoplankton and microorganisms, accounts for around 3 PgC [2].

As anticipated in the previous subsection, the net ocean uptake of anthropogenic carbon is about  $2.5 \text{ Pg yr}^{-1}$  (Figure 1.3, Figure 1.4). The coastal areas occupy about 10% of the global ocean [5], thus the associated net uptake is  $0.25 \text{ Pg yr}^{-1}$ . The exchange of CO<sub>2</sub> between the atmosphere and the ocean is governed by the difference in partial CO<sub>2</sub> pressure between air and seawater. In a simple model of air-sea gas exchange, where the interface is assumed to be composed by two stagnant interfaces where no turbulent effects are considered, it can be found that fluxes  $\Phi$  are directly proportional to the difference in partial pressure between air ( $p_a$ ) and sea ( $p_w$ ) [6]:

<sup>1</sup>*Dissolved* organic matter is defined as that fraction of total organic matter that passes through a  $0.45 \mu\text{m}$  pore size; conversely, *particulate* matter refers to the coarser fraction of total organic matter that is retained by a pore of the same size. [4]

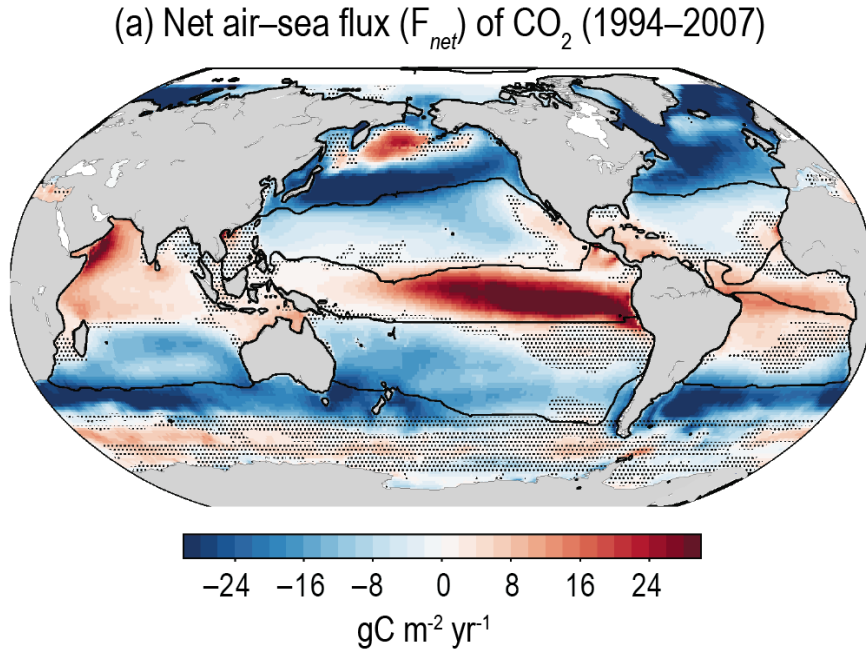


Figure 1.5: Net fluxes of  $\text{CO}_2$  derived from the ensemble of six observation-based products. Figure reproduced from sixth IPCC assessment report [1].

$$\Phi = -k_w \cdot S_{\text{CO}_2}(p_a - p_w) \quad (1.2)$$

where  $k_w$  is a parameter proportional to molecular diffusivity and thickness of layers and  $S_{\text{CO}_2}$  is the solubility of  $\text{CO}_2$  in seawater. Depending on the sign of  $p_a - p_w$ , the direction of fluxes changes: if  $p_a > p_w$ , the overall flux is negative and  $\text{CO}_2$  moves from the air into the seawater, and vice versa.

Solubility is directly related to the concentration of dissolved  $\text{CO}_2$  in seawater. This parameter depends on physical variables such as temperature and salinity: specifically, increased temperature and salinity lead to a decrease in solubility. It is estimated that the solubility of colder, polar waters is twice as that of warmer, equatorial waters [7].

The variability of these physical features of water masses across global regions has an overall effect on intensity and direction of  $\text{CO}_2$  fluxes, as shown in Figure 1.5. The majority of carbon uptake occurs at high latitudes, where the air-sea  $\text{CO}_2$  fluxes are strongly negative. On the other hand, the equatorial region is characterized by positive fluxes, resulting in a release of carbon dioxide from seawater to air.

The dissolved  $\text{CO}_2$  undergoes a series of chemical reactions with seawater, dissoci-

ating in bicarbonate and carbonate ions, and becoming part of the DIC reservoir. Although CO<sub>2</sub> uptake occurs at the air-sea interface, DIC concentration increases with depth (Figure 1.6). This gradient is maintained by the so-called *ocean carbon pumps* [8], that are responsible for carbon sequestration from surface layer to the bottom, and whose role is fundamental in regulating the partition of carbon between the ocean and the atmosphere. A *physical* and a *biological* pump can be identified, based on the nature of the processes that drive them.

The physical pump (or solubility pump) refers to carbon sequestration and burial that is driven by the enhanced solubility of deep, cold bottom waters. These waters form in the polar regions of North Atlantic and Southern Ocean (Figure 1.6, panels a and b), sink to depth and are meridionally transported to the Equator as part of the thermohaline circulation, ultimately upwelling in subtropical regions and outgassing CO<sub>2</sub> as their temperature increases. The effectiveness of the solubility pump relies on the contrast between the large volume of the deep ocean that is ventilated by cold polar water masses and the relatively smaller volume of the oceanic thermocline that is ventilated by the wind-driven subtropical circulation [7]. Figure 1.6 (panel c) shows that the abiotic component of DIC increases from 2000  $\mu\text{mol kg}^{-1}$  to about 2080  $\mu\text{mol kg}^{-1}$  in the first 1000 m of depth and it is estimated that the solubility pump accounts for only about 10% of the observed vertical gradient of DIC [4, 6].

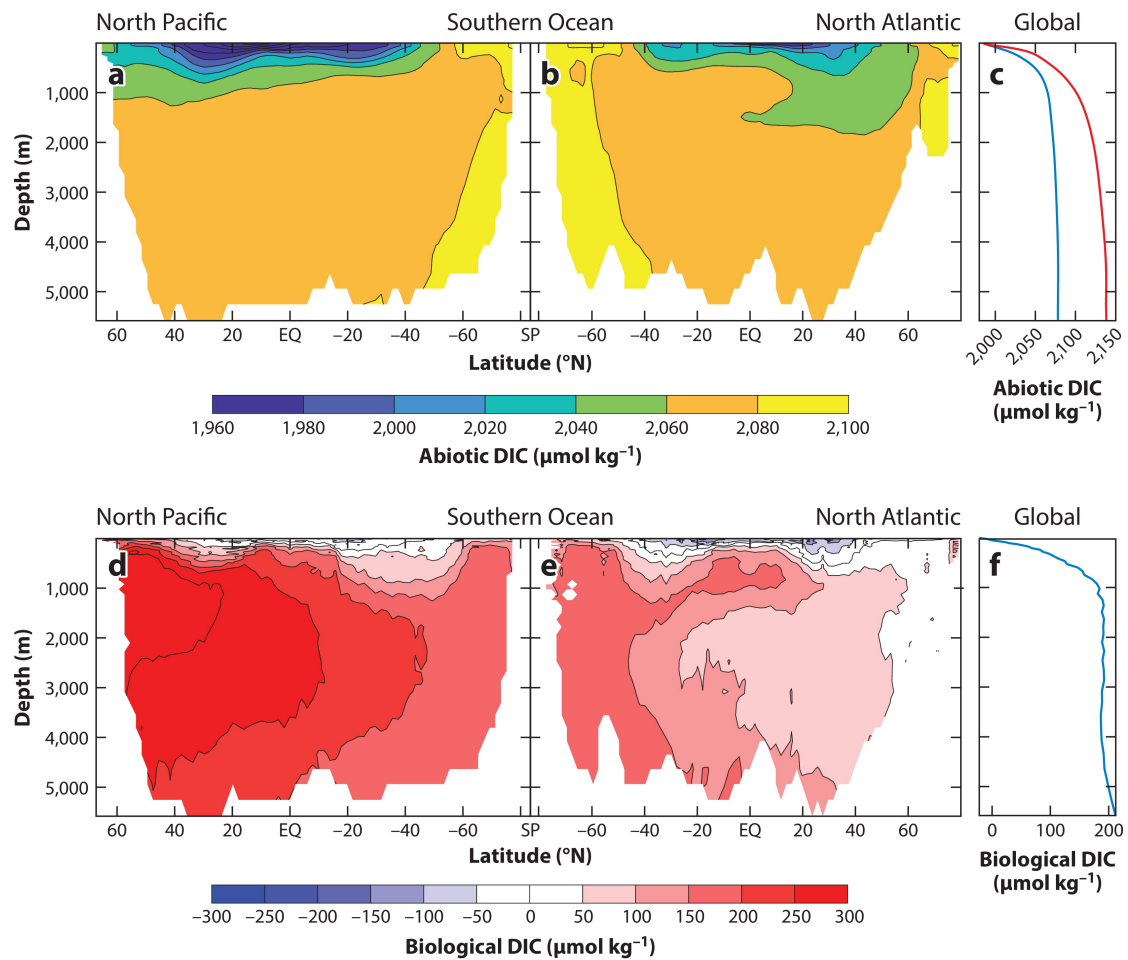
The remainder contribution to DIC gradient is to be attributed to biological processes. As it can be seen from Figure 1.6 (panels d and e), the presence of the biological pump leads to a deficit of DIC at the surface, and an excess at higher depths, significantly increasing the vertical concentration of DIC (panel f) of almost 200  $\mu\text{mol kg}^{-1}$  in the first 1000 m of depth.

The biological carbon pump is based on the net ecosystem productivity of the surface ocean, exporting organic carbon into the deep ocean via various mechanisms to be covered in the next subsection.

### 1.1.3 Biological carbon pump and carbon export

Marine organisms can be classified by their trophic status, that is their position in the food chain [6]. The primary distinction in marine ecosystems is between *autotrophs*, that can create organic matter directly from nutrients (mainly nitrogen and phosphorus) using photosynthesis, and *heterotrophs*, that gain energy and nutrition only from preexisting organic matter. Autotroph organisms include microscopic floating marine plants and bacteria, and they are commonly known as *phytoplankton*. The heterotroph group refers to small floating animals and bacteria known as *zooplankton*.

The biological activity of these organisms increases the total amount of DIC stored in the ocean of roughly 2,800 PgC, making up about 8% of the total DIC inventory,



DeVries T. 2022  
*Annu. Rev. Environ. Resour.* 47:317–41

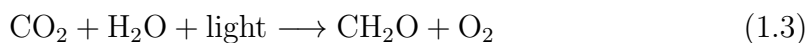
Figure 1.6: Profiles of dissolved inorganic carbon (DIC) concentrations from outputs of a global ocean carbon cycle model (preindustrial values). Panels a and b show a latitude-depth profile of abiotic DIC (without biological processes taken into account), in Pacific and Atlantic ocean respectively. Panel c is a meridionally-averaged vertical profile of DIC, where the red curve represents the vertical profile for full air-sea partial pressure equilibrium. Panels d, e and f represent the distribution profiles of DIC concentration derived from effect of biological processes, obtained by subtracting the abiotic profile from observational data. Figure reproduced from DeVries (2022) [7].

and it is capable to export about 10.2 PgC every year into the ocean's interior [7]. Two types of subprocesses (or pumps) can be identified in the biological carbon pump:

- The *organic carbon* pump, also referred to as the *soft-tissue* pump, results from the transformation of DIC into dissolved and particulate organic carbon through photosynthesis by phytoplankton;
- The *carbonate* pump (*hard-tissue* pump) is a process of conversion of DIC into calcium carbonate ( $\text{CaCO}_3$ ) performed by certain species of phytoplankton.

The contributions of soft-tissue and hard-tissue pumps to the maintaining of the vertical DIC gradient are 70% and 20%, respectively [4, 6]. This suggests that the soft-tissue pump is the dominating factor controlling the distribution of DIC in the water column, and it has the potential of extracting and exporting larger amounts of  $\text{CO}_2$  from the atmosphere compared to hard-tissue and solubility pump.

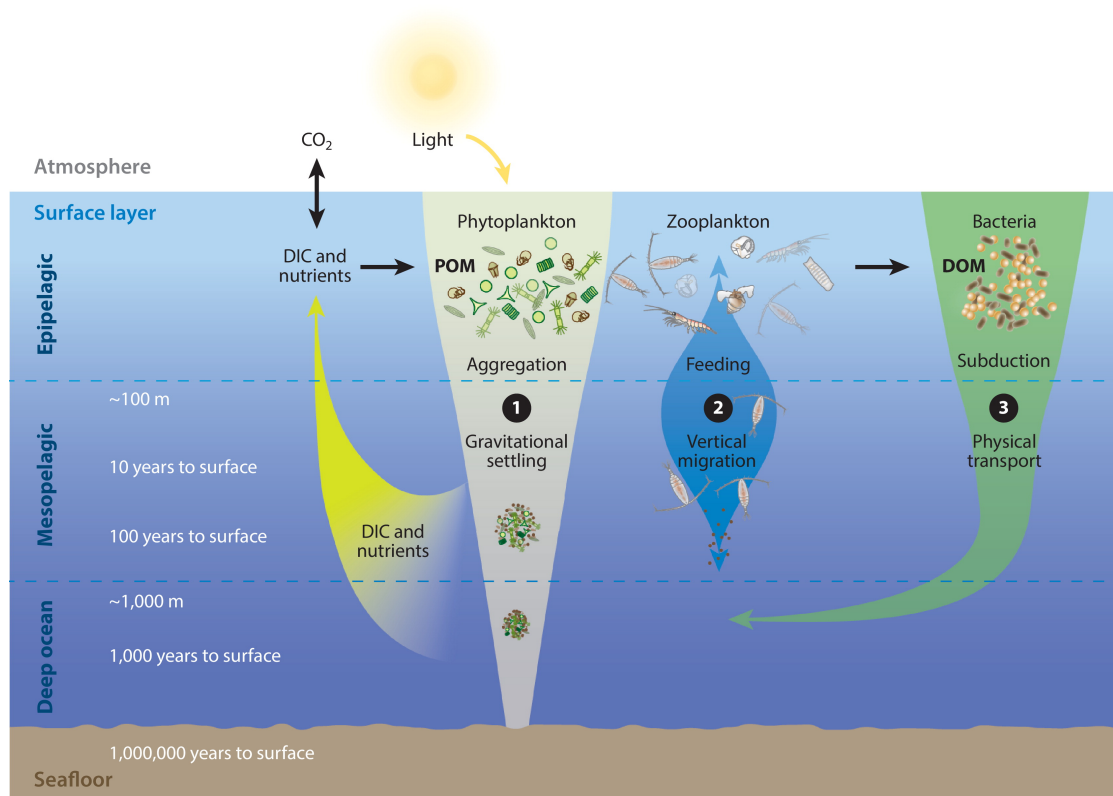
Figure 1.7 shows the processes that characterize the cycle of carbon fixation and export to depth in the organic carbon pump. The first step is primary production, that is the fixation of  $\text{CO}_2$  and nutrients by phytoplankton in the uppermost layer, known as euphotic depth (30-150 m), where light can support photosynthesis:



The majority of this fixation (about 80%) is carried out in the open ocean, while the remaining amount occurs in the very productive upwelling regions in the eastern boundary of oceanic basins [5]. Primary production is able to synthesize both dissolved and particulate organic carbon: DOC is mostly recycled in the surface layer by bacteria, and remineralized into dissolved inorganic carbon; conversely, the produced POC is further processed by microbes, zooplankton and other heterotrophs organisms into fecal pellets, that aggregate with dead phytoplankton and microbes, inorganic material or sand. These aggregates, known as *marine snow*, gravitationally sink to depth (process number 1 in Figure 1.7). Marine snow is the primary fraction of organic matter that escapes remineralization in the upper 100 m of the water column, thus being exported [4]. Another type of export occurs through sinking of fecal pellets produced by diurnal and seasonal vertical migration of zooplankton (process number 2 in Figure 1.7).

The variability of the velocity field in the ocean can have a contribution in carbon export through the physical subduction of dissolved organic matter (process number 3 in Figure 1.7). Physical subduction processes occur on a wide range of space and time scales, and are strictly linked to the local dynamics.






 Iversen MH, 2023  
*Annu. Rev. Mar. Sci.* 15:357–81

Figure 1.7: Diagram of the processes of carbon export in the soft-tissue component of biological carbon pump, with depths and expected time scales of export. Figure reproduced from Iversen (2022) [4].

A study in the Eastern Mediterranean basin has proved that the presence of strong vertical, downward velocities can amplify organic matter fluxes in the water column at 500 meters, although a direct effect in determining high sedimentation rates at greater depths cannot be found [9]. This suggests that physical subduction could be more relevant at climatological scales, and of secondary importance in relatively fast carbon export.

Lateral advection of biomass from the shelf domain to the open ocean is another physical mechanism that might be particularly relevant in the yearly budget of carbon export. Primary production, that is more intense in shallower ocean regions, produces pellets and marine snow that can sink to greater depths when transported to the open ocean, enhancing the vertical fluxes of matter. It is estimated that this type of transport from the shallow shelf could contribute up to 80% of the open ocean carbon export off the coast of Cape Blanc, Mauritania [4]. Cross-shelf currents and their associated transport of tracers can develop through many mechanisms. This work focuses on a particular type of circulation that is generated by mesoscale and submesoscale instabilities. The physical framework of this type of dynamics is presented in the next subsection.

## 1.2 Mesoscale and submesoscale dynamics

Figure 1.8 shows the complex dynamics of the ocean, characterized by structures and currents that evolve in a very large range of spatial and temporal scales. A mathematical description of the motion of water masses can be derived from the principles of momentum, energy and mass conservation for a viscous, gravity-forced fluid that flows in a rotating reference system: the result is a set non-linear partial differential equations, whose solution cannot be analytical in most cases and requires numerical treatment [10, 11].

These equations can be simplified through a spatial and temporal scale analysis that allows the separation of the motion in different dynamical regimes, through the definition of the Rossby number:

$$\epsilon = \frac{U}{Lf_0} \quad (1.4)$$

where  $U$  represents the order of magnitude of horizontal velocity field,  $L$  is the horizontal scale of the phenomenon and  $f_0 = 2\Omega \cos \varphi$  is the Coriolis parameter, dependent on Earth's rotation velocity  $\Omega = 7.29 \cdot 10^{-5} \text{ s}^{-1}$  and latitude  $\varphi$ . At mid latitudes,  $f_0 \sim 10^{-4} \text{ s}^{-1}$ . The Rossby number expresses the ratio between the effect of local processes and the planetary effect induced by Earth's rotation. A high Rossby number describes dynamical systems where planetary rotation is not relevant and local effects are more important in the momentum balance. Con-

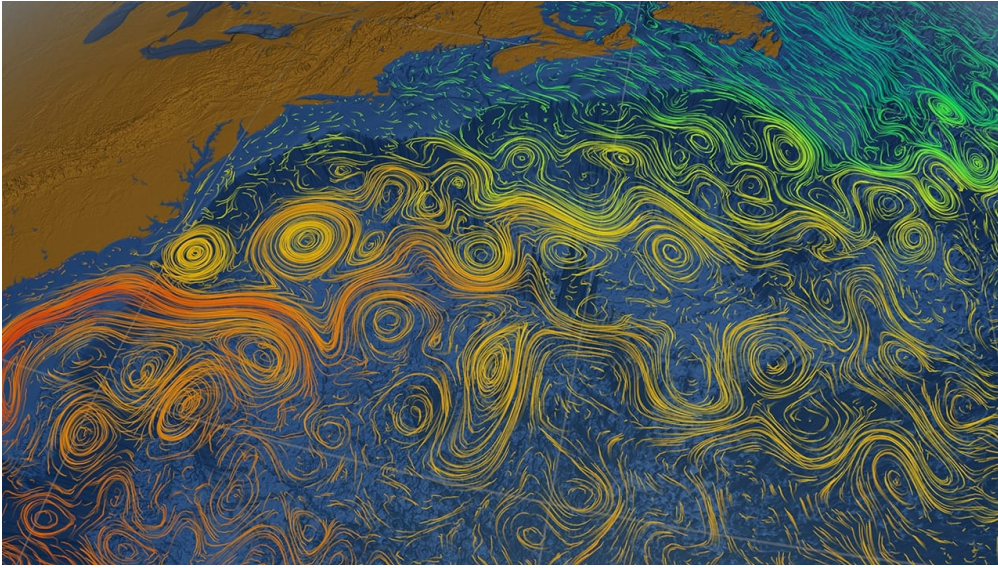


Figure 1.8: Gulf Stream sea surface currents and temperatures generated from the output of a model. The main branch of the Gulf stream is clearly visible throughout the whole image, and its meandering generates eddies (roughly-circular structures) and other, lower-scale structures. Figure reproduced from [12, 13].

versely, small values of  $\epsilon$  characterize scales where planetary effects drive the fluid motion. The overall ocean dynamics results from a combination of small-scale and large-scale processes, that can be grouped in four main regimes [11]:

1. The large-scale circulation ( $\epsilon \ll 1$ ) acts on spatial scales of the order of thousands of kilometers, that is the extension of ocean basins, and time scales greater than one year. Large-scale processes include the surface wind-driven circulation, that form subtropical and subpolar gyres, and the thermohaline circulation, that acts on climatological scales.
2. Mesoscale circulation ( $\epsilon < 1$ ) occurs at scales of the order of 10 to 100 kilometers and time scales of months. This type of dynamics is primarily governed by the geostrophic balance between two fundamental forces: the Coriolis effect and the horizontal pressure gradient. Within this category of circulation, there are oceanic eddies that hold significant importance in the transport of heat and various tracers within the ocean.
3. Submesoscale processes ( $\epsilon > 1$ ) occur at typical horizontal scales of 1 to 10 kilometers and time scales of the order of days, representing an intermediate regime of transition from large-scale, purely horizontal processes towards isotropic turbulence. In this regime, the effect of vertical processes assumes

importance in the momentum equation, weakening the geostrophic balance and introducing non-linear instabilities.

4. Microscales ( $\epsilon \gg 1$ ) are associated with extremely small length scales, typically less than 1 kilometer down to millimeters, and very short time scales on the order of days to seconds. In this regime, energy is dissipated through friction, ultimately transforming into heat.

These processes effectively interact through momentum and energy exchanges, that result in an energetic cascade from the planetary, wind-forced scale to the microscale, where dissipation occurs through friction.

As anticipated in Subsection 1.1.2, large-scale circulation drives the solubility pump through the sinking of cold, carbon-rich polar water masses and their transport towards equatorial latitudes. Mesoscale and submesoscale processes have an influence on local biomass transport, affecting the ecosystem productivity and the effectiveness of the biological carbon pump.

### 1.2.1 Eddies and filaments

Mesoscale dynamics emerges at large scales where planetary rotation dominates the motion of water masses. One of the most used mathematical models to study medium to large-scale ocean dynamics is the shallow water system in small amplitude approximation for a homogeneous fluid, where neither viscous nor turbulent effects are considered [10]:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + fv = -g \frac{\partial \eta}{\partial x}; \quad (1.5)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} - fu = -g \frac{\partial \eta}{\partial y}; \quad (1.6)$$

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (H\vec{u}) = 0 \quad (1.7)$$

where  $u$  and  $v$  represent the two components of horizontal velocity field  $\vec{u}$ ,  $\eta$  is the free-surface deviation from a flat reference surface  $H$  ( $\eta \ll H$ ),  $f$  is the Coriolis parameter and  $g$  is the gravitational acceleration ( $g = 9.81 \text{ m s}^{-2}$ ). As previously seen, large-scale motion is characterized by a small value of the Rossby number, therefore a perturbative approach can be used to expand  $u, v$  and  $\eta$  in a series of components that are multiplied by increasing powers of  $\epsilon$ :

$$\begin{aligned} u &= u_0 + \epsilon u_1 + \epsilon^2 u_2 + \dots \\ v &= v_0 + \epsilon v_1 + \epsilon^2 v_2 + \dots \\ \eta &= \eta_0 + \epsilon \eta_1 + \epsilon^2 \eta_2 + \dots \end{aligned} \quad (1.8)$$

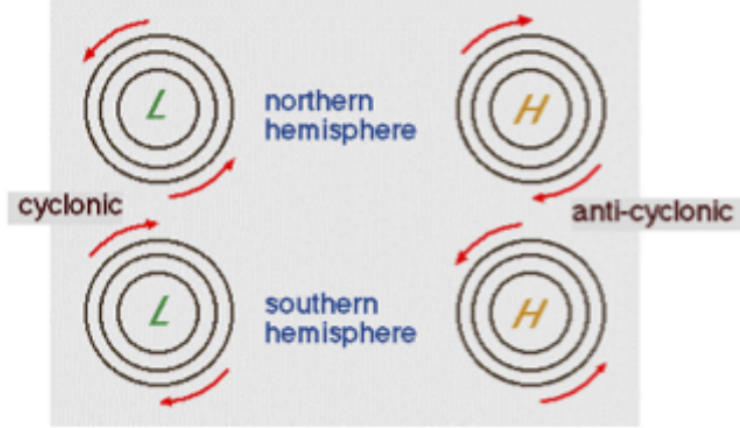


Figure 1.9: Diagram of eddy sense of rotation depending on the hemisphere. Figure reproduced from Pinardi (2022) [10].

The Coriolis parameter can be expanded by admitting small variations of latitude  $y$  in the flow, and the result is:

$$f = f_0 + \beta y \quad (1.9)$$

where  $\beta = 2\Omega \cos \varphi_0 / R$ , with  $R$  as the Earth radius and  $\varphi_0$  as a fixed latitude. This allows to obtain a new series of equations, each one describing a specific dynamical mode associated to a certain power of  $\epsilon$ . The zero-order equations simplify to the geostrophic balance between Coriolis and horizontal pressure gradient forces:

$$\begin{cases} f_0 v_g = g \frac{\partial \eta}{\partial x} \\ f_0 u_g = -g \frac{\partial \eta}{\partial y} \end{cases} \quad (1.10)$$

The geostrophic velocity field  $(u_g, v_g)$  obtained from Equation 1.10 depends on the gradient of sea level and its direction is perpendicular to the isolines of free surface  $\eta$ . This balance is only diagnostic, as it does not have explicit time dependence. Given a static  $\eta$  field, it is possible to derive the corresponding instantaneous geostrophic currents, but not the evolution of the fields over time.

The first-order mode obtained from the expansion of shallow water equations results in the *quasi-geostrophic* equations:

$$\frac{\partial u_0}{\partial t} + u_0 \frac{\partial u_0}{\partial x} + v_0 \frac{\partial u_0}{\partial y} - \beta y v_0 - v_1 = -\frac{\partial \eta_1}{\partial x} \quad (1.11)$$

$$\frac{\partial v_0}{\partial t} + u_0 \frac{\partial v_0}{\partial x} + v_0 \frac{\partial v_0}{\partial y} + \beta y u_0 + u_1 = -\frac{\partial \eta_1}{\partial y} \quad (1.12)$$

These two momentum equations can be combined in a single conservation equation:

$$\left[ \frac{\partial}{\partial t} + \vec{u}_0 \cdot \nabla \right] \underbrace{(\zeta_z - F\eta_0 + \beta y)}_{\Pi} = 0 \quad (1.13)$$

where the conserved quantity,  $\Pi$ , is called *quasi-geostrophic potential vorticity*.  $\Pi$  is the sum of three terms: the vertical relative vorticity of the flow field  $\zeta_z$ , a term associated to water column stretching  $F\eta_0$ , and the planetary vorticity induced by Earth's rotation  $\beta y$ .

Equation 1.13 is a prognostic equation, as it explicitly contains the dependence on time, and it can be used to predict mesoscale features such as Rossby waves. Perturbations greater than first order will generate further time-dependent imbalances in the geostrophic field, that can result in local minima or maxima in the free surface field. Following the geostrophic balance described in Equation 1.10, currents will flow in a roughly circular motion around these centers, generating an eddy. The sense of rotation depends on the sign of  $f_0$ , which changes with the chosen Earth hemisphere (Figure 1.9).

Submesoscale currents (SMCs) are intermediate-scale flow structures in the form of density fronts and filaments, generated from mesoscale instabilities. These patterns are typically visible in satellite images of sea surface temperature or tracers concentration and characterized by linear sharp edges or horizontally elongated features. Being associated with strong, ageostrophic currents, their essential dynamics is advective, hence it includes nonlinear terms that make it difficult to treat theoretically. However, because of their abundance in the ocean (as visible in Figure 1.8), their role in energy transfer and transport of tracers is of great importance.

McWilliams (2016) [14] summarizes several mechanisms of generation of SMCs:

- a. *Mixed-layer instability*: the mixed layer is the uppermost ocean layer, characterized by nearly uniform temperature and density due to turbulent mixing. Its depth can vary from tens of meters up to 500 m, depending on geographical position and seasonality. Baroclinic instability can generate density gradients on spatial scales that can be estimated from the baroclinic deformation radius  $\ell_s$ :

$$\ell_s \sim \frac{N_s h_b}{f} \quad (1.14)$$

where  $N_s$  is the Brunt-Väisälä frequency, associated to the stability of the density stratification,  $h_b$  is the surface layer thickness and  $f$  is the Coriolis parameter. With typical values of mixed layer,  $N_s \sim 10^{-3} \text{ s}^{-1}$ ,  $h_b \sim 10^2 \text{ m}$  and  $f \sim 10^{-4} \text{ s}^{-1}$ , a deformation radius of  $\sim 1 \text{ km}$  is obtained, accordingly to the

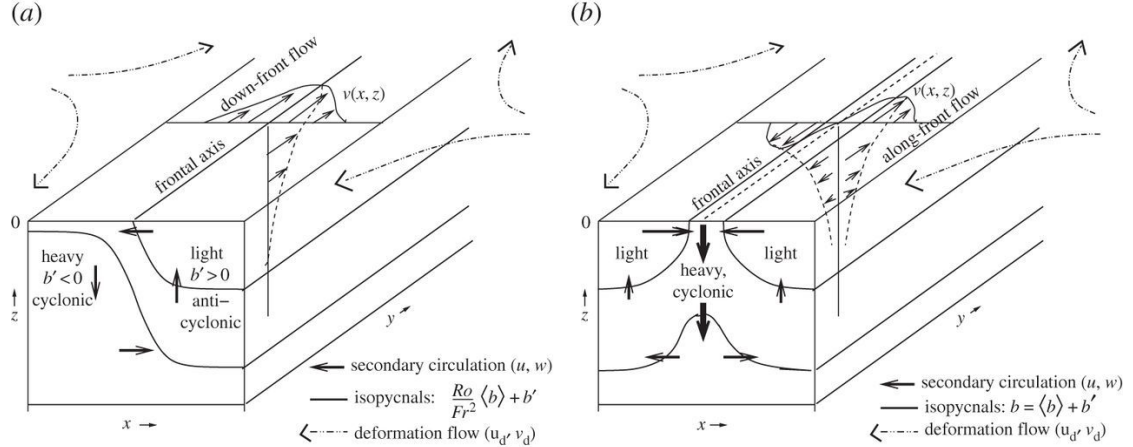


Figure 1.10: Schemes of surface-layer frontogenesis generated by a large-scale flow deformation for a front (panel a) and a dense filament (b). Secondary overturning circulations in the  $xz$  plane are highlighted. Figure reproduced from McWilliams (2016) [14].

typical scales of SMCs. This local perturbation of density field in the mixed layer, that can be induced by sea surface temperature gradients, generates a secondary ageostrophic circulation whose intensity is considerably greater than the surrounding geostrophically-balanced field.

- b. *Strain-induced frontogenesis*: frontogenesis is a classical dynamical process in meteorology, generating horizontal buoyancy gradients that intensify in presence of a favourable background strain mechanism. Frontogenesis occurs also in the ocean, where the primary background strain is from mesoscale currents and eddies. Figure 1.10 shows the processes of frontogenesis for a buoyancy front (panel a) and a dense filament at the surface (panel b). Both these two configurations generate an ageostrophic circulation in the cross-front and vertical plane ( $u, w$ ). In the case of a buoyancy front, this secondary circulation is an overturning cell with the upwelling of light water and downwelling of dense water. Conversely, the dense filament presents two secondary circulation cells that converge in a central, downwelling branch. This type of dynamics concentrates tracers in thin, elongated structures that can propagate for several hundreds of kilometers. Analogous frontogenesis can occur in the case of light-water filaments: however, a crucial distinction arises due to the presence of a surface divergence line, making this variant of filaments inherently weaker and seemingly less prevalent in natural occurrences. The combination of mixed-layer instability and frontogenesis explains the generation of SMCs, the former generating the horizontal buoyancy gradient and the latter characterizing the

advective evolution of these structures.

- c. *Turbulent thermal wind*: both mixed-layer instability and frontogenesis require the the assumption of a conservative flow, which is a rough approximaton for the oceanic surface layer that is subjected to strong turbulence induced by wind stress. The previous frameworks can be integrated with the assumption of turbulent flow by introducing an approximate linear momentum balance called turbulent thermal wind. This theory is a composite generalization of geostrophic, hydrostatic, conservative balance and an Ekman boundary-layer balance that accounts for the effect of wind stress. Its solution shows the same overturning-cell structure described in strain-induced frontogenesis for both the front and dense filament cases.
- d. *Topographic wakes*: it is well known that the presence of slopes and topographic variations of ocean bottom have an impact on the vorticity balance of the flow, generating secondary effects such as shear instabilities and mixing.

These mechanisms of generation of SMCs rely on the deformation of a background, mesoscale flow, that shows significant variations in size and amplitude depending on the geographic region and seasonality. While the global variability of mesoscale eddies has been well investigated with both models and observations, at the moment there is no comparable global measurement system for SMCs, as they arise only from regional-scale simulations that require very fine grid resolutions.

Between the SMCs, cold-water filaments take on particular significance. Theoretical research has shown that these structures are linked to a more pronounced intensification of the ageostrophic circulation when compared to their warm counterparts [15]. An illustration of the process of cold filament formation is depicted in Figure 1.11. In this example, the filament is created through the straining of two warm-core anticyclonic eddies. These warm-core eddies generate a secondary ageostrophic circulation that results in the convergence of cold waters from the south. This convergence occurs within an extremely narrow structure and is accompanied by a relatively high surface velocity anomaly directed northward. Consequently, this process is associated with the transport and injection of cold waters into the warmer ocean domain. These structures, often referred to as *streamers*, typically originate at the boundary between the shelf and open ocean. They play an important role in oceanic biogeochemistry by significantly influencing the transport of various tracers, including nutrients, phytoplankton, and zooplankton larvae.

In the upcoming section, the biogeochemical aspect of this process of transport of biomass from the continental shelf to the open ocean will be further investigated.



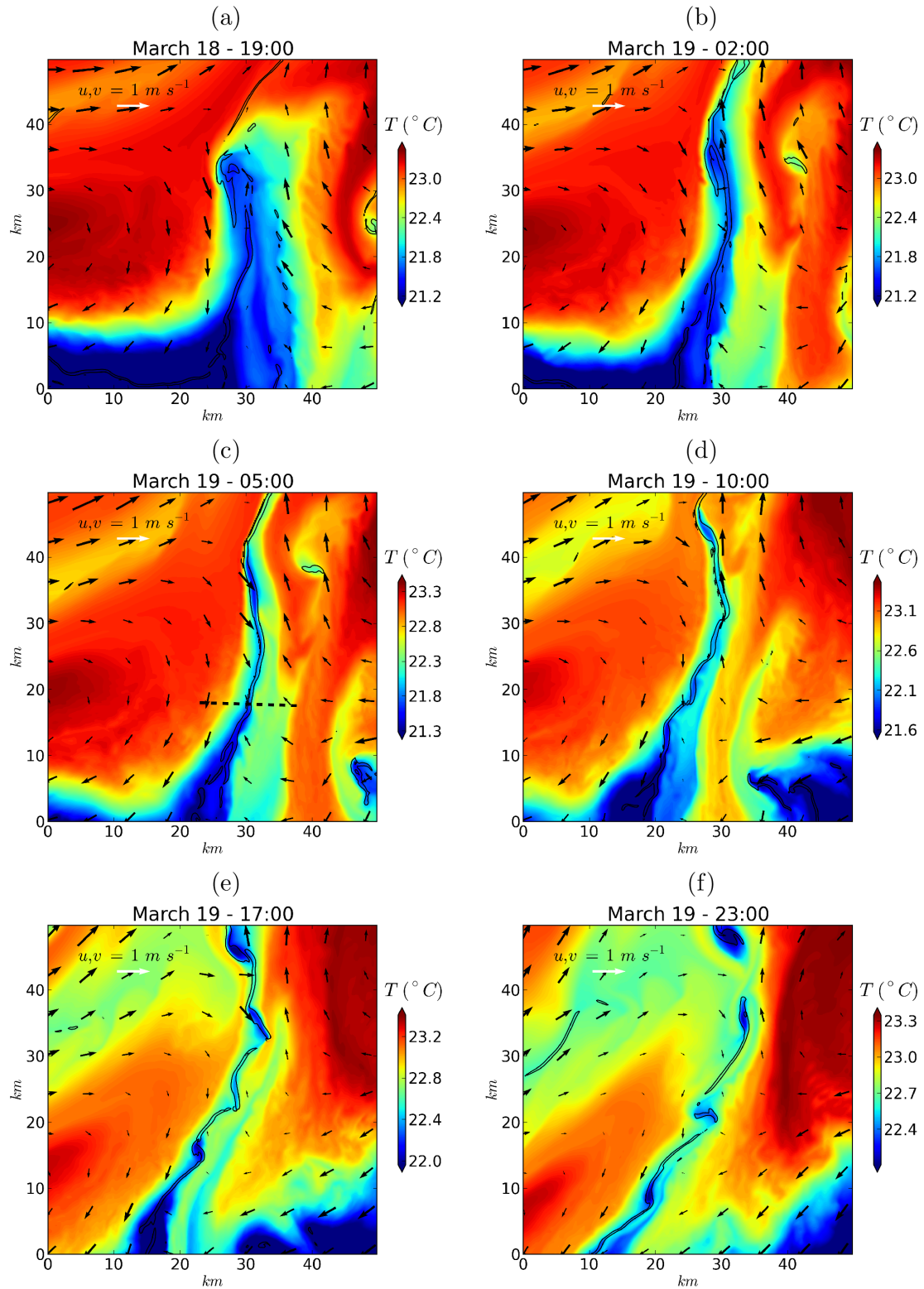


Figure 1.11: Generation and evolution of a cold-water filament. The figure displays the sea surface temperature (in colors), the surface relative vorticity (black contours) and the anomaly of surface horizontal velocity field (arrows). Figure reproduced from Gula et al. (2014) [16].

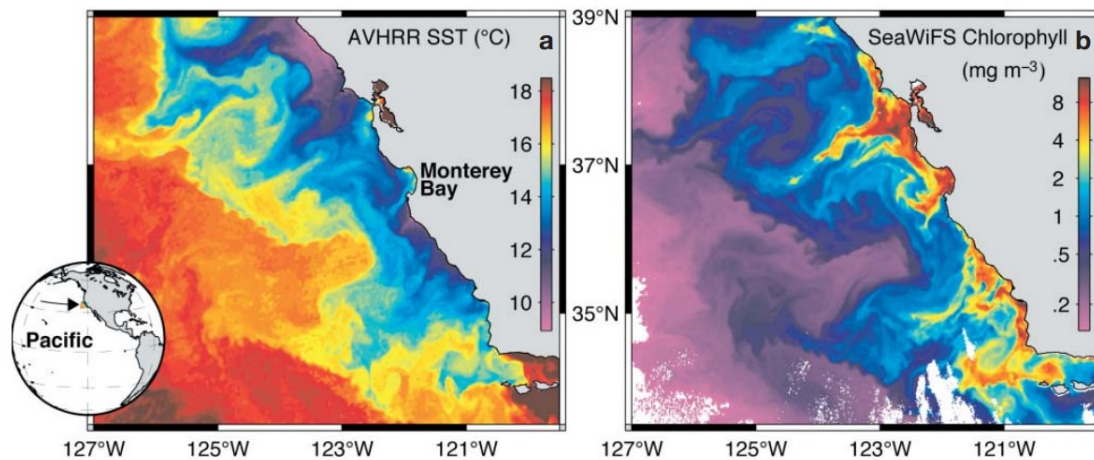


Figure 1.12: Upwelling near the coast of California seen in satellite images of sea surface temperature (left panel, August 14, 2000) and chlorophyll-*a* concentration (right panel, August 16, 2000). Coastal waters expand towards the open ocean through filaments and submesoscale structures. Figure reproduced from Ryan et al (2005) [17].

### 1.3 Lateral export or recycling?

Although the majority of atmospheric carbon fixation by primary production occurs in the open ocean, coastal regions are characterized by much higher productivity that results in enhanced chlorophyll concentrations near land, as clearly visible in satellite images (Figure 1.12, right panel). This increased productivity is due to a combination of several processes related to the shallowness of the water column [18]. Upwelling is an additional mechanism that characterizes the most productive oceanic regions, causing the cold, nutrient-rich deep waters to move towards the surface, enhancing productivity.

Coastal waters are typically separated from the warmer, nutrient-depleted surroundings by a front that exhibits baroclinic instability. This instability leads to the formation of mesoscale and submesoscale structures, including streamers, which propagate towards the open ocean. As discussed in earlier sections, these cold filaments create a secondary overturning circulation that has the ability to concentrate tracers, such as phytoplankton, within submesoscale-wide channels and transport them laterally over distances of several hundreds of kilometers.

Streamers exhibit distinctive characteristics in satellite images, transporting cold and chlorophyll-rich coastal waters into open ocean regions in thin, elongated shapes, as clearly visible in Figure 1.12.

This mechanism of transport of biological material from the coastal region towards

the open ocean could be important in the ocean carbon cycle, as the marine snow and pellets that are ejected by primary production can sink to greater depths in these regions. The first hypothesis that a fraction of primary production is exported from the continental shelf to the deep ocean basins was formulated by Walsh et al. (1981) [19]. This paper estimated that the carbon sink flux due to these advection processes could account for about  $1.5 \text{ PgC yr}^{-1}$  in a global, annually balanced carbon budget for food chains in continental shelves. Several experiments have been carried out to confirm this hypothesis. Malone et al. (1983) [20] verified the association between an increase in biomass export and intense lateral currents induced by wind events in the New York Bight. However, a following research by Falkowski et al. (1988) [21] in the same region during the spring of 1984 found that the export of shelf-derived production to the deep ocean was small.

The contribution of streamers to this lateral export has been also investigated in the productive Eastern Boundary upwelling regions. A simulation experiment in the Cape Blanc region has been conducted by Lovecchio et al. (2018) [22], showing that 80% of the total flux of organic carbon at 100 km offshore is transported by upwelling filaments, and that their contribution can extend up to 1000 km offshore. The role of filamentary-induced biomass transport in global carbon export is still uncertain, due to the difficulties in modelling submesoscale currents, and it has been assessed only in regional frameworks. High-resolution satellite observations provide global and continuous time coverage of the state of the ocean and do not rely on mathematical approximations required by simulations. However, a systematic monitoring of these structures is still a challenge due to their complex features and shapes, and many studies on submesoscale currents require first human identification.

## 1.4 Thesis objectives

The aim of this work is the estimation of the contribution of streamers to the lateral transport of biomass from the shelf domain to the open ocean. Only observational data are employed, consisting of satellite products of sea surface temperature and ocean color.

The work is structured as follows:

1. an image analysis method has been developed for various off-shelf areas in the world's oceans, including Mauritania, California, and the South China Sea. This method is based on the K-means clustering of satellite-derived data products of chlorophyll-*a* concentration and sea surface temperature. This algorithm is capable to detect and classify different dynamical regimes, including streamers.

2. an estimation of the chlorophyll content of these filamentary regions is provided, and its evolution over time is analysed;
3. an estimation of the lateral export of carbon from the coastal areas through this mechanism is given.

The datasets are described in Chapter 2, where an introduction on remote sensing data collection and processing is given. The employed unsupervised learning framework and statistical methods are presented in Chapter 3. The results are shown in Chapter 4, and conclusions are summarized in Chapter 5.

# Chapter 2

## Satellite data

As anticipated in Chapter 1, the goal of this work is understanding the contribution of streamers to lateral biomass transport from the shelf break to the open ocean, by using only observational data from satellite remote sensing products. This chapter gives a general overview of these datasets. Section 2.1 presents the basis of remote sensing and its applications. The employed sea surface temperature product is presented in Section 2.2. The processing of ocean colour satellite data, including chlorophyll and an estimation of the euphotic depth, is described in Section 2.3. Finally, the sea bottom topography product is described in Section 2.4.

### 2.1 Introduction to remote sensing

Modern Earth observing sensors are installed on satellites that are placed in several types of orbits around the Earth, and provide a global, continuous monitoring of the Earth system components. Remote sensing of the ocean is particularly important in the field of climate sciences, as it can provide data for understanding and monitoring various aspects of the marine environment and ocean circulation. This technology allows to access and observe areas that are otherwise difficult or impossible to reach due to their remote or hazardous nature, such as open ocean regions and polar seas. [23]

Earth observing sensors use the electromagnetic radiation to sense the environment. Radiation propagates in the atmosphere and in the vacuum of space in the form of waves. Each wave, identified by a certain wavelength, carries a certain amount of energy that characterizes a spectrum. The shape of energy spectra varies depending on many factors, such as the composition of the source of radiation, and the processes that electromagnetic waves undergo when propagating in a medium. [23]

Radiation undergoes absorption and scattering as it propagates through the atmo-

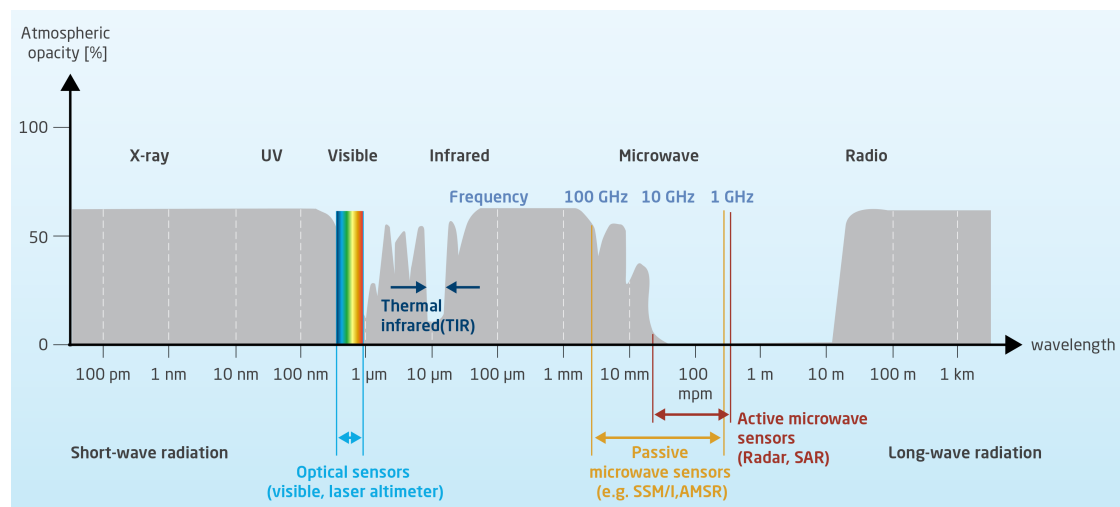


Figure 2.1: Diagram of atmospheric opacity as a function of wavelength. The typical bands employed by different remote sensing instruments are highlighted. Figure reproduced from SEA ICE Portal website [24].

sphere. It is then reflected back by the Earth’s surface and captured by sensors. The final spectrum of this radiation is shaped by various factors, including the concentration and type of atmospheric components and the properties of the surface being sensed. Figure 2.1 shows the atmospheric opacity to electromagnetic radiation: as it can be seen, the atmosphere is transparent to electromagnetic radiation only in certain ranges of wavelengths, limiting the unperturbed observation of the Earth’s surface to these windows [23]. Sensors are usually designed to detect radiation in specific ranges of wavelengths, called *bands*, that identify and characterize different physical or biological properties of water masses. Sensors can be grouped into two families [23]:

- *passive* sensors only gather radiation that is emitted or reflected by the covered area;
- *active* sensors emit a signal and measure the reflected or backscattered radiation from the target.

Ocean satellite data products are derived from both passive and active sensors, depending on the feature of interest. Raw data obtained from sensors consist of radiance measurements  $I_\lambda(\Omega, t)$ , that is the energy carried by an electromagnetic wave of wavelength  $\lambda$  in a solid angle  $\Omega$  at time  $t$ , per unit of time and surface. In order to obtain the final geographically mapped and interpretable product, these measurements must undergo atmospheric correction and several other processing steps. Satellite product levels span from 0 to 4 [25]:

- *Level 0* products represent the raw data at full instrument resolution, with any and all communications artifacts or duplicate data removed;
- *Level 1* data are processed to sensor units, time-referenced and annotated with ancillary information including georeferencing parameters;
- *Level 2* data are processed to derive the geophysical variable at the same resolution and location annotated in Level 1 data;
- *Level 3* data are statistically interpolated on regular space-time grids, also combining data from different sensors, and validated for quality assurance;
- *Level 4* data are similar to Level 3 data, but in this case, statistical extrapolation or modeling is also used to fill data gaps.

In the next sections, the characteristics of the employed remote sensing data are explored. These datasets include sea surface temperature, ocean color, and bathymetry.

## 2.2 Sea surface temperature

Sea surface temperature (SST) is the water temperature of the uppermost layer of the ocean. It is a very important parameter for understanding and monitoring various oceanic processes, including climate patterns, marine ecosystems, and ocean circulation. On local scales, SST is used operationally to identify eddies, fronts and upwelling regions.

This work employs the daytime sea surface skin SST Level 3 data product derived from the *Moderate-Resolution Imaging Spectroradiometer* (MODIS) instrument mounted on NASA's Aqua (EOS PM) satellite, launched in 2002. MODIS is a passive sensor collecting data across 36 spectral bands spanning the visible and infrared range (0.4-14.4  $\mu\text{m}$ ), providing information on the atmospheric column and the surface of the ocean [26]. This product is made available by the NASA *Ocean Colour Web* on a daily basis, mapped on an uniform grid with a 4 kilometers resolution [27].

Planck's law expresses the spectral radiance emitted by a black body as a function of temperature and wavelength [28]:

$$I_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \quad (2.1)$$

where  $h$  is Planck constant,  $c$  is the speed of light in vacuum, and  $k_B$  is Boltzmann constant. By inverting this equation, it is possible to derive a *brightness temperature*  $T_\lambda$ :

Band number	Band center [ $\mu\text{m}$ ]	Bandwidth [ $\mu\text{m}$ ]
20	3.750	0.1800
22	3.959	0.0594
23	4.050	0.0608
31	11.030	0.5000
32	12.020	0.5000

Table 2.1: Bands for MODIS infrared SST determination. The bands used for the computation of the used SST product in this work are highlighted.

$$T_\lambda = \frac{hc}{k_b\lambda} \ln^{-1} \left( 1 + \frac{2hc^2}{I_\lambda\lambda^5} \right) \quad (2.2)$$

SST is derived from MODIS measurements within specific bands in the mid-far infrared range (3-14  $\mu\text{m}$ ), as outlined in Table 2.1, utilizing a dual-window technique that makes the correction of atmospheric absorption and scattering straightforward. This method combines brightness temperatures derived from radiance measurements within these bands, and distinct algorithms have been developed depending on the specific combination employed. In this study, the SST product derived from the thermal infrared window (11-12  $\mu\text{m}$ ) has been used [26, 27, 29]:

$$\begin{aligned} \text{SST} = & c_0 + c_1 \cdot T_{31} + c_2 \cdot (T_{32} - T_{31}) \cdot T_{ref} \\ & + c_3 \sec(\theta - 1)(T_{32} - T_{31}) + \text{corrections}(\theta) \end{aligned} \quad (2.3)$$

where  $T_{31}$  and  $T_{32}$  represent the brightness temperatures computed from radiance measurements in band 31 and 32, respectively.  $T_{ref}$  denotes a reference SST,  $c_i$  are empirical coefficients derived and continuously verified based on match-ups between the satellite brightness temperatures and field measurements, and  $\theta$  is the sensor zenith angle.

## 2.3 Ocean color

The color of the ocean, while typically appearing blue, can exhibit variations, ranging from blue to green or even yellow in certain circumstances. These variations are primarily influenced by factors such as dissolved organic matter, living phytoplankton containing chlorophyll pigments, and non-living particles like degraded phytoplankton, other plant material, marine snow and mineral sediments.

The study of ocean color is essential for understanding the composition of suspended materials in the water column. It serves various purposes, including exam-



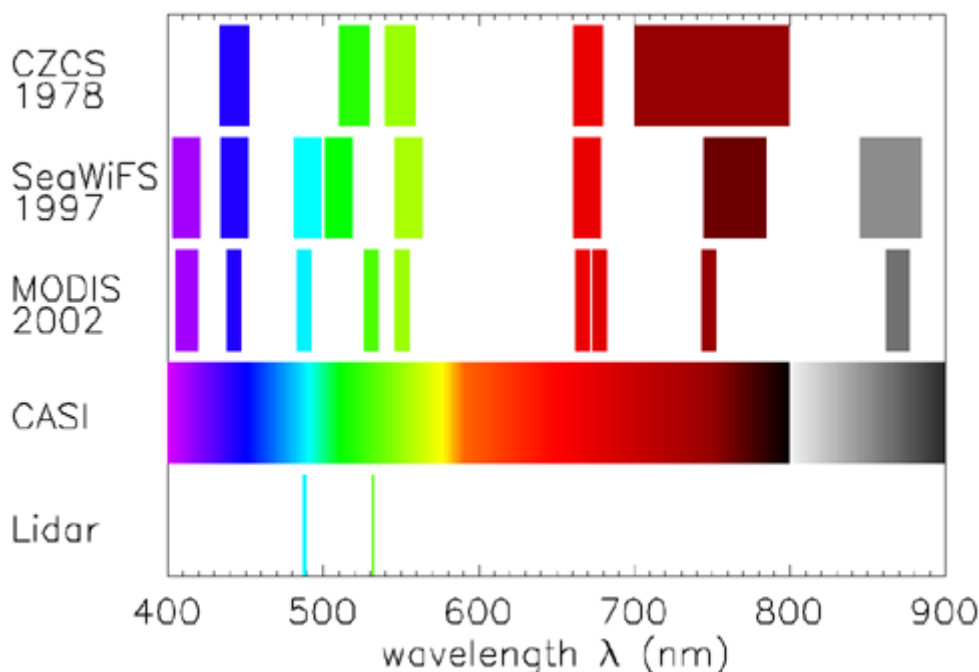


Figure 2.2: Visible range bands employed by different ocean color remote sensors. Figure reproduced from Mobley (2022) [28].

ining carbon fixation and cycling, monitoring ecosystem changes resulting from climate shifts, assessing the health of coral reefs, and detecting harmful algal blooms and pollution events [28].

Ocean color remote sensing instruments typically use bands within the visible wavelength range, spanning from 400 to 800 nm, as shown in Figure 2.2. These sensors passively collect sunlight reflected by water bodies and measure its energy spectrum in this wavelength range.

Radiative transfer equations have been developed to link the collected radiances to the biogeochemical properties of water masses. These equations include various terms that account for different scattering and absorption processes, which are defined by the *inherent optical properties* (IOPs) of the water. The types and concentrations of substances in the water modify these IOPs, affecting the spectral characteristics of the collected radiation.

The relationship between the measured radiances and the biogeochemical properties of seawater is based on solving an inverse problem. Given a measured spectrum and information on the boundaries, the IOPs of the water body from which the radiance originates can be determined, and its biogeochemical properties subsequently. The solution of this inverse problem is extremely complex, and it presents several limits related to the uniqueness of the solution and its sensitivity to mea-

surement errors, numerical modelling and approximations [28].

IOPs are generally derived from another type of optical properties known as *apparent optical properties* (AOPs) that represent the quantities typically observed by remote sensing instruments. AOPs are dependent on the inherent properties of water bodies (i.e. the IOPs), but also on external agents such as the angle of observation or sky conditions. Two important AOPs are [28]:

- the *remote sensing reflectance*  $R_{rs}(\lambda)$  parameter, that is the spectral radiance upwelling from beneath the ocean surface, normalized by the downwelling solar irradiance, and measured at each band wavelength of the sensor;
- the *diffuse attenuation coefficient for downwelling irradiance*  $K_d(\lambda)$ , typically measured at the 490 nm wavelength band, representing the rate at which the downwelling solar irradiance decreases with depth in a water column due to scattering and absorption processes.

The generation of a long time series of ocean color measurements for climate studies, or simply for monitoring limited regions, is not straightforward, as the quality of this type of data is affected by several factors. First of all, cloud cover creates a serious limitation to the availability of ocean color data, since clouds strongly reflect light in the visible range. Furthermore, satellite missions have a finite duration, and the spectral characteristics of sensors evolve in successive developments, with implications for consistency in data [30]. Multiple datasets derived from different sensors are combined to improve the quality of ocean color datasets, as explained below.

### 2.3.1 Satellite chlorophyll-*a*

Chlorophyll-*a* (Chl-*a*) is the primary pigment found in all plants, including phytoplankton. Unlike other pigments that serve either photoprotective or light-harvesting functions, Chl-*a* plays a central role in the conversion of sunlight into chemical energy through photosynthesis, ultimately leading to the production of carbon compounds. The concentration of Chl-*a* in aquatic environments serves as an indicator of primary production intensity. This molecule exhibits a distinctive absorption spectrum with characteristic absorption lines in the violet, blue, and red wavelengths.

The Chl-*a* concentration dataset used in this study is a multi-sensor Level 3 product distributed by the *Ocean Colour Climate Change Initiative* (OC-CCI) project of ESA, with daily time resolution and uniform space resolution of 4 kilometers, made available by the *Copernicus Marine Service* (CMEMS) platform [30, 31]. This product is derived from measurements of  $R_{rs}(\lambda)$  of several sensors, including MODIS, spanning different time period ranges, as shown in Figure 2.3. Differ-

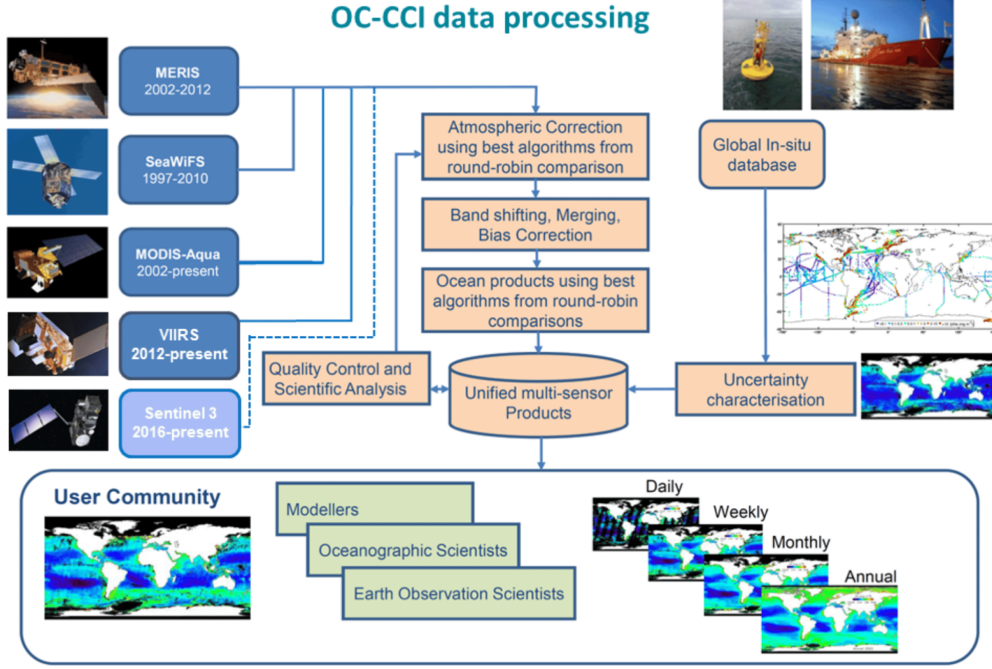


Figure 2.3: Flowchart of multi-sensor OC-CCI satellite product processing. Figure reproduced from Sathyendranath et al. (2019) [30]

ent sensors measure radiances in different bands, which is an obstacle for inter-comparison and merging. A band-shifting scheme is employed in order to map all the  $R_{rs}$  values to the same set of wavelengths, and the datasets produced in each mission  $i$  are corrected for a multi-annual bias of the average  $R_{rs}$  values,  $\langle R_{rs} \rangle$ , defined as:

$$\delta_i^{ref}(\lambda, b) = \frac{\langle R_{rs}^i(\lambda, b) \rangle}{\langle R_{rs}^{ref}(\lambda, b) \rangle} \quad (2.4)$$

for each wavelength  $\lambda$  and region  $b$ . The reference spectral reflectance  $R_{rs}^{ref}$ , employed also for the wavelengths mapping, is the one from the SeaWiFS mission. The final corrected product has the form:

$$R_{rs}^i(\lambda, b) = \frac{R_{rs}^i(\lambda, b)}{\delta_i^{ref}(\lambda, b)} \quad (2.5)$$

The merged product is then obtained by averaging the available data for a given pixel [30].

Chl- $a$  concentration is then derived from a fourth-order polynomial relationship of the ratio of  $R_{rs}$  measured in different bands [30, 32, 33]:

$$\log_{10}(\text{Chl}a) = c_0 + \sum_{i=1}^4 c_i \left[ \log_{10} \left( \frac{R_{rs}(\lambda_{\text{blue}})}{R_{rs}(\lambda_{\text{green}})} \right) \right]^i \quad (2.6)$$

where the numerator,  $R_{rs}(\lambda_{\text{blue}})$  is the greatest remote sensing reflectance computed between 443, 489 and 510 nm, and  $R_{rs}(\lambda_{\text{green}})$  is computed at 555 nm [34]. The coefficients  $c_i$  are empirically estimated by fitting Equation 2.6 to in-situ measurements. The usage of different in-situ dataset impacts the value of the coefficients, identifying multiple algorithms. The best algorithm is chosen based on a classification of water types through clustering and subsequent weighted error estimation [30, 35].

### 2.3.2 Satellite euphotic depth

The euphotic zone refers to the uppermost layer of the water column where there is sufficient sunlight for phytoplankton to carry out photosynthesis at a rate greater than the rate of respiration. The depth and extent of the euphotic zone can vary significantly and depend on factors such as the amount of light attenuation in the water column, which is influenced by its turbidity and other optical properties. In this study, the euphotic depth is directly derived from the diffuse attenuation coefficient for downwelling irradiance  $K_d(\lambda)$  measured at 490 nm, from the multi-sensor Level 3 CMEMS GlobColour product [36]. The main difference of this merged product with the previously-introduced OC-CCI is that CMEMS GlobColour Chl-*a* concentration is computed for each sensor using its specific characteristics, and subsequently resampled and merged [37].  $K_d(490 \text{ nm})$  is directly derived from Chl-*a* concentration by the following relation:

$$K_d(490 \text{ nm}) = A + B \cdot (\text{Chl } a)^C \quad (2.7)$$

where the coefficients  $A$ ,  $B$ , and  $C$  are obtained by regression analysis [38, 39]. An estimation of the thickness of the euphotic zone is then given by an empirical relationship with  $K_d(490 \text{ nm})$  [40]:

$$\Delta Z = \frac{4.6}{K_d(490 \text{ nm})} \quad (2.8)$$

## 2.4 Sea bottom topography

The average depth of the oceans is about 4 kilometers, but it is highly variable and can range from shallow coastal areas to extreme depths in ocean trenches. This variability is due to numerous geological processes of different time scales.

Bathymetry is the science that studies underwater depths, particularly the measurement and characterization of the ocean floor. A comprehensive mapping of the global seabed is a challenge, since most of the oceanic regions, especially the deepest ones, remain unexplored. Remote sensing emerges as a very important tool in this field, allowing to gather data from these largely inaccessible areas.

The topography of the seabed strongly influences marine ecosystems and oceanic circulation patterns. A comprehensive global mapping of the seafloor can help distinguishing different oceanic domains, and understanding where the carbon can be efficiently exported and buried to depth.

In this study, the General Bathymetric Chart of the Oceans 2022 (GEBCO) product is employed to distinguish open ocean and continental shelf regions [41]. GEBCO is created from a combination of several data gathered through ship-based surveying and airborne and satellite altimetry.

Ship-based surveys is the primary source of high-resolution bathymetric data: research vessels equipped with specialized sonar systems and echo sounders traverse the world's oceans, emitting sound waves that bounce off the seafloor and return depth measurements. This type of measurements, however, covers only specific regions, and it is expensive and slow [42]. Satellite measurements involve the use of on-board radar altimeters, active sensors that emit microwave pulses in the radar frequency domain ( $\lambda \sim 1$  m) and collect the reflected radiation, measuring the round travel time of the signal. Variations in gravity anomalies can be highly correlated with seafloor topography. Satellite altimetry provides global coverage at a fraction of costs and time required by ship-based surveying, but it lacks of spatial resolution [42].

The combination of this two types of data can provide a global, high-resolution product. GEBCO 2022 product is provided with 15 arcseconds resolution, corresponding to about 450 meters at the Equator.

# Chapter 3

## Methods

Chapter 2 described the characteristics and acquisition methods of the employed satellite products of sea surface temperature, ocean color and sea bottom topography. This chapter presents the unsupervised learning framework with which these images are analysed. Section 3.1 introduces the concepts of machine learning and its applications in oceanography, with a focus on unsupervised clustering tasks and K-means. Section 3.2 gives an overview of the literature on image analysis techniques for the detection of upwelling regions and filaments, and it presents the employed clustering technique. Section 3.3 presents how the biomass and carbon contents of given clusters are computed from satellite ocean color data.

### 3.1 Introduction to Machine Learning

Machine learning represents a broad field of algorithms that automatically detect patterns (thus *learn*) from a great number of observations, and then use these patterns to make predictions on unseen data. Three types of machine learning can be identified [43]:

- *Supervised learning* tries to approximate a complex, unknown mapping  $f$  from inputs  $\mathbf{x}$  to outputs  $y$  with a function  $\tilde{f}_\theta(\mathbf{x}_i)$ , dependent on a set of parameters  $\theta$ , given a previously-labelled *training* set  $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ . Supervised learning methods include classification and regression tasks, depending on the categorical or real-valued nature of the label  $y$ .
- *Unsupervised learning* is used to discover patterns in unlabelled data that contain only inputs,  $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$ . More specifically, the learned function  $\tilde{f}_\theta(\mathbf{x}_i)$  is an approximation of the function  $f(\mathbf{x}_i)$  that generated the data. Common unsupervised learning tasks include dimensionality reduction, generative modeling and clustering.

- *Reinforcement learning*, less commonly used, is a class of algorithms that learn how to act or behave when given occasional rewards or punishment signals.

Every machine learning algorithm deals with an optimization problem. The set of parameters  $\theta$  that defines the approximation function  $\tilde{f}_\theta(\mathbf{x}_i)$  is determined by minimizing an *objective function*  $\mathcal{L}$ . The nature of this objective function varies according to the specific task: in supervised learning, it quantifies the discrepancy between predictions and the original labels, denoted as  $\mathcal{L}(y, \{\mathbf{x}, \theta\})$ ; in unsupervised learning tasks, this objective function  $\mathcal{L}(\{\mathbf{x}, \theta\})$  describes the similarity with the original data (for generative modeling and dimensionality reduction) or measures the coherence of clusters (for clustering tasks). Loss functions are generally dependent on many parameters and can have very complex shapes, making the determination of the global minimum a significant challenge.

Due to the substantial amount of geophysical data continuously provided by model outputs, reanalysis, and satellite products, machine learning frameworks are being investigated in Earth System science across various fields. In ocean and atmospheric modeling, supervised learning is employed for the parameterization of subgrid processes. This allows for a faster and more accurate inclusion of turbulent processes, often necessitating strong approximations to manage computational intensity.

Machine learning is also widely applied in remote sensing, both in preprocessing steps to enhance consistency with in situ data and operationally for monitoring and analyzing the evolution of ocean regions. The training of artificial neural networks on geometrical contour-based models for detecting and classifying eddies from sea surface height satellite data has shown promising results, significantly reducing computation and resource time [44].

In line with the objectives of this work, a similar approach could be considered for the detection and classification of submesoscale filaments. However, the intricate dynamics and variability in their shapes present a challenge for developing a systematic detection framework based on traditional geometry-based methods. Without access to ground truth labeled data, the use of supervised learning becomes impractical. The potential of unsupervised learning in this task hinges on the ability to discover patterns in the data without the need for labels.

### 3.1.1 Unsupervised learning and K-means clustering

Unsupervised learning is broadly used in data analysis to characterize the distribution of datasets and discovering structures in data without needing pre-existing labels. This makes unsupervised learning more widely applicable than supervised learning, since it does not require a human expert to manually label the data;

however, this type of problem is more difficult and subtle as it is not well-defined and there is no background truth that can be used to verify the performances of the method [43, 45].

Clustering is a field of unsupervised learning tasks that characterizes algorithms that group unlabeled data into clusters according to their similarity or distance measure. Different clustering types are defined, based on the distribution of clusters, the geometry of the data, the cluster size distribution, the dimensionality of data, and their computational efficiency [46]:

- *Partitional* methods, such as *K-means* [47], relocate data points by moving them from one cluster to another, given an initial starting point. These methods require the definition of an initial number of clusters, whose distribution is optimized by minimizing a distance-based loss.
- *Hierarchical* methods [48] construct clusters by recursively partitioning data points. *Agglomerative* hierarchical clustering organizes points in a bottom-up fashion, where each object represents a cluster of its own and then clusters are successively merged. The opposite approach is adopted by *divisive* hierarchical clustering, where clusters are obtained by recursively splitting a first single cluster of all objects.
- *Density-based* methods assume that the points that belong to each cluster are drawn from a specific probability distribution, thus allowing to construct clusters of arbitrary shape and dimension. Furthermore, this type of approach does not require specifying a number of clusters for partitioning. The most famous clustering algorithm belonging to this family is DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [49], which finds clusters by searching the neighborhood of each data point in a pre-specified radius and checking if it contains more than a defined number of objects.
- *Model-based* methods optimize the fit between the dataset and some mathematical model or distribution. *Gaussian Mixture* models [50] belong to this family: this type of clustering method assumes that all data points are generated from a mixture of a finite number of Gaussian distributions with unknown mean and variance.
- *Soft clustering* methods, such as the *Fuzzy C-means* [51], assign each data point a membership value to all the clusters rather than assuming that every object belongs to one and only cluster.

K-means is one of the simplest and most commonly used clustering algorithms because of its simple interpretability, implementation, and better computational



performances. It is a partitioning procedure that starts with an initial set of cluster centers (*centroids*), the number of which must be specified in advance, that are iteratively updated based on an optimization problem.

Given a dataset of  $N$  samples  $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$  and a pre-defined positive integer  $K$ , K-means partitions  $\mathcal{D}$  into  $K$  disjoint clusters  $\{C_k\}_{k=1}^K$ . Each cluster is defined by a centroid  $\mu_k$ , which is initially randomly initialized and subsequently computed as the mean of the samples belonging to that cluster in subsequent iterations:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{x}_n \quad (3.1)$$

where  $N_k$  is the cardinality of the cluster  $C_k$ .

Each observation  $\mathbf{x}_i$  is assigned to the nearest centroid, determined by Euclidean distance. Consequently, a cluster  $C_k$  at iteration  $t$  is defined as:

$$C_k^{(t)} = \{\mathbf{x}_p : \|\mathbf{x}_p - \mu_k\|^2 \leq \|\mathbf{x}_p - \mu_j\|^2 \forall j, 1 \leq j \leq K\} \quad (3.2)$$

The centroids are then recomputed by applying Equation 3.1 to the updated cluster points. The algorithm iterates until convergence, which is achieved when either the assignments no longer change or when the difference between centroids determined in two consecutive iterations is smaller than a given threshold.

This procedure is equivalent to minimizing the *within-cluster sum-of-squares* (or *inertia*):

$$\mathcal{L}(\mathbf{x}, \{\mu_k\}) = \sum_{k=1}^K \sum_{n=1}^{N_k} \|\mathbf{x}_n - \mu_k\|^2 \quad (3.3)$$

The primary drawback of the K-means algorithm is its sensitivity to the randomly-determined initial partition, which can lead to the algorithm converging to a local minimum in the optimization function  $\mathcal{L}$ . To mitigate its sensitivity to initial conditions, K-means is typically run multiple times under the same conditions and the solution with the smallest final inertia value is chosen.

In addition, K-means is sensitive to noisy data and outliers, and it is more suitable for datasets having isotropic, spherical, and well-separated clusters, which are often challenging to find in real-world problems.

The number of clusters is typically determined through a preliminary study of the algorithm's performances. The inertia, often displayed as a function of  $K$ , can be visualized using the so-called *elbow plot*, and the optimal value of  $K$  is selected as the point where the plot exhibits an inflection point [47].

## 3.2 Coastal ocean image analysis

This work presents an image analysis algorithm based on clustering of SST and Chl-*a* satellite products that is capable of identifying automatically submesoscale filaments. These structures transport cold, highly-productive coastal waters into the warmer and nutrient-depleted open ocean in the form of thin and elongated structures, and are clearly visible in both SST and Chl-*a* satellite data products. As mentioned earlier, submesoscale filaments emerge from the instability of coastal currents, resulting in highly variable shapes in both time and space. Due to their turbulent nature, modeling these structures and characterizing them using traditional equation-based frameworks is a challenging task.

### 3.2.1 State of the art

Several methods have been developed for the automatic detection of cold, chlorophyll rich upwelling waters from satellite products or model outputs, given their importance in assessing coastal ocean productivity and their implications for the fishing industry. Some of these methods are based on classical image segmentation tools. Cayula et al. (1992) and Nieto et al. (2012) developed a filament and front segmentation tool that detects edges based on gradient values in sea surface temperature satellite images [52, 53]. Artal et al. (2019) combined thresholding and edge detection for the systematic identification of filaments in the Chile upwelling system using numerical model outputs of sea surface temperature and currents [54]. The primary aim of these methods is the automatic detection of upwelling filaments and the study of their geographical distribution rather than precise delimitation of the area occupied by these structures. Marcello et al. (2002) proposed a segmentation algorithm for filament detection based on coarse segmentation through histogram-based thresholding, followed by a fine-detailed growing process in SST satellite images [55]. However, these image processing-based techniques are strongly affected by data quality and cloud cover, requiring several complex and time-consuming preprocessing steps that make them less suitable for long time-series studies.

Clustering is widely employed in this type of image segmentation tasks for its adaptability to every input. Clusters are generated depending only on the distribution of the input data, therefore not taking the position of the pixels into account. This is both an advantage, as much less preprocessing of cloud-covered points is required, but it lacks geographical information that could be important in localizing shelf waters.

Clustering-based image segmentation methods have been employed for the detection of upwelling regions and filaments, based on the distribution of SST (or Chl-*a*) satellite data. Nascimento et al. (2012) have developed an upwelling detec-

tion method based on fuzzy clustering applied to sea surface temperature satellite images on the Portuguese coast [56]. The same method has been applied in the Moroccan coast region by El Abidi et al. (2021), by using information on both SST and Chl *a* data [57]. Geographical information on pixel position is added in further work by Nascimento et al. (2015) and Nascimento et al. (2020), where clusters are created by choosing a first seed (the coldest pixel in the image) and expanding it iteratively [58, 59]. The common feature of these algorithms is that the clustering is performed on one image at a time, thus losing the *learning* feature of this type of unsupervised tasks. Furthermore, this increases significantly the number of parameters to be tuned, having often very abstract meanings that are difficult to interpret.

Upwelling detection through a clustering algorithm that is fitted on multiple images at a time has been proposed by Hammond et al. (2022), in the Western Indian Ocean [60]. K-means is fitted on a random subset of a time series of Level 4 satellite images of SST and Chl-*a* between 2007 and 2020, from May to September, and it is capable of effectively delimitate shelf waters.

### 3.2.2 Proposed clustering-based algorithm

The technique employed in this work was inspired by the paper by Hammond et al. (2022) [60], presented in the previous subsection. A similar framework based on K-means clustering is developed for the analysis of an 18-year time series (2003-2021) of satellite images from MODIS sea surface temperature (SST) and OC-CCI chlorophyll-*a* (Chl-*a*) for three different world coastal ocean regions. A further description of the data products, their acquisition, and processing is given in Chapter 2.

The aim of this work differs from the original paper, as it focuses on assessing the lateral export from shelf regions towards the open ocean. Coastal regions are filtered based on the associated water column depth provided by GEBCO bathymetry, and clustering is performed on time and space anomalies rather than the raw product. Furthermore, the employed satellite products are of the Level 3 type, as it has been observed that the interpolation of Level 4 processing can smooth data and filter submesoscale features captured by satellite images. One drawback of Level 3 data is that cloud cover is not corrected, significantly reducing the availability of sea surface data, especially in the optical wavelength, which affects the quality of satellite images. To improve the number of available pixels in each image, 8-day time averages are computed. This reduces the number of images in the time series but enhances their quality while still capturing the majority of submesoscale dynamical features.

The unsupervised learning framework is applied to three different world regions (displayed in Figure 3.1): two of them are Eastern Boundary Upwelling regions



Figure 3.1: Position of the three selected regions for this study.

(EBUs), California and Mauritania, where productivity and carbon export are naturally enhanced at the coast [61, 62]. The South China Sea is a semi-enclosed sea subject to high seasonal variability, and the high occurrence of this type of filamentary dynamics has been assessed previously [63].

The steps followed by the algorithm can be grouped into two units, as shown in Figure 3.2:

1. The *preprocessing* unit is developed to mask off-shelf data and compute time and space anomalies.
2. The *clustering* unit subsets and normalizes the data, fits the K-means algorithm and returns a cluster partitioning which is then employed on the remaining time series.

In the next subsections, a more specific overview of the algorithm and its steps is presented. This algorithm has been entirely developed in the Python programming language, utilizing the NumPy [64] and Scikit-learn libraries [65]. Data manipulation and analysis are performed through Pandas [66] and xarray [67] libraries, and visualization is done using Matplotlib [68] and Cartopy [69].

### Preprocessing

This initial unit is required to preprocess the 8-day average time series of Chl-*a* and SST. For mathematical purposes, from now on these two datasets will both be referred to as  $\phi_{ij}(t_{yd})$ , where the indices  $i = 1, \dots, N$  and  $j = 1, \dots, M$  identify

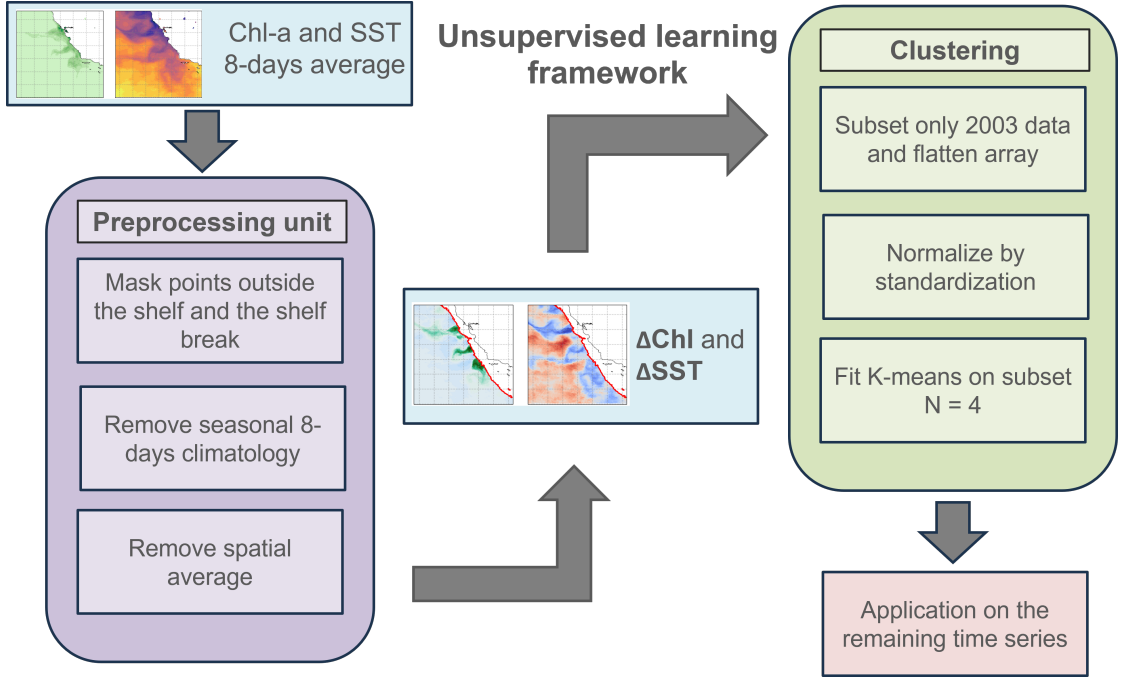


Figure 3.2: Workflow of the unsupervised learning framework.

the longitude and latitude of each pixel. The three chosen regions are squares of  $10^\circ \times 10^\circ$ , that correspond to  $N = M = 240$ . The time coordinate  $t_{yd}$  is indexed by the year  $y = 1, \dots, 18$  and  $d = 1, \dots, 46$  representing the day of the year that identifies each 8-day window.

The two time series are preprocessed as follows:

1. The first step is the selection of open ocean regions in the image. Pixels are masked based on their associated depth retrieved from GEBCO bathymetry, that has been interpolated to the satellite products resolution of 4 kilometers. Bathymetry profiles are highly complex and vary depending on the region, as shown in Figure 3.3. Different reference depths have been visually selected, depending on the region: 2500 m for California and 2000 m for Mauritania and the South China Sea. Every pixel associated with shallower depths is discarded from the image.
2. After identifying the off-shelf areas in each region, the seasonal 8-days climatology is removed from the images. The seasonal signal is defined as:

$$\bar{\phi}_{ij}(t_d) = \frac{1}{T} \sum_{y=1}^T \phi_{ij}(t_{yd}) \quad (3.4)$$

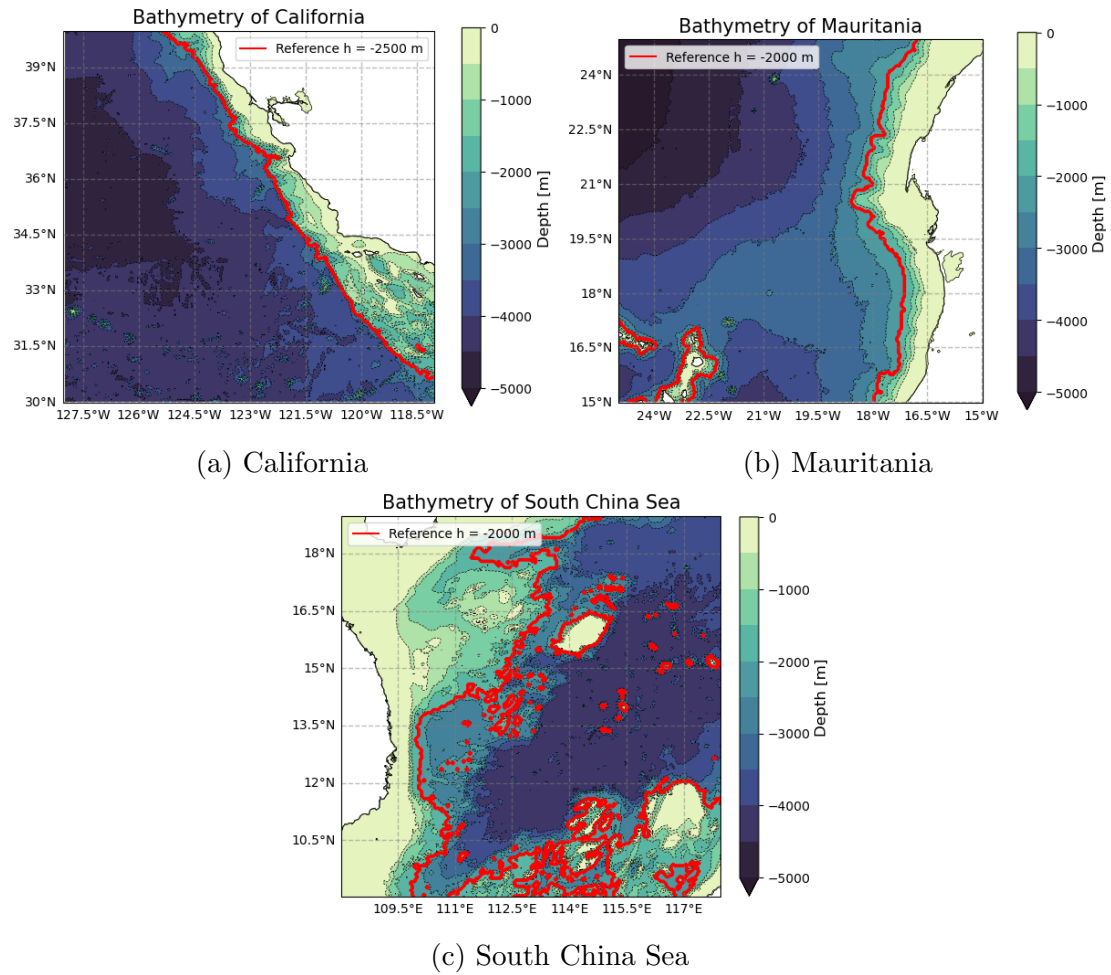


Figure 3.3: Bathymetry profiles of the three regions included in the study. The red line highlights the chosen reference depth to distinguish open ocean regions from continental shelf.

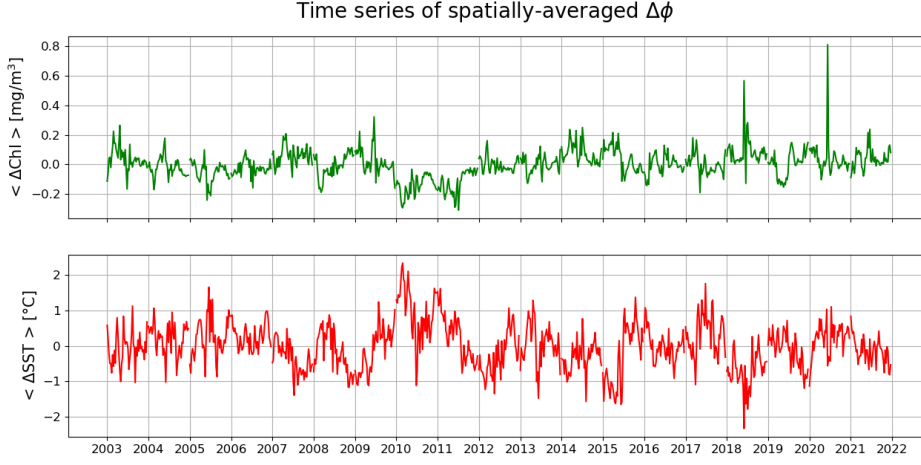


Figure 3.4: Example of time series of the spatially-averaged  $\Delta\phi_{ij}(t_{yd})$  in Mauritania region.

An anomaly  $\Delta\phi_{ij}(t_{yd})$  is then obtained as:

$$\Delta\phi_{ij}(t_{yd}) = \phi_{ij}(t_{yd}) - \bar{\phi}_{ij}(t_d) \quad (3.5)$$

3. The spatially-averaged anomaly of each image in the time series is defined as follows:

$$\langle \Delta\phi(t_{yd}) \rangle = \frac{1}{N} \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M \Delta\phi_{ij}(t_{yd}) \quad (3.6)$$

As it can be seen from Figure 3.4, the average anomaly time series still presents a multi-year variability that is related to climatological heating or cooling processes that also affect productivity. Since the turbulent dynamical features that characterize submesoscale processes are assumed to be stochastic and not having any connection with the seasonal cycle, the final anomaly is computed by also removing its spatial average [70]:

$$\Delta\phi'_{ij}(t_{yd}) = \Delta\phi_{ij}(t_{yd}) - \langle \Delta\phi(t_{yd}) \rangle \quad (3.7)$$

## Clustering

The preprocessing phase produces a  $\Delta\phi'_{ij}(t_{yd})$  anomaly product for SST and Chl-*a* that highlights turbulent patterns. K-means clustering is applied to a subset of

<b>Region</b>	<b>Chl-<i>a</i> [%]</b>	<b>SST [%]</b>
California	$85 \pm 22$	$94 \pm 15$
Mauritania	$81 \pm 23$	$88 \pm 18$
South China Sea	$69 \pm 29$	$84 \pm 24$

Table 3.1: Average percentage of available points of off-shelf satellite data. The error is computed as one standard deviation over the time series.

these data, while the remaining time series is used to verify its performance and conduct further studies. This particular decision is influenced by the fact that it yields results nearly identical to using the entire dataset, while significantly reducing the computational time required for the K-means fitting process.

To maintain the continuity of the time series for subsequent analyses, this subset is chosen to be the data belonging to the year 2003 and transformed to a tabular dataset of two features, Chl-*a* and SST anomalies, where each row corresponds to the values of a specific pixel located at position  $(i, j)$  on a given day  $d$  within the year 2003. The time and space information will be disregarded by the algorithm, thus focusing solely on the data’s inherent characteristics.

Any pair with at least a null value, which represents cloud-covered pixels, is removed. A more in-depth analysis of data availability, as presented in Table 3.1, indicates that cloud cover has a greater impact on the chlorophyll-*a* product compared to sea surface temperature (SST). It is noteworthy that the South China Sea is the region most affected by cloud cover among the studied areas, with less than 70% of data available on average. The high standard deviations observed are primarily a result of the inherent seasonality of cloud cover in the studied regions. Before applying the K-means algorithm, the extreme outliers outside the 0.01 and 99.9% quantiles are eliminated and the two feature columns are normalized by standardization, centering them around zero and scaling them to have a standard deviation of one. This standardization process is a common procedure preceding the application of any machine learning algorithm, as it ensures uniformity and comparability of the features.

The determination of the optimal number of clusters  $K$  is made through a visual analysis of the elbow plot, as illustrated by an example in Figure 3.5. This type of approach yielded an optimal  $K$  equal to 4 clusters in all the three considered regions. K-means has been fitted to the subset with 10 random initializations in order to reduce its sensitivity to initial conditions, a maximum number of iterations equal to 300, and a tolerance of  $10^{-4}$  as the minimum difference in the cluster centers of two consecutive iterations to declare convergence.

K-means determines a partitioning of the Chl-*a* and SST space that can be used



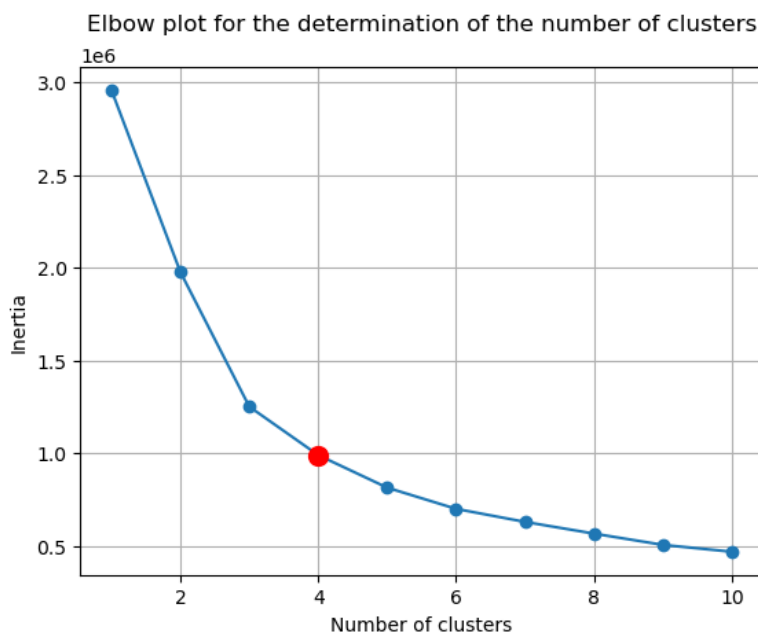


Figure 3.5: Example of elbow plot (Mauritania region). The red dot shows the optimal number of clusters, equal to 4. This optimal value is found also in California and South China Sea.

for the analysis of single images. Each pair of Chl-*a* and SST images for a given day, year and region is fed to the algorithm, that preprocesses it computing the time and space anomalies, reduces it into the tabular form and returns the cluster labels for each pixel.

The analysis of the clusters and their physical interpretations are presented in the next chapter (Section 4.1).

### 3.3 Biomass content calculation

Chlorophyll primarily accumulates in the uppermost layer of the ocean, where sufficient light for photosynthesis is available. Satellite remote sensing products provide surface Chl-*a* distributions as a concentration, typically measured in  $\text{mg m}^{-3}$ . To estimate the total mass of chlorophyll in the surface layer using satellite data, each pixel's chlorophyll concentration has to be multiplied by a volume.

The image analysis algorithm introduced earlier assigns each Chl-*a* and SST pair to one of the four clusters  $C_k$ , each representing a different dynamical regime. At a given time  $t_{yd}$ , the mass content of chlorophyll in a specific cluster  $C_k$  is calculated as:

$$M_{Chl}(C_k, t_{yd}) = \sum_{i,j \in C_k} Chl_{ij}(t_{yd}) \cdot \Delta Z_{ij}(C_k, t_{yd}) \cdot S_{pixel} \quad (3.8)$$

where  $Chl_{ij}(t_{yd})$  represent the satellite Chl-*a* concentration of a pixel in position  $(i, j)$ ,  $\Delta Z_{ij}(t_{yd})$  is the associated euphotic depth (derived from the diffuse attenuation coefficient with Equation 2.8), and  $S_{pixel}$  is the area corresponding to the grid resolution, in this case 16 km<sup>2</sup>.

The carbon content of each cluster  $C_k$  can be estimated from the Chl-*a* mass content by using a carbon-to-chlorophyll ratio  $R$ :

$$M_C(C_k, t_{yd}) = R \cdot M_{Chl}(C_k, t_{yd}) \quad (3.9)$$

It is important to note that the value of  $R$  may vary based on factors such as light and nutrient availability and phytoplankton composition, and it can be specific to different regions [71]. For this initial estimation, a conservative constant value of  $R = 50$  has been chosen.

A carbon content of per unit of area can also be computed. The area occupied by the cluster  $C_k$  at each time  $t_{yd}$  is determined as:

$$A_k(t_{yd}) = \sum_{i,j \in C_k} S_{pixel} = N_k(t_{yd}) \cdot S_{pixel} \quad (3.10)$$

where  $N_k(t_{yd})$  is the cardinality of the cluster. Finally, the carbon mass content per unit of area for each image is computed as:

$$MA_C(C_k, t_{yd}) = \frac{M_C(C_k, t_{yd})}{A_k(t_{yd})}. \quad (3.11)$$

The results of these estimations are presented in the next chapter, more specifically in Section 4.2.

# Chapter 4

## Results and discussion

The previous chapter presented the K-means clustering-based partitioning of off-shelf SST and Chl-*a* images. This chapter shows the results of the application of these methods to the three considered regions. Section 4.1 presents the clustering analysis on the fitting subset, and its overall performance in the remaining time series. One of the clusters identifies the shelf-water streamers presented in previous chapters. Section 4.2 analyses the obtained estimations of the lateral export of biomass and carbon by streamers in the three regions.

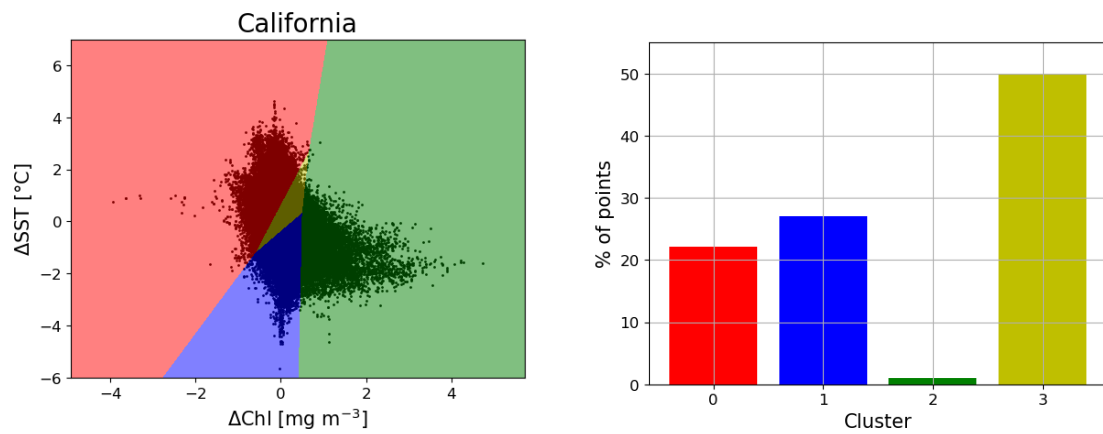
### 4.1 K-means fitting

K-means is fitted on the subset of data belonging to the year 2003, in order to guarantee continuity of the time series for further studies, and its performances are evaluated on the remaining time series for the estimation of biomass and carbon content.

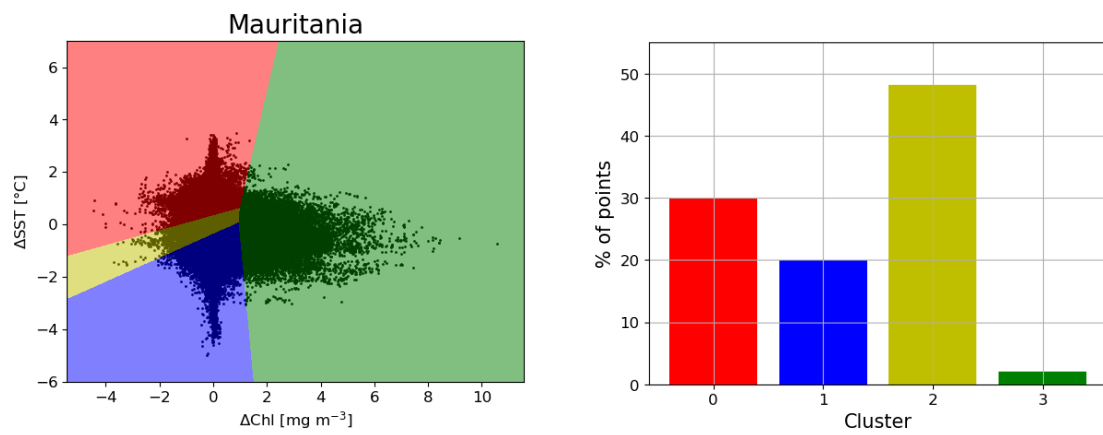
#### 4.1.1 Results on 2003 subset

The representation of the 2003 fitting dataset within the SST and Chl-*a* anomalies space and its segmentation into clusters for the three distinct regions is shown in Figure 4.1, along with the proportion of each cluster's population that denotes the percentage of data points assigned to each region.

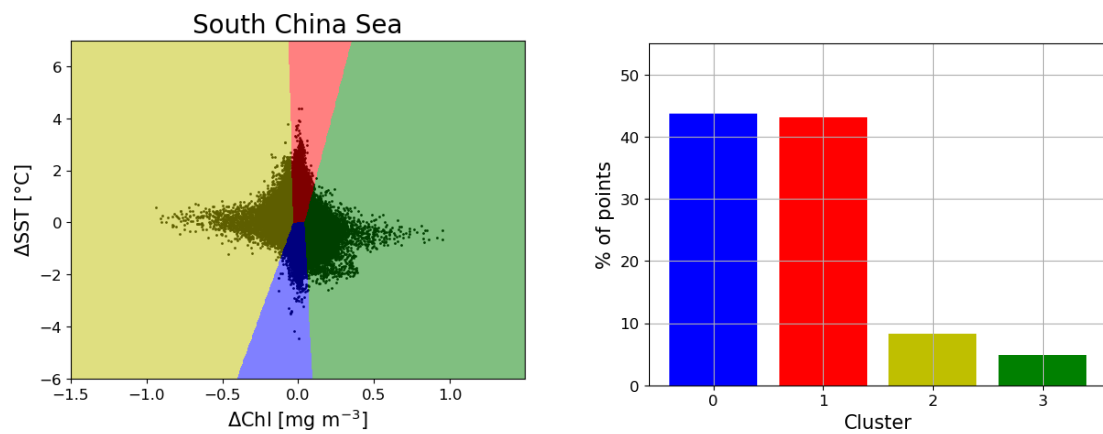
The shape of all three distributions does not lead to the straightforward identification of clearly separated clusters. It is important to note that K-means is designed for the identification of isotropic, spherical, and equivalently sized clusters: in this context, it is rather used to partition such distributions through automatically-determined and more intricate linear thresholds that would be challenging to discern manually. These threshold values differ by region, yet they exhibit common



(a) California



(b) Mauritania



(c) South China Sea

Figure 4.1: Clustering results on the 2003 subset for California (top panel), Mauritania (middle panel) and South China Sea (bottom panel). The left panels show the distribution of points in Chl-*a* and SST anomaly space in the *x* and *y* axes respectively, and the cluster partitioning found by K-means. The right panels show the percentage population of each cluster, accordingly to their color.

features that warrant further analysis.

California and Mauritania point distributions are remarkably similar, as shown in Figure 4.1a and 4.1b: they share a lengthy and broad tail in the positive Chl-*a* space (greater than about  $0.5 \text{ mg m}^{-3}$  for California and  $1 \text{ mg m}^{-3}$  for Mauritania), all assigned to the same cluster (indicated in green). The majority of points within this space are characterized by negative SST anomalies. Therefore, this green cluster, that is also the least populated (1-2 % of points), comprises pixels that are particularly colder and richer in chlorophyll than their surroundings. The attributes of data points belonging to this cluster, combined with the fact that this study exclusively considers open ocean data, suggest a potential association between this cluster and the presence of an intrusion of cold, highly-productive shelf waters in the respective region.

The remaining clusters segment the space into regions with weakly positive or negative Chl-*a* anomalies. Red and blue regions separate warm and cold anomalies, and have a similar population between 20 and 30%. However, in both California and Mauritania, the red cluster also occupies a portion of space that contains negative SST anomaly regions, making this distinction less clear. The final cluster, in yellow, is the most populous despite its smaller size and characterizes pixels belonging to an intermediate regime, encompassing about 50% of all the data points.

The partitioning of the South China Sea subset mirrors the pattern observed in the other two regions regarding the green, red, and blue clusters (as seen in Figure 4.1c). The separation of the warm and cold anomalies in the red and blue cluster is more distinct than California and Mauritania. The overall chlorophyll anomalies are considerably smaller in this region, as it is not a productive upwelling area, thus the threshold defining the green cluster is much closer to zero than the other two counterparts. The green cluster, encompassing cold, high-chlorophyll pixels, exhibits a higher population compared to California and Mauritania regions, comprising nearly 5% of the data points. The yellow cluster in this context has a different meaning, as it is one of the least populated clusters, accounting for approximately 8% of the data points. This cluster identifies a region that opposes the dynamical regime identified by the green cluster, as the distribution exhibits a symmetrical tail towards warm and low-chlorophyll anomalies that is not present in the two other upwelling regions.

Examples of the clustering performances on singles images are proposed in the next subsection.

### 4.1.2 Application on 2004-2021 time series

The fitted K-means clustering method has been applied to the subsequent time series starting from 2004. In Figure 4.2, examples of the clustering results on

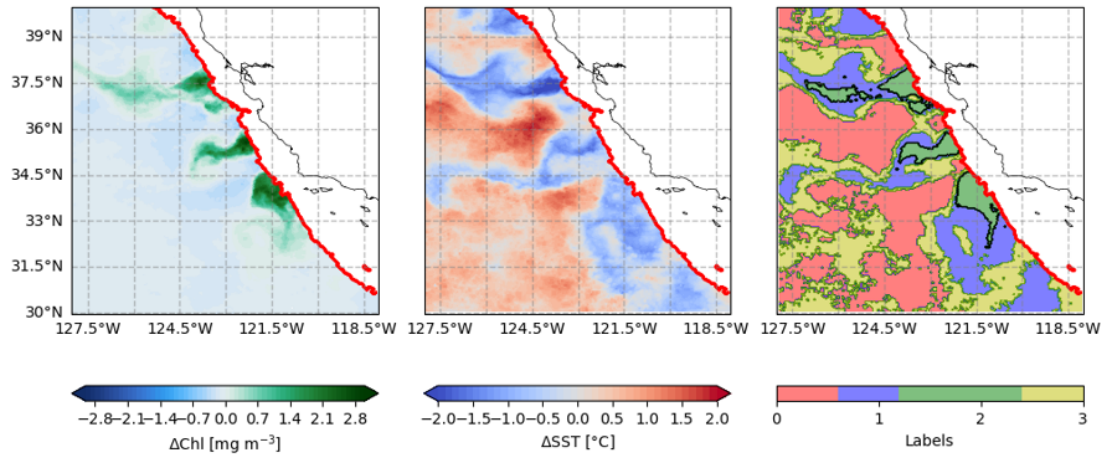
individual images for each respective region are presented.

Examining the Chl-*a* and SST anomalies in the California example, presented in Figure 4.2a, reveals the presence of multiple cold, high-chlorophyll filamentary structures that extend from the reference line towards deeper regions. Three of these structures are visible in both Chl-*a* and SST images, roughly located at 34°N, 35°N, and 37.5°N, with two more visible only in SST above 39°N and 30°N. These regions are separated by warm, low-chlorophyll areas, which likely correspond to anticyclonic mesoscale eddies around which submesoscale filaments form. As anticipated from the previous subsection's cluster analysis, the green cluster groups the three prominent high-chlorophyll filamentary structures, which are either extended or surrounded by the blue cluster, representing cold, low-chlorophyll regions. Warm anomalies are consistently grouped within the red cluster, as expected, and yellow regions are predominantly distributed between cold and warm anomalies, representing an intermediate dynamical regime.

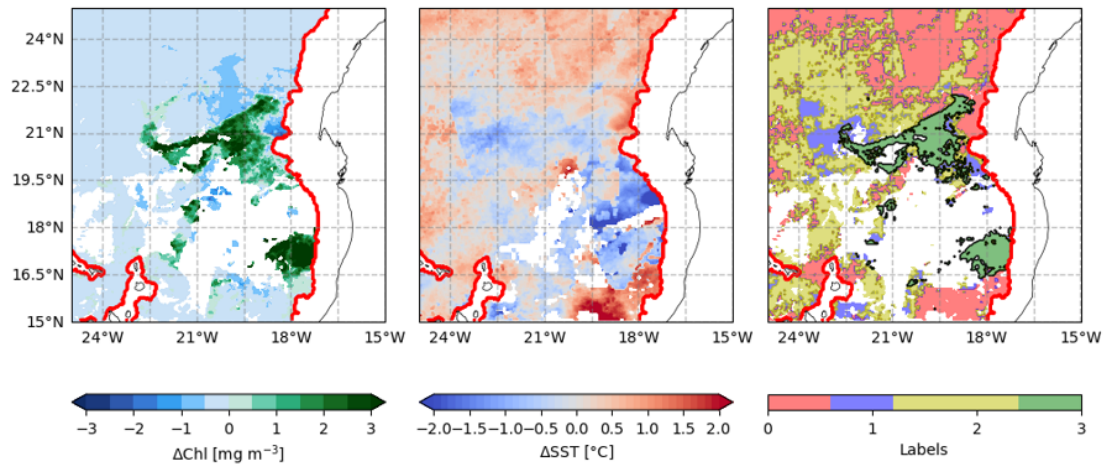
A comparable structure is observable in the Chl-*a* anomaly image of Mauritania (Figure 4.2b), featuring a single broader area characterized by very high anomalies extending between 18 and 21°W. While it is less distinct in the SST anomaly space, this region is mostly associated with colder water. Analogous to the California case, the green cluster groups the extensive high-chlorophyll filament, terminating with a brief extension between 22 and 24°W and 21 and 22.5°N, which is labeled as blue. The yellow cluster labels the majority of points in the image, consistently with the population distribution shown in Figure 4.1, and surrounding the filament region. The red cluster groups warm anomalies.

Similarly to the prior two regions, the South China Sea example (Figure 4.2c) reveals a shorter yet still thin structure of positive Chl-*a* and negative SST anomalies near 12°N and 109.5°E, once again grouped in the green cluster and surrounded by cold water. However, in contrast to California and Mauritania, an interesting feature emerges from the yellow-labelled region between 14 and 19°N, that presents a similar elongated structure propagating from the shelf to deeper regions. As mentioned in the previous subsection, this region groups warm and negative-chlorophyll regions in space, indicating that the yellow cluster possesses a distinct physical interpretation from its other two counterparts, suggesting the presence of warm, low-chlorophyll filaments in the South China Sea.

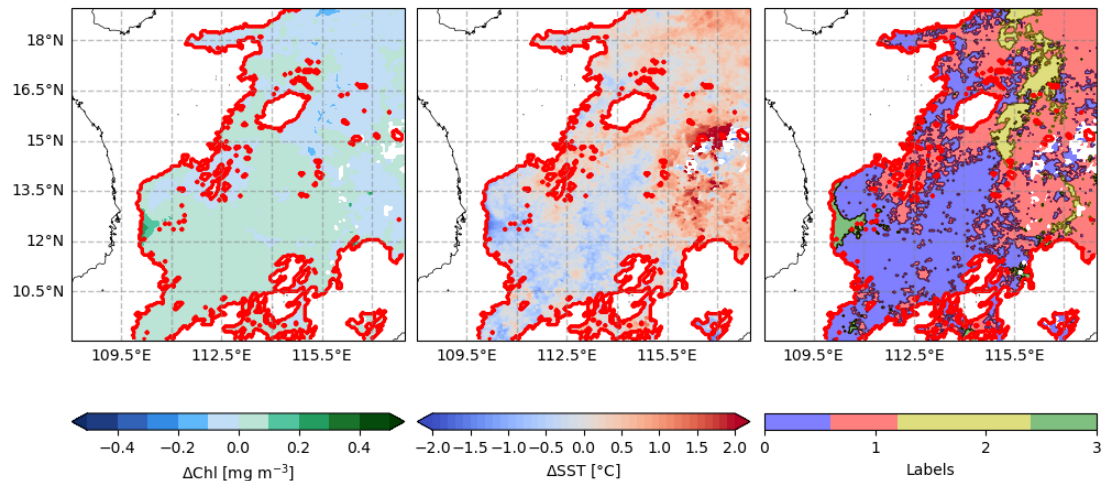
In all the three examples, the green cluster predominantly contains data points with the highest chlorophyll anomalies. The majority of these pixels are situated in areas characterized by negative temperature anomalies, implying higher productivity and lower temperatures than the surrounding areas. When displayed in individual images, these pixels define regions with an elongated shape, extending from the shelf reference line towards the open ocean. This characteristic recurs consistently throughout the entire time series, illustrating the formation, evolution



(a) California, 24 - 31 October 2019



(b) Mauritania, 26 June - 3 July 2018



(c) South China Sea, 17 - 25 May 2005

Figure 4.2: Examples of clustering results on 8-day average images in the time series for California (panel a), Mauritania (panel b) and South China Sea (panel c). The left panel shows the computed Chl-*a* anomalies, the middle panel the SST anomalies and the right panel shows the clusters derived from K-means. The red line delimits the slope reference depth.

and dissipation of these structures over subsequent time steps as shown Figure 4.3. Figure 4.4 presents the frequency of occurrence of green-labelled pixels relative to the total available data. In all three regions, the highest occurrence is observed along the reference depth line. Mauritania exhibits an overall lower frequency, consistently smaller than 15%. California displays a slightly higher percentage, reaching 15% particularly north of 33°N, where the shelf region is narrower. The South China Sea exhibits the highest frequencies (greater than 15%), with non-negligible values even in open ocean regions, reflecting also the greater cluster population described in Figure 4.1c.

Given all these features, it can be inferred that the green cluster systematically represents waters originating from the relatively cold, highly productive shelf regions and transported by submesoscale filaments in the open ocean region. Therefore, from now the green cluster will be referred to as the *filament* (or *streamer*) cluster  $C_{fil}$ .

## 4.2 Lateral transport estimation

The previous section described the cluster partitioning performed using K-means and highlighted a particular cluster associated with streamers responsible for transporting shelf water into open ocean regions. This section is dedicated to estimating the content of biomass and carbon within these structures and the associated lateral export.

### 4.2.1 Chlorophyll content time series

The lateral transport of biomass by streamers can be estimated by quantifying the mass content of chlorophyll within the points marked in their associated cluster. This is calculated using Equation 3.8. For each pixel classified as belonging to the filament cluster ( $C_{fil}$ ), the Chl-*a* concentration from the original 8-day product has been extracted and then multiplied by the pixel's surface resolution and the estimated euphotic depth in that pixel. Figure 4.5 displays the resulting time series of chlorophyll content for each region covering the years from 2004 to 2021. To enhance clarity and remove noise, a Gaussian filter has been applied to the data. In all three regions, the time series exhibit irregularities, featuring the superposition of multiple signals with distinct characteristic periods. These variations include seasonal fluctuations in productivity within the respective oceanic regions and multi-year variabilities linked to factors such as temperature, river discharge, and light availability. Seasonal cloud cover patterns may also influence the data by reducing the number of available pixels in the images, potentially leading to an underestimation of biomass content in each image.



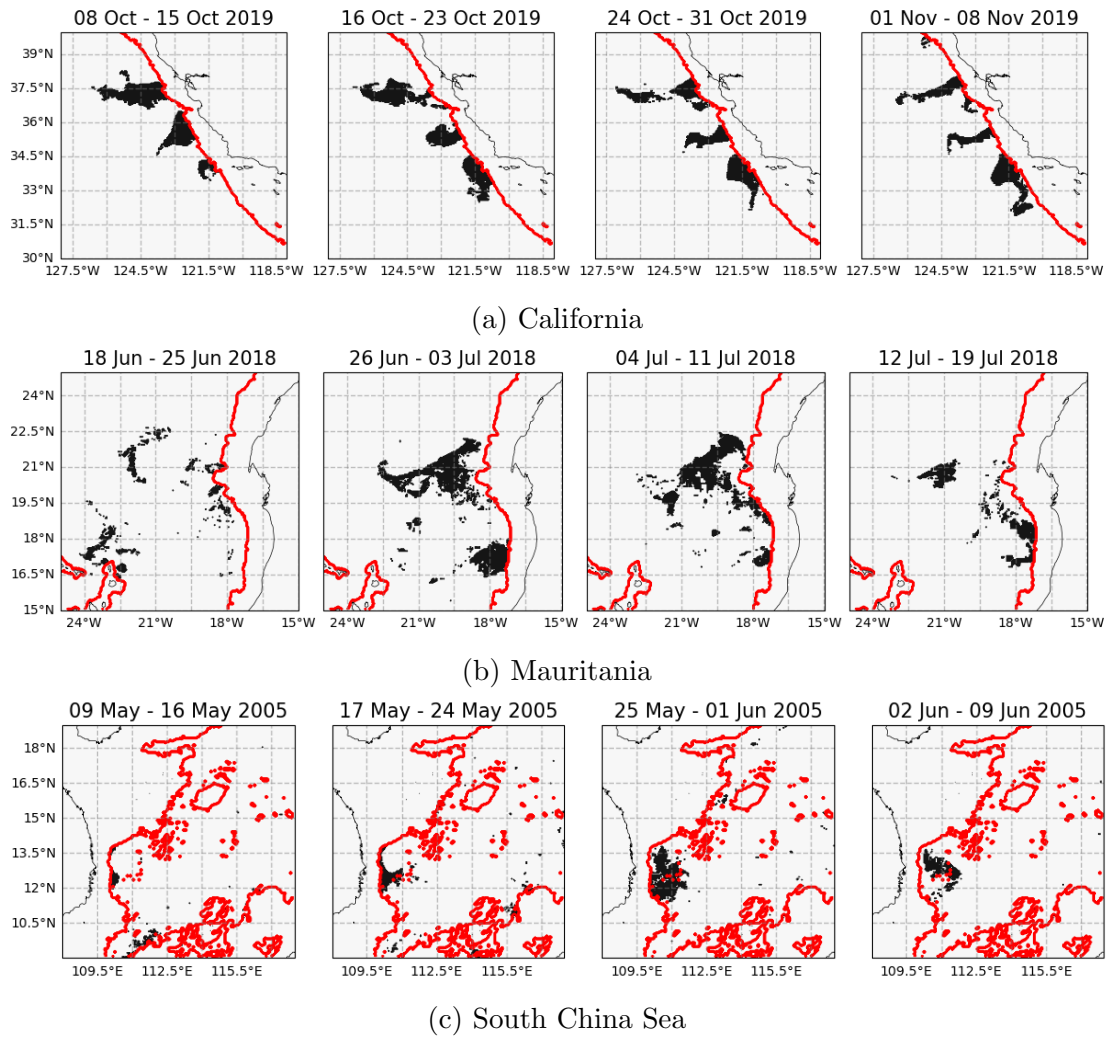


Figure 4.3: Example of the evolution of the green cluster segmentation over subsequent steps for the three regions. The red line delimits the slope reference depth.

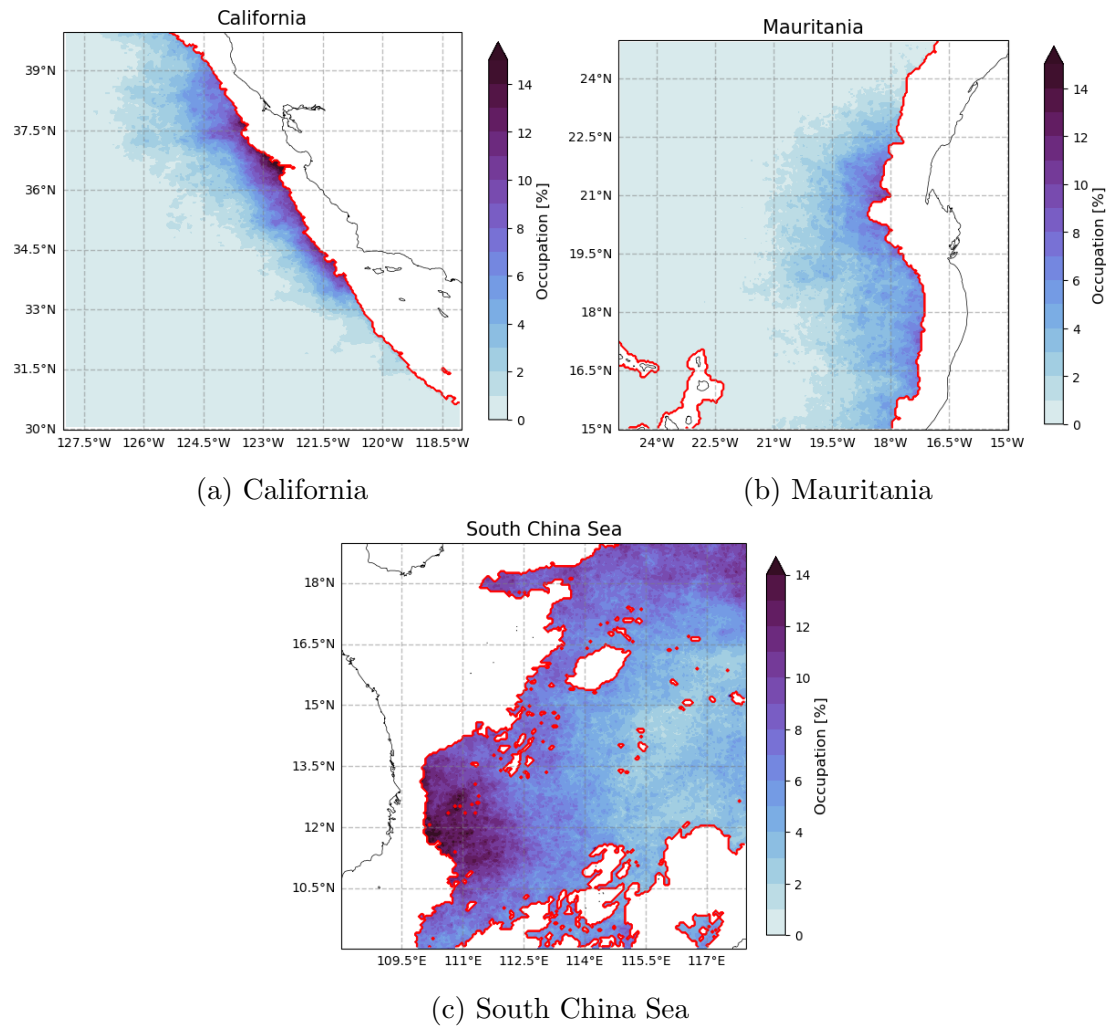
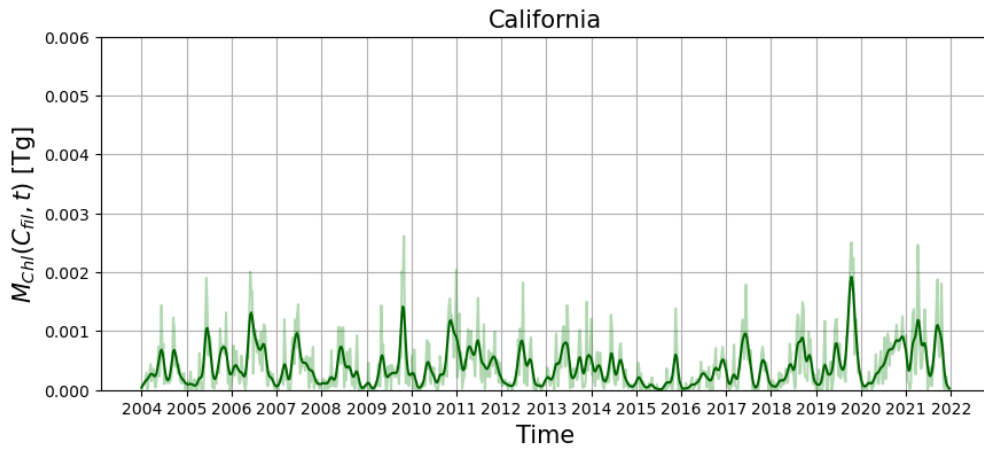
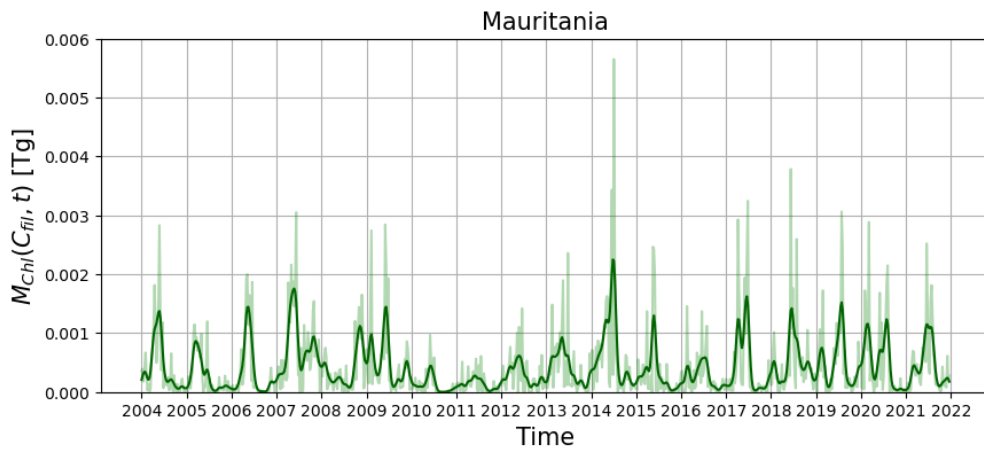


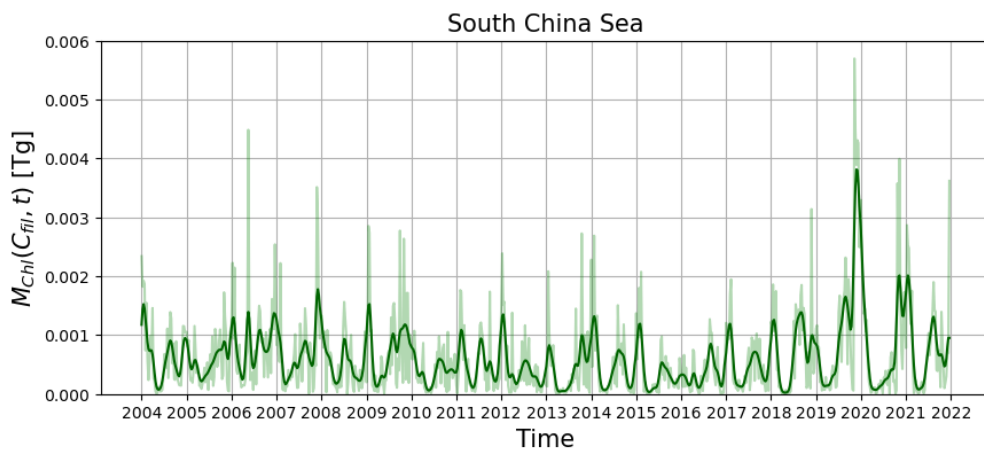
Figure 4.4: Percentage occupation of the green cluster over the available (non-NaN) pixels for the three regions. The red line delimits the slope reference depth.



(a) California



(b) Mauritania



(c) South China Sea

Figure 4.5: Time series (2004-2021) of Chl-*a* filament cluster mass content. A Gaussian filter with standard deviation equal to 2 has been applied to the original time series (in lighter color) to remove noise. Filtered data are colored with a darker shade of green. One teragram (Tg) corresponds to  $10^{12}$  g.

Region	Total $M_C$ [Tg]	Carbon off-shelf export [Tg yr <sup>-1</sup> ]	Carbon off-shelf flux [g m <sup>-2</sup> yr <sup>-1</sup> ]
California	16.0 ± 1.8	0.94 ± 0.11	150 ± 16
Mauritania	16.7 ± 0.5	0.98 ± 0.04	149 ± 5
South China Sea	24.6 ± 1.5	1.45 ± 0.09	65 ± 4

Table 4.1: Carbon content in filaments, export per year and associated flux for the three regions of the study. One teragram (Tg) corresponds to  $10^{12}$  g.

California (Figure 4.5a) shows the lowest amplitudes in the time series, with oscillations that mostly have a peak placed between spring and autumn, when the upwelling conditions are favourable [61]. A weakening of filamentary chlorophyll content is evident between late 2014 and late 2016, coinciding with an anomalous warming of the region and a decrease in productivity documented in 2014 and 2015 [72], which could have extended its impact into the following year.

A similar pattern is observed in the Mauritania region, as depicted in Figure 4.5b. A weakening trend is present between 2010 and 2013, also associated with an anomalous warming and a substantial decrease in primary production that affected the entire Canary Current upwelling system during those years [73].

Conversely, the chlorophyll content variability in the South China Sea, shown in Figure 4.5c, lacks a specific seasonal signal, with peaks occurring throughout the year over the 17-year period. The time series is characterized by high-frequency variability, suggesting that the overall signal may be associated with purely turbulent dynamics. This differs from the two upwelling regions, where the biomass content time series is primarily modulated by the seasonality of primary production.

### 4.2.2 Estimation of carbon export and fluxes

The total carbon content within the streamers cluster,  $M_C(C_{fil}, t_{yd})$ , has been determined using Equation 3.9. Table 4.1 presents the total carbon content for the years 2004-2021 in each region, and the carbon content per year calculated as the total  $M_C$  divided by the number of years in the time series (17). Since streamers transport shelf water into the open ocean, the carbon within these structures can also be interpreted as a laterally exported mass from the shallower regions.

To estimate the error in these calculations, a detailed analysis of the distribution of  $M_C(C_{fil}, t_{yd})$  over the same time span (2004-2021) has been conducted. Figure 4.6 illustrates this distribution using the Mauritania case as an example. This distribution can be fitted with a Weibull distribution, known for its fat-tailed and

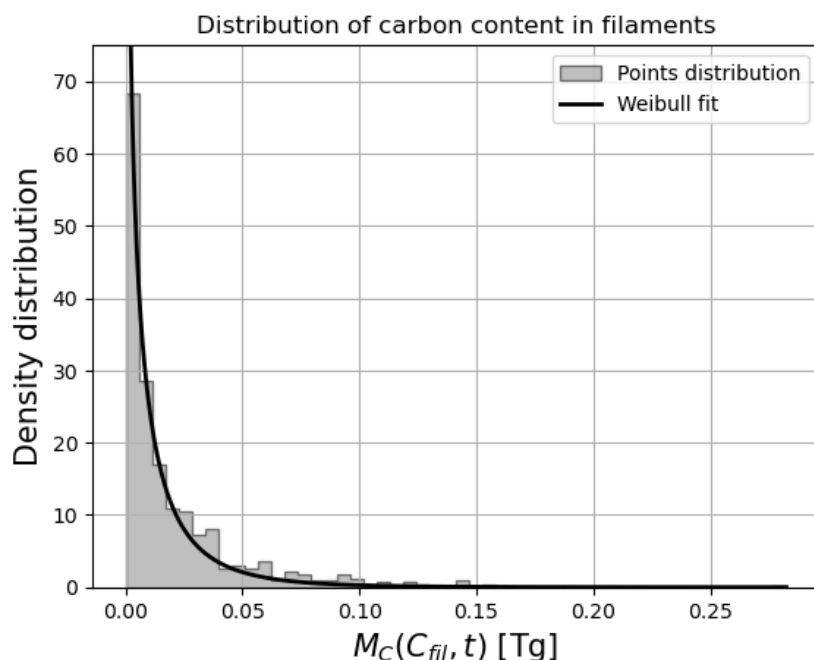


Figure 4.6: Distribution of carbon content over the time series (Mauritania example). One teragram (Tg) corresponds to  $10^{12}$  g.

positively-defined shape. The Weibull distribution is commonly used to describe various variables in atmospheric and oceanic flow fields, such as wind speed and currents, which, in turn, influence the distribution of tracers [74, 75]. The relative error for each region has been estimated as one minus the value of the cumulative Weibull probability distribution calculated at the 95% quantile of the original data. Both California and Mauritania display comparable values in terms of lateral export, approximately  $1 \text{ Tg yr}^{-1}$ . Surprisingly, the South China Sea exhibits a value more than 50% higher than these two highly productive upwelling regions. Several factors could account for this unexpected difference.

The analysis conducted in 2003 (Section 4.1) revealed that the filament cluster in the South China Sea comprises a significantly higher population of points. This phenomenon might be linked to the region's semi-enclosed basin nature, featuring more extensive shelf areas, including the southern part of the square and the majority of the western portion. Consequently, streamers can form in a larger region. In contrast, the two upwelling regions only have the eastern portion of the image as shelf areas.

These export values are comparable to the net ocean uptake of atmospheric carbon occurring in global coastal regions, as discussed in Subection 1.1.2, which is estimated to be approximately  $0.25 \text{ Pg yr}^{-1}$ . Given that this study was conducted

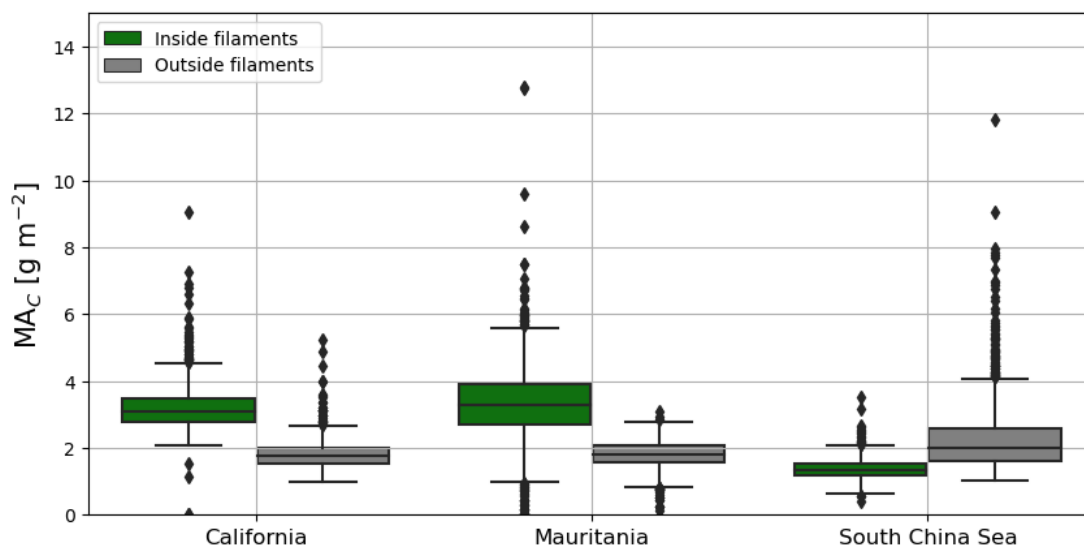


Figure 4.7: Carbon content per area inside and outside the filament cluster for the three regions.

on three  $10^\circ \times 10^\circ$  squares, the potential to extend this estimate to the global coastal ocean is significant.

Given that the streamers' cluster has a smaller population compared to the other clusters, it is insightful to compare the carbon content per unit area occupied,  $MA_C$ , as reported in Equation 3.11. Figure 4.7 shows the distributions of carbon content per unit area inside and outside the filaments. In the cases of California and Mauritania, the streamers exhibit higher values (median of approximately  $3\ g\ m^{-2}$ ) compared to their surroundings (less than  $2\ g\ m^{-2}$ ). Conversely, the South China Sea streamers are characterized by even lower values than the background. This reaffirms that the previously discussed observed increase in chlorophyll transport in this region may primarily generate from the distinct partitioning of the clustering algorithm, which tends to group a higher average number of pixels as streamers. The carbon flux, quantified as mass per unit area and unit time, as reported in Table 4.1, is computed by summing the carbon content per unit area across the entire time series and then dividing by the 17-year duration. Once again, it highlights a remarkable similarity between the two upwelling regions, amounting to  $150\ g\ m^{-2}\ yr^{-1}$ , and a significantly lower flux (less than 50%) for the South China Sea. These fluxes are comparable to the net air-sea fluxes of  $CO_2$  depicted in Figure 1.5, which can reach absolute values of approximately  $30\ g\ m^{-2}\ yr^{-1}$ .

# Chapter 5

## Conclusions and future work

This study has introduced an automated image analysis tool based on K-means clustering, applied to an 18-year time series of spatial and temporal anomalies of chlorophyll-*a* concentration and sea surface temperature Level 3 satellite data. The algorithm is capable of detecting and classifying various dynamical regimes, including shelf-water streamers. These streamers are submesoscale structures formed through stirring and straining by mesoscale eddies, capable of transporting cold, chlorophyll-rich water from the shelf region, where productivity is enhanced, to the open ocean. This type of transport is particularly important because it could increase the vertical fluxes of carbon in deeper regions, efficiently sequestering atmospheric carbon in the ocean depths.

This methodology was tested in three distinct coastal ocean regions worldwide: California and Mauritania, which are productive Eastern boundary upwelling regions, and the South China Sea. The algorithm consistently produced similar results for California and Mauritania, demonstrating its ability to effectively handle analogous dynamical regions. In contrast, the results for the South China Sea displayed notable differences, reflecting the region's unique dynamical characteristics, while still maintaining internal consistency.

The development of an unsupervised learning framework for detecting and classifying off-shelf areas in coastal ocean regions represents pioneering work, with substantial room for improvement. For instance, now that a label map has been generated, opportunities for supervised learning have opened up. Advanced methods can enhance the identification and segmentation of streamers by considering factors such as the aspect ratio of identified regions or pixel positions in the image. Supplying the algorithm with additional information, such as high-resolution sea level anomaly data, could improve the partitioning of the image.

After detecting and classifying areas occupied by streamers, a following study has been conducted to assess their contribution in the lateral transport of biomass from the shelf region to the open ocean.

The examination of chlorophyll content time series for streamers revealed the diverse dynamical regimes in the regions of study. While biomass export in upwelling regions follows the seasonality of primary production, the South China Sea exhibits high-frequency variability attributed to turbulent processes.

The study estimated that the combined yearly lateral export of biomass through streamers in these three regions totals approximately  $3.5 \pm 0.7 \text{ Pg yr}^{-1}$ . The comparison of carbon mass fluxes between the three region shows a remarkable difference: the upwelling regions are characterized by higher fluxes ( $150 \text{ g m}^{-2} \text{ yr}^{-1}$ ) than the South China Sea ( $65 \text{ g m}^{-2} \text{ yr}^{-1}$ ), suggesting that streamers have different areas and carbon content in these regions.

These initial estimates involve certain assumptions, such as a constant carbon-to-chlorophyll ratio, which may not hold universally. A region-specific ratio should be determined through phytoplankton composition analysis, as different species contribute differently to primary production and carbon synthesis. Cloud cover could potentially reduce the number of classified streamer pixels, leading to underestimations.

These results provide significant insights into this type of lateral transport of organic carbon in different dynamical regimes, and the next step in this research is to estimate the global contribution of coastal ocean regions worldwide. The extent to which this mechanism contributes to actual carbon export to depth remains uncertain, as the majority of surface ocean carbon typically undergoes remineralization or is released back into the atmosphere. To comprehensively investigate the correlations between lateral transport and carbon burial to depth, an integrated approach that incorporates data from in situ campaigns to capture vertical fluxes is necessary. This integration would contribute to a deeper understanding of the natural atmospheric carbon removal process and how this mechanism can be preserved, with important implications for climate change mitigation.



# Acknowledgments

I want to express my sincere gratitude to my supervisor, Prof. Nadia Pinardi. She has been an incredible mentor, always there to support and guide me with her contagious passion for science and oceanography, and opening up so many opportunities for me that I never even imagined. She will always be a role model to me.

I would also like to express my gratitude to my co-supervisors, Prof. Silvia Biancinci and Prof. Astrid Bracher, for their invaluable feedbacks and support throughout this research.

A portion of this thesis was conducted during a three-month guest stay at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven. I am thankful to the entire PhytoOptics group and Physical Oceanography section for warmly welcoming me. I would also like to extend my thanks to Dr. Wilken-Jon von Appen, Prof. Morten Iversen, and Dr. Nasrollah Moradi for fruitful discussions. These three months provided me with a deeper understanding of what it means to be a scientist, and I am truly appreciative of all that I have learned during this experience.

Ringrazio di cuore i miei genitori, Cristina e Gianni, per avermi dato tutte queste possibilità e per avere sempre creduto in me. Devo tutto a voi.

I hope that this is only the beginning!

# Bibliography

- [1] J.G. Canadell, P.M.S. Monteiro, M.H. Costa, L. Cotrim da Cunha, P.M. Cox, A.V. Eliseev, S. Henson, M. Ishii, S. Jaccard, C. Koven, A. Lohila, P.K. Patra, S. Piao, J. Rogelj, S. Syampungani, S. Zaehle, and K. Zickfeld. Global carbon and other biogeochemical cycles and feedbacks. In V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, page 673–816. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021. doi: 10.1017/9781009157896.007.
- [2] P. Ciais, C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R.B. Myneni, S. Piao, and P. Thornton. Carbon and Other Biogeochemical Cycles. In T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, page 465–570. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- [3] Trends in Atmospheric Carbon Dioxide - Global Monthly Mean CO<sub>2</sub>. <https://gml.noaa.gov/ccgg/trends/global.html>, 2023. Accessed: 2023-08-04.
- [4] Morten H. Iversen. Carbon export in the ocean: A biologist’s perspective. *Annual Review of Marine Science*, 15(1):357–381, 2023. doi: 10.1146/annurev-marine-032122-035153. URL <https://doi.org/10.1146/annurev-marine-032122-035153>. PMID: 36055975.
- [5] C.L. De La Rocha and U. Passow. 8.4 - The Biological Pump. In Heinrich D. Holland and Karl K. Turekian, editors, *Treatise on Geochemistry (Second Edition)*, pages 93–122. Elsevier, Oxford, second edition edition, 2014.

- ISBN 978-0-08-098300-4. doi: <https://doi.org/10.1016/B978-0-08-095975-7.00604-5>. URL <https://www.sciencedirect.com/science/article/pii/B9780080959757006045>.
- [6] Jorge L. Sarmiento and Nicolas Gruber. *Ocean Biogeochemical Dynamics*. Princeton University Press, 2006. ISBN 9781400849079. doi: 10.2307/j.ctt3fgxqx.
- [7] Tim DeVries. The Ocean Carbon Cycle. *Annual Review of Environment and Resources*, 47:317–341, 2022. doi: 10.1146/annurev-environ-120920. URL <https://doi.org/10.1146/annurev-environ-120920->.
- [8] Tyler Volk and Martin I. Hoffert. *Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO<sub>2</sub> Changes*, pages 99–110. American Geophysical Union (AGU), 1985. ISBN 9781118664322. doi: <https://doi.org/10.1029/GM032p0099>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/GM032p0099>.
- [9] L. Patara, N. Pinardi, C. Corselli, E. Malinverno, M. Tonani, R. Santoleri, and S. Masina. Particle fluxes in the deep Eastern Mediterranean basins: the role of ocean vertical velocities. *Biogeosciences*, 6(3):333–348, 2009. doi: 10.5194/bg-6-333-2009. URL <https://bg.copernicus.org/articles/6/333/2009/>.
- [10] Nadia Pinardi. Physical Oceanography - Notes for Master Degree program in Earth System Science, 2022.
- [11] Francesco Trotta. Lecture slides of Numerical Laboratory of the Atmosphere and Ocean, 2022.
- [12] Woods Hole Oceanographic Institution. Currents, Gyres, & Eddies. <https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-circulation/currents-gyres-eddies/>, 2023. Accessed: 2023-08-15.
- [13] NASA Scientific Visualization Studio. Gulf Stream Sea Surface Currents and Temperatures. <https://svs.gsfc.nasa.gov/3913>, 2023. Accessed: 2023-08-15.
- [14] James C. McWilliams. Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189):20160117, 2016. doi: 10.1098/rspa.2016.0117. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2016.0117>.

- [15] J. C. McWilliams, F. Colas, and M. J. Molemaker. Cold filamentary intensification and oceanic surface convergence lines. *Geophysical Research Letters*, 36(18), September 2009. doi: 10.1029/2009gl039402. URL <https://doi.org/10.1029/2009gl039402>.
- [16] Jonathan Gula, Jeroen J. Molemaker, and James C. McWilliams. Submesoscale cold filaments in the Gulf Stream. *Journal of Physical Oceanography*, 44:2617–2643, 2014. ISSN 15200485. doi: 10.1175/JPO-D-14-0029.1.
- [17] JP Ryan, FP Chavez, and JG Bellingham. Physical-biological coupling in Monterey Bay, California: topographic influences on phytoplankton ecology. *Marine Ecology Progress Series*, 287:23–32, 2005. doi: 10.3354/meps287023. URL <https://doi.org/10.3354/meps287023>.
- [18] Mathis P. Hain Daniel M. Sigman. The Biological Productivity of the Ocean. *Nature Education Knowledge*, 3(10), 2012.
- [19] John J. Walsh, Gilbert T. Rowe, Richard L. Iverson, and C. Peter McRoy. Biological export of shelf carbon is a sink of the global CO<sub>2</sub> cycle. *Nature*, 291(5812):196–201, May 1981. doi: 10.1038/291196a0. URL <https://doi.org/10.1038/291196a0>.
- [20] Thomas C. Malone, Thomas S. Hopkins, Paul G. Falkowski, and Terry E. Whittedge. Production and transport of phytoplankton biomass over the continental shelf of the New York Bight. *Continental Shelf Research*, 1(4): 305–337, April 1983. doi: 10.1016/0278-4343(83)90001-8. URL [https://doi.org/10.1016/0278-4343\(83\)90001-8](https://doi.org/10.1016/0278-4343(83)90001-8).
- [21] Paul G. Falkowski, Charles N. Flagg, Gilbert T. Rowe, Sharon L. Smith, Terry E. Whittedge, and Creighton D. Wirick. The fate of a spring phytoplankton bloom: export or oxidation? *Continental Shelf Research*, 8(5-7):457–484, May 1988. doi: 10.1016/0278-4343(88)90064-7. URL [https://doi.org/10.1016/0278-4343\(88\)90064-7](https://doi.org/10.1016/0278-4343(88)90064-7).
- [22] Elisa Lovecchio, Nicolas Gruber, and Matthias Münnich. Mesoscale contribution to the long-range offshore transport of organic carbon from the canary upwelling system to the open north atlantic. *Biogeosciences*, 15(16):5061–5091, August 2018. doi: 10.5194/bg-15-5061-2018. URL <https://doi.org/10.5194/bg-15-5061-2018>.
- [23] NASA EarthData. What is Remote Sensing? <https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing#data-processing-interpretation-and-analysis>, 2023. Accessed: 2023-08-31.

- 
- [24] SEA ICE Portal. Measurements from space. <https://www.meereisportal.de/en/learn-more/sea-ice-measuring-methods/measurements-from-space>, 2023. Accessed: 2023-09-01.
- [25] NASA EarthData. Data Processing Levels. <https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels>, 2023. Accessed: 2023-09-01.
- [26] O. B. Brown, P. J. Minnett, R. Evans, E. Kearns, K. Kilpatrick, A. Kumar, R. Sikorski, and A. Závody. MODIS Infrared Sea Surface Temperature Algorithm Algorithm Theoretical Basis Document Version 2.0. Technical report, NASA and University of Miami, 1999. URL [https://modis.gsfc.nasa.gov/data/atbd/atbd\\_mod25.pdf](https://modis.gsfc.nasa.gov/data/atbd/atbd_mod25.pdf).
- [27] NASA/JPL. MODIS Aqua Level 3 SST Thermal IR Daily 4km Daytime V2019.0, 2020. URL [https://podaac.jpl.nasa.gov/dataset/MODIS\\_AQUA\\_L3\\_SST\\_THERMAL\\_DAILY\\_4KM\\_Daytime\\_V2019.0](https://podaac.jpl.nasa.gov/dataset/MODIS_AQUA_L3_SST_THERMAL_DAILY_4KM_Daytime_V2019.0).
- [28] Curtis D. Mobley. *The Ocean Optics Book*. International Ocean Colour Coordinating Group (IOCCG), 2022. doi: 10.25607/OBP-1710. URL <http://dx.doi.org/10.25607/OBP-1710>.
- [29] NASA Ocean Color. Long-Wave Sea Surface Temperature (SST) . <https://oceancolor.gsfc.nasa.gov/resources/atbd/sst/>, 2023. Accessed: 2023-09-01.
- [30] Shubha Sathyendranath, Robert Brewin, Carsten Brockmann, Vanda Brotas, Ben Calton, Andrei Chuprin, Paolo Cipollini, André Couto, James Dingle, Roland Doerffer, Craig Donlon, Mark Dowell, Alex Farman, Mike Grant, Steve Groom, Andrew Horseman, Thomas Jackson, Hajo Krasemann, Samantha Lavender, Victor Martinez-Vicente, Constant Mazeran, Frédéric Mélin, Timothy Moore, Dagmar Müller, Peter Regner, Shovonlal Roy, Chris Steele, François Steinmetz, John Swinton, Malcolm Taberner, Adam Thompson, André Valente, Marco Zühlke, Vittorio Brando, Hui Feng, Gene Feldman, Bryan Franz, Robert Frouin, Richard Gould, Stanford Hooker, Mati Kahru, Susanne Kratzer, B. Mitchell, Frank Muller-Karger, Heidi Sosik, Kenneth Voss, Jeremy Werdell, and Trevor Platt. An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative (OC-CCI). *Sensors*, 19(19):4285, October 2019. doi: 10.3390/s19194285. URL <https://doi.org/10.3390/s19194285>.
- [31] E.U.Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). Global Ocean Colour Plankton and Reflectances MY L3 daily ob-

- servations. [https://data.marine.copernicus.eu/product/OCEANCOLOUR\\_GLO\\_BGC\\_L3\\_MY\\_009\\_107/description](https://data.marine.copernicus.eu/product/OCEANCOLOUR_GLO_BGC_L3_MY_009_107/description), 2023. Accessed: 2023-09-04.
- [32] John E. O'Reilly and P. Jeremy Werdell. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, 229:32–47, August 2019. doi: 10.1016/j.rse.2019.04.021. URL <https://doi.org/10.1016/j.rse.2019.04.021>.
- [33] Shubha Sathyendranath. ESA Ocean Colour Climate Change Initiative Product User Guide for v6.0 Dataset. Technical report, Plymouth Marine Laboratory, 2022.
- [34] NASA Ocean Color. Chlorophyll a (chlor a). [https://oceancolor.gsfc.nasa.gov/resources/atbd/chlor\\_a/](https://oceancolor.gsfc.nasa.gov/resources/atbd/chlor_a/), 2023. Accessed: 2023-09-04.
- [35] Thomas Jackson, Shubha Sathyendranath, and Frédéric Mélin. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. *Remote Sensing of Environment*, 203:152–161, December 2017. doi: 10.1016/j.rse.2017.03.036. URL <https://doi.org/10.1016/j.rse.2017.03.036>.
- [36] E.U.Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). Global Ocean Colour (Copernicus-GlobColour), Bio-Geo-Chemical, L3 (daily) from Satellite Observations (1997-ongoing). [https://data.marine.copernicus.eu/product/OCEANCOLOUR\\_GLO\\_BGC\\_L3\\_MY\\_009\\_103/description](https://data.marine.copernicus.eu/product/OCEANCOLOUR_GLO_BGC_L3_MY_009_103/description), 2023. Accessed: 2023-09-05.
- [37] Philippe Garnesson, Antoine Mangin, Odile Fanton d'Andon, Julien Demaria, and Marine Bretagnon. The CMEMS GlobColour chlorophyll *a* product based on satellite observation: multi-sensor merging and flagging strategies. *Ocean Science*, 15(3):819–830, June 2019. doi: 10.5194/os-15-819-2019. URL <https://doi.org/10.5194/os-15-819-2019>.
- [38] S. Colella, E. Böhm, C. Cesarini, P. Garnesson, J.Netting, and B. Calton. Copernicus Marine Product User Manual for Ocean Colour Products. Technical report, Mercator Ocean International, 2022. URL <https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-OC-PUM.pdf>.
- [39] André Morel, Yannick Huot, Bernard Gentili, P. Jeremy Werdell, Stanford B. Hooker, and Bryan A. Franz. Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sensing of Environment*, 111(1):69–88, November 2007. doi: 10.1016/j.rse.2007.03.012. URL <https://doi.org/10.1016/j.rse.2007.03.012>.

- 
- [40] John T. O. Kirk. *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, 1994. doi: 10.1017/cbo9780511623370. URL <https://doi.org/10.1017/cbo9780511623370>.
- [41] GEBCO Compilation Group. GEBCO 2022 Grid. <https://download.gebco.net/>, 2022. Accessed: 2023-09-06.
- [42] J. J. Becker, D. T. Sandwell, W. H. F. Smith, J. Braud, B. Binder, J. Depner, D. Fabre, J. Factor, S. Ingalls, S-H. Kim, R. Ladner, K. Marks, S. Nelson, A. Pharaoh, R. Trimmer, J. Von Rosenberg, G. Wallace, and P. Weatherall. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30\_PLUS. *Marine Geodesy*, 32(4):355–371, November 2009. doi: 10.1080/01490410903297766. URL <https://doi.org/10.1080/01490410903297766>.
- [43] Kevin P. Murphy. *Machine Learning: A probabilistic perspective*. MIT Press, 2012. URL [http://noiselab.ucsd.edu/ECE228/Murphy\\_Machine\\_Learning.pdf](http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf).
- [44] Redouane Lguensat, Miao Sun, Ronan Fablet, Evan Mason, Pierre Tando, and Ge Chen. EddyNet: A Deep Neural Network For Pixel-Wise Classification of Oceanic Eddies. 2017. doi: 10.48550/ARXIV.1711.03954. URL <https://arxiv.org/abs/1711.03954>.
- [45] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124, May 2019. doi: 10.1016/j.physrep.2019.03.001. URL <https://doi.org/10.1016/j.physrep.2019.03.001>.
- [46] Lior Rokach and Oded Maimon. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*, pages 321–352. Springer-Verlag, 2005. doi: 10.1007/0-387-25465-x\_15. URL [https://doi.org/10.1007/0-387-25465-x\\_15](https://doi.org/10.1007/0-387-25465-x_15).
- [47] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [48] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, March 1963. doi: 10.1080/01621459.1963.10500845. URL <https://doi.org/10.1080/01621459.1963.10500845>.

- [49] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, volume 96, pages 226–231, 1996. URL <https://cdn.aaai.org/KDD/1996/KDD96-037.pdf>.
- [50] scikit learn. Gaussian Mixture models. <https://scikit-learn.org/stable/modules/mixture.html#mixture>, 2023. Accessed: 2023-09-17.
- [51] J. C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, January 1973. doi: 10.1080/01969727308546046. URL <https://doi.org/10.1080/01969727308546046>.
- [52] Jean-François Cayula and Peter Cornillon. Edge detection algorithm for SST images. *Journal of Atmospheric and Oceanic Technology*, 9(1):67–80, February 1992. doi: 10.1175/1520-0426(1992)009<0067:edafsi>2.0.co;2. URL [https://doi.org/10.1175/1520-0426\(1992\)009<0067:edafsi>2.0.co;2](https://doi.org/10.1175/1520-0426(1992)009<0067:edafsi>2.0.co;2).
- [53] K. Nieto, H. Demarcq, and S. McClatchie. Mesoscale frontal structures in the Canary Upwelling System: New front and filament detection algorithms applied to spatial and temporal patterns. *Remote Sensing of Environment*, 123:339–346, August 2012. doi: 10.1016/j.rse.2012.03.028. URL <https://doi.org/10.1016/j.rse.2012.03.028>.
- [54] Osvaldo Artal, Héctor H. Sepúlveda, Domingo Mery, and Christian Pieringer. Detecting and characterizing upwelling filaments in a numerical ocean model. *Computers & Geosciences*, 122:25–34, January 2019. doi: 10.1016/j.cageo.2018.10.005. URL <https://doi.org/10.1016/j.cageo.2018.10.005>.
- [55] J. Marcello, F. Maques, and F. Eugenio. Automatic feature extraction from multisensorial oceanographic imagery. In *IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2002. doi: 10.1109/igarss.2002.1026585. URL <https://doi.org/10.1109/igarss.2002.1026585>.
- [56] Susana Nascimento, Pedro Franco, Fátima Sousa, Joaquim Dias, and Filipe Neves. Automated computational delimitation of SST upwelling areas using fuzzy clustering. *Computers & Geosciences*, 43:207–216, June 2012. doi: 10.1016/j.cageo.2011.10.025. URL <https://doi.org/10.1016/j.cageo.2011.10.025>.
- [57] Zineb El Abidi, Khalid Minaoui, Anass El Aouni, Ayoub Tamim, and Hicham Laanaya. An Efficient Detection of Moroccan Coastal Upwelling Based on



- Fusion of Chlorophyll-a and Sea Surface Temperature Images With a New Validation Index. *IEEE Geoscience and Remote Sensing Letters*, 18(8):1322–1326, August 2021. doi: 10.1109/lgrs.2020.3002473. URL <https://doi.org/10.1109/lgrs.2020.3002473>.
- [58] Susana Nascimento, Sérgio Casca, and Boris Mirkin. A seed expanding cluster algorithm for deriving upwelling areas on sea surface temperature images. *Computers & Geosciences*, 85:74–85, December 2015. doi: 10.1016/j.cageo.2015.06.002. URL <https://doi.org/10.1016/j.cageo.2015.06.002>.
- [59] Susana Nascimento, Sayed Mateen, and Paulo Relvas. Sequential self-tuning clustering for automatic delimitation of coastal upwelling on SST images. In *Lecture Notes in Computer Science*, pages 434–443. Springer International Publishing, 2020. doi: 10.1007/978-3-030-62365-4\_41. URL [https://doi.org/10.1007/978-3-030-62365-4\\_41](https://doi.org/10.1007/978-3-030-62365-4_41).
- [60] Matthew Lee Hammond, Fatma Jebri, Meric Srokosz, and Ekaterina Popova. Automated detection of coastal upwelling in the Western Indian Ocean: Towards an operational “Upwelling Watch” system. *Frontiers in Marine Science*, 9, August 2022. doi: 10.3389/fmars.2022.950733. URL <https://doi.org/10.3389/fmars.2022.950733>.
- [61] David M. Checkley and John A. Barth. Patterns and processes in the California Current System. *Progress in Oceanography*, 83(1-4):49–64, December 2009. doi: 10.1016/j.pocean.2009.07.028. URL <https://doi.org/10.1016/j.pocean.2009.07.028>.
- [62] C. Lathuilière, V. Echevin, and M. Lévy. Seasonal and intraseasonal surface chlorophyll-a variability along the northwest African coast. *Journal of Geophysical Research: Oceans*, 113, 5 2008. ISSN 21699291. doi: 10.1029/2007JC004433.
- [63] Jiaxun Li, Guihua Wang, and Xiaoming Zhai. Observed cold filaments associated with mesoscale eddies in the South China Sea. *Journal of Geophysical Research: Oceans*, 122(1):762–770, January 2017. doi: 10.1002/2016jc012353. URL <https://doi.org/10.1002/2016jc012353>.
- [64] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke,

- and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585 (7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- [65] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [66] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- [67] S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *In revision, J. Open Res. Software*, 2017.
- [68] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- [69] Met Office. *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon, 2010 - 2015. URL <https://scitools.org.uk/cartopy>.
- [70] Abraham H. Oort and José P. Peixóto. Global angular momentum and energy balance requirements from observations. In *Advances in Geophysics*, pages 355–490. Elsevier, 1983. doi: 10.1016/s0065-2687(08)60177-6. URL [https://doi.org/10.1016/s0065-2687\(08\)60177-6](https://doi.org/10.1016/s0065-2687(08)60177-6).
- [71] S Sathyendranath, V Stuart, A Nair, K Oka, T Nakane, H Bouman, MH Forget, H Maass, and T Platt. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Marine Ecology Progress Series*, 383:73–84, May 2009. doi: 10.3354/meps07998. URL <https://doi.org/10.3354/meps07998>.
- [72] Katherine D. Zaba and Daniel L. Rudnick. The 2014-2015 warming anomaly in the Southern California Current System observed by underwater gliders. *Geophysical Research Letters*, 43(3):1241–1248, February 2016. doi: 10.1002/2015gl067550. URL <https://doi.org/10.1002/2015gl067550>.
- [73] Markel Gómez-Letona, Antonio G. Ramos, Josep Coca, and Javier Arístegui. Trends in primary production in the canary current upwelling system—a regional perspective comparing remote sensing models. *Frontiers in Marine Science*, 4, November 2017. doi: 10.3389/fmars.2017.00370. URL <https://doi.org/10.3389/fmars.2017.00370>.

- [74] Yongyun Hu and Raymond T. Pierrehumbert. The advection–diffusion problem for stratospheric flow. part II: Probability distribution function of tracer gradients. *Journal of the Atmospheric Sciences*, 59(19):2830–2845, October 2002. doi: 10.1175/1520-0469(2002)059<2830:tadpfs>2.0.co;2. URL [https://doi.org/10.1175/1520-0469\(2002\)059<2830:tadpfs>2.0.co;2](https://doi.org/10.1175/1520-0469(2002)059<2830:tadpfs>2.0.co;2).
- [75] Yosef Ashkenazy and Hezi Gildor. On the probability and spatial distribution of ocean surface currents. *Journal of Physical Oceanography*, 41(12):2295–2306, December 2011. doi: 10.1175/jpo-d-11-04.1. URL <https://doi.org/10.1175/jpo-d-11-04.1>.