Biological Sciences

UNIVERSITY OF CAPE TOWN

BIOLOGICAL SCIENCE DEPARTMENT

Quantifying Phytoplankton Biomass and Sediment in River Plumes along the Agulhas Bank Using Remotely Sensed Data with Deep Learning Techniques

Author:
Humeshni Pillay

Supervisors:
Natasha Karenyi
Lisl Robertson Lain
Marié Smith

Thesis presented in fulfilment of the requirements for the degree of Master of Science (Biology) in the Faculty of Science at the University of Cape Town

August 2023

Abstract

River plumes play a major role in land-sea connectivity by providing essential nutrients, sediment, and organic matter to marine environments, and maintaining ecosystem function and habitat structure. Phytoplankton communities' reliance on riverine nutrient deposition to build up their biomass and suspended sediments released from terrestrial systems have been identified as major products of riverine outputs that impact pelagic and benthic ecosystems. Quantifying their concentrations in river plumes could improve our understanding of how anthropogenic activities, seasonality and climate change affect their concentrations in marine habitats and ecosystems over time. The need to monitor phytoplankton biomass and sediment loads in coastal systems has encouraged the use of remotely sensed data obtained from satellites. Remote sensing capabilities have advanced over the past few decades, along with increased computational efficiency, and sizeable open-access data pipelines. However, current satellite data products are not always appropriate for highly turbid, optically complex waters, along coastal regions where Atmospheric Correction (AC) is known to be challenging. The aim of this project was to use four rivers along the Agulhas Bank of South Africa to test satellite-derived Total Suspended Matter (TSM) and chlorophyll-a concentration ([Chl-a]) data products using a suite of available AC options against *in situ* [Chl-a] and TSM. The *in situ* data were collected through filtrations for fluorometric ([Chl-a]) and gravimetric (TSM) analyses. The performances of geophysical algorithms for TSM and [Chl-a] were assessed using remote sensing reflectance (R_{rs}) derived from standard Sentinel-3 Level-2 files, as well as two alternative approaches to AC, namely the Case-2 Regional CoastColour (C2RCC) processor, and the POLYnomial-based for MERIS (POLYMER). Thereafter, a regionally parameterised deep learning neural network (Multi-Layer Perceptron; MLP) model for retrieving TSM and [Chl-a] was developed and evaluated in the context of its application to Sentinel-3 satellite data. The MLP model is trained on a synthetic dataset of R_{rs} parameterised using the *in situ* ranges of TSM and [Chl-a]. The MLP model was evaluated using the three mentioned ACs as the model requires R_{rs} uncontaminated by the atmosphere. The regional MLP model was separated into a model containing the full range of TSM *in situ* values and a constrained version. POLYMER's [Chl-a] and C2RCC's TSM were the best geophysical products over the region when assessing the accuracy of retrievals to the match-up *in situ* dataset. However, AC is a major concern in regions where land-based fynbos biogenic burning is common. As a result, the current [Chl-a] satellite data product performances were not optimal, especially when mapped over the open ocean region. For the developed [Chl-a] MLP algorithm, the constrained and full MLP models were similar, with C2RCC's AC producing the best results. However, the constrained TSM MLP model's results vastly outperformed the full MLP model, with no clear delineation for the best-performing AC. The use of deep learning models shows promising results over the optically complex Agulhas Bank river plumes, however, the effectiveness of the MLP technique may be dependent on the variability of *in situ* data (i.e. phytoplankton size) used to create the training synthetic dataset,

as well as the quality of R_{rs} applied to the MLP models. Thus, more *in situ* data is required to develop flexible, regional algorithms for highly turbid waters along South Africa's atmospherically variable coastlines.

0.1 Declaration

I hereby declare that the work presented in this thesis, submitted in partial fulfilment of the requirement for the award of the degree of Master of Science in Biology to the University of Cape Town, is an original piece of research under the guidance of Dr. Natasha Karenyi, Dr. Lisl Robertson Lain and Dr. Marié Smith. The information and methodologies used in this thesis have been duly acknowledged in the text of references provided. Furthermore, no part of this dissertation has been previously presented for another degree at this or any other institution.

Signature: Signed by candidate

Date: 12/08/2023

0.2 Acknowledgments

I am thankful for the National Research Foundation's support towards the ACEP Agulhas Bank Project as this work would not have been possible without the funding received for the greater project. I am immensely grateful for my supervisors who were always supportive and kind throughout this difficult journey, and the love and care they invested in my personal growth, as well as this thesis. Thank you for being excellent role models for women in science and teaching me to put my mental health first among all other things. I would also like to thank Pieter, Andrea and Chelsea for their tremendous contribution to my sampling trips, and their company throughout the long nights of filtrations. Our Spiderman movie night accompanied by whale watching at the Breede the next day made this whole experience that much lighter and more fun. I would also like to thank my lab members and other friends for their wisdom, guidance, company, and laughter you all brought me while in lab meetings or gym sessions, and for making this whole process that much easier. To my younger cousins Sarah, Dylan and Mishaylin, thank you for always making holidays eventful and light, and putting up with me working through most of our time together this year. And to my brother Koi, parents Rajen and Ella, and late grandmother, Veena, for their never-ending love and support, I would not be the person I am today without you.

0.3 List of Notations

0.3.1 Symbols

a_{gd} Absorption of Gelbstoff and Detritus

AU Absorbance Units

Chl-a Chlorophyll-a pigment

[Chl-a] Chlorophyll-a concentration

CHL$\_nn$ Chlorophyll-a Neural Network

ρ_ω Water Leaving Reflectances

R_{rs} Remote-sensing Reflectance

TSM Total Suspended Matter

TSM$\_nn$ Total Suspended Matter Neural Network

TSS Total Suspended Sediment

0.3.2 Acronyms

AAC Alternative Atmospheric Correction

AC Atmospheric Correction

AOP Apparent Optical Properties

AOT Aerosol Optical Thickness

BAC Baseline Atmospheric Correction

C2RCC Case-2 Regional Coast Colour Processor

CDOM Coloured Dissolved Organic Matter

CZCS Coastal Zone Colour Scanner

ESA European Space Agency

EUMETSAT European Organisation for the Exploitation of Meteorological Satellites

FDA Functional Data Analysis

HAB Harmful Algal Bloom

IMP Inverse Modelling Technique

IOP Inherent Optical Properties

MAPE Median Absolute Percentage Error

MERIS MEdium Resolution Imaging Spectrometer

MLP Multi-Layer Perceptron

NASA National Aeronautics and Space Administration

NIR Near-Infared

NOMAD NASA bio-Optical Marine Algorithm Dataset

OLCI Ocean and Land Colour Instrument

POLYMER POLYnomial for MERis

RFU Raw Fluorescence Units

RMSLE Root Mean Square Log Error

TOA Top-of-Atmosphere

WLR Water-leaving Reflectance

Contents

Chapter 1

Introduction

1.1 Biological importance of river plumes

River plumes play a crucial role in land-sea connectivity by transporting 36 000 km^3 of freshwater annually into the global ocean, along with 20 billion tons of solid and dissolved material (Milliman & Farnsworth, 2013). The geographical extent of river plumes ranges from warm tropical waters to cold polar regions, accumulating various watersheds, along with their unique combinations of land-based matter. The outputs of such can be utilized by a multitude of biological systems which often rely on the consistent nutrient and sediment supply from river plumes to maintain their population structure and biomasses (Hermand et al., 2008; Lohrenz et al., 2008). Therefore, improving our understanding of river plume discharge is vital for protecting and conserving both pelagic and benthic marine ecosystems.

Coastal and pelagic marine ecosystems, being the first to receive outputs from rivers, are directly affected by changes in riverine flow (Zhou et al., 2008). Phytoplankton forms the base of aquatic and marine trophic food webs, providing carbon-based organic matter which other organisms can readily use to carry out their metabolic activities in both pelagic and benthic ecosystems (Jassby et al., 2003; Jones, 1998). Since river plume phytoplankton communities rely on land-derived nutrients to build up their population biomasses through the process of photosynthesis, their biomass is constantly in flux (Liu & Dagg, 2003; Wurtsbaugh et al., 2019). Thus, variability in phytoplankton biomass attributed to changes in riverine flow, and the associated nutrient alterations, will have a cascading impact on pelagic marine organisms, such as zooplankton, marine fish, and seabirds (Diaz & Rosenberg, 2008; Dybas, 2005), whose populations rely on phytoplankton as a food source, either

directly or indirectly (Frederiksen et al., 2006).

In addition, phytoplankton contributes significantly to global marine CO_2 sequestration, fixing between 30-50 billion metric tonnes of atmospheric CO_2 into organic carbon, which is removed by organisms in the food web, eventually deposited along the seafloor as material bulky described as either labile or refractory matter (Basu & Mackey, 2018; Mikaloff Fletcher et al., 2006; Volk & Hoffert, 1985). Labile material is organic matter which can be readily remineralised by microbial activities in marine environments (Ittekkot, 1988), while refractory matter is the oldest and most abundant pool of organic matter (i.e. dissolved organic carbon) with a high degree of resistance to biological remineralisation and does not contribute significantly to the nutrient flux in the ecosystem (Baltar et al., 2021; Ogawa et al., 2001). Labile material, such as the minute remains of phytoplankton and riverine detritus, could be further broken down by marine microbes (Ittekkot, 1988), acting as a source of nutrients which can be recycled and inserted back into pelagic surface waters through coastal upwelling or cyclonic eddies (Bernard et al., 2011; Corredor et al., 2004). New phytoplankton communities can further proliferate from nutrients deposited by historical river plumes during seasons of heightened water column mixing, such as spring bloom events (Imbeau et al., 2021; Moran et al., 2022), adding to the present phytoplankton biomass induced by riverine nutrient deposits along coastlines.

Benthic marine habitats will also be affected by changes in river flow, due to their reliance on riverine sedimentary discharge for habitat structuring (Levin et al., 2001) and phytodetritus as a food source in some regions (Goineau et al., 2012). Benthic marine systems often rely on various external sources for nourishment due to the lack of primary producers, unlike the surface, where there is abundant light available for photosynthesis (Iken et al., 2001; Kiørboe, 1993; Whitfield, 1989). Previous seasonal surface phytoplankton blooms that sink rapidly to the seafloor contribute to the pool of organic matter benthic organisms can metabolise (Grebmeier et al., 1988). Nutrient input into marine surface waters from river plume discharges has been found to cause such blooms, as highlighted in previous studies (Danovaro et al., 2000; Li et al., 2022; Wang et al., 2019), where evidence of seasonal meiofauna density shifts in response to river plume frontal systems migrations was observed. In addition to phytoplankton remains, detritus from rivers can also be utilized as a land-based food source for benthic organisms along the seafloor and submarine canyons (Heezen et al., 1964; Mann, 1988). Isotope studies have found a relationship in surface phytoplankton populations with organisms along the seafloor, substantiating the importance of sinking phytoplankton as a food resource for benthic communities where sources of nutrition are limited (Mäkelä et al., 2017; Newell et al., 1995). Furthermore, benthic mudbelt habitat structuring has been strongly associated with river plume sedimentary discharge (Lee & Chu, 2001; Summerhayes et al., 1978), stemming from fine-grained sediment of terrigenous origins settling on the seafloor, forming an amalgamation of

organic-rich matter and muds. As a result, unique assemblages of benthic fauna were found in such substrata, such as foraminifera forming species-related zonations along the mudbelt (Aller & Aller, 1986).

The role of river plumes in connecting the land to the sea is vita connecting the land to the sea is vital for both pelagic and benthic ecosystems, through nutrient deposition and sediment transport. However, water quality degradation through anthropogenic effects (Best, 2019), and the impact of unexpected droughts and floods from climate change (Stagl & Hattermann, 2016) may have effects the marine ecosystems that rely on riverine outputs. Monitoring river plume phytoplankton biomass and sediment discharge along fine temporal and spatial scales could assist efforts to protect and conserve pelagic and benthic ecosystem functioning. The main constituents of river plumes, namely nutrients fueling phytoplankton biomass and sediment will be considered further in the next section

1.2 Particulate and dissolved constituents in river plumes

1.2.1 Phytoplankton and nutrients

The nutrients collected from land masses fuel the growth of phytoplankton in rivers, estuaries and marine systems, converting inorganic compounds into a food source for other organisms (Jassby et al., 2003; Jones, 1998). Ecosystems have different concentrations of nutrients available, in turn, disparate habitats have their own unique phytoplankton assemblages (Cloern & Jassby, 2010; Weithoff & Beisner, 2019). In areas with high, consistent nutrient concentrations, such as river plumes, phytoplankton is adapted to utilizing the large quantities of nutrients and can grow much larger, with higher biomass, than phytoplankton communities in oligotrophic waters (Dufour et al., 1999; Hutchings et al., 1995; Marañón et al., 2001). Thus, phytoplankton biomass has a direct relationship with the nutrients available in a given system, along with the seasonality governing the nutrient concentrations and distribution.

Complicated dynamics of coastal upwelling, nutrient deposition from land masses into riverine channels, coupled with the retention and diffusion of nutrients from estuaries into marine systems, all have an impact on the horizontal and vertical extent of the nutrient transfer in marine systems (Han et al., 2012; Lohrenz et al., 2008). The nutrients available in river plumes are governed by both coastal and climatological changes throughout the year (Bargu et al., 2011; Liu & Dagg, 2003). Seasonal changes in rainfall are strongly associated with a region's historical precipitation patterns, as well as the added effects of El Niño and Southern Oscillation Index events (Deng et al., 2018; Schmidt et al., 2001). During seasons of high rainfall, nutrients from both natural and anthropogenic sources combine with remineralised detritus and plant matter, producing rivers enriched in bioavailable nutrients

for phytoplankton primary production (Han et al., 2023; Paerl et al., 2020). Moreover, seasons with low precipitation were found to have lower phytoplankton growth rates in both the estuarine system and the adjacent sea surface (Fong et al., 2020; Yin et al., 2000). In addition to influences from precipitation, river plumes and the associated coastlines are also subjected to biannual nutrient variability induced by seasonal changes in temperature, such as transitions from warmer to colder months, and vice versa (Ducklow & Kirchman, 1983; Lohrenz et al., 1999). Spring blooms are a common feature in productive oceanic water bodies, where the increased light availability is accompanied by nutrients accumulated during unproductive winter months and encourages phytoplankton biomass growth (Chiswell et al., 2015; Sambrotto et al., 1986). Autumn months may experience a slight bloom in some regions, mainly attributed to increased water column mixing, and the upwelling of nutrients from depth (Pingree et al., 1976; Wihsgott et al., 2019). In turn, phytoplankton populations along river plumes respond to changes in nutrient availability throughout the year, through shifting in population structure (Ault et al., 2000), sizes (Mei et al., 2009), and densities (Agawin et al., 2000).

Nutrients (i.e. nitrogen, phosphorus, silicate etc.) are essential for phytoplankton growth, however, in excess, phytoplankton biomass can proliferate to extreme extents, potentially causing harmful algal blooms (HABs), whereby algae not grazed by zooplankton within a sufficient time frame leads to eutrophication. HABs can occur naturally in ecosystems, however, over the past several decades the world's oceans have experienced an increase in the frequency of toxic blooms attributed to nutrient supply from human settlement sewage waste (Brewton et al., 2022), factory discharge (Zhang et al., 2006) and agricultural activities (Kiørboe, 2008). HABs cause harm through the production of toxins, the physical structure of cells or frequent occurrences of monumental population biomass accumulations along the water column, affecting the other organism residing in the habitat by means of mechanical damage (e.g. clogging of gills) or oxygen depletion (Anderson et al., 2002; Van der Lingen et al., 2016). The effects of HABs have led to mass mortalities of fish and shellfish, human illness and the death of many mammals and seabirds associated with the affected water body (Broadwater et al., 2018; Landsberg, 2002; Shumway, 1990). Li et al. (2014) have found that extreme nutrient loads from the Changjiang River have led to HABs in the East China Sea, where macrozoobenthic biomass was found to decrease as a consequence of the increased occurrences of eutrophication states (Wang, 2006). Furthermore, remnants from HABs can be transported into the benthos, leading to the bioaccumulation of the harmful toxins released from the bloom (Kankaanpää et al., 2005). An example of a well-known 'dead zone' is the Gulf of Mexico, where excess nutrients from the Mississippi River have led to an extremely hypoxic region as microbial organic matter decomposition removed much of the oxygen along the benthos, making the system inhabitable for benthic organisms and commercially important fisheries, which in turn, were negatively impacted by the oxygen-depleted substratum (Rabalais et al., 2002; Rabotyagov et al., 2014).

Monitoring the frequency and magnitude of phytoplankton blooms in regions where large nutrient discharge is common, such as river plumes, will be pertinent to manage marine ecosystems, and human health in the future where changes in river discharge from anthropogenic activities are anticipated. Satellites have been proposed as an inexpensive and routine method to study phytoplankton communities with a wide spatial coverage of Earth's water bodies (Joint & Groom, 2000). The use of remote sensing to monitor phytoplankton populations along coastlines dates to the launch of the Coastal Zone Colour Scanner (CZCS) sensor onboard the National Aeronautics and Space Administration (NASA) historic Nimbus-7 satellite in October 1978. The primary goal was to investigate whether ocean colour properties could be measured usefully for the quantification of in-water constituents, primarily phytoplankton's chlorophyll-a (Chl-a) pigment (Gordon et al., 1980; Hovis et al., 1980), and its interaction with light along varying magnitudes of Chl-a concentrations ([Chl-a]). As mentioned, this thesis will only describe [Chl-a] as a concentration to delineate phytoplankton biomass proxy, [Chl-a], that satellites measure from the pigment Chl-a present in phytoplankton cells. In the last four decades, an array of incoming ocean colour products has led to a plethora of freely available [Chl-a] satellite algorithms, each with their own merits and downfalls (Giannini et al., 2021; Kratzer & Plowey, 2021) . The creation and validation of such algorithms is an ongoing process, especially in optically complex systems such as river plumes.

1.2.2 Sediment

River plumes transport 13.5-22 billion metric tons of sediment from land mass into the marine continental shelves every year (Milliman & Farnsworth, 2013). Sediments in river plumes are either transported along the river bottom or in a suspension state (Paola et al., 2011). Suspended sediments are solids that are typically derived from eroded lithogenic material from weathering stone, with discharge rates measured by mass per unit time (m/s) and concentrations as g/m^3. Riverine sedimentary discharge is a major feature of river plumes, which alters both the sea surface colour and turbidity upon entering the adjoining marine environment. The eventual destination of such material is marine benthic ecosystems, which rely on riverine discharge for habitat structuring, as well as a source of food in the form of attached organic material. Thus, understanding the mechanisms influencing the concentrations, and eventual spatial distribution of riverine sediment in marine systems is essential for protecting and conserving marine benthic habitats.

Sediment loads have been attributed to the drainage basin area and morphology, where higher sediment discharge rates are usually measured from larger rivers (Hovius, 1998). However, Milliman and Syvitski (1992) noted that smaller rivers provide a sediment flux disproportionate to their size, attributed to steep gradients associated with the smaller watersheds of tectonically active margins relative to the gentle relief of the large watersheds on passive margins. In some river systems,

an increase in discharge introduces coarser sediment into suspension because of increased bottom shear stress (Harris & Wiberg, 2002). In others, however, suspended sediment sizes become finer as discharge increases, likely due to enhanced delivery from hill slopes or side channels (Walling & Moorehead, 1989). Estuarine retention of sediment may also occur as a result of the enclosed geomorphology of the basin, inducing sediment to accumulate until contents of the basin are flushed into the marine system (Kench, 1999). When suspended in the marine water column, sediment type may be a mixture of both marine resuspended coarse sediments, from tidal forcing and marine sediment intrusion along the estuarine basin, and fine-grained sands from land deposits (*ibid*). Seasonality also plays a major role in determining the concentrations of the Total Suspended Sediment (TSS), the inorganic component of matter present in the riverine and plume water bodies. High rainfall seasons, especially in regions where indigenous vegetation has been removed, observe relatively high sediment yields into river channels due to the lack of soil stability (Thomas et al., 2018). Land erosion is a common occurrence in farming and agricultural activities, where mass quantities of unnaturally deposited sediment contribute to estuarine sediment overloading, resulting in out-of-phase river mouth closures (Freeman et al., 2019). Other activities such as damming have the opposite effect, where natural sediment loads into marine systems are severely reduced as a consequence of freshwater discharged along the connecting river system (Collier et al., 1996; Dethier et al., 2022). In turn, other organic matter such as detritus, bacteria and phytoplankton remains do not enter the ocean leading to reductions in organic matter released to marine systems (Hay, 1998).

While suspended in the river and estuary water columns, sediment interacts with the encompassing organic matter deposited and formed in those systems through flocculation (Guo & He, 2011). Flocculation is the process where the materials in suspension forms large, loosely packed agglomerations of mineral grains and organic material (Milligan et al., 2007). Biologically mediated removal of fine sediments has been observed in some plumes, where the process of flocculation has led to organic matter aggregating around fine sediments, resulting in an increase in the particles size (Ayukai & Wolanski, 1997). Through flocculation, the settling velocity of small particles bound within flocs can effectively increase by several orders of magnitude (Kranck, 1980). These agglomerations of particles sink many times faster than their component grains and deliver a poorly sorted mixture of grain sizes both to riverbeds and floodplains and to discharge plumes that meet the sea (Walling & Moorehead, 1989). This has been known to occur in estuaries with high phytoplankton and detrital material, such as regions where mangroves were prominent (Wolanski, 1995). Anoxia is prone to the regions where the fine-grained suspended matter settles and accumulates. The accumulation of fine-grained sediment and flocculation deposits along the seafloor forms terrigenous mudbelt habitats (Gratiot & Anthony, 2016; Summerhayes et al., 1978). A diverse array of various benthic organisms rely on such deposits for both habitat structuring and as a source of labile organic matter (Aller & Aller, 1986; Barbieri et al., 2019). With anthropogenic activities such as damming, land erosion, and the

added effects of climate change, mudbelt ecosystems are at risk due to their reliance on land-based sediment and riverine discharges (Mertes & Dunne, 2007). Protecting these crucial, understudied ecosystems will require advanced river plume monitoring techniques to reinforce our understanding of these dynamic effects.

Sediment loads are rarely captured regularly, especially in countries where resources for consistent ecological research are limited, resulting in a lack of data. In addition, river plume flow rates and discharge volumes, are subjected to various environmental and anthropogenic changes over time, making our observations and subsequent conclusions difficult to rely on, an artefact of the lack of data collected over the necessary temporal and spatial scales (Ostendorf, 2011). Remote sensing could be used to combat such drawbacks as distinct sedimentary backscatter reflectance signals received by satellites can be used to quantify suspended sediment (Dekker et al., 2002; Pitarch et al., 2019), and the mechanism driving the transport and distribution of the inorganic particles (Froidefond et al., 1993; Ritchie et al., 1990). The use of remote sensing could assist the management of riverine systems, and their associated estuaries and marine biological systems, which are interlinked.

1.2.3 Coloured Dissolved Organic Matter

Coloured dissolved organic matter (CDOM; *gelbstoff* or humic substance) is a biologically derived constituent, contributing to Earth's large pool of organic matter in both marine and aquatic systems (Osburn & Morris, 2003). There is great diversity in the origin of CDOM, as it can be released from both terrestrial and marine photosynthetic activities, as well as from bacterial decomposition. River channels can accumulate CDOM in high concentrations through encroaching vegetation releasing dissolved matter into the water column or as a by-product from microbial processing of detrital organic matter (Mulholland, 2003). Phytoplankton in the open ocean release CDOM as a waste product from the process of photosynthesis, however, in river plumes the large complex fulvic and humic acid molecules are primarily released through planktonic leakage, senescence, grazing, and microbial degradation (Bauer & Bianchi, 2011). As a result, CDOM contains large concentrations of carbon leached from the mentioned processes, subsequently contributing to the global carbon budget (Guéguen et al., 2005). Although CDOM plays an important role in carbon sequestration, in large concentrations it can become a photoinhibitor to phytoplankton in the water column. Contrastingly, CDOM has also been found to be photo-protective to planktonic cells, as CDOM can absorb ultraviolet light efficiently. Thus, quantifying CDOM in water bodies will assist our understanding of the role of CDOM from biogeochemical and microbial activities in carbon sequestration. Remote sensing has been proposed as a technique to quantify CDOM in aquatic and marine systems, however, the effectiveness of such methods requires significant improvements (Zhu et al., 2014).

1.3 Remote sensing to study river plumes

Routinely available open-source, remote sensing technologies can help to supplement a lack of *in situ* data and offer a synoptic view of the effects of river plume sediment loads and phytoplankton communities in marine and estuarine systems. Radiometric sensors can retrieve the bulk optical signal from the water's surface, and optically significant constituents (i.e. [Chl-a], TSM, and CDOM) can be retrieved by a variety of empirical or semi-analytical methodologies. It is noted that Total Suspended Matter (TSM) and TSS are at times interchangeable in the remote sensing community. However, TSS in this study is used to describe the inorganic component of sediment. In contrast, TSM is used to describe satellite retrievals, as most satellite products incorporate both organic (i.e. bacterial and algal communities, detritus, extracellular polymers) and inorganic (i.e. clays and silts) matter when measuring sediment concentrations. Additionally, TSS was used for the bio-optical modelling description, as models used for sediment are specific to the inorganic component alone. Satellites are imperative for the future of marine and aquatic conservation as they can provide the user with accessible and inexpensive means to quantify aquatic and marine constituents using optical sensors. Onboard Sentinel-3's satellite is the Ocean and Land Colour Instrument (OLCI) sensor, which was released in April of 2016, with spatial resolutions of both 300 m and 1 km and a spectral extent of 21 bands between 400-1020 nm. The mentioned sensor is a branch of the historic MERIS mission, with the sensor collecting radiometric data with the appropriate signal-to-noise ratio for ocean colour remote sensing, generated by the large spectral resolutions over broad areas.

1.3.1 Marine Optics

When light from the sun reaches the waterbody surface, photons either get scattered or absorbed characteristically in magnitudes dependent on the concentrations and type of constituents present (Mobley, 2001). The mentioned constituents' interactions with light are termed the Inherent Optical Properties (IOPs), which are dependent on the physical and chemical makeup of in-water substances. Specialised satellite sensors are designed to measure electromagnetic radiation (light) emerging from the water mass surface and transmitted through the atmosphere. Top-of-Atmosphere (TOA) reflectances are thus made up of the combined interactions of light with the atmosphere, i.e. gases, aerosols and clouds, as well as the light leaving the water body. This signal from the atmosphere has to be removed in order to obtain the water-leaving reflectance, representative of the combination of IOPs of constituents within the water environment. The observable signals satellite detect are termed the Apparent Optical Properties (AOPs), which is representative of the IOPs' interaction with the marine light environment (Figure 1.1).

The four primary optically active constituents in aquatic environments are generally considered to be water, phytoplankton, CDOM, and mineral particles. Water absorbs strongly in the red-NIR

region (Morel, 1974). Chl-a is a proxy used to estimate phytoplankton biomass as it is the primary photosynthetic pigment found in their cells (Huot et al., 2007). Chl-a has strong absorbing qualities in the blue (443 nm) and red (675 nm) spectral region and these can be identified in water-leaving reflectances when phytoplankton are the dominant optical component in the water body (O'Reilly et al., 1998). Multiple studies on the optical properties of CDOM have described the exponentially decaying absorption spectrum with its maximum absorption located in the blue region of visible light (400 nm) (Bricaud et al., 1981). TSS absorbs light in a similar decay to that of CDOM (Schalles, 2006), although TSS backscatters strongly whereas CDOM does not as it contains no particulate matter. IOPs of sediment are highly variable, attributed to the diverse size ranges, lithogenic compositions, and flocculation occurring naturally in real water environments (Binding et al., 2005; Bowers & Binding, 2006).

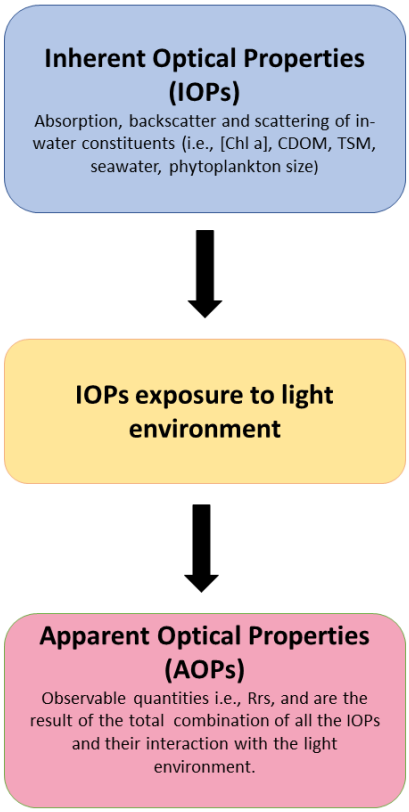

Figure 1.1: A schematic of constituent IOPs' interaction with the light environment to form observable AOPs.

Example remote sensing reflectances (R_{rs}) are shown in Figure 1.2, where the spectral shape of each component dominating a given water body is distinctive. The [Chl-a] dominant curve is typically classified as an example of 'Case 1' waters, where phytoplankton and their co-varying degradation products (i.e. CDOM) dominate the R_{rs} signal, allowing for elementary empirical relations between

changes in R_{rs} signals due to changes in [Chl-a] (Morel & Prieur, 1977). While much open ocean and nutrient-poor waters can be well described using generalized Case 1 constituent algorithms, Case 2 waters are regionally variable with unique and complex optical characteristics. River plumes comprise of elevated concentrations of multiple constituents varying independently to that of phytoplankton, generating the 'Case 2' water type. Flocculation of fine sediment particles, along with organic compounds in riverine water columns, results in bio-aggregates, further complicating the signals received by satellite (Hatcher et al., 2001). The *in situ* characteristics of such waters are difficult to sample as flocs are readily broken down by the common water sampling protocols (*ibid*).

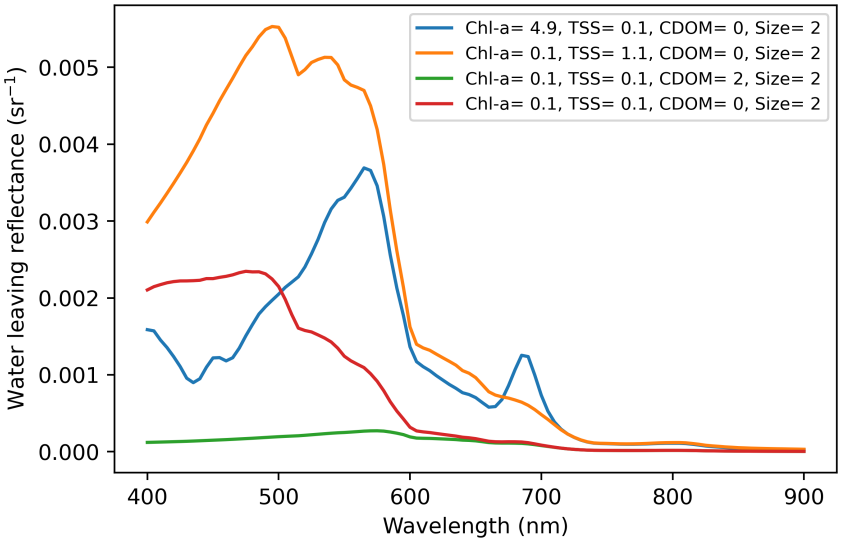

Figure 1.2: Example reflectances (modelled) showing characteristic spectra resulting from the dominance of different constituents. Blue= [Chl-a] (mg/m^3) dominant, orange= TSS (g/m^3) dominant, green= CDOM (m^{-1}) dominant, and red= water dominant. The models used to generate the spectra are described in section 2.2.

The lack of *in situ* bio-optical and radiometric measurements of inland and coastal regions has propelled the development of synthetic datasets that seek to represent the primary drivers of optical complexity found in such water bodies. River plumes are commonly known to be dominated by terrigenous sediments which vary in concentration depending on seasonality, geographical position, riverbank slope and climate change (Chen et al., 2011; De Rose et al., 2005; Walling, 2006). Mixed phytoplankton assemblages proliferate at the interface between riverine and marine systems (Carreto et al., 2008; Lunven et al., 2005). However, phytoplankton absorption features may be masked in the bulk optical R_{rs} by the high degree of backscatter from sediment particles. Further increasing the optical complexity is the presence of various CDOM contributors. Thus, the tendency for water bodies to be optically diverse regionally requires flexible algorithms for geophysical product retrievals.

1.3.2 Case 2 Ocean Colour Algorithms

The first ocean colour algorithms used the empirical spectral relationship between the maximum R_{rs} retrieved in the blue ($R_{rs}443 < R_{rs}490 < R_{rs}520$) to the green bands ($R_{rs}555$), which continues to work well in Case 1 open ocean waters, where the optical effects of phytoplankton dominate the resulting R_{rs} signals (O'Reilly et al., 1998). The OC4ME standard Case 1 data product for Sentinel-3 OLCI's sensor used the Maximum Band Ratio approach, which is derived from the classical algorithm developed by O'Reilly et al. (1998).

Due to the optical complexity and variability of Case 2 waters, several semi-analytical algorithms have been developed to supplement the simplistic historic empirical algorithms, by addressing each constituent explicitly. The approach is demonstrated in Figure 1.1, where the relationship between AOPs (i.e. R_{rs}) and IOPs are used with radiative transfer theory to retrieve biogeochemical parameters, such as [Chl-a]. A numerical ocean colour model developed by Carder et al. (2004) estimates cell size and chloroplast counts using parameters for sea surface temperature, nitrogen depleted temperature, and CDOM absorption to quantify phytoplankton biomass in optically complex waters (Carder et al., 2004). Evers-King et al. (2014) used a semi-analytical inversion algorithm to assess the sensitivity of R_{rs} to phytoplankton cell size along the Benguela current using a modelled synthetic dataset. However, semi-analytical approaches are greatly dependent on accurate knowledge of the IOPs, which may be regionally specific (Aurin & Dierssen, 2012). The interaction of IOPs in optically complex water bodies may result in ambiguity in R_{rs} signals (Defoin-Platel & Chami, 2007), and in turn, impaired geophysical product retrieval accuracy.

Another approach to Case 2 biogeophysical parameter retrievals is the application of artificial intelligence solutions for in-water constituent estimations. Previous studies in the late 1990s (Atkinson & Tatnall, 1997; Keiner, 1999), when general deep learning methodologies were being established, have proposed the use of artificial neural networks in remote sensing to predict Chl-a pigments from R_{rs} signals in optically complex and highly turbid waters. More recently, the use of Multi-Layer Perceptron (MLP) models has allowed for multiple constituents predictions with singular R_{rs} retrievals, and additionally, the ability to identify and remove the ambiguous influence of the atmosphere (e.g. Kravitz et al., 2021). Some models are trained using a synthetic dataset composed of concentrations of the constituents and R_{rs} pairs, such as the Oceana and Land Colour Instrument (OLCI) Neural Network Swarm (Hieronymi et al., 2017; Kravitz et al., 2021), where the added variability produced by the widely parameterised synthetic dataset makes the model functional for most optically complex water types. The Inverse Modelling Technique (IMP) is another deep learning approach used for OLCI's standard [Chl-a] and TSM neural network data products, which adopts inverse neural network modelling trained on a substantially large synthetic dataset (Doerffer & Schiller, 2007).

1.3.3 Atmospheric correction

In addition to the complicated additive interactions of in-water constituents, molecules and aerosols in the atmosphere also contribute to the signals satellites retrieve. As energy from the sun enters the atmosphere, a portion of the light can be absorbed along the blue wavebands, or scattered, through variations in molecular scattering (i.e. Rayleigh, Mie, etc.), in multiple directions, one possibility being towards the satellite sensor itself. The atmosphere contributes approximately 90% of the R_{rs} signals satellites detect (Wang, 2010). Developing an algorithm to remove the convoluted effects of the atmosphere from R_{rs} retrievals is paramount in remote sensing research.

Coupled oceanic and atmospheric semi-analytical models have been proposed to combat the effects of sun glint and scattering aerosols over optically complex waters, primarily using equations that are regionally parameterised (Steinmetz et al., 2011), such as POLYnomial for MERis (POLYMER) for Northern Hemisphere water bodies. Others, such as the Case 2 Regional CoastColour (C2RCC), use a neural network approach to correct for the atmosphere, using the Hydrolight (Numerical Optics Ltd.) radiative transfer models for the in-water light field and MODTRAN-generated look-up tables, for atmospheric influences on satellite retrieved TOA radiances (Brockmann et al., 2016). Regional atmospheric conditions are variable and can impact satellite geophysical product retrieval algorithm performances.

1.4 Case study: Agulhas Bank rivers

The study region is situated along the south coast of South Africa, 33 km to the south-west of the city of Mossel Bay (Figure 1.3) where four diverse rivers drain into the Southern Benguela and Indian Ocean over an extended continental shelf region known as the Agulhas Bank. The Breede, Duiwenhoks, Goukou and Gouritz rivers, each have distinct geomorphological attributes and catchment ranges which result in a wide variety of river mouth dynamics and estuarine-marine interactions in the study region. The rivers are situated above the Agulhas point of retroflection, which subjects the system to spontaneous warm, saline eddies branching off from South Africa's dynamic western boundary Agulhas point current (Penven et al., 2001; Swart & Largier, 1987). The combination of cyclonic (upwelling) eddies, and coastal upwelling, coupled with seasonal marine water column stratification and mixing may result in unique biogeochemical signatures observed over various temporal and spatial scales (Penven et al., 2001; Swart & Largier, 1987), influencing the nutrient characteristic of the region (Lutjeharms et al., 1996). As a result, the river plume region is subjected to a variety of both riverine and oceanographic forcings, as well as unique primary production attributes (Jacobs et al., 2022; Poulton et al., 2022), creating an energetic and complicated system for remote monitoring.

The climate of the region follows a Mediterranean regime of high rainfall commonly occurring during the austral winter months, i.e. May to August, and a dry summer season from October to January. All of the rivers are permanently open, flowing through the Bredasdorp Limestone aeolianite ridges, which substantially increased the calcium carbonate concentrations of the river plumes (Flemming & Martin, 2021). Historical *in situ* measurements made from 1978-1980 were aimed at assessing the spatial and temporal distribution of [Chl-a] along South African coastlines, particularly between the western reaches (St. Helena Bay) to the south coast (Shannon et al., 1984). [Chl-a] along the Agulhas Bank region were described as moderate to high, where phytoplankton [Chl-a] ranged from 3 to 10 mg/m^3 (*ibid*). These [Chl-a] were similar to those measurements throughout the year, however, there have not been many historical studies on the individual rivers and their corresponding TSM loads conducted.

The Breede River mouth (Figure 1.3) is located at 34°.483' South and 20°50.833' East, with an estimated catchment length of 313 km, the river passes three biomes, namely the Fynbos, Succulent Karoo and the Nama Karoo, mostly consisting of agricultural lands. The Breede has the fourth highest mean annual runoff in South Africa at approximately 1873 x 10^6 m^3, and transports an estimated sediment bulk density of 1.5 t/m^3 per year (Flemming & Martin, 2021). Additionally, significant concentrations of epipelic diatoms were found in the Breede estuary (Bate, 2013). These findings suggest that the nutrient content of the estuary may be substantially high as diatoms have a relatively large nutrient uptake rate during the process of photosynthesis to build up their biomass. While in comparison to smaller species, such as dinoflagellates and cyanobacteria, whose relative diameter dictates the amount of [Chl-a], and subsequent need for nutrients, larger cells such as diatoms require more nutrients for phytoplankton growth.

The Duiwenhoks River (Figure 1.3), adjacent to the Breede River, is encompassed by a range of elevated land sloping into the river with fynbos vegetation hinging alongside the corresponding hills. Based on field observations during this study, the Duiwenhoks River had a very brown-stained estuary. This observation may lend room for the assumption that tannins (i.e. CDOM) were seemingly released from the plants, seeped into the ground watershed, and later accumulated in the river, causing such colouration. The river is 54 km long, with the estuary mouth located at 34°21'54.31" South and 21°0'0.51" East, collecting agricultural and pastoral discharge from pesticides, herbicides and fertilizers used by farmers along the river channel (DWS, 2014). Accompanying the agricultural chemical runoff is a significant increase in land erosion, in turn, the sediment yield was detected (*ibid*). Extensive farming practices along the river beds caused the natural vegetation and associated soil microbes, which held terrestrial sediment in place, to decline. As a result, more dried topsoil entered the aquatic water body, increasing riverine sediment loads (*ibid*). In addition to higher sediment loads, the reduced impact of floods in the province has decreased the flushing time of

matter in the estuary. The reduced flow rates have led to a higher retention of sediments and an increase in the intrusion of marine sediments in the Duiwenhoks estuary (*ibid*).

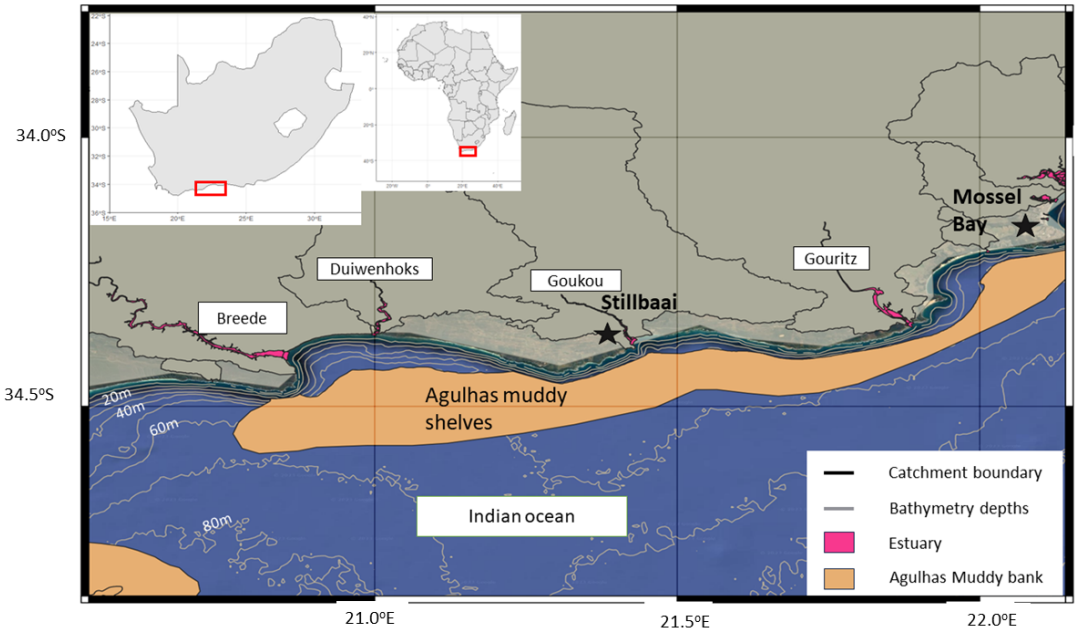

Figure 1.3: The geographical locations of the Breede, Duiwenhoks, Goukou and Gouritz, and their respective estuaries. Marine depth contour and substratum are included and were mapped by SANBI (2018). The bathymetry depths are denoted in the image, with 10m internals.

The Gouritz River is the third largest river catchment in South Africa, measured to be 1045 km long, the river flows past a multitude of ecoregions, into the Cape Langeberg Mountain ranges. The river mouth, situated at 34°20'37.3" South and 21°53'7.21" East, is the closest river to Mossel Bay (33 km away), with a shallow dynamic opening which shifts according to regular tidal and freshwater regimes (DWS, 2015). Major changes in the river-estuary system from historical conditions are related to catchment discharge reductions, with a comparatively minor effect by increased sediment concentrations (*ibid*). The Gouritz River catchment has however been largely unimpacted by high nutrient inflows, in comparison to its adjacent counterparts (*ibid*). The Goukou River (hereafter named Goukou) is situated in the lightly populated town of Stillbaai (300 km east of Cape Town) with the river mouth located at 34°22'43.36" South and 21°25'22.19" East, and a total length of 64 km. The Goukou delivers a low volume of fine sediment which largely dissipates when entering the marine realm (*ibid*). However, land erosion from agricultural practices has increased the sediment yield in the estuary over time (*ibid*). The river mouth is described as permanently open, even though the shallowest point of the estuary is < 2.5 m, fairly close to the river mouth delta (*ibid*). However, river floods, in conjunction with dynamic tidal shear, flush out both riverine and marine sediment,

ensuring the mouth remains permanently open throughout the year (*ibid*).

Estuaries form the interface between aquatic and marine systems, acting as a refugia for organisms escaping predators or as a nursery ground for juveniles. Changes in turbidity maxima and phytoplankton populations within estuarine ecosystems may impact young survivorship and subsequent growth rates. A study conducted by Lamberth et al. (2008) on fish migration within estuaries along the Agulhas Bank found that the multiple fish species tagged at the Goukou were also detected in the Breede and Gouritz, highlighting the connectivity between the estuaries. Organisms which rely on visibility to detect prey will be negatively influenced by increased sediment loads retained in estuaries, as a result, population structures may shift amongst estuaries (Kemp et al., 2011; Lowe et al., 2015). This was observed by Murray et al. (2018) in the Goukou, where fish populations tended to migrate during the high rainfall winter season into other neighbouring estuaries. Some of the more adverse effects may be declining populations as hypothesised by Murray et al. (2018). This in turn may have a far greater impact on the south coast recreational, subsistence, and inshore fisheries (Lamberth & Turpie, 2003). Ignoring the importance of monitoring river plume discharge, and not setting appropriate precautions to minimize river flow reductions, as well as increased run-off from agricultural activities and land pollutants, may have dire effects on the organism in estuaries and the adjacent marine system within the coming years.

Unlike the coastal ocean along the West and East coasts of South Africa - which has a considerable collection of literature concerning the efficiency of available satellite geophysical products and newly developed algorithms (Bernard et al., 2014; Evers-King et al., 2014; Smith et al., 2013) - the marine optical properties of the Agulhas Bank region are not well characterised. A historical study by Shannon et al. (1984) used satellite images from NASA's CZCS to estimate near-surface [Chl-a] in the region during the late 1970s; notably, they mentioned the importance of appropriate atmospheric correction, which has historically been (and remains) a major issue along South African coastlines that will be further investigated in this study. The Agulhas Bank is an ideal study area to conduct new research, quantify in-water constituents, and assess atmospheric correction processes in optically complex waters.

1.5 Aims and Objectives

1.5.1 Motivation

River plumes are dynamic regions along the coastline that are subject to biological and physical forcings, supporting both estuarine and marine ecosystem functioning. The ramifications of ignoring the importance of riverine outputs may have a cascading impact on water quality, economic pursuits, and environmental health (Drinkwater, 1986; Lathrop & Lillesand, 1989; Rudorff et al., 2018). [Chl-a] and TSS, proxies to phytoplankton biomass and sediment loads are underscored as vital riverine runoff yields in the present study. Improving our ability to monitor such distributions along coastal systems may be the most effective method to understand the consequences of anthropogenic activities, seasonality and climate change. Developing a tool that can frequently gather information along fine temporal and spatial scales is essential when approaching water management planning and legislative actions. Remote sensing capabilities have advanced over the past twenty decades, along with our computational abilities and large open-access data pipelines, allowing for more cost-effective methods of quantifying phytoplankton biomass and sediment runoff from land-based watersheds. However, deriving optimal methodologies to quantify such concentrations may be challenging as river plumes are one of the most optically complex water bodies on Earth. The amalgamation of mixed phytoplankton assemblages, various size classes of riverine and marine sediment, flocculation and unique CDOM concentrations, depends on both land and marine biological and physical processes, with the added effects from seasonality, climate change, and human impact. Standard Sentinel-3 geophysical data products for [Chl-a] and TSM have rarely been validated over highly turbid, optically complex water bodies, such as river plumes. In addition, accurate atmospheric correction is difficult to achieve over coastal regions where atmospheric conditions are known to be variable and dependent on land-based activities. Mapping [Chl-a] and TSS concentrations may require a regional approach, by developing an ocean colour product which captures the full extent of the possible optical characteristics of the prescribed study area possesses, with diverse AC techniques.

1.5.2 Research Aim

To evaluate the current standard Sentinel-3 satellite data products for [Chl-a] and Total Suspended Matter (TSM), with the understanding that atmospheric correction (AC) is difficult in the Agulhas Bank region. Thereafter, test a regionally parameterised MLP model trained on synthetic (AC free) R_{rs} to try and determine the impact of 1) AC issues and 2) biogeophysical product algorithm uncertainties.

1.5.3 Research Objectives

1. Evaluate the performance of the default ocean colour biogeochemical products ([Chl-a] and TSM) from the standard, C2RCC and POLYMER satellite atmospheric correction processors in the region using *in situ* data.

2. Use *in situ* measurements for [Chl-a] and TSS to parameterise the IOP models that will generate a synthetic dataset of the expected reflectances over the selected Agulhas Bank river plumes.

3. Apply a state-of-the-art neural network to the training synthetic dataset to develop a regional ocean colour product for TSM and [Chl-a].

4. Examine the efficiency of the regional model using the standard, C2RCC and POLYMER atmospheric correction processors to assess the effectiveness of different AC techniques in correcting for the Agulhas Bank region.

Chapter 2

Methods

2.1 Sample collection and processing

2.1.1 Water collection and filtration

Water samples were collected in 2022 between the 5^{th} of June and the 7^{th} of June for the winter season data set, as well as the 29^{th} of September to the 1^{st} of October, which represented the spring season. Sampling locations were spaced approximately 1 km (300 m resolution x 3 pixels) or more meters away from one another to avoid pixels overlapping when extracting satellite match-ups. During the first sampling trip, the water samples were collected along a transect from the marine system up to 2 km inside the estuary for each of the sites. The aim was to collect information on the concentration ranges of in-water constituents that could be expected in these systems. During the second sampling trip in October, the sites were moved further away from the coastline and estuary in order to improve the number of high-quality satellite match-ups. The sampling locations from the two trips can be seen in Figure 2.1, while the *in situ* concentrations of Chl-a and TSM locations can be seen in Appendix A.1 for all of the rivers and seasons.

Sea surface water samples were collected in 5 L blacked-out bottles, to minimize light penetration and subsequent primary production from the phytoplankton in the sample. At the sampling locations, the bottles were rinsed in the seawater sample at the site three times before they were filled with the sample. The bottles were sealed, covered and kept in a blackened large container before they were transported to the laboratory. The samples for both the gravimetric and fluorometric analysis were filtered less than 12 hours after collection. When in the laboratory, all of the apparatus were rinsed

with Milli-Q, three times, before the filtration began. This included the filtration flask, filter funnel, glass vials, squeezy bottles and measuring cylinders. From the 5 L of surface seawater collected, a 500 ml measuring cylinder was used to separate 250 ml of the sampled water, which was later used for the fluorometric filtration. The remaining sample was kept for the gravimetric component. After both the fluorometric and gravimetric filtrations were complete, the unused seawater was discarded and the samples were stored.

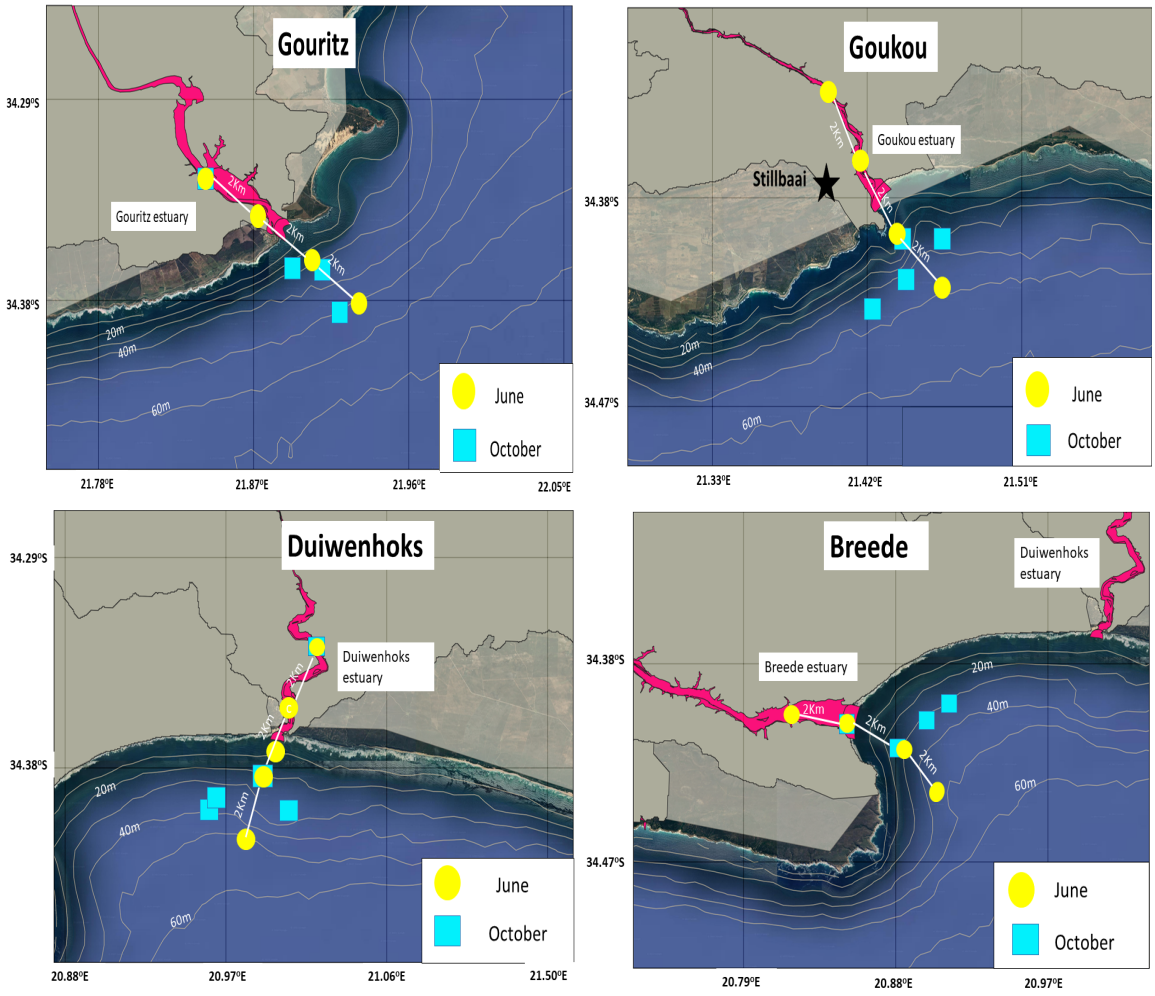

Figure 2.1: The June (yellow circles) and October (blue squares) sample locations along the Gouritz, Goukou, Duiwenhoks and Breede rivers in 2022.

2.1.2 TSM gravimetric analysis

Determining sample volume

The filtration for the Total Suspended Matter (TSM) followed the protocols developed by the Environmental Protection Agency (1983), which can be applied to both chemical and sedimentary analysis. Before the field sampling, the filter pads that were going to be used were enclosed in tin foil and were placed in a furnace for 1.5 hours at 400 °C, to burn off all the organic matter on the filters prior to sampling. The filter pad weights were recorded in milligrams and were placed in closed petri-disks inside a desiccator to be taken to the field campaign. The amount of water filtered per site was determined by the concentration of matter collected on the filter pad, whereby the filtration process would be halted when the filtration time was extended by 10 minutes. The volume of water decanted into the measuring cylinders was meticulously recorded in litres on a bench sheet during the process to ensure future TSM calculations were as accurate as possible.

The volume of seawater filtered ranged between 0.5 to 2.5 L. The graduated cylinder and filter were rinsed with 60 mL volume of deionized water, to allow for complete drainage between washings, and to ensure the salts were displaced and removed from the filter pad. The suction ran for 3 minutes after the filtration was completed for each sample. The filter pad was folded, sample side in, and placed in labelled foil pouches. The samples were stored in a -20 °C freezer before the drying and weighing process took place. Samples were processed within 2-3 days after the match-up *in situ* samples to the satellite retrievals were collected.

Furnace and weighing

The filters were placed on foil tins in an oven at 104 ± 1 °C and were set to dry for a minimum of 1 hour. Thereafter, filters were removed from the oven and were cooled to room temperature in a clean laboratory. The samples were weighed to the nearest 0.1 mg, to retrieve the first mass of the filters. The filters were placed in the oven for a second time under the same conditions and weighed again to ensure that the mass decrease was less than 0.5 mg, ensuring that the filters have been dried fully and no water contributed to the filter pad mass during the first weighing. The dried filter pads were individually weighed in grams, to obtain the total organic and inorganic components of the sample, namely TSM. After weighing these filter pads were placed into aluminium foil and labelled with a pencil on the outside of the sample's crucible, the samples were again placed in a 450 °C furnace for 1.5 hours, to burn off all of the organic matter in the sample. Thereafter, the furnace was turned off and the samples were removed after 24 hours to be individually reweighed, achieving the inorganic component of the sample, or Total Suspended Sediment (TSS), which will be used for the sediment

mineral component of the bio-optical model in section 2.2.

2.1.3 Fluorometric analysis

During the June sampling trip, the fluorometer was available to analyse samples within the day of collection after the filtration process. The October fluorometric samples were folded in aluminium foil and were placed in a dewar containing liquid nitrogen, to conserve the photosynthetic pigments and avoid further primary production taking place. When the samples were ready for analysis, the filters (47 mm, 0.7 μm nominal pore size) were individually placed in glass vials and submerged in 8 ml of 90% acetone. The glass vials were thereafter placed in a -20 °C freezer for 24 hours.

The next day, the glass vials were removed from the freezer and stored in a dark drawer for 45 minutes, to ensure that the sample equilibrated to room temperature. Before each session of analysis, a blank (8 ml of 90% acetone) raw fluorescence was recorded on a log sheet. The glass vial with the filter pad sample was inverted gently a few times to lightly extract the Chl-a pigments from the sample. The sample was poured into a glass test tube which was later placed in the fluorometer. The outside of the test tube was dried with laboratory tissue before the fluorometric analysis commenced. The raw fluorescence was measured and recorded on a log sheet. The test tube was rinsed 3 times with Milli-Q between samples to avoid cross-contamination. The used filters were thrown away and the acetone was discarded into an acetone waste jar. The measurement in raw fluorescence units (RFU) was later converted into [Chl-a], where equation 2.1 corrects for a blank measurement. Thereafter, equation 2.2 gives [Chl-a] (a proxy for phytoplankton biomass), with the calibration set as a constant gain factor of 0.240001482 for the June samples and 0.11878639 for the October analysis. The calibrations of the fluorometer varied between seasons as different fluorometers were used to carry out the analysis.

$$\textbf{Corrected RFU} = \textbf{Measured RFU - Blank} \tag{2.1}$$

$$\textbf{[Chl-a] (mg/m}^3\textbf{)} = \textbf{Correct RFU} \times \textbf{Calibration} \times (\textbf{Volume water filtered} \div \textbf{Acetone volume}) \tag{2.2}$$

2.2 Synthetic dataset

Various in-water constituent IOP models were selected to represent likely occurring ranges of Inherent Optical Property (IOP) variability in the study area. Ranges of concentrations were parameterised based on a combination of *in situ* measurements and reported literature values, with some assumptions about expected interseasonal variability and phytoplankton size variability. The parameterisation of

the [Chl-a] and TSM ranges of the IOP models was larger than the *in situ* measurements ranges, to allow for more variability in the water types which may be present in the study sites. These IOP combinations were inputted into Hydrolight-Ecolight's (Numerical Optics Ltd.) four component model, with generalized solar geometry, atmospheric conditions etc., as described in Lain et al. (2014). In this way a dataset of 18 000 simulated hyperspectral R_{rs} was produced, over a wavelength range from 400 to 900 nm, at 5 nm resolution, intended to encompass a reasonable range of natural R_{rs} observed in the region.

2.2.1 Inherent Optical Properties models

Phytoplankton component

Locally developed and validated (Bernard et al., 2009; Lain et al., 2014) Equivalent Algal Populations model of phytoplankton IOPs was used to represent generalized coastal phytoplankton communities likely to occur in the study region. The model considers the effective diameter of the chloroplast, as well as the encapsulating cytoplasm. From a bio-optical perspective, the total effects of surface area play a major role in the absorbency of light, as a larger surface area results in a higher magnitude of light absorbed. The capabilities of varying the equivalent size distributions with respect to the [Chl-a] allow for both biomass and phytoplankton functional types parametrisations of the synthetic dataset. *In situ* measurements of [Chl-a] collected during this study, as the dominant light-harvesting pigment of most phytoplankton types, were used to parameterise phytoplankton concentrations. Mixed assemblages with varying proportions of small (2 μm) and larger (8 μm) cells were simulated as a simple representation of cell size variability to be expected in coastal waters of South Africa (Lemley et al., 2015).

Coloured dissolved organic matter

The spectral absorption of Coloured Dissolved Organic Matter (CDOM) can be represented as a simple decaying exponential shape (Bricaud et al., 1981). The rate of CDOM absorption decay is frequently approximated together with the absorption of detrital matter which follows a similar curve (Carder et al., 1991). The term used to describe the absorption of CDOM in the following equations is denoted by a_{gd} and represented by the following equation:

$$a_{gd}(\lambda) = a_{gd}(400)exp[-S(-400)] \tag{2.3}$$

where a_{gd} (λ) is calculated from the absorption at 400 nm, with a regionally variable choice of slope, S. The parameterisation of $a_{gd}(400)$ was accomplished in the same manner as the Lain et al. (2014)

study, based on coastal observations.

Mineral

In the absence of appropriate optical measurements of river plume sediment, the ”average mineral” IOPs provided with Hydrolight were used as a generalised representation of mixed assemblages of commonly found sedimentary particles (see Ahn (1990)). The concentrations of such particles were parameterized using the *in situ* TSS measurements from the gravimetric analysis. The difference between Total Suspended Matter (TSM) and TSS concentrations is acknowledged and is considered unavoidable in this context. It is discussed in the results section how a further constrained parameterisation of TSS concentration was required and applied.

2.3 Satellite data processing and extraction

2.3.1 Satellite data

For the purposes of this study ocean colour data from the Ocean and Land Colour Instrument (OLCI) on board the Sentinel 3A and 3B satellite constellation were used; OLCI was chosen due to its near-daily overpass revisit cycle, as well as the retrievals of high spectral resolution (21 bands in the visible range) data gathered at 300 m spatial resolution. Level-1b and Level-2 OLCI ocean colour products were accessed and downloaded through the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) data store (https://data.eumetsat.int/). OLCI Level-2 data products are provided with bio-optical variables which are derived using algorithms trained on global datasets aimed at quantifying constituents in marine water bodies (i.e. OC4ME, [Chl-a] and TSM neural networks). In addition to various geophysical products, Normalized Water Leaving Reflectances (ρ_ω) are also provided with Level-2 files. However, the synthetic dataset used in this study models remote sensing reflectance (R_{rs}) and not ρ_ω. Thus, R_{rs} was derived from ρ_ω as follows:

$$R_{rs} = \frac{\rho_\omega}{\pi} \tag{2.4}$$

2.3.2 Atmospheric correction and product descriptions

The R_{rs} outputs from three different Atmospheric Correction (AC) processors were used in this study. These were either obtained as Level-2 files from EUMETSAT, or from ACs applied directly to the Level-1b files. Each AC method also provides outputs from one or more ocean colour products (i.e.[Chl-a], TSM etc.). Descriptions of the ACs and their associated ocean colour product derivation are mentioned below, and the labels used hereafter are summarised in Table 2.1.

Table 2.1: Summary table of the models either used or developed in the present study to quantify TSM and [Chl-a].

Model	Constituent	AC	Algorithm
Chl_{L2-MLP}	[Chl-a]	Standard R_{rs}	Regional MLP
TSM_{L2-MLP}	TSM	Standard R_{rs}	Regional MLP
$\text{Chl}_{C2RCC-MLP}$	[Chl-a]	C2RCC R_{rs}	Regional MLP
$\text{TSM}_{C2RCC-MLP}$	TSM	C2RCC R_{rs}	Regional MLP
$\text{Chl}_{POLY-MLP}$	[Chl-a]	POLYMER R_{rs}	Regional MLP
$\text{TSM}_{POLY-MLP}$	TSM	POLYMER R_{rs}	Regional MLP
$\text{Chl}_{L2-OC4ME}$	[Chl-a]	Standard R_{rs}	OC4ME
Chl_{L2-CHL_NN}	[Chl-a]	Standard R_{rs}	IMP
TSM_{L2-TSM_NN}	TSM	Standard R_{rs}	IMP
$\text{Chl}_{C2RCC-CHL\_NN}$	[Chl-a]	C2RCC R_{rs}	C2RCC
$\text{TSM}_{C2RCC-TSM\_NN}$	TSM	C2RCC R_{rs}	C2RCC
$\text{Chl}_{POLY-POLY}$	[Chl-a]	POLYMER R_{rs}	POLYMER

Standard Baseline Atmospheric Correction (BAC)

This is the standard AC applied to all Level-2 OLCI data downloaded from EUMETSAT. The algorithm involved three major steps: glint and white cap correction, Case 2 Near-Infrared (NIR) reflectance estimations, and aerosol and Rayleigh correction (ESA, 2023b). The glint and white caps correction removes pixels of high glint, such as ice and aerosol-loaded regions, as well as pixels not corrected for sun glint. The Case 2 NIR reflectance estimation for highly turbid waters first removes estimates for the effects of aerosol scattering in the NIR bands, where the remainder is identified as the estimate for TSM. Thereafter, a dedicated sediment-dominated Case 2 flag is applied to the TSM estimate, which switches from the standard BAC algorithm to the Alternative Atmospheric Correction (AAC) procedure (ESA, 2023a). The AAC provides water-leaving reflectances using a neural network approach for turbid Case 2 waters and for regions contaminated by sun glint. However for waters not classified as turbid Case 2, aerosol and Rayleigh correction is carried out in the NIR bands, using predefined models and tabulated radiative transfer computations to determine their contribution to received reflectances for atmospheric correction to take place.

The ocean colour products used from EUMETSAT used the reflectance after BAC correction for their OC4ME, [Chl-a] and TSM neural networks. The OC4ME [Chl-a] product uses a Maximum Band Ratio semi-analytical algorithm for typical Case 1 waters, by Morel et al. (2007) adapted from O'Reilly et al. (1998) findings. OC4ME's iterative semi-analytical approach is described in ESA (2023d). The neural network [Chl-a] and TSM were retrieved using an Inverse Modelling Technique (IMP) by an inverse radiative transfer model-neural network trained on an extensive synthetic dataset consisting of approximately 550 000 (Doerffer & Schiller, 2007), and described implicitly in ESA (2023c). In addition, the atmospheric conditions were robustly collected over the greater study area, by using the Aerosol Optical Thickness (AOT) and Angstrom exponent products for selected days near or during both sampling trips. These parameters were by-products of the previously mentioned AC approaches. The AOT is a measure of the density of the components in the atmosphere, increasing during high precipitation and smoke. The Angstrom exponent is inversely proportional to the size of particles in the atmosphere, where smaller particles induce a higher exponent value. Both metrics are dimensionless and were used to describe the type of aerosols present along the intervening atmosphere.

Case-2 Regional CoastColour Processor (C2RCC)

The SNAP-ESA Sentinel Application Platform v9.0.0 (http://step.esa.int) C2RCC v2.1 thermatic water processor for OLCI was applied to Level-1 Top-of-Atmosphere (TOA) radiances. C2RCC takes on a neural network approach, by using a radiative transfer derived synthetic dataset comprised of approximately 5 million water-leaving reflectance and TOA pairs, parametrized with a combination of various atmospheric and biological conditions (Brockmann et al., 2016). The neural network aimed to remove the effects of the atmosphere from reflectances, as well as derive IOPs for algal absorption (IOP_a) and sedimentary backscatter (IOP_btot), with covarying CDOM influences in the blue region of the visible over highly turbid and optically complex waters. The IOPs were later converted into mass concentrations, using parameterisation retrieved from a regression analysis using NASA's bio-Optical Marine Algorithm Dataset collection (Werdell & Bailey, 2005). The resulting conversions for TSM and [Chl-a] are as follows:

$$\textbf{Chl-a} = 21 \times \textbf{IOP\_a}^{1.04} \tag{2.5}$$

$$\textbf{TSM} = 1.06 \times \textbf{IOP\_btot}^{0.942} \tag{2.6}$$

where IOP_a is the IOP derivation for algal absorption, and IOP_btot is that of total suspended solid backscatter. All of the IOPs were defined at 443 nm wavelengths when training C2RCC's neural network. Both CHL_nn and TSM_nn products derived from the standard products were subjected to

the above NOMAD parameterisation.

POLYnomial-based approach MERis (POLYMER)

POLYnomial-based approach MERis (POLYMER) is a model derived by Steinmetz et al. (2011) to assist with the removal of atmospheric effects and sun glint, a major problem for most satellites, such as the historic MERIS sensor. Python was used to apply POLYMER v4.15 to OLCI Level-1 TOA radiances. A coupled atmospheric sun glint and Rayleigh scattering aerosol model was optimized along with an oceanic bio-optical [Chl-a] and non-covarying particle backscatter terms, to correct for the effects of the atmosphere, as well as derive water-leaving reflectances and a logged [Chl-a] product. After each spectral fitting, a loss function was calculated based on how well the atmospheric and oceanic model fits the spectra, after which the atmospheric contribution can be characterized by the parameters retrieved from the optimization process. As a result, a robust method to remove the effects of the atmosphere from ρ_ω retrievals is produced. A [Chl-a] product is derived by using the model's best fitting [Chl-a] parameter to the retrieved ρ_ω spectral shape. The atmospheric component, formulated primarily for Rayleigh scattering and sun glint, does not account for aerosol absorption, as well as the oceanic bio-optical model for the effects of CDOM, throughout the spectral matching process and [Chl-a] derivation.

2.3.3 Pixel extraction techniques

All Level-2 ocean colour outputs were geographically subsetted to the region 33 to 35 °South and 20 to 22 °East and saved as netCDF4 data files. Satellite pixel extractions were performed over the pixel closest to the station location in addition to the eight surrounding pixels, as demonstrated in Figure 2.2. This 3x3 megapixel was used to obtain the mean and standard deviation for all satellite-derived variables.

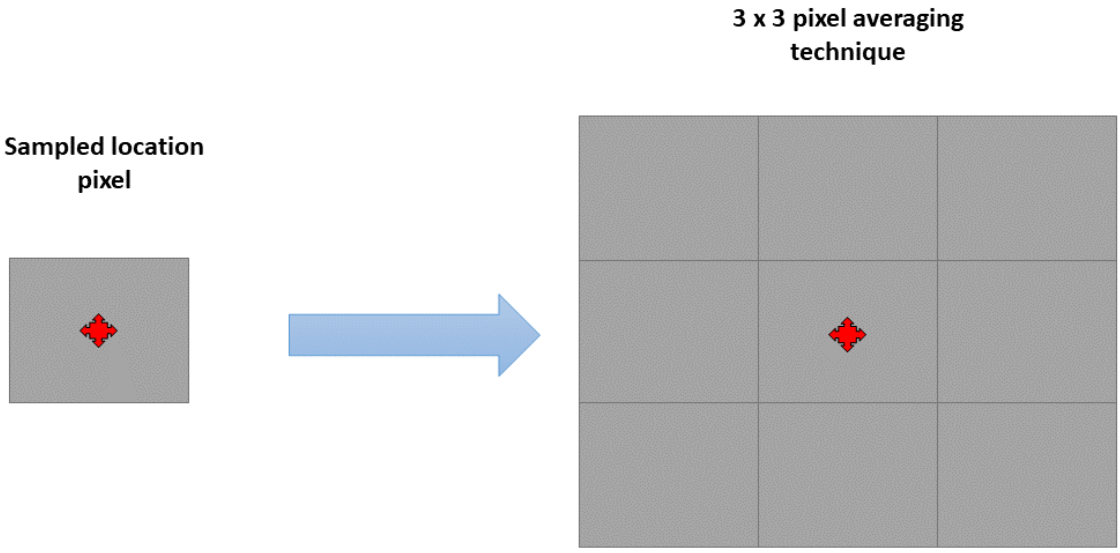

Figure 2.2: The pixel averaging technique over the sampled area, using a 3x3 pixel area procedure.

Averaging and obtaining the standard deviation for the images R_{rs} over the 3x3 pixel box for each of the study locations in Figure 2.1, were calculated as follows:

$$\mu = \frac{1}{n}\sum_{i=i}^{n} x_i \tag{2.7}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{2.8}$$

where μ is the mean, n is the sample size, x_i is the ρ_ω per a waveband and $\overline{x}$ is the population mean calculated prior, i.e. μ. The following wavebands were extracted: 400, 412.5, 442.5, 490, 510, 560, 620, 665, 674, 681, 709 and 755 nm. The maximum allowable time difference between a match-up *in situ* station and satellite overpass was <7 hours, which is higher than what is typically used for temporally variable systems such as estuaries. Time and budget constraints when sampling both the marine and estuarine sites created large time windows between the match-up *in situ* sample collection and the single Sentinel-3 overpass each day. All extracted pixels were subjected to quality assessment using the various quality and science flags provided with the Level-2 output files. The labels and type of flags differ according to the type of AC processor used, resulting in a difference in the flags used (Table 2.2).

The final quality assessment involved using the coefficient of variance, formulated by:

$$CV = \frac{\sigma}{\mu} \tag{2.9}$$

which measures the dispersion of the standard error over the mean of the R_{rs} data set. In the case of the 3x3 pixel box, the averaged ρ_{ω} and ocean colour products megapixel and their corresponding standard error were used to compute the CV along each waveband per site. Thereafter, sites with a median dispersion $>$0.15, were removed from the match-up dataset to minimize the effect of outliers (Bailey & Werdell, 2006). Furthermore, OLCI Level-2 atmospheric condition parameters, AOT, and Angstrom exponent were robustly averaged over the area map, and the standard deviations were computed before the fynbos fire (6^{th} June 2022), after (8^{th} June 2022) and for the 1^{st} of October 2022.

Table 2.2: The flags used when extracting R_{rs} and biogeophysical satellite data products from the respective ACs.

AC	Flags
Standard	'LAND','CLOUD','CLOUD_AMBIGUOUS', 'CLOUD_MARGIN','INVALID','COSMETIC','SATURATED', 'SUSPECT','HISOLZEN','HIGHGLINT','SNOW_ICE','AC_FAIL','WHITECAPS','ADJAC', 'RWNEG_O2','RWNEG_O3',,'RWNEG_O4','RWNEG_O5','RWNEG_O6','RWNEG_O7','RWNEG_O8'
C2RCC	'land', 'bright', 'invalid', 'cosmetic', 'sunlight_risk', 'dubious', 'sun_glint_risk', 'dubious', 'duplicated'
POLYMER	'cloudMask', 'landMask', and 'anomMask'

2.4 Neural Network Algorithm development

Developing an algorithm which captures the subtle variations in reflectance spectra (R_{rs}) produced by the spectral variability as described above for each constituent, as well as by varying constituent concentrations, requires state-of-the-art machine learning techniques. Deep learning was found to be the most effective method to train the multi-output R_{rs} dataset that Hydrolight produces (Kravitz et al., 2021), as opposed to other algorithms such as random forests, K-nearest neighbour and XGBoost. Like other neural networks, Multi-Layer Perceptron (MLP) models require training in order to gain the competency to determine the relationships between constituent concentrations and the bulk optical signal of the water body. The simulated dataset consisting of a unique spectral R_{rs} corresponding to discrete values for [Chl-a], phytoplankton size, CDOM concentration, and mineral (sediment), satisfies the requirements for training of MLP models.

To replicate the type of input received from satellites, the hyperspectral synthetic data (400 to 900 nm at 5 nm resolution - see Figure 2.3) were resampled to the OLCI wavebands (mentioned in section 2.3.3.). OLCI's wavebands were designed to cover optically significant regions such as the 400 to 675 nm ranges, which encapsulate phytoplankton and CDOM absorption, and values in the NIR bands, (above 700 nm), where sediment backscatter signals are observed (Figure 2.3). Thus, the MLP model training used interpolated hyperspectral synthetic data that matched the selected OLCI's R_{rs} wavebands used in this study.

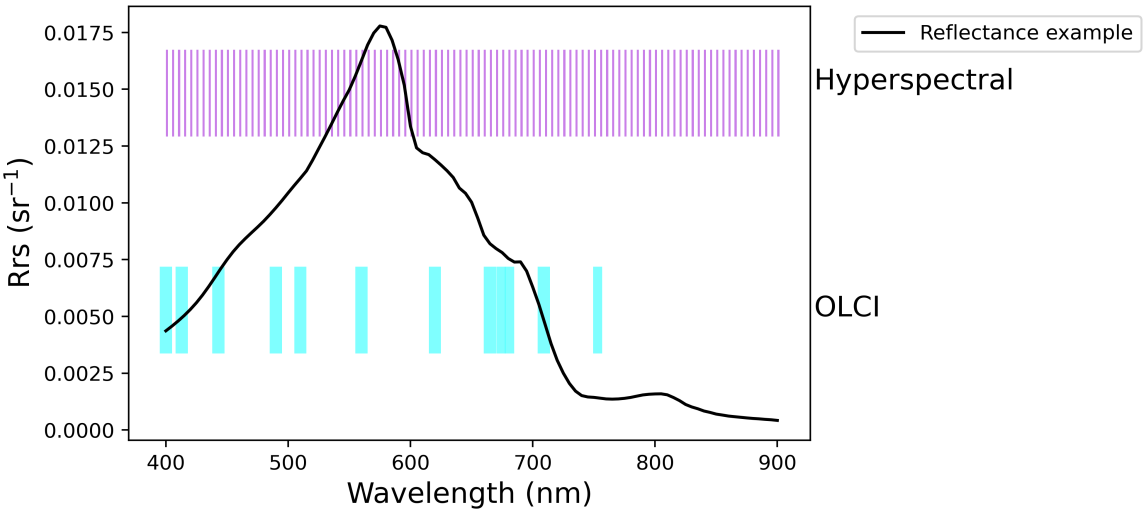

Figure 2.3: The range of R_{rs} data available from the synthetic hyperspectral dataset denoted by the purple vertical bands, and R_{rs} averaged along the selected bandwidths centred around the 12 wavelengths OLCI sensors retrieve.

The input R_{rs} was log-transformed, which is a convention in machine learning as signals become enhanced through the log transformation process. A floating point number of 0.001 was added to the output array to ensure constituents with concentrations modelled as zero did not tend to infinity when log-transforming the output variables, as well as it being unlikely any constituent would actually be zero. Thereafter, Python's scikit-learn *Standard Scalar* method was applied to both input and output variables to ensure that the data is similar across all records to accelerate the process of gradient descent and convergence during model training.

The MLP model was created by using the *sequential dense layer model* provided by TensorFlow (Abadi et al., 2016) and Keras (Chollet et al., 2015). These are common deep-learning repositories that house functions for most deep learning activities from image recognition to neuro-linguistic programming applications. The ranges of constituent concentrations in the simulated dataset were carefully parameterised to comfortably encompass those expected in nature, resulting in a wider

range of R_{rs} than that expected from the satellite. The complexity of interactions between constituent optical properties sometimes resulting in subtle signals in the generated synthetic dataset, makes a deep learning neural network appropriate for the identification of constituents and their concentrations from the extracted match-up R_{rs} to the *in situ* dataset.

The requirements for training the model included a 90%:10% dataset split. The scikit-learn (Pedregosa et al., 2011) random splitting *shuffle* method, was used to separate the training dataset from the cross-validation one (Figure 2.4), as well as for choosing the target variables upon which the MLP will make predictions (i.e. Chl-a, size, CDOM or mineral concentrations). The training dataset was further divided into four unique datasets using the scikit-learn *KFold* function, where each of the smaller training datasets was used to iteratively train the MLP model until the final model was devolved (Figure 2.4). This step adds a higher degree of randomness in the dataset used, allowing for a more stable final model, with cross-validation predictions relatively close to the modelled constituents concentrations. Hyper-parameter tuning was performed according to Kravitz et al. (2021) bottom-of-atmosphere (ρ_ω) model training (Table 2.3), where the parameters set followed the initial paper. However, the learning rate for the MLP model was adjusted based on the level of smoothness of the loss vs epochs functions. Besides the adjusted parameters, the following process remains a 'black box', in terms of the mathematical and computational mechanics that produce the cross-validation outputs, as well as those to be extracted from the satellite R_{rs}, as most of the computation was carried out by pre-built functions in Python.

Table 2.3: MLP model hyper-parameterisation adjustments.

Parameters	**Selection**
Dense layers	(500, 200, 50)
Activation function	ReLu
Optimizer	Adam
Loss function	Mean Absolute Error (MAE)
Learning rate	0.0001
Kfold batches	4
Training split	1:9

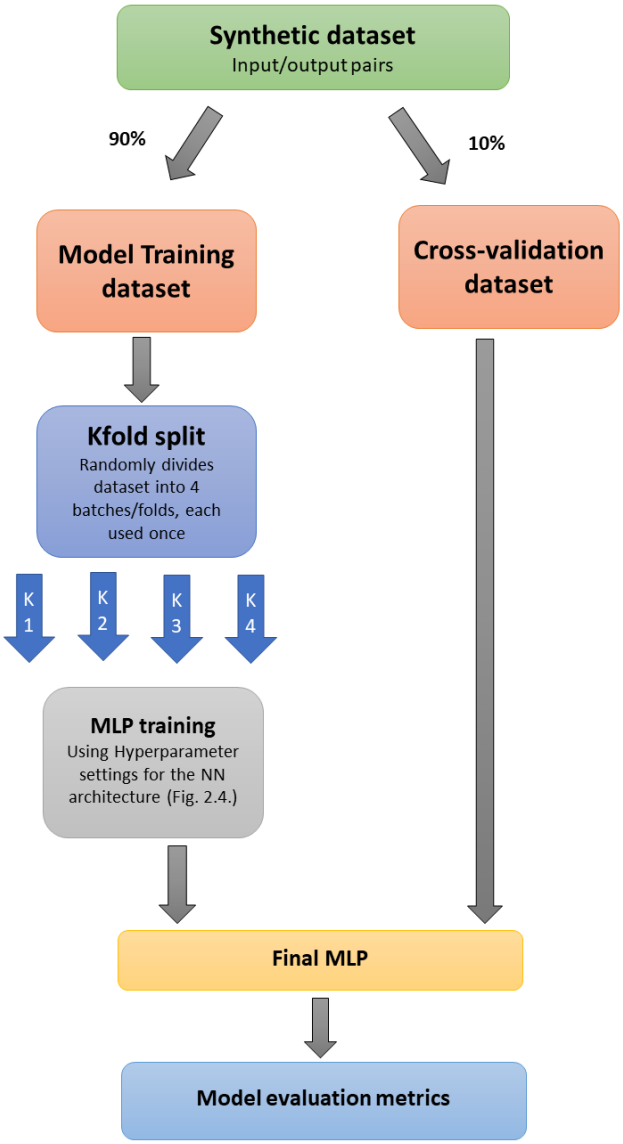

Figure 2.4: Diagram showing the procedure followed when training the MLP model.

The model training for each *KFold* batch used the neural network architecture in Figure 2.5. The process of training the MLP model involves Forward propagation (supervised learning), Error checking (with loss function MAE), and Backpropagation (unsupervised learning). Forward propagation entailed the process of feeding the MLP model with batches of reflectance spectra, with the important, unique features detected in each spectrum. The loss function, mean absolute error (MAE), was used to evaluate how close the output predicted constituent concentration variables are to the given input synthetic R_{rs} spectra. Corrections of the weights/nodes were accomplished during backpropagation, resulting in weights being adjusted according to the results from the error evaluation step. The model was trained in batch sizes of 200 spectra, with the hyper-parameter settings in Table 2.3, and the training was repeated 32 times (epochs), to ensure that the weights in the nodes were adjusted

accurately to minimize the prediction error of the MLP model. After the series of KFold batch training iterations, the model was evaluated using the cross-validation dataset, where the performance of the predictive capabilities of the model was assessed.

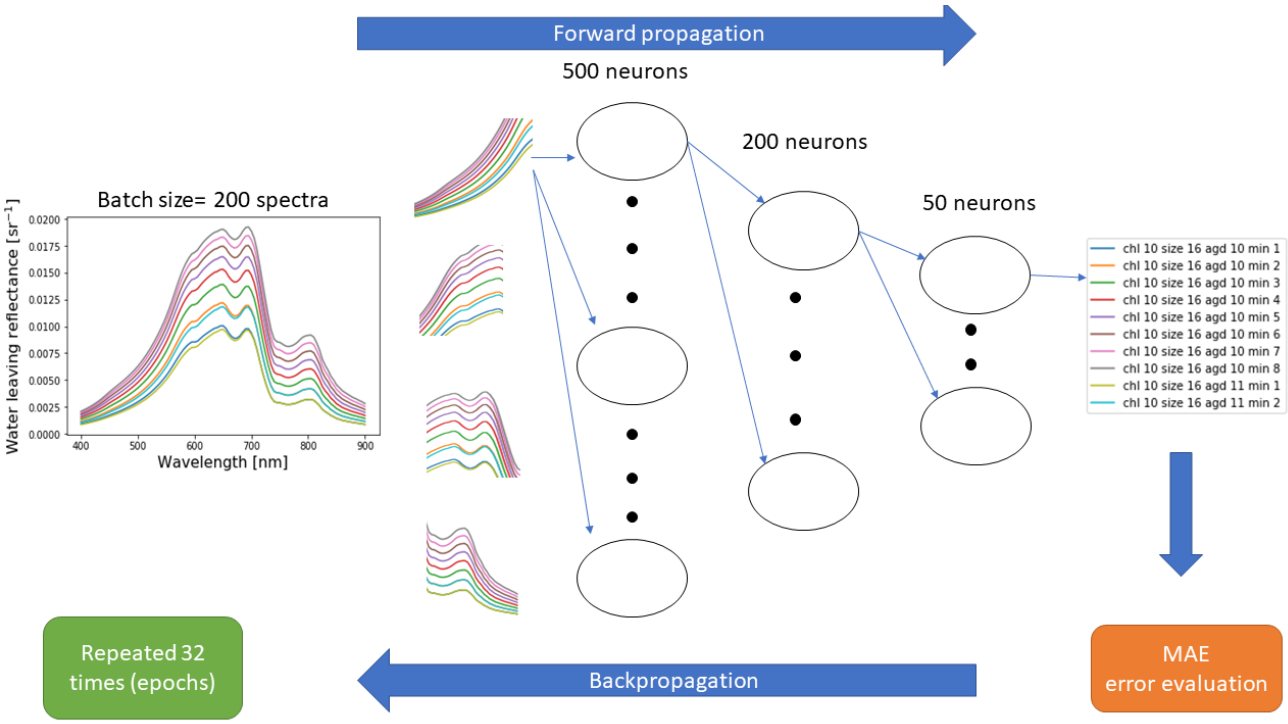

Figure 2.5: The architecture of the neural network used in the study.

After cross-validation, the selected model was saved, which included the sequential model (JSON file), the weight for the model (h5 file), and the output scalars of the data set saved as a Python pickle file. Saving the model resulted in standardized results when applying the MLP models to the match-up datasets and the regional plots, making statistical and visual evaluations comparable. Two models were created, with the first consisting of the full synthetic dataset of both estuarine and marine TSM and [Chl-a] over all of the rivers, and the other represents a constrained version of the full dataset using lower TSM concentration ranges (see section 3.2 for exact ranges).

2.5 Model evaluation

When applying the model to satellite R_{rs} retrievals, the R_{rs} dataset was first log-transformed, and standard scaled. The outputs from the MLP models were scaled back using the previously calculated scalar saved and were exponentiated to convert the values into standard measuring quantities for bio-optical constituents. Negative values, an artefact of the 'black pixel' method that Level-2 data products may be subjected to, were changed to 1e-5 to minimize the chances of ambiguous MLP model predictions.

Finally, model evaluations were performed to assess the competency of the developed MLP models in this study, with other Sentinel-3 OLCI ocean colour products. The MLP models were applied to the R_{rs} from the three different ACs to assess the best-performing AC for the selected study region. Neural network products for [Chl-a] and TSM provided by EUMETSAT's standard procedure were evaluated next to the internally SNAP-processed C2RCC [Chl-a] and TSM. Additionally, EUMETSAT's 'Case 1' water type [Chl-a] product, OC4ME, and POLYMER's spectral matching derived [Chl-a] estimations were assessed. The notations used to denote variations in the model type and AC for each product can be seen in Table 2.1.

This study follows the statistical techniques suggested by Seegers et al. (2018), which emphasises the notion that no matter how large the match-up data set is, the *in situ* to predicted fitted linear regression model distributions will very unlikely obey the requirements for the Gaussian distribution, needed for Ordinary Least Square regression statistical metrics such as R^2 and the associated slope. Seegers et al. (2018) instead suggest the use of median absolute percentage error (MAPE) for the accuracy of the model prediction and bias to describe the spread of error along the one-to-one line, by using the following equations:

$$\text{MAPE}(y, \hat{y}) = \text{median}(\mid \frac{y_i - \hat{y}_i}{y_i} \mid, \ldots, \mid \frac{y_n - \hat{y}_n}{y_n} \mid) \times 100 \tag{2.10}$$

$$\text{Bias}(y, \hat{y}) = \sum_{i=0}^{N-1} \frac{(\log y_i - \log \hat{y}_i)}{N} \tag{2.11}$$

Where N is the sample size of the match-up data set, $\hat{y}$ is the model outputs and y represents the *in situ* measured values. The method of reporting the bias in this study, a bias value of 1 will represent 'unity', while values below one are a negative bias and values above 1 represent a positive bias. This is attributed to the process of logging the data points, which complicated the reporting process as a

log-transformed bias of 0.3 does not represent 30% uncertainty, but rather a 100% bias though the back conversion $10^{0.3} = 1.995$, as mentioned by Seegers et al. (2018). Thus, the bias metric was treated in the following way,

$$\text{Bias}(y, \hat{y}) = 10 \exp(\sum_{i=0}^{N-1} \frac{(\log y_i - \log \hat{y}_i)}{N}) \tag{2.12}$$

Taking the exponent of the bias to base 10 allowed for easier reporting as values in the linear space. An additional reporting statistic, classically used in remote sensing, is the root mean square error (RMSE) metric, which aims to quantify the standard distribution of prediction errors amongst themodelledd value and the *in situ* measurements. However, RMSE tends to be highly sensitive to outliers, inflating the value, and making it difficult to compare model performances, especially with small datasets. Root mean square log error (RMSLE) was selected, as the nature of logarithms allows for an error metric of the proportional difference between the *in situ* and predicted values, as opposed to the usual difference calculated in RMSE (Iqbal et al., 2021). RMSLE was defined as follows:

$$\text{RMSLE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (\log(\hat{y}_i + 1) - \log(y_i + 1))^2}{N}} \tag{2.13}$$

where N is the sample size of the match-up data set, $\hat{y}$ is the model outputs and y represents the *in situ* measured values. However, RMSLE penalizes underestimations more so than overestimations, an attribute of logarithms right-skewed distribution, although not as greatly as RMSE accounts for larger errors from outliers (Ashraf et al., 2022).

Visual assessments using scatter plots and constituent predictions over all four river plumes were done to observe the general spread of the small match-up dataset (Appendix **??**). The models with the best-performing statistics for [Chl-a] and TSM estimations were applied to corresponding R_{rs} to assess the performance over the full spatial extent of the study region. Data flagging for the area plots followed the same procedures as for the pixel extractions (as described in section 2.3.3).

Chapter 3

Results

3.1 *In situ* data description

Across all seasons and all rivers, the chlorophyll-a concentration ([Chl-a]) ranged from 0.24 to 2.32 mg/m^3, while Total Suspended Sediment (TSS) extended from 0.52 to 13.41 g/m^3 (Table 3.1). June measurements for [Chl-a] and TSS were below average for most rivers, while the October samples were mainly above the mean (Figure 3.1). There were no obvious trends between TSS and [Chl-a] *in situ* values which delineated the estuarine samples from the marine (Figure 3.1). The June TSS measurements were more variable than in October when TSS within the estuaries was generally higher than in the marine environments (see Appendix A.1). [Chl-a] mean amongst the rivers was similar, however, the Gouritz TSS mean deviated from the other rivers, with the highest TSM concentrations recorded for both seasons (Figure 3.1). Contrastingly, the Breede River did have a lower TSS average than the other rivers, accompanied by the highest [Chl-a] values (Figure 3.1).

Table 3.1: Statistical descriptors of the *in situ* measurements collected during both June and October sampling trips in 2022, for [Chl-a] and TSS. The minimum (Min), 25th percentile (25th), median, 75th percentile (75th), maximum (Max), mean, standard deviation (StDev) and sample size (N) are reported in the following table.

In situ	**Min**	**25th**	**Median**	**75th**	**Max**	**Mean**	**StDev**	**N**
[Chl-a] (mg/m^3)	0.24	0.69	1.15	1.68	2.32	1.21	0.58	30
TSS (g/m^3)	0.52	2.01	2.46	4.32	13.41	3.56	2.81	30

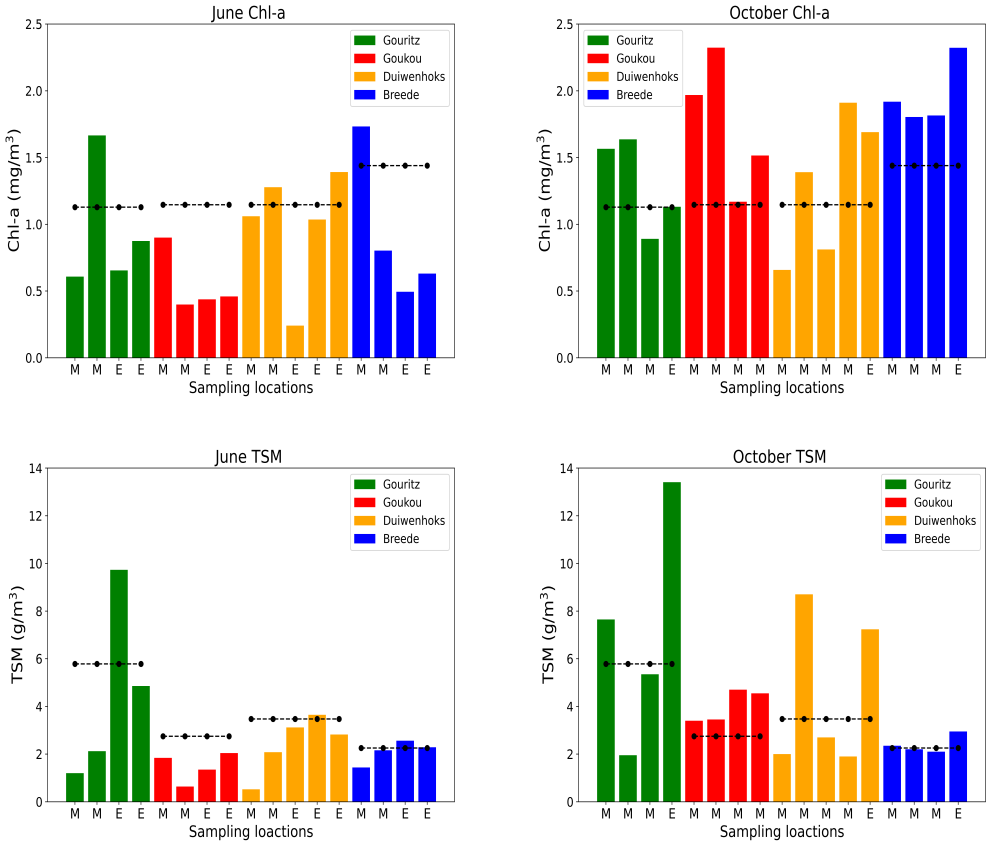

Figure 3.1: Measured [Chl-a] (mg/m^3) and TSS (g/m^3) of the water samples taken in June and October 2022. The rivers are Gouritz, Goukou, Duiwenhoks and Breede, each with their respective colour bars representing either E= estuarine sites or M= marine sites. The black line across each river colour bar denotes the average constituent concentration for individual rivers, over both sampling seasons. The exact geographical locations are in Appendix A.1 for both June and October samples.

3.2 Synthetic R_{rs} versus satellite R_{rs}

The maximum and minimum ranges in R_{rs} of the satellite match-up pixels from the three Atmospheric Corrections (ACs) were captured within the range of the synthetic dataset (Figure 3.2). When the full dataset MLP model was applied to satellite R_{rs}, the evaluation metrics were weaker than expected. On closer expectation, it was noted that while the TSS ranges from the *in situ* values were between 0.5 and 13 g/m^3, the median measurement was situated at 2.56 g/m^3 (Table 3.1), being approximately 12.5 g/m^3 lower than the maximum TSS concentration parameterised for the TSS IOP model (Table 3.2). Thus, the full dataset was constrained by removing spectra with a TSS above 8.1 g/m^3; this excludes elevated measurements from the Gouritz estuary, known to have a particularly high retention of particulate matter.

Table 3.2: The parameter ranges from [Chl-a], TSM, and CDOM absorption at 400 nm and the selected size ranges, along with the steps between iterations when generating the initial full synthetic dataset and then the constrained dataset.

Parameter	**Full dataset**	**Constrained dataset**
[Chl-a] (mg/m^3)	0.1 - 5 (0.2 step)	0.1 - 5 (0.2 step)
TSS (g/m^3)	0.1-15 (0.5 step)	0.1-8.1 (0.5 step)
a_{gd}(400) (m^{-1})	1-12 (1 step)	1-12 (1 step)
Size (μm)	2 and 8	2 and 8
Dataset size	18 000	9 600

However, the full dataset does extend past the maximum of the other retrievals, to 0.0175 sr^{-1} (Figure 3.2). The constrained dataset, which has a similar maximum range as the satellite retrievals R_{rs}, only extends to 0.0125 sr^{-1}, with a similar spectral shape to the full dataset. C2RCC and the synthetic datasets' 400-450 nm R_{rs} followed a similar gradual ascending slope starting at 0.005 sr^{-1}, opposed to the standard and POLYMER R_{rs} retrievals, starting at 0.006 sr^{-1} and 0.0075 sr^{-1} for the maximum retrievals (Figure 3.2). However, POLYMER does have a similar spectral peak at 550 nm and a slight trough in the 675 nm waveband compared to the synthetic dataset, whereas the standard and C2RCC ACs do not (Figure 3.2). The minimum synthetic R_{rs} signal is much lower than the retrievals, with virtually no reflectance along any waveband (Figure 3.2). The lowest satellite-derived retrievals were between 400 and 440 nm for the standard algorithm; some of these match-ups showed negative reflectances in the blue region and were excluded. A fluorescence peak is visible in both the minimum and maximum R_{rs} of POLYMER. The full set of R_{rs} retrievals for both sampling seasons can be seen in Appendix A.2.

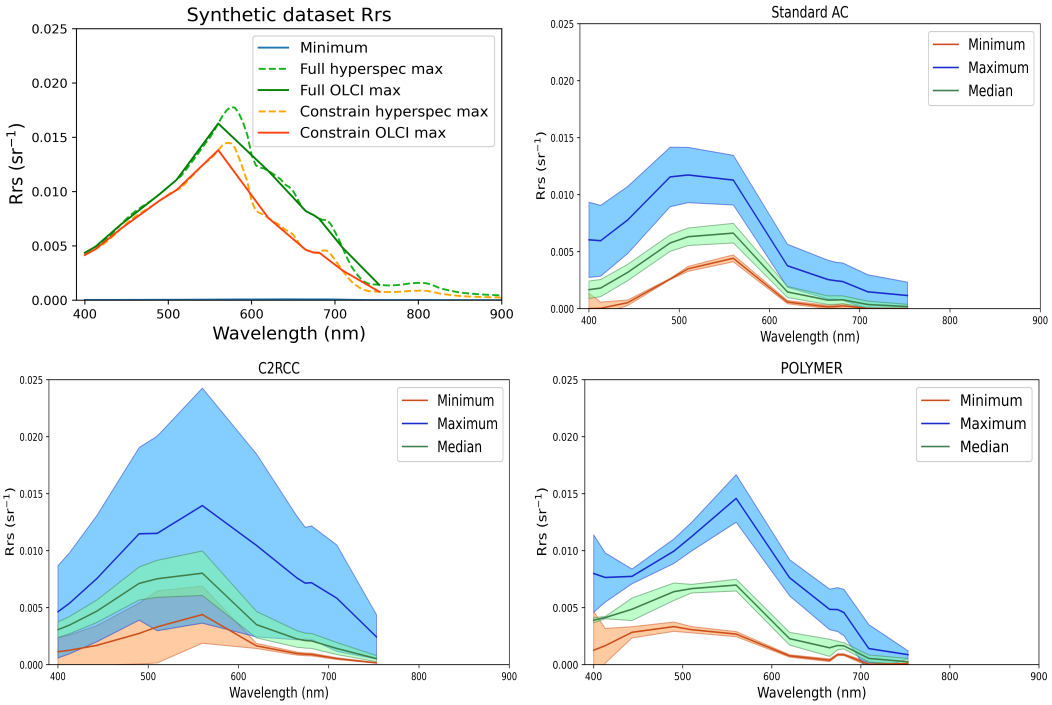

Figure 3.2: The minimum (blue) and maximum (green/orange) ranges for the synthetic R_{rs} produced by the model as well as R_{rs} retrieved by standard AC, C2RCC and POLYMER (minimum= orange, median= green, maximum= blue) for all of the sites sampled from the study area and after applying the prescribed pixel quality assessments. The standard deviations for the ACs after pixel retrievals are shaded for each line.

3.3 MLP model training performance

The efficiency of the hyperparameter selection for the neural network used to train the synthetic dataset R_{rs} and concentration pairs have been examined in Figures 3.3, 3.4, and 3.5. There was a gradual improvement for the training mean absolute error (MAE) and cross-validation dataset when applied to the model after each training session (epoch) (Figure 3.3). From the first model training, at an epoch of 1, the training dataset had an MAE of 65% and 70% for the full and constrained dataset, respectively (Figure 3.3). Both MAE decreased by approximately 10% during the first cross-validation training session, before error rates converged (Figure 3.3). The full dataset had more ease in training the neural network, as the error margins steeply declined, while there was a more gradual decline to the constrained dataset MLP model training (Figure 3.3). However, after 20 training sessions, the constrained dataset was able to reach a training error rate similar to the full dataset, where both final cross-validation evaluations were $<$25%, with the constrained neural network having a lower final error rate.

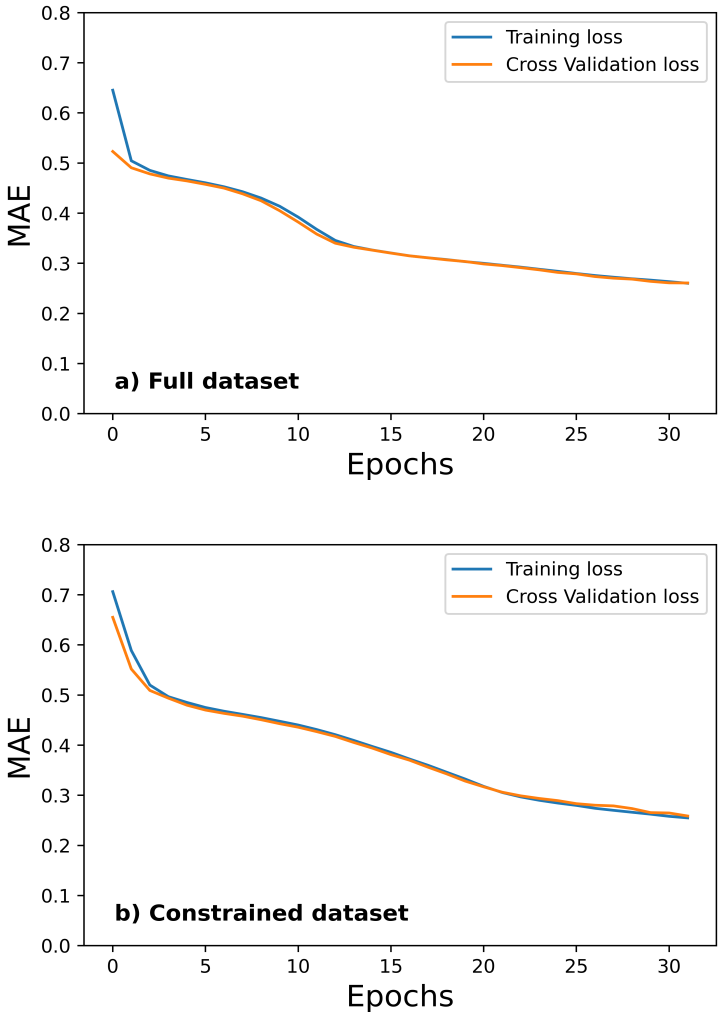

Figure 3.3: The calculated MAE from the final cross-validation performance test after each of the 32 training iterations (epochs) for the full and the constrained MLPs.

Subsequently, the final cross-validation evaluation performed exceptionally well for the most important biologically active components in this study, being that of [Chl-a] and TSM. TSM predictions were nearly perfect across the full parameterized ranges, with virtually no dispersion from the actual concentrations. TSM predictions were the most accurate with a bias close to unity and a low error margin for both the full (bias= 1.007, MAPE= 3.8%, n= 1800) and constrained (bias= 1.011, MAPE= 3.8%, n= 960) datasets (Figure 3.4 and 3.5 respectively). However, [Chl-a] predictions began to falter, especially $>$2 mg/m^3, where values were overestimated for both the full and constrained datasets. Furthermore, [Chl-a] errors for the full dataset were homoscedastic (Figure 3.4), while the constrained dataset errors were truncated (Figure 3.5). The full MLP model's final [Chl-a] cross-evaluation did produce lower errors than the constrained model. Along with TSM, CDOM for both the full and constrained dataset performed ideally with biases and MAPEs with relatively

similar error margins (Figure 3.4), where the former neural network underestimated CDOM as concentrations increased. Size performed similarly unsatisfactory for both the full and constrained datasets, with MAPE 98% and biases of approximately 0.2 (Figure 3.4 and 3.5 respectively).

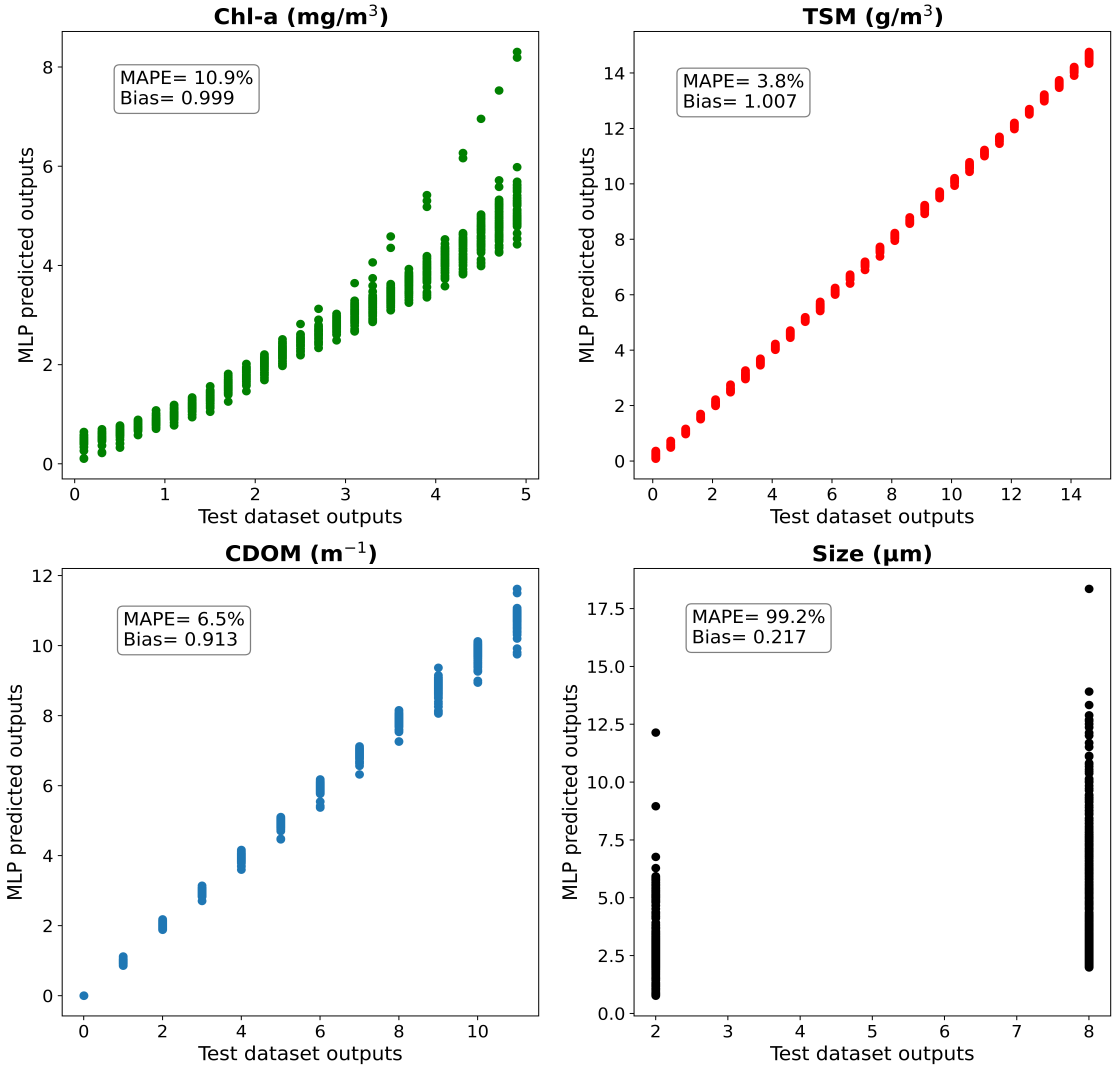

Figure 3.4: Full dataset MLP model cross-validation results showing the predictive ability of the trained MLP neural network model on 10% of the full synthetic dataset, which was randomly selected and removed from the initial full synthetic dataset. The bias and MAPE of each target's performance are denoted in the legend box associated with each target.

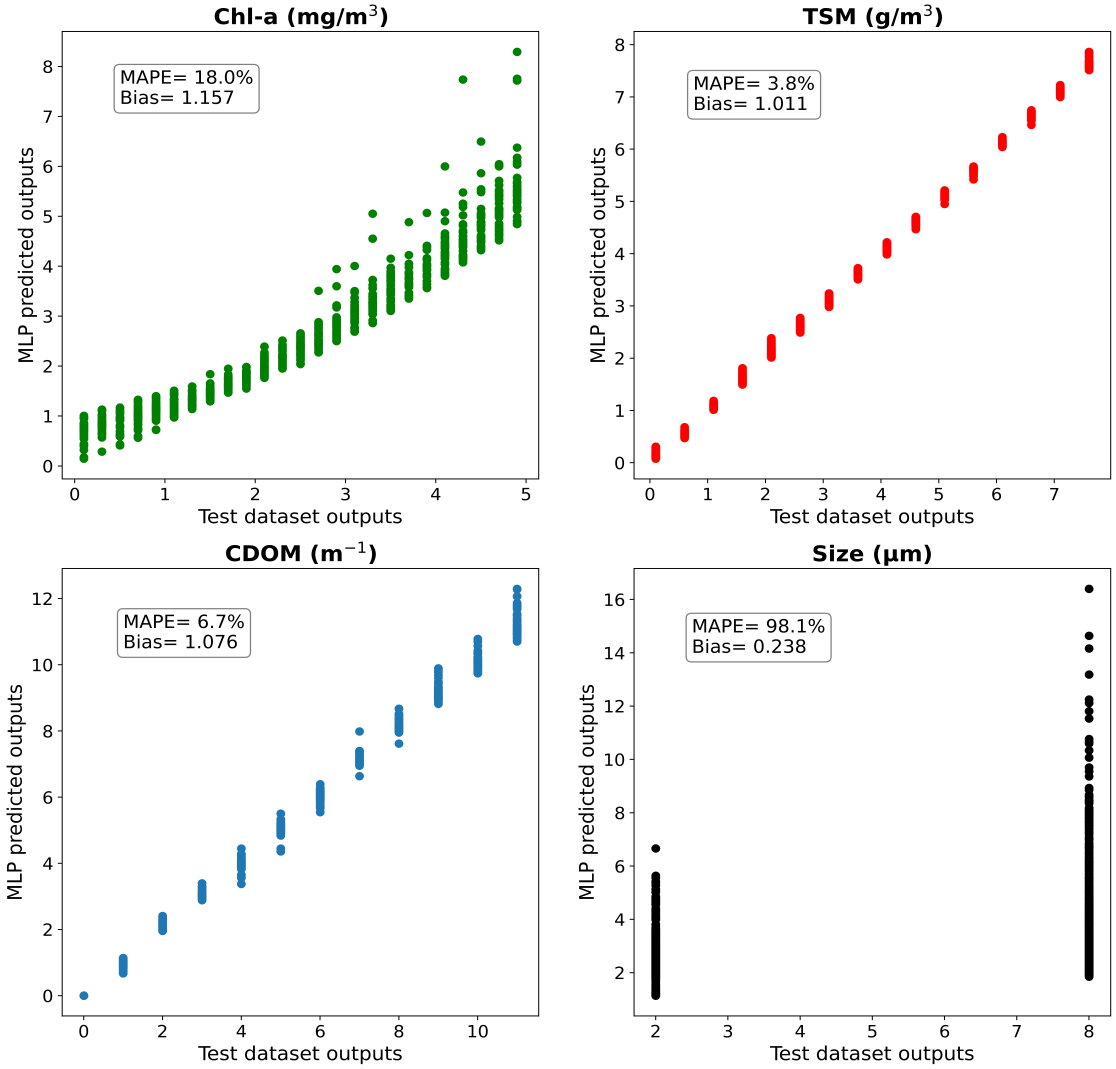

Figure 3.5: Constrained dataset MLP model cross-validation results showing the predictive ability of the trained MLP neural network model on 10% of the constrained synthetic dataset, which was randomly selected and removed from the initial full synthetic dataset. The bias and MAPE of each target's performance are denoted in the legend box associated with each target.

3.4 Geophysical Data Products

3.4.1 Standard satellite data products

POLYMER [Chl-a] product performed the best overall, with metrics for bias, RMSE and MAPE being the lowest, and closest to unity (bias= 1.05, RMSLE= 0.26, MAPE= 28%), as well as having one of the highest pixel retrievals, at 19 (Table 3.3). $\text{TSM}_{C2RCC-TSM\_NN}$ performed optimally for the TSM products, where RMSE and MAPE were minimized (RMSE= 0.54, MAPE= 41%, n=16). [Chl-a] algorithms using the standard AC, such as that of $\text{Chl}_{L2-OC4ME}$ and Chl_{L2-CHL_NN}, performed better than C2RCC's $\text{Chl}_{C2RCC-CHL\_NN}$, which produced an overestimation of 55% and a relatively large MAPE of 41%, in comparison to the 33-36% from the standard Level-2 products (Table 3.3). $\text{Chl}_{L2-OC4ME}$ did however overestimate [Chl-a] predictions by 28%, as opposed to the slight 2% overestimation from Chl_{L2-CHL_NN} (Table 3.3).

Table 3.3: Ocean colour products available with the different AC processors.

Model	Bias	RMSLE	MAPE (%)	n
$\text{Chl}_{L2-OC4ME}$	1.28	0.27	33.6	15
Chl_{L2-CHL_NN}	1.02	0.25	36.3	15
TSM_{L2-TSM_NN}	0.56	0.61	50.4	15
$\text{Chl}_{C2RCC-CHL\_NN}$	1.55	0.38	41.4	16
$\text{TSM}_{C2RCC-TSM\_NN}$	0.97	0.54	41.2	16
$\text{Chl}_{POLY-POLY}$	1.05	0.26	28.1	19

3.4.2 MLP model retrieval products

TSM estimates from satellite R_{rs} processed with the different AC's showed a distinct improvement after constraining the synthetic dataset (Table 3.4). The full model tended to overestimate predictions by >120% with median errors >76%, whereas the constrained MLP outputs were close to unity, with biases of 1-1.1, and with median errors <70%. The unconstrained MLP TSM outputs for all three ACs had very similar evaluation metrics, although $\text{TSM}_{C2RCC-MLP}$ had the lowest median error, distinguishable from the other ACs with a MAPE of 77% . The median error of the standard and POLYMER decreased by 63% and 41% respectively after constraining the model, while C2RCC estimations improved by 21%. TSM predictions from the constrained model were greatly improved with C2RCC (bias= 1.09, RMSLE= 0.87, MAPE= 55%, n=19) and standard (bias= 1,

RMSLE= 0.88, MAPE= 47%, n=15) ACs, which had estimates close to unity, with the lowest RMSLE and median errors. However, note that some of the pixels were flagged in the standard Level-2 product pre-processing; although the remainder would theoretically have a higher level of confidence, it does result in lower pixel retrievals along the river mouth-marine interface region.

Table 3.4: The evaluation statistics for the MPL models derived from the full and constrained datasets, applied to R_{rs} outputs from three ACs, namely the standard (Level-2), C2RCC and POLYMER

Model	Bias	RMSLE	MAPE (%)	n
Full dataset				
Chl_{L2-MLP}	0.70	0.94	77.8	15
TSM_{L2-MLP}	2.20	1.19	110.3	15
$\text{Chl}_{C2RCC-MLP}$	0.80	0.33	28.2	19
$\text{TSM}_{C2RCC-MLP}$	2.22	1.14	76.5	19
$\text{Chl}_{POLY-MLP}$	1.16	0.84	114.6	19
$\text{TSM}_{POLY-MLP}$	2.28	1.11	111.2	19
Constrained dataset				
Chl_{L2-MLP}	1.38	0.82	108.7	15
TSM_{L2-MLP}	1.00	0.88	47.0	15
$\text{Chl}_{C2RCC-MLP}$	1.26	0.28	25.4	19
$\text{TSM}_{C2RCC-MLP}$	1.09	0.87	55.3	19
$\text{Chl}_{POLY-MLP}$	1.82	0.71	113.4	19
$\text{TSM}_{POLY-MLP}$	1.07	0.85	69.5	19

Estimates for the [Chl-a] products did not change as much as the TSM predictions did after constraining the MLP model (Table 3.4). C2RCC's AC performed the best for both the full and constrained MLP models, where the median error was mostly >50% lower than all other [Chl-a] predictions (Table 3.4). Application of the constrained MLP model to the C2RCC resulted in a 26% average overestimation, differing from the 20% underestimate from the full MLP model; however, both the median error and the RMSLE decreased by 3% and 0.05 respectively, suggesting that the constrained

model performs marginally better than the full MLP model. The [Chl-a] MLP model biases did show that the constrained dataset tended to overestimate values, while the full dataset generally underestimated concentrations (Table 3.4); the RMSLE decreased slightly for all ACs when constraining the MLP model(Table 3.4), acknowledging that RMSLE penalises biases below 1 more so than overestimations. Conversely, the median error estimate for the standard AC was approximately 40% higher for the constrained MLP model compared to the full model. POLYMER's [Chl-a] outputs were unsatisfactory for both MLP models, with high error margins, for the full (bias= 1.16, RMSLE= 0.84, MAPE= 114.6%, n= 19) and constrained (bias= 1.82, RMSE= 0.71, MAPE= 113.4%) datasets respectively; there was a trade-off between the average error dispersion metrics (RMSLE and bias), although both MLP models produced high MAPEs >113% (Table 3.4).

3.5 Regional Comparison

Since the MLP models applied to C2RCC R_{rs} showed the best performance in Table 3.4, these data were used for regional mapping of the MLP model outputs. Chl$_{C2RCC-MLP}$ for the full MLP model exhibited very low [Chl-a] (<0.1 mg/m^3) along the coastlines, while the outputs from the constrained MLP model were more variable, with estimates ranging from 1 to 5 mg/m^3 (Figure 3.6). However, both MLP models showed similar synoptic patterns over the open ocean region, with [Chl-a] not extending past approximately 5 mg/m^3 (Figure 3.6). The full MLP TSM$_{C2RCC-MLP}$ predictions were higher than the constrained, although with similar synoptic patterns to one another (Figure 3.6). The constrained TSM$_{C2RCC-MLP}$ was more variable with much higher concentrations but also analogous spatial patterns to that of the best performing TSM satellite product deduced from Table 3.3 (Figure 3.7), i.e. the standard C2RCC TSM product.

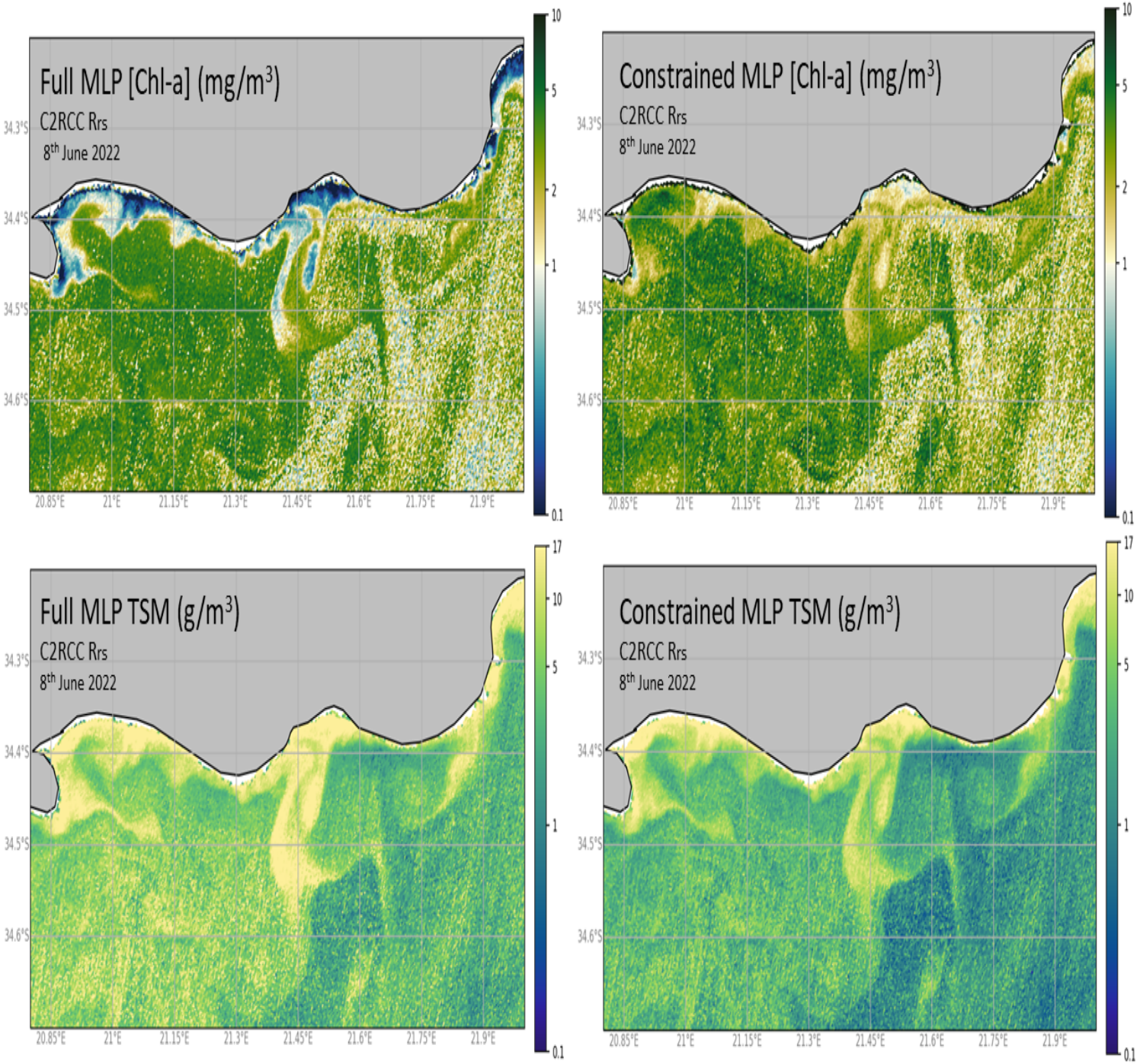

Figure 3.6: Maps of Chl-a (top) and TSM (bottom) for the full (left) and constrained (right) MLP models applied to C2RCC R_{rs} from 8 June 2022.

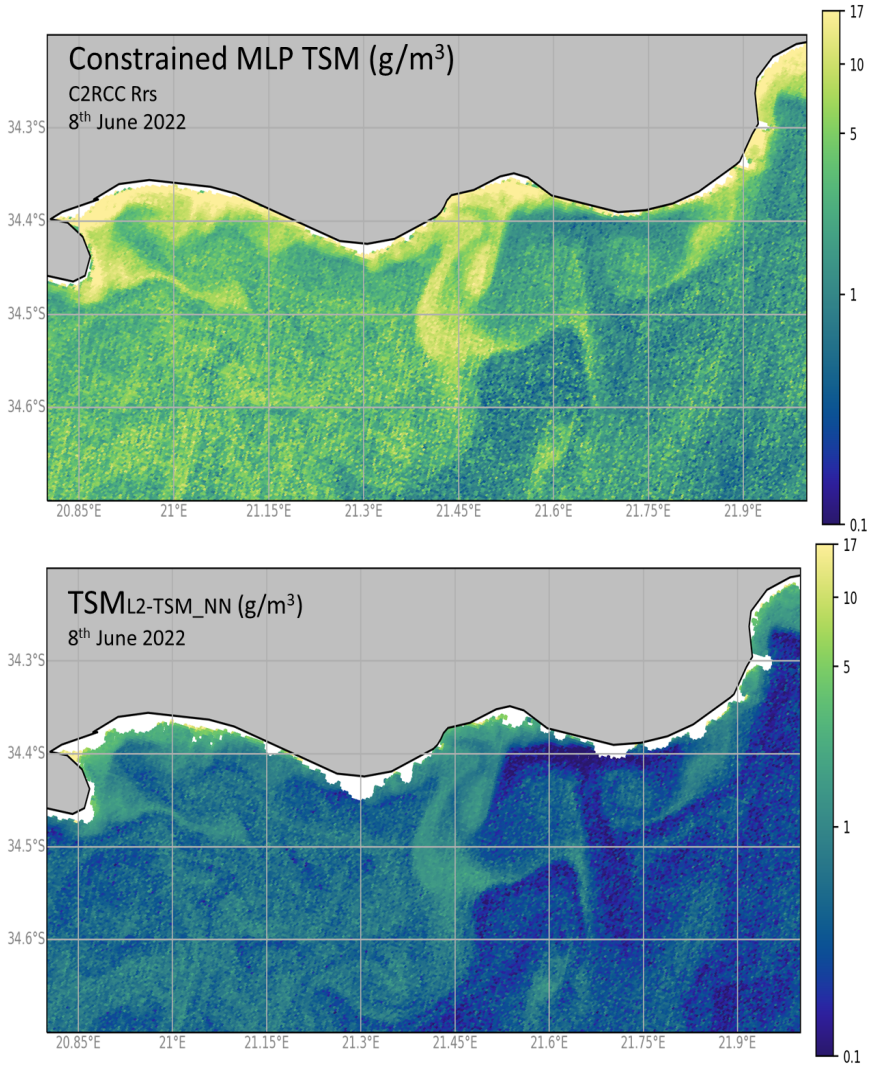

Figure 3.7: TSM outputs from the constrained MLP model when applied to C2RCC R_{rs} (top) compared to the standard C2RCC TSM product (bottom), for 8 June 2022.

When sampling on the 7^{th} of June, it was evident that a biomass burning event was affecting atmospheric conditions. To contextualize potential effects on AC procedures, standard Level-2 AOT and Angstrom exponent products from the day before and after the fire event were compared. Differences between AOT from before the fire and after the fire show a 0.035 increase and an Angstrom exponent rise of 0.8, from June 6^{th} to June 8^{th} (Table 3.5). The October 1^{st} AOT mean was much higher than the June values at 0.11 and a relatively high standard deviation of 0.058 (Table 3.5). The corresponding Angstrom exponent was the lowest out of all days assessed at 0.49 with a low standard deviation of 0.13 (Table 3.5).

Table 3.5: Aerosol Optical Thickness (AOT) and Angstrom exponent before the June fynbos fire, after and in October 2022 over the river plume region in the Agulhas Bank.

Month	Day	AOT mean	AOT std	Angstrom mean	Angstrom std
June	**6**	0.037	0.013	0.82	0.31
June	**8**	0.073	0.009	1.59	0.13
October	**1**	0.11	0.058	0.49	0.13

The two best performing standard [Chl-a] satellite products as deduced from Table 3.3, i.e. $\text{Chl}_{POLY-POLY}$ and Chl_{L2-CHL_NN}, were used to observe the effects of atmospheric conditions on [Chl-a] estimates. $\text{Chl}_{POLY-POLY}$ variations from before and after the June fire showed very striking differences, where the 6^{th} of the June map exhibited high [Chl-a] readings over most of the region, with the concentrations drastically increasing within the next two days (Figure 3.8). Chl_{L2-CHL_NN} also had a similar [Chl-a] increase after the fire event, where concentrations increased from 0.5-2 mg/m^3 to an upper range of almost 5 mg/m^3 dispersed further away from the coastline (Figure 3.9). Notably, none of the October plots exhibited the same degree of chlorophyll spatial variability as the June sites, especially when compared to the after fire plot. The October plots of both [Chl-a] products were strikingly similar.

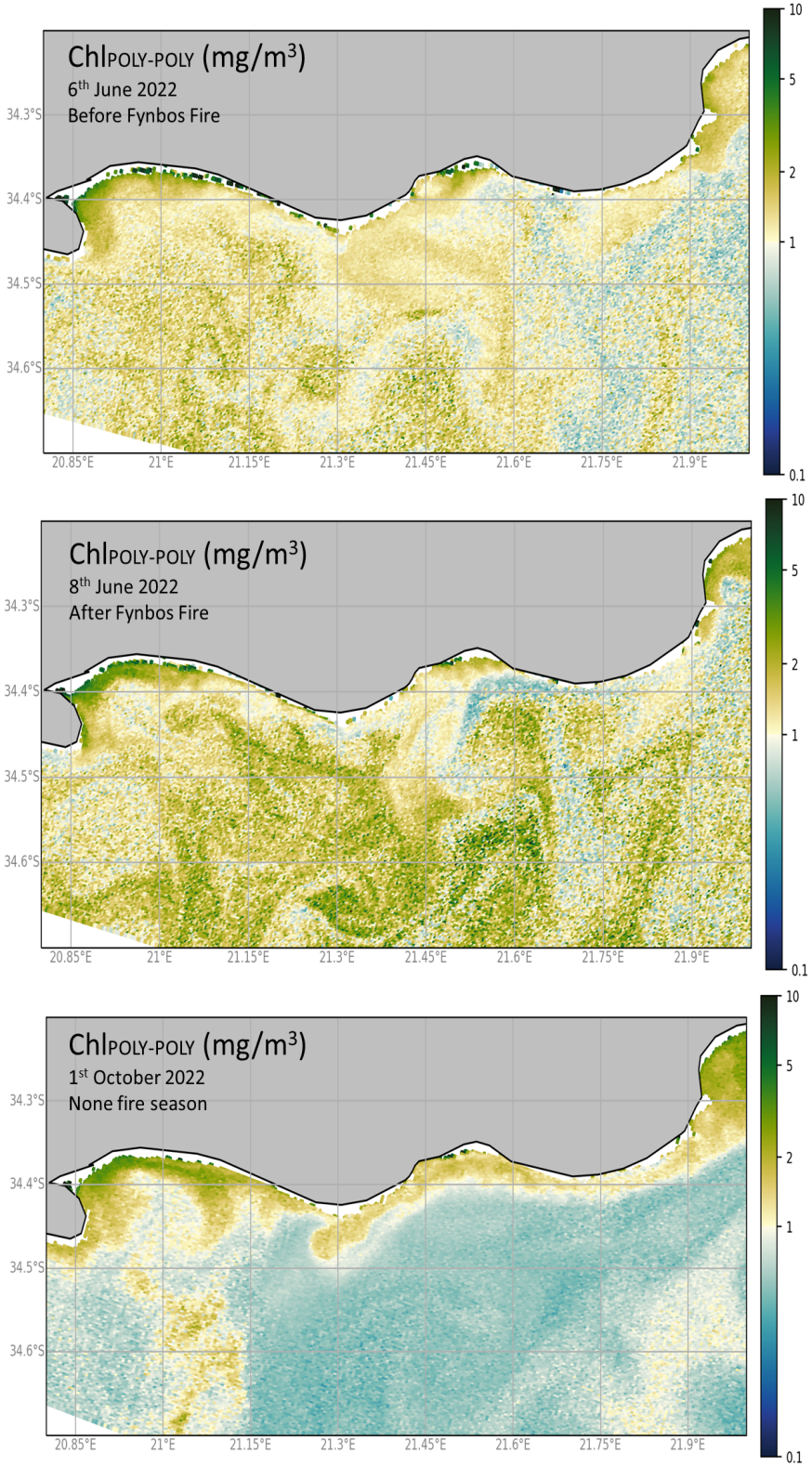

Figure 3.8: Maps of POLYMER [Chl-a] on the 6^{th} of June before a June fynbos burning event (top), on the 8^{th} of June after the biogenic aerosol release (middle), and on the 1^{st} of October (bottom).

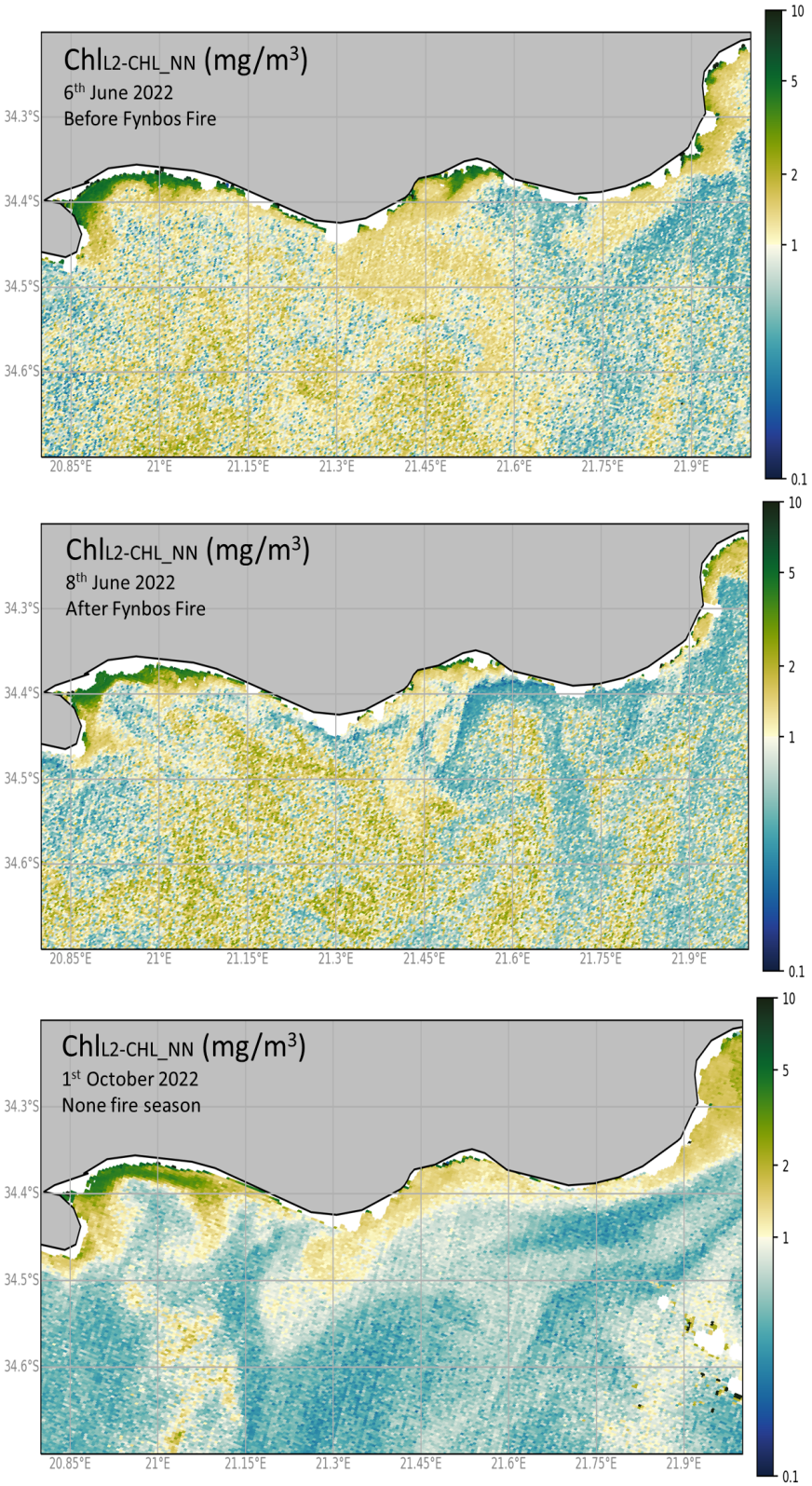

Figure 3.9: Maps of the standard Level-2 NN [Chl-a] on the 6^{th} of June before a fynbos burning event (top), on the 8^{th} of June after the biogenic aerosol release (middle), and on the 1^{st} of October (bottom).

Chapter 4

Discussion

Marine optics combined with data satellites sensors measure could be used as a tool to monitor river plume discharge. However, river plumes are particularly challenging to study as the highly turbid conditions induced by high sediment loads create optically complex waters. In addition, atmospheric correction is regionally variable and not well understood along most coastlines. This is the first study to investigate the utility and accuracy of existing satellite biogeophysical products and develop a new regional model for the Agulhas Bank river plumes for the Gouritz, Goukou, Duiwenhoks and Breede Rivers and test their performance against the match-up *in situ* dataset.

4.1 Interpreting *in situ* measurements

The outputs from rivers and estuaries, in general, are dependent on geographical, climatic, and physical features such as seasonality, estuarine geomorphology (Shi, 2010), catchment size (Neil et al., 2002), surrounding vegetation (Vundavilli et al., 2021), and river mouth tidal dynamics (McCabe et al., 2009). *In situ* chlorophyll-a concentration ([Chl-a]) and Total Suspended Sediment (TSS) measurements revealed that June experienced low TSS and low surface [Chl-a], while the October samples had much higher [Chl-a] detected (Figure 3.1), with intermediate sediment measurements (Appendix A.1) in comparison to the previous sampling season. The heightened [Chl-a] measurements in October may have been caused by increased water column mixing through the warming of surface waters, also seen in Mitchell-Innes et al. (1999) along the western Agulhas Bank during the spring months (October and/or September). Furthermore, the Agulhas Bank exhibits features of both an upwelling and temperate shallow shelf system (Hutchings et al., 2009); with Cape Agulhas (Breede) displaying periodic wind-induced upwelling of nutrient-rich waters during the summer

months (Largier et al., 1992; Schumann et al., 1982), which may have also contributed to the seasonal variability in [Chl-a].

However, the sedimentary (TSS) increase in October may have been attributed to the heightened rainfall experienced during the winter months, June and August. Research in the Western Cape (Schulz, 2001) and other international studies have identified a strong correlation between rainfall and sediment discharged into river channels (Hicks et al., 1996; Langbein & Schumm, 1958), later deposited in the adjacent marine system (Osadchiev & Korshenko, 2017; Shi & Wang, 2009). Additionally, TSS measurements were higher in the estuarine section of the region than in the marine sites, noticeably so in the Gouritz River (Figure 3.1), where the highest suspended sediment concentrations were measured, a common feature of sediment retention in estuaries (Yang, 1998; Zwolsman et al., 1993).

4.2 Synthetic dataset parameterisation

4.2.1 Augmenting the *in situ* dataset

River plumes are optically complex water bodies, whereby the in-water constituents are regionally and seasonally variable, making it difficult to derive optimal geophysical ocean colour satellite products accurately. Such regions require flexible, robust algorithms that capture the range of possible resulting remote sensing reflectance (R_{rs}) from known constituent (and Inherent Optical Property; IOP) variability in the region. Augmenting the *in situ* measurements towards a regionally parameterised synthetic dataset is the initial step when developing an appropriate machine learning algorithm for constituent retrievals in the absence of a comprehensive set of *in situ* measurements.

However, it is not possible to assess the accuracy and precision of synthetic datasets formulated in this way. The scarcity of *in situ* measurements by definition prohibits the validation of the synthetic dataset parameterised in this study. So while the synthetic dataset is presented as a 'best guess' based on recorded concentration ranges, the IOP models used are generalised. This is a major source of uncertainty in the synthetic dataset, and these uncertainties will be propagated throughout the Multi-Layer Perceptron (MLP) model constituent retrievals as a consequence of training the MLP models on synthetic data.

Although [Chl-a] and Total Suspended Matter (TSM) were identified as the most crucial biogeochemical parameters to measure, both the short temporal scales captured by the sampling trips and the lack of other important parameters necessitated input from previous studies in the study region to ensure an appropriate range of constituents were represented in the synthetic dataset. Phytoplankton

species identification and [Chl-a] were measured in a study conducted by Lemley et al. (2015) throughout 2013 into 2014 along three of the river-estuarine systems investigated in this study: the Duiwenhoks, Goukou and Gouritz. Phytoplankton species collected included cyanobacteria, dinoflagellates, flagellates, and diatoms, where size ranges of even the smallest amongst them can vary from $<$ 1 to 100 μm (Mehdizadeh Allaf & Peerhossaini, 2022). Lemley et al. (2015) found that the median [Chl-a] ranged from 0.6 to 6.7 mg/m^3 near river mouth sites from all three relevant estuaries. Recent oceanographic observations in the Agulhas Bank marine system by Jacobs et al. (2022) and Poulton et al. (2022) found that surface [Chl-a] ranged from 0.3 to 5.1 mg/m^3, with a subsurface chlorophyll maximum of 1.7 to 10.3 mg/m^3, substantiating historic measurements made by Shannon et al. (1984) in the late 1970s. The marine system does have a larger range of [Chl-a] than the estuaries observed in Lemley et al. (2015); however Poulton et al. (2022) did capture [Chl-a] concentrations over a larger extent into the general Agulhas Bank region, which may have exaggerated the upper marine ranges. Furthermore, no phytoplankton identification studies have been conducted in the river plumes, nor have TSM concentration measurements been made within the estuary, river plume, or adjacent marine systems. This makes hypothesizing the full range of TSM and phytoplankton class variability along the river plume present in the study region difficult. It is known that river plume constituent composition and concentration are affected by fluxes when substances are flushed out of the estuary, as well as the natural climatological (i.e. natural flood cycles) and seasonal (i.e. temperature fluxes) variability occurring in the study region, which may influence the river plume's TSM concentrations and phytoplankton biomass.

4.2.2 Constructing the synthetic dataset

Notwithstanding the aforementioned constraints, the main objective when constructing the regional synthetic dataset was to include R_{rs} that capture not only the range of concentrations of constituents within the river plume system but also appropriate IOP parameters for each of them. The equivalent Algal Population model was chosen for this study (Lain et al., 2014). The parameterization for phytoplankton concentration was decided based on the *in situ* [Chl-a] measurements in the present study, and previous studies by Lemley et al. (2015) and Shannon et al. (1984), where the maximum value was set to 5 mg/m^3 (the *in situ* upper measurement for the current study was $<$ 2.5 mg/m^3). The selected parameterisation allows for predictions exceeding the sampled values but does not account for the possibility of sporadic phytoplankton blooms that could occur above these concentrations (Mitchell-Innes et al., 1999). Furthermore, a generalised mixed diatom/dinoflagellate assemblage model - containing Chl-a as the dominant light-absorbing pigment in phytoplankton - was used, since no species identification was conducted in this study, nor was the necessary detail found in Lemley et al. (2015) and Poulton et al. (2022) for pigment delineations.

In conjunction with [Chl-a], phytoplankton size was also estimated, with variable populations of small (2 μm) and larger (8 μm) cells represented in the modelled assemblages. Generally, when low [Chl-a] are measured, pico- and nanophytoplankton populations dominate population structure, attributed to the scarcity of bioavailable nutrients (Gin et al., 2000). The lack of diversity in the modelled cell sizes may result in large prediction discrepancies in the model predictions. In addition, some species identified by Lemley et al. (2015), such as cyanobacteria, are known for being efficient backscatterers, attributed to air vacuoles within the cell (Matthews & Bernard, 2013). These species are not accounted for in the phytoplankton IOP model used in this study.

Due to the lack of previous TSS or TSM studies in the Agulhas Bank rivers and estuaries, the concentration ranges selected for sediment were based solely on the measurements collected during June and October 2022, thus the maximum value was set to 15 g/m^3, which is 2 g/m^3 above the highest *in situ* value. Generalised IOPs were used for this constituent ("average mineral" component as per Hydrolight) which is unlikely to be a true representation of the actual physio-chemical composition and/or particle size distribution found in the river plume. River plume sediment types are generally diverse, where marine continental shelf sediment can be uplifted when there are strong tidal forcings (Shi, 2010), as well as fine sedimentary particles combining with organic matter through the process of flocculation (Milligan et al., 2007). Size identification may help parameterize available TSS models, resulting in a more accurate representation of the optical complexity induced by varied size fractions (Stumpf & Pennock, 1989).

In the absence of Coloured Dissolved Organic Matter (CDOM) *in situ* measurements, postulating the concentration of this constituent within the river plumes is difficult. Thus, the IOP parameterisation relied on river plume CDOM measurements from international studies. CDOM estimations were gathered from previous literature on other riverine systems, such as that of Shao and Wang (2020) where CDOM measurements spanned 0.19 - 10.80 m^{-1} along the Yiluo River in China. In the Mississippi and Atchafalaya river plume system a_{gd}(440) for selected samples ranged from 0.192 - 7.068 m^{-1} (Zhu et al., 2011). The higher ranges may be representative of the Duiwenhoks River site, which appeared saturated in tannins during both sampling trips. This study used a range of CDOM absorption at 400 nm of between 1 m^{-1} to 12 m^{-1}. The resulting dataset of R_{rs} each represents a unique combination of constituent concentrations and is only as accurate as the IOP inputs, whose constraints have been discussed.

4.3 Satellite R_{rs} versus synthetic R_{rs}

It was uncertain which atmospheric correction algorithms were optimal for satellite R_{rs} retrievals over the region, thus three common and effective algorithms (Brockmann et al., 2016; Steinmetz & Ramon, 2018), were chosen based on merits from previous studies (Giannini et al., 2021; Liu et al., 2021), to assess the AC performances with the developed MLP model. No radiometric *in situ* data were collected throughout this study, thus, accuracy assessments between the corrected Top-of-Atmosphere (TOA) radiances and the ground truth water-leaving radiances cannot be accomplished. The synthetic dataset R_{rs} was treated as an ideal retrieval instance, where no atmospheric and environmental contaminators disrupted the target constituent IOP signals emitted from the water body. Thus, the MLP models were trained on idealised R_{rs} retrieval instances, and broadly parameterized IOPs.

The maximum POLYMER R_{rs} produces a spectral shape very similar to that of the synthetic dataset, especially from $>$500 nm, where the second Chl-a minima at 675 nm can be observed, withstanding the strong sedimentary backscatter and molecular Rayleigh scattering signals (Figure 3.2. Yet, POLYMER's correction for the 400-500 nm bands is likely to be poor, as the atmospheric-oceanic model does not account for the effects of CDOM and absorbing aerosols (Steinmetz et al., 2011). Both CDOM and absorbing aerosols strongly absorb energy in the blue-green visible light regions and may have resulted in inflated R_{rs} along those wavelengths. It can be seen in Figure 3.2 that the maximum POLYMER (400 nm) is much larger than the maximum synthetic dataset at 400 nm. Wang et al. (2020) attempted to compensate for POLYMER's lack of absorbing aerosol term by using a novel two-angle atmospheric correction algorithm scheme, termed TAACA, to remove the atmospheric contributions by aerosols, but this was out of the scope of this present study. The lack of *in situ* CDOM and R_{rs} measurements do not allow for a deeper analysis into the effectiveness of POLYMER's AC procedure in the presence of high CDOM as the river plumes CDOM concentrations are unknown. It is well known that many ACs tend to perform poorly in the 400-500 nm ranges (Ahmad et al., 2007; Clark et al., 1997).

The standard AC R_{rs} do not extend as high as the POLYMER and C2RCC maximum signal (Figure 3.2). This can be simply explained by the masking of near-land and estuarine water body pixels by the quality flags applied to the standard AC (n= 15), potentially exacerbated by the effects of adjacency to land. Land adjacency effects are the influences from neighbouring terrestrial surfaces onto top-of-atmosphere radiances, common to inland and estuarine waterbodies (Palmer et al., 2015; Sterckx et al., 2011). Moreover, there were instances where the standard AC retrieved negative R_{rs} in the 400 nm range, an artefact of the classical 'black pixel hypothesis' AC method R_{rs}, where absorption from 750 nm onwards is considered negligible. In Case 2 waters, this assumption may

not hold, resulting in an overcorrection in the retrieved R_{rs} signal. POLYMER and C2RCC were also able to retrieve more data (n= 19) within the estuary than the standard AC, in common with other studies (Giannini et al., 2021; Uudeberg et al., 2019). C2RCC's R_{rs} were the only satellite-derived R_{rs} to fit entirely within the synthetic R_{rs} range.

The synthetic dataset's minimum signal along each wavelength may be representative of the combination of backscatter and absorption which produce extremely low R_{rs} signals along the visible light wavelengths, such as instances with negligible TSM, and very low phytoplankton biomass, existing in heavily CDOM dominated water types. Boreal and arctic regions with large concentrations of CDOM are extreme environments where the R_{rs} values are negligible in almost the entire visible part of the spectrum (400–700 nm) due to the strong absorption by CDOM (Kutser et al., 2016). However, this water type does not correspond to the satellite's minimum retrievals, where all of the R_{rs} after atmospheric correction produced backscatter driven R_{rs} signals far greater than the synthetic dataset's minimum values. Yet, the model was not constrained to accommodate the minimum ranges of the satellite retrievals as the *in situ* measurements for TSM and [Chl-a] were low, and there were no CDOM concentration and phytoplankton size class data collected.

Using the highest TSM measurement collected further into the estuary may have been the cause of the heightened R_{rs} signal from the unconstrained synthetic dataset, as the median measurement was almost 11 g/m^3 lower than the maximum estuarine *in situ* TSM sample (Table 3.1). The synthetic dataset was then constrained to a more reasonable maximum of 8.1 g/m^3 since the estuarine measurement was likely flagged by land masks and poor quality retrieval controls (Jiang et al., 2023; Urquhart & Schaeffer, 2020). The resulting synthetic backscatter signal decreased to reach a similar maximum of the R_{rs} retrievals. Thereafter, the full and constrained synthetic datasets were each used to train an MLP model, creating two ocean colour products that do not include the effects of the atmosphere.

4.4 MLP performance evaluation

The cross-validation results for both the full and constrained datasets for CDOM and TSM predictions show that both MLP models performed ideally (see Figures 3.4 and 3.5). CDOM marine optical properties have been studied for decades (Bricaud et al., 1995; Bricaud et al., 1981). Modelling CDOM alone is relatively simple, however, with added effects of phytoplankton diversity, aerosol absorption and lack of satellite sensor instrumental work done to improve R_{rs} retrievals in the 400 nm bands, makes quantifying CDOM in the real system arduous (Kutser et al., 2016; Toming et al., 2016). TSM's strong backscatter signal was well predicted during the final model testing, despite the wide optical variability captured in the synthetic dataset. Thus, the MLP models were

near-perfect predictors for the synthetic dataset, especially so for TSM and CDOM. However, [Chl-a] was dependent on the number of spectra used to generate the MLP models, where the full dataset containing more variability performed substantially better than the constrained MLP model's cross-validation. It should be noted that the MLP retrieval performance relates only to its capability to distinguish R_{rs}/concentrations relationships from within the synthetic dataset. In other words, this provides no information on how closely the synthetic dataset may represent the actual *in situ* conditions. Thereafter, these well-trained MLP models were applied to the R_{rs} from the three ACs, as the synthetic dataset represents atmospherically corrected R_{rs}. The match-up evaluations in section 4.5.3 further show the above-mentioned discrepancies in R_{rs} retrievals from the three ACs to the synthetic datasets, as well as how the full versus constrained MLP models might have impacted the MLP model performances.

4.5 Biogeophysical data product Retrievals

4.5.1 Standard data products

When comparing the MLP outputs to the other ocean colour products, POLYMER's [Chl-a] product performed the best statistically, although Chl$_{C2RCC-MLP}$ had a much better dispersion along the one-to-one line, whereas POLYMER tended to overpredict the smaller [Chl-a] measurements (Appendix A.6 and A.8). Opposed to the neural network approach, POLYMER's bio-optical product constrains [Chl-a] to a specified atmospheric-ocean coupled model, which uses all wavebands given to the model to spectrally match [Chl-a] and a co-varying backscatter term (Steinmetz et al., 2011). This approach works well for the sampled region, as the backscatter term has been found to assist [Chl-a] predictions along turbid coastal waters, as well as add insight into the relationship between the mean particulate backscatter-to-[Chl-a] relationships along the 510 nm bands (Loisel & Morel, 1998), resulting in optimal [Chl-a] model predictions.

The standard Level-2 neural network's [Chl-a] product (Chl$_{L2-CHL_{NN}}$) was the second best performing [Chl-a] geophysical product after POLYMER's [Chl-a] algorithm. [Chl-a] OC4ME algorithm (Chl$_{L2-OC4ME}$) performed relatively well, which is surprising since it is a simple ratio empirical algorithm best suited for Case 1 waters (O'Reilly et al., 1998) and was not expected to provide optimal results in optically complex non-phytoplankton dominate waters. The relatively good performance of standard algorithms when applied to the standard Level-2 R_{rs} suggests that the turbid-flag switching method within the standard AC processor is useful in regions where a combination of coastal and open ocean water sites are present. This is supported by a recent study by Zhou et al. (2023) which found that both Sentinel-3's OC4ME and neural network algorithms performed similarly along inland turbid water surfaces where phytoplankton and TSM dominate the optical signals, with the neural network

[Chl-a] performing slightly better than the empirical approach. Based on the mentioned study and results found in this work, OC4ME may also be a robust, simple means to measure [Chl-a], without deploying more complicated, time-consuming neural networks and their associated synthetic dataset parameterisations.

C2RCC's neural network-derived TSM product was identified marginally as the best TSM retrieval model when compared to the *in situ* values. However, C2RCC products for both TSM and [Chl-a] had difficulties in predicting concentrations near the coastlines and the estuary, resulting in lower retrievals (n= 16) than the corresponding R_{rs} (n= 19). This was a result of the model overestimating [Chl-a] and TSM, which were consequently flagged during the coefficient of variation quality assessment. Land adjacency effects may have inflated the R_{rs} signals, resulting in heightened TSM and [Chl-a] retrievals, which have also been observed in Kyryliuk and Kratzer (2019). If retrievals within the estuary and close to the mouth are a user requirement to carry out sediment load and phytoplankton biomass estimations, a large portion of the dataset will be flagged, making it difficult to accurately estimate near estuary-river plume, point-specific sediment or [Chl-a] (i.e., nutrient enrichment proxy) concentrations. Although, the constrained MLP model did follow a similar pattern to C2RCC standard TSM distributions, where the MLP model systematically overestimated sediment concentrations when the constrained MLP model was applied to C2RCC R_{rs}. This result may be promising, revealing that the simple action of adapting the MLP model to the receiving R_{rs} and constraining the MLP model could require fewer training data (9600 as opposed to 5 million or 550 000 spectra), and improve TSM retrievals over the areas of diverse sediment concentrations with lower processing power required. More *in situ* data and IOP models, accompanied with functional data analysis and clustering for the large synthetic dataset (Kravitz et al., 2021) could improve estimates, making the model more generally applicable to other plume regions as well (as discussed further in section 4.5.3).

However, it should be noted that more marine sites were captured in the model evaluation statistics, especially for the standard and C2RCC [Chl-a] and TSM algorithms, as land flags (i.e. land adjacency effects) and retrieval quality checks removed most of the estuarine sites. This may have led to a bias in the model performances detected by the evaluation metrics, as sites known to be highly turbid, were removed from the dataset, which may have resulted in improved evaluation statistics. Models such as OC4ME were an example of such statistical improvements from land flagging in the estuary, as the model is typically used for Case 1 waters and not the turbid Case 2 conditions present in such habitats. This will impact the applicability of the algorithms to the river plumes themselves as monitoring changes in riverine discharge will require algorithms which accurately quantify TSM and [Chl-a] close to land, or within the estuary themselves. Thus, there is a trade-off between good quality retrievals and useful information, which will impact the decision of the best-performing model for the region.

4.5.2 Atmospheric correction evaluation

[Chl-a] estimations for the best-performing ocean colour products were highly variable, and inconsistent when compared to one another. The $\text{Chl}_{POLY-POLY}$ and Chl_{L2-CHL_NN} October estimates were almost identical, while the June days did not show any corresponding patterns to one another. There was a substantially large Fynbos fire near the Duiwenhoks River on the second day of sampling (7^{th} June 2022). The added effects of natural and anthropogenic Fynbos biomass-burning events, common to the Western Cape Fynbos biome (Kruger & Bigalke, 1984), may have complicated atmospheric correction procedures, and impede [Chl-a] model predictions. Averaged aerosol optical parameters were retrieved over the river plume and open ocean regions mapped in Figures 3.8 and 3.9 to establish the extent of absorbing aerosols released from the biogenic burning event before, after, and in comparison to the atmospheric conditions in October.

Biogenic burnings are a common feature of most parts of the world, be it through forest clearing or traditional agricultural activities. Many studies have attempted to use satellite-derived atmospheric products (i.e., the Angstrom exponent and Aerosol Optical Thickness; AOT), as modes to track seasonal variability and intensity of biogenic fires (Waquet et al., 2013). South Africa, being a semi-developed country, has a coastline with both biogenic and industrial atmospheric components that may alter the composition of scattering and absorbing particles in the atmosphere, as well as their interactions with clouds and sun glint, creating another optically complex system, in addition to the water body constituents. It was apparent that after the fire the atmospheric conditions did change as both remotely sensed atmospheric parameters, the Angstrom exponent and AOT, increased (Table 3.5). The Angstrom exponent, inverse to particle size, shows that there were more smaller particles present, with a high aerosol optical thickness, which is a common biogenic aerosol emission, noted by Sano et al. (2011). Research by Pace et al. (2005) found that biogenic-burning events through forest fires in the Mediterranean resulted in Angstrom exponents $>$1.5, and does coincide with the after Fynbos fire Angstrom exponent detected in this study (Table 3.5). Notably, the pre-fire Angstrom exponent was relatively high in comparison to the October measurement, which may suggest that the atmosphere over the region was already saturated in absorbing aerosols from prior burning events in the study region. The heightened AOT in October may be attributed to higher precipitation and humidity as October is spring for the region. However, the AOT increase from June 6^{th} to June 8^{th} may be solely attributed to the increase in biogenic aerosols released from vegetation burning. In this study, all [Chl-a] algorithms were found to be the most impacted by the above-mentioned atmospheric variability.

POLYMER's atmospheric model does account for the effects of sun glint and Rayleigh scattering by molecules, however, the effects of absorbing aerosols were ignored during the model development

phase. This may be a small source of uncertainty over more urbanised regions, but not in developing countries where biomass burning and other aerosol particulate pollution significantly contribute to atmospheric composition. POLYMER's [Chl-a] estimates plotted over the study region show a striking difference between June and October predictions, with June open ocean estimation being much higher than is expected, while October's measurements are realistic (Figure 3.8). A manifestation of the lack of an absorbing aerosol term in POLYMER's model can be observed when comparing the June concentration maps to $\text{Chl}_{L2-CHL_{NN}}$ estimations (Figure 3.9), where the standard [Chl-a] neural network was not as susceptible to absorbing aerosols effects. Heightened concentrations of absorbing aerosols, not modelled in POLYMER's atmospheric model may have led to inflated R_{rs} retrievals along the 400-510 nm range (Appendix A.5), resulting in the bio-optical model erroneously overestimating the open ocean [Chl-a]. As mentioned before, POLYMER R_{rs} retrievals in the 400-500 nm ranges were much higher than the synthetic datasets (see Figure 3.2) and were discussed in more detail in section 4.2. POLYMER's atmospheric component parametrization was also based on European Centre of Medium Range Weather Forecast data (Steinmetz et al., 2011) and not South African weather parameters. Although POLYMER's [Chl-a] product performs satisfactorily when taking into account the near river mouth match-up evaluations, open ocean observations were unrealistic as heightened [Chl-a] is expected along the coastline where coastal upwelling and nutrients from river runoff fuel phytoplankton biomass growth. Although a recent study by Wang et al. (2020) did attempt to improve POLYMER's R_{rs} retrievals in the presence of absorbing aerosols, which may assist the [Chl-a] predictions. However, further developing POLYMER's semi-analytical AC model for South African coastal waters may be arduous as the atmospheric, aquatic and oceanic variables are in constant flux throughout the year and are highly variable.

In theory, a well-parameterised regionally specific coupled ocean-atmosphere model could optimally retrieve both atmospheric and biogeophysical data products. However, as discussed, the success of such approaches depends heavily on the availability of *in situ* measurements for model validation. Adapting atmospheric IOPs to account for the regional aerosol type and seasonal variability could be a possible solution to assist South Africa's ocean colour bio-optical model retrievals.

4.5.3 Machine learning algorithms applied to R_{rs}

The synthesised full dataset MLP and constrained dataset MLP models were applied to corrected R_{rs} retrieved from the various ACs to discern the bio-optical constituent concentrations within the river plume. The elected neural network hyper-parameterization, as per Table 3.2, produced near-perfect cross-validation results after the 32 epoch model performance evaluations for both synthetic datasets (see section 3.2). Size cross-validation failures may be attributed to the oligotrophic [Chl-a] ranges, where previous research by Evers-King et al. (2014) along the Benguela current of South Africa has

shown when [Chl-a] is $<$10 mg/m^3, discerning size ranges from R_{rs} retrievals is difficult to achieve. The added effects of sediment backscatter may further confuse the size component of the neural network. Failures in size prediction do not impact [Chl-a] predictions, which may be promising when using the model later on for application purposes, as size and [Chl-a] predictions are not dependent one on another in the neural network. [Chl-a] cross-validation performance relied on the dataset size, where the full synthetic dataset of 18 000 outperformed the constrained dataset of 9600 spectra.

TSM evaluations were similar across all three ACs, with C2RCC performing marginally better, and with higher pixel retrievals (Figure 3.4). The constrained dataset's MLP models performed substantially better than the full dataset (Figure 3.4), where a distinct improvement can be observed when comparing the predictions made by the constrained dataset to the full. This could be an attribute of subsetting the data, which created a MLP that was more closely suited to the optical water type present in the river plume region itself, and not the estuary. As a result, the MLP model's synthetic R_{rs} and TSM training data, with added variabilities in [Chl-a], phytoplankton size and CDOM, did not overestimate TSM concentrations as much. However, [Chl-a] predictions did not improve as extensively after constraining the model. Although, estimates did tend to overpredict while using the constrained dataset and conversely underpredict with the full dataset, for all ACs. The overestimated TSM and underestimated [Chl-a], can be explained further when observing the model's application to the whole river plume system area plots (Figure 3.6), and will be discussed in more detail in this section. Studies using image recognition have found that constraining the contents of an image and highlighting features of significance has shown promising results, accurately detecting characteristics in complex realistic scenarios (Bayar & Stamm, 2018; Oktay et al., 2017). Although different from image recognition exercises, the MLP model used in this study did also experience a similar form of improvement after constraining the full synthetic dataset's R_{rs} to the reflectance retrievals from the different AC algorithms.

Both versions (full and constrained) of the MLP models produced better [Chl-a] results when applied to C2RCC R_{rs} data compared to the standard and POLYMER ACs. This may be an indication of the C2RCC processer producing the closest water-leaving radiances when given highly contaminated top-of-atmosphere reflectances. The actual C2RCC match-up R_{rs} retrievals, plotted in Appendix A.5, show that along the 400-500 nm waveband, R_{rs} signals were more stable and were a closer representative to the synthetic dataset's spectral shape. Research done by Uudeberg et al. (2019) found that C2RCC produced the most reliable and accurate estimates of R_{rs}, for the eventual goal of optical water type classification along 24 Estonian lakes and coastal waters of the Baltic Sea, the Wadden Sea, the southeastern part of the North Sea. Poor standard and POLYMER R_{rs} retrievals along the blue-green wavebands may have negatively impacted the neural network as the MLP models use all wavebands to estimate [Chl-a] and TSM pairs.

Unlike [Chl-a], there was no defining best AC for the TSM models, when assessing the performance of the neural network over the river plume waters collected for the *in situ* dataset. This could be attributed to TSM's prominent backscatter signal reflected in the >500 nm region of the R_{rs} spectrum, being very broad and not as specific as Chl-a, which relied on two Chl-a absorption troughs (445 nm and 675 nm) and a backscattering maximum in between, which can readily be contaminated by other in-water body constituents, as well at atmospheric absorbing and scattering aerosols. Chen et al.(2013) provides a more comprehensive explanation of the problems associated with [Chl-a] retrievals along coastal water.

Using merits from the river mouth statistical evaluations, C2RCC was chosen to be used when generating the MLP's regional area maps. Additional uncertainties in the other ACs can be seen in Appendix A.9, where TSM estimates using the constrained MLP model with the standard and POLYMER ACs show highly variable results, unlike C2RCC's constrained MLP model TSM retrievals, as seen in Figure 3.7. C2RCC was the closest to the best performing TSM algorithm, which was C2RCC IOP sediment mass conversion output, $\text{TSM}_{C2RCC-TSM\_NN}$ (Figure 3.7). Another major improvement made by constraining the MLP can be seen in the panels of Figure 3.6, where the previous model tended to switch between high TSM and low [Chl-a] along coastlines and high [Chl-a] and low TSM in the open ocean. This was possibly a result of the full synthetic dataset containing more values outside of the range of possibilities for the *in situ* measurements in both the coastlines and open ocean, adding variability that adversely affected the MLP model.

Using C2RCC R_{rs}, the constrained MLP TSM and [Chl-a] were mapped relatively well, not drastically switching to extremely low [Chl-a] in the open ocean (Figure 3.7). The [Chl-a] plots did start showing high [Chl-a] along coastlines, which is common in ocean colour [Chl-a] maps, but [Chl-a] was still being overpredicted in the open ocean, not exceeding the upper bounds of the synthetic dataset (5 mg/m^3), but especially so when there were intermediate TSM estimations. The initial dataset was developed for both the river plume region, where phytoplankton communities are proliferating from the released nutrients, and the offshore region of the ocean where nutrient concentrations are limited, inhibiting phytoplankton growth, as [Chl-a] was set to range from 0-5 mg/m^3. However, [Chl-a] collected within the immediate river plume region were accurately quantified by the MLPs (Table 3.4), but not in the offshore waterbody (Figure 3.7). The synthetic dataset used in this study was heavily approximated, as only TSM and [Chl-a] biogeochemical *in situ* parameters were measured and incorporated into the IOP models. Mixed sediment size distributions, variable phytoplankton assemblages, and CDOM concentrations were not incorporated into the IOP models. Thus, the synthetic dataset may not be an accurate representation of the in-water constituent concentration ranges of the study region, as the R_{rs} spectral variability might not have incorporated the true spectra received by satellites. Without validation of the synthetic dataset, it is not possible to identify the

causes of error in the MLP models. Increasing the complexity of the MLP model using additional computational and statistical techniques (Kravitz et al., 2021; Soriano-González et al., 2022) is unlikely to improve retrievals where confidence in the synthetic dataset cannot be determined.

Other studies have tried to combat the poor performance of current ocean colour neural network algorithms over optically complete waters by creating smaller models, specific to the range of R_{rs} received. Studies by Soriano-González et al. (2022) and Giannini et al. (2021) have employed the use of C2RCC alternative neural networks, specified for optically complex, highly turbid waters, to quantify TSM concentrations and retrieve reflectances. Like the constrained MLP model in this study, the alternative neural networks are constrained to the possible optical characteristics of the selected study region. Subsequent retrieval predictions for R_{rs} and TSM concentration improvements by using different versions of C2RCC neural networks developed for different types of turbid waters, as opposed to the full generic C2RCC neural network. A more nuanced method applied by Kravitz et al. (2021) used k-means clustering to characterize distinct water types present among the suite of synthetic R_{rs} generated for the inland waterbody. Clustering was highlighted as a common technique used for both the creation and application of biogeophysical remote sensing optical models, which improves our ability to discern the underlying optical relationships and variability amongst generated spectra. Functional data analysis (FDA) techniques were further used; FDA is the field of statistics where individual continuous functions are treated as data points, thereby simplifying the application of statistical techniques, such as k-means clustering, which in a normal instance uses single data values as opposed to functions. Thereafter, individual MLP models for each of the water types robustly defined in Kravitz et al. (2021) input R_{rs} and output constituent concentration pairs were created, and performances of the match-up retrievals for each water type neural network were evaluated. Developing ocean colour neural network algorithms in the future may incorporate water type identification, using FDA approaches, and state-of-the-art deep learning techniques with regionally adapted, water body-specific synthetic datasets.

4.6 Limitations and recommendations

The lack of *in situ* measurements such as CDOM concentration, phytoplankton species identification and size classification, as well as sediment type delineation and size distributions, made parameterising the synthetic dataset difficult. Thus, the synthetic dataset used in this study was heavily approximated as only Chl-a and TSS concentrations were collected. Additionally, the dataset size was small, and the seasonal variability was only represented by two sampling trips. Thus, the MLPs were broadly regionally and seasonally parameterized, which may have led to their subsequent failures. It is recommended that the future *in situ* [Chl-a], TSM, CDOM, phytoplankton size and sedimentary particle size distributions should be collected from river plumes and adjacent open nearshore pelagic

ecosystems to ensure that the synthetic dataset captures the full range of variables.

The modelled synthetic dataset R_{rs} spans 101 wavebands across the visible light spectrum, from 400 nm to 900 nm, unlike Sentinel-3's OLCI bands, which only capture reflectances over 12 wavelengths from 400 to 754 nm. The regional neural networks will be applied to satellite data, thus, creating a model that fits OLCI's 12 bands, as opposed to extrapolating OLCI Sentinel-3 reflectance to match the hyperspectral bands, will be of optimal efficiency when in use in the future. However, during the process of interpolation, information was lost in the regions where OLCI does not collect R_{rs} data, as well as in the NIR bands, beyond 754 nm, where OLCI's wavebands are treated as regions for atmospheric correction extrapolation. The constituent most affected by the removal of NIR AC bands is TSM, which even with concentrations as low as 0.2 g/m^3, backscatter can be detected in the NIR wavelengths (Doxaran et al., 2002). The synthetic dataset generated in this study does produce R_{rs} signals relating to TSM backscatter in the NIR, although much of it was removed during the interpolation process. The potential for hyperspectral R_{rs} to improve retrievals is promising, but not yet fully understood. The neural network approach is similar to analytical and semi-analytical algorithms, which require full spectral data and very effective AC methodologies. However, simplistic empirical band ratio algorithms, such as the Maximum Band Ratio technique, as described in O'Reilly et al. (1998), where the loss of blue bands to poor aerosol correction is less critical, may at times be a more appropriate algorithm for [Chl-a] estimations.

The accuracy of the ACs used in this study was not fully assessed, as *in situ* R_{rs} data collection was beyond the scope of the study; thus the most appropriate AC cannot be conclusively determined in the present study. Without *in situ* R_{rs} data we cannot accurately delineate the effects of poor AC performance from suboptimal MLP model architecture, thus, it is not clear that improving the MLP's structure will ensure better results for R_{rs} from any given AC. Future studies aimed at improving remote sensing applications in the region should endeavour to collect more *in situ* measurements, including a broad range of biogeochemical, bio-optical and radiometric measurements with coincident satellite data. Land adjacency effects may have also added to the satellite's signal, creating heightened reflectances retrieved in the estuarine match-up sites. Land adjacency is a major problem in remote sensing as there are very few measures in place to correct for such effects, besides quality flags along the coastlines; as a result, some standard Level-2 pixels were flagged out near the estuary mouth. Land adjacency processors such as SIMilarity Environment Correction (Sterckx et al., 2015) and Improved Contrast between Ocean and Land (Kratzer & Vinterhav, 2010) processor are available for application to Sentinel-3 OLCI data; while assessing their effectiveness was beyond the scope of this study, future research should endeavour to include such analyses.

4.7 Conclusion

Appropriate atmospheric correction, particularly over relatively dark targets such as the marine environment, is paramount in remote sensing. Unfortunately, a lack of coincident *in situ* radiometric data makes it nearly impossible to comprehensively validate the uncertainty of the output R_{rs} from different ACs, or how these may propagate through to their derived biogeophysical data products. [Chl-a] estimates were the most affected by atmospheric effects, as Fynbos biogenic burning events are common in the region. Appropriate quality flagging of pixels near land masses is important for statistical evaluation metrics, where performances can be misleading. For example, if the OC4ME algorithm for [Chl-a] has a small error margin, but the remaining match-up dataset represented mostly offshore stations; this does not imply that it is broadly appropriate for application over the entire region of interest (i.e. including the estuarine environment).

The use of satellite machine learning algorithms trained on a synthetic dataset is only as appropriate as the synthetic data the models are trained on. Synthetic datasets are used to compensate for the lack of IOP and radiometric data from the study region; thus, abundant *in situ* data for the assessed biogeochemical products is needed to create an appropriate representation of the water-leaving reflectances. Improved performance of MLP models may be achieved after appropriate validation of the input dataset's IOP models as accurate stand-alone representations of the study region. More detailed *in situ* data can be used to create highly parameterized synthetic datasets and for validation of the developed MLP models. Thereafter, other statistical and computational techniques can be applied to the large and diverse synthetic dataset to further improve the MLP performances. Constraining models/algorithms appears to be the most effective method to improve product retrievals. There may be a balance between representing variability in the synthetic dataset and constraining the degrees of freedom in the MLP to achieve accurate biogeophysical retrievals, especially for the most dominant constituent to R_{rs}, such as TSM in highly turbid, optically complex river plume waters. It is the hope that the findings in this thesis will help pave the way toward increased bio-optically focused research and improved satellite-derived models for biogeophysical products over the Agulhas Bank region in the future.

Although the MLP approach produced suboptimal results, this study was the first to validate ocean colour satellite data products for river plumes in the Agulhas Bank. The findings show that some standard ocean colour products are very much fit for use over the study area. These preliminary results provide a very useful performance assessment and guide to optimal standard ocean colour products for the atmospherically variable region.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J. M., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P. A., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: a system for large-scale machine learning. *Operating Systems Design and Implementation*, 265–283. https://doi.org/10.5555/3026877.3026899

Agawin, N. S. R., Duarte, C. M., & Agustı, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnology and Oceanography*, *45*(3), 591–600.

Ahmad, Z., McClain, C. R., Herman, J. R., Franz, B. A., Kwiatkowska, E. J., Robinson, W. D., Bucsela, E. J., & Tzortziou, M. (2007). Atmospheric correction for NO 2 absorption in retrieving water-leaving reflectances from the SeaWiFS and MODIS measurements. *Applied Optics*, *46*(26), 6504–6512.

Ahn, Y.-H. (1990). *Propriétés optiques des particules biologiques et minérales présentes dans l'océan. application: Inversion de la réflectance* (Doctoral dissertation). Paris 6.

Aller, J. Y., & Aller, R. C. (1986). General characteristics of benthic faunas on the Amazon inner continental shelf with comparison to the shelf off the Changjiang River, East China Sea. *Continental Shelf Research*, *6*(1-2), 291–310.

Anderson, D. M., Glibert, P. M., & Burkholder, J. M. (2002). Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries*, *25*, 704–726.

Ashraf, M., Iqbal, M. F., Rauf, M., Ashraf, M. U., Ulhaq, A., Muhammad, H., & Liu, Q.-f. (2022). Developing a sustainable concrete incorporating bentonite clay and silica fume: Mechanical and durability performance. *Journal of Cleaner Production*, *337*, 130315.

Atkinson, P. M., & Tatnall, A. R. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Rensing*, *18*(4), 699–709.

Ault, T., Velzeboer, R., & Zammit, R. (2000). Influence of nutrient availability on phytoplankton growth and community structure in the Port Adelaide River, Australia:[2pt] bioassay assessment of potential nutrient limitation. *Hydrobiologia*, *429*, 89–103.

Aurin, D. A., & Dierssen, H. M. (2012). Advantages and limitations of ocean color remote sensing in CDOM-dominated, mineral-rich coastal and estuarine waters. *Remote Sensing of Environment*, *125*, 181–197.

Ayukai, T., & Wolanski, E. (1997). Importance of biologically mediated removal of fine sediments from the Fly River plume, Papua New Guinea. *Estuarine, Coastal and Shelf Science*, *44*(5), 629–639.

Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, *102*(1-2), 12–23.

Baltar, F., Alvarez-Salgado, X. A., Arıstegui, J., Benner, R., Hansell, D. A., Herndl, G. J., & Lønborg, C. (2021). What is refractory organic matter in the ocean? *Frontiers in Marine Science*, *8*, 327.

Barbieri, G., Rossi, V., Vaiani, S. C., & Horton, B. P. (2019). Benthic ostracoda and foraminifera from the North Adriatic Sea (Italy, Mediterranean Sea): A proxy for the depositional characterisation of river-influenced shelves. *Marine Micropaleontology*, *153*, 101772.

Bargu, S., White, J. R., Li, C., Czubakowski, J., & Fulweiler, R. W. (2011). Effects of freshwater input on nutrient loading, phytoplankton biomass, and cyanotoxin production in an oligohaline estuarine lake. *Hydrobiologia*, *661*, 377–389.

Basu, S., & Mackey, K. R. M. (2018). Phytoplankton as key mediators of the biological carbon pump: Their responses to a changing climate. *Sustainability*, *10*(3), 869.

Bate, G. C. (2013). Epipelic diatoms in the estuaries of South Africa. *Water SA*, *39*(1), 105–118.

Bauer, J., & Bianchi, T. (2011). 5.02—dissolved organic carbon cycling and transformation. *Treatise on Estuarine and Coastal Science*, *5*, 7–67.

Bayar, B., & Stamm, M. C. (2018). Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, *13*(11), 2691–2706.

Bernard, C. Y., Dürr, H. H., Heinze, C., Segschneider, J., & Maier-Reimer, E. (2011). Contribution of riverine nutrients to the silicon biogeochemistry of the global ocean–a model study. *Biogeosciences*, *8*(3), 551–564.

Bernard, S., Probyn, T. A., & Quirantes, A. (2009). Simulating the optical properties of phytoplankton cells using a two-layered spherical geometry. *Biogeosciences Discussions*, *6*(1), 1497–1563.

Bernard, S., Pitcher, G., Evers-King, H., Robertson, L., Matthews, M., Rabagliati, A., & Balt, C. (2014). Ocean colour remote sensing of harmful algal blooms in the Benguela system. *Remote Sensing of the African Seas*, 185–203.

Best, J. (2019). Anthropogenic stresses on the world's big rivers. *Nature Geoscience*, *12*(1), 7–21.

Binding, C. E., Bowers, D. G., & Mitchelson-Jacob, E. G. (2005). Estimating suspended sediment concentrations from ocean colour measurements in moderately turbid waters; the impact of variable particle scattering properties. *Remote Sensing of Environment*, *94*(3), 373–383.

Bowers, D. G., & Binding, C. E. (2006). The optical properties of mineral suspended particles: A review and synthesis. *Estuarine, Coastal and Shelf Science*, *67*(1-2), 219–230.

Brewton, R. A., Kreiger, L. B., Tyre, K. N., Baladi, D., Wilking, L. E., Herren, L. W., & Lapointe, B. E. (2022). Septic system–groundwater–surface water couplings in waterfront communities contribute to harmful algal blooms in Southwest Florida. *Science of The Total Environment*, *837*, 155319.

Bricaud, A., Babin, M., Morel, A., & Claustre, H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. *Journal of Geophysical Research: Oceans*, *100*(C7), 13321–13332.

Bricaud, A., Morel, A., Prieur, L., et al. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, *26*(1), 43–53.

Broadwater, M. H., Van Dolah, F. M., & Fire, S. E. (2018). Vulnerabilities of marine mammals to harmful algal blooms. *Harmful algal blooms: A compendium desk reference*, 191–222.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., & Ruescas, A. (2016). Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. *Living Planet Symposium*, *740*, 54.

Carder, K. L., Chen, F. R., Cannizzaro, J. P., Campbell, J. W., & Mitchell, B. G. (2004). Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. *Advances in Space Research*, *33*(7), 1152–1159.

Carder, K. L., Hawes, S. K., Baker, K. A., Smith, R. C., Steward, R. G., & Mitchell, B. G. (1991). Reflectance model for quantifying chlorophyll a in the presence of productivity degradation products. *Journal of Geophysical Research: Oceans*, *96*(C11), 20599–20611.

Carreto, J. I., Montoya, N., Akselman, R., Carignan, M. O., Silva, R. I., & Colleoni, D. A. C. (2008). Algal pigment patterns and phytoplankton assemblages in different water masses of the Rıo de la Plata maritime front. *Continental Shelf Research*, *28*(13), 1589–1606.

Chen, J., Zhang, M., Cui, T., & Wen, Z. (2013). A review of some important technical problems in respect of satellite remote sensing of chlorophyll-a concentration in coastal waters. *IEEE journal of selected topics in applied earth observations and remote sensing*, *6*(5), 2275–2289.

Chen, S., Huang, W., Chen, W., & Chen, X. (2011). An enhanced MODIS remote sensing model for detecting rainfall effects on sediment plume in the coastal waters of Apalachicola Bay. *Marine Environmental Research*, *72*(5), 265–272.

Chiswell, S. M., Calil, P. H., & Boyd, P. W. (2015). Spring blooms and annual cycles of phytoplankton: A unified perspective. *Journal of Plankton Research*, *37*(3), 500–508.

Chollet, F., et al. (2015). *Keras*. https://github.com/fchollet/keras

Clark, D. K., Gordon, H. R., Voss, K. J., Ge, Y., Broenkow, W., & Trees, C. (1997). Validation of atmospheric correction over the oceans. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 17209–17217.

Cloern, J. E., & Jassby, A. D. (2010). Patterns and scales of phytoplankton variability in estuarine–coastal ecosystems. *Estuaries and Coasts*, *33*, 230–241.

Collier, M., Webb, R. H., & Schmidt, J. C. (1996). *Dams and rivers: A primer on the downstream effects of dams* (Vol. 1126). US Department of the Interior, US Geological Survey.

Corredor, J. E., Morell, J. M., LoPEZ, J. M., Capella, J. E., & Armstrong, R. A. (2004). Cyclonic eddy entrains Orinoco River plume in eastern Caribbean. *Eos, Transactions American Geophysical Union*, *85*(20), 197–202.

Danovaro, R., Gambi, C., Manini, E., & Fabiano, M. (2000). Meiofauna response to a dynamic river plume front. *Marine Biology*, *137*(2), 359–370.

De Rose, R., Wilson, D. J., Bartley, R., & Wilkinson, S. (2005). Riverbank erosion and its importance to uncertainties in large scale sediment budgets. *Sediment Budgets Proceedings of 7th IAHS Scientific Assembly, Foz do Igunzu, Brazil*, 85–92.

Defoin-Platel, M., & Chami, M. (2007). How ambiguous is the inverse problem of ocean color in coastal waters? *Journal of Geophysical Research: Oceans*, *112*(C3).

Dekker, A. G., Vos, R. J., & Peters, S. W. M. (2002). Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. *International Journal of Remote Sensing*, *23*(1), 15–35.

Deng, S., Chen, T., Yang, N., Qu, L., Li, M., & Chen, D. (2018). Spatial and temporal distribution of rainfall and drought characteristics across the Pearl River basin. *Science of the Total Environment*, *619*, 28–41.

Department of Water and Sanitation (DWS), (2014). Reserve Determination Studies for Surface Water, Groundwater, Estuaries and Wetlands in the Gouritz Water Management Area: Estuaries RDM Report – Intermediate Assessment, Volume 1 (Duiwenhoks Estuary). Prepared by the Council for Scientific and Industrial Research (CSIR) for Scherman Colloty and Associates cc. Report no. RDM/WMA16/04/CON/0813, Volume 1. (n.d.).

Department of Water and Sanitation (DWS), (2015). Reserve Determination Studies for Surface Water, Groundwater, Estuaries and Wetlands in the Gouritz Water Management Area: Estuaries RDM Report – Intermediate Assessment, Volume 2 (Gouritz Estuary). Prepared by the Council for Scientific and Industrial Research (CSIR) for Scherman Colloty and Associates cc. Report no. RDM/WMA16/04/CON/0813, Volume 2.. (n.d.).

Department of Water and Sanitation (DWS), (2015). Reserve Determination Studies for Surface Water, Groundwater, Estuaries and Wetlands in the Gouritz Water Management Area: Estuaries RDM Report – Intermediate Assessment, Volume 3 (Goukou Estuary). Prepared by the Council

for Scientific and Industrial Research (CSIR) for Scherman Colloty and Associates cc. Report no. RDM/WMA16/04/CON/0813, Volume 3. (n.d.).

Dethier, E. N., Renshaw, C. E., & Magilligan, F. J. (2022). Rapid changes to global river suspended sediment flux by humans. *Science*, *376*(6600), 1447–1452.

Diaz, R. J., & Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science*, *321*(5891), 926–929.

Doerffer, R., & Schiller, H. (2007). The meris case 2 water algorithm. *International Journal of Remote Sensing*, *28*(3-4), 517–535.

Doxaran, D., Froidefond, J.-M., & Castaing, P. (2002). A reflectance band ratio used to estimate suspended matter concentrations in sediment-dominated coastal waters. *International Journal of Remote Sensing*, *23*(23), 5079–5085.

Drinkwater, K. F. (1986). On the role of freshwater outflow on coastal marine ecosystems—A workshop summary. In *The role of freshwater outflow in coastal marine ecosystems* (pp. 429–438). Springer.

Ducklow, H. W., & Kirchman, D. L. (1983). Bacterial dynamics and distribution during a spring diatom bloom in the Hudson River plume, USA. *Journal of Plankton Research*, *5*(3), 333–355.

Dufour, P., Charpy, L., Bonnet, S., & Garcia, N. (1999). Phytoplankton nutrient control in the oligotrophic South Pacific subtropical gyre (Tuamotu Archipelago). *Marine Ecology Progress Series*, *179*, 285–290.

Dybas, C. L. (2005). Dead zones spreading in world oceans. *BioScience*, *55*(7), 552–557.

ESA. (2023a). *Alternative Atmospheric Correction (Neural network) - Sentinel online*. https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-3-olci/level-2/alternative-atmospheric-correction

ESA. (2023b). *Baseline Atmospheric Correction- Sentinel Online*. https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-3-olci/level-2/baseline-atmospheric-correction

ESA. (2023c). *IMT Neural Net - Sentinel Online*. https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-3-olci/level-2/imt-neural-net

ESA. (2023d). *OC4Me Chlorophyll - Sentinel Online*. https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-3-olci/level-2/oc4me-chlorophyll

Evers-King, H., Bernard, S., Lain, L. R., & Probyn, T. A. (2014). Sensitivity in reflectance attributed to phytoplankton cell size: Forward and inverse modelling approaches. *Optics Express*, *22*(10), 11536–11551.

Flemming, B., & Martin, K. (2021). The Breede River estuary (Cape Province, South Africa): A historical perspective on hydrology, geomorphology, and sedimentology. *Geo-Marine Letters*, *41*(1), 15.

Fong, C. R., Gaynus, C. J., & Carpenter, R. C. (2020). Extreme rainfall events pulse substantial nutrients and sediments from terrestrial to nearshore coastal communities: A case study from French Polynesia. *Scientific Reports*, *10*(1), 2955.

Frederiksen, M., Edwards, M., Richardson, A. J., Halliday, N. C., & Wanless, S. (2006). From plankton to top predators: Bottom-up control of a marine food web across four trophic levels. *Journal of Animal Ecology*, *75*(6), 1259–1268.

Freeman, L. A., Corbett, D. R., Fitzgerald, A. M., Lemley, D. A., Quigg, A., & Steppe, C. N. (2019). Impacts of urbanization and development on estuarine ecosystems and water quality. *Estuaries and Coasts*, *42*, 1821–1838.

Froidefond, J. M., Castaing, P., Jouanneau, J. M., Prud'Homme, R., & Dinet, A. (1993). Method for the quantification of suspended sediments from AVHRR NOAA-11 satellite data. *International Journal of Remote Sensing*, *14*(5), 885–894.

Giannini, F., Hunt, B. P., Jacoby, D., & Costa, M. (2021). Performance of OLCI Sentinel-3A satellite in the Northeast Pacific coastal waters. *Remote Sensing of Environment*, *256*, 112317.

Gin, K. Y.-H., Lin, X., & Zhang, S. (2000). Dynamics and size structure of phytoplankton in the coastal waters of Singapore. *Journal of Plankton Research*, *22*(8), 1465–1484.

Goineau, A., Fontanier, C., Jorissen, F., Buscail, R., Kerhervé, P., Cathalot, C., Pruski, A., Lantoine, F., Bourgeois, S., Metzger, E., et al. (2012). Temporal variability of live (stained) benthic foraminiferal faunas in a river-dominated shelf–Faunal response to rapid changes of the river influence (Rhône prodelta, NW Mediterranean). *Biogeosciences*, *9*(4), 1367–1388.

Gordon, H. R., Clark, D. K., Mueller, J. L., & Hovis, W. A. (1980). Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: comparisons with surface measurements. *Science*, *210*(4465), 63–66.

Gratiot, N., & Anthony, E. J. (2016). Role of flocculation and settling processes in development of the mangrove-colonized, Amazon-influenced mud-bank coast of South America. *Marine Geology*, *373*, 1–10.

Grebmeier, J. M., McRoy, C. P., & Feder, H. M. (1988). Pelagic-benthic coupling on the shelf of the northern Bering and Chukchi Seas. I. food supply source and benthic biomass. *Marine Ecology Progress Series*, 57–67.

Guéguen, C., Guo, L., & Tanaka, N. (2005). Distributions and characteristics of colored dissolved organic matter in the western Arctic Ocean. *Continental Shelf Research*, *25*(10), 1195–1207.

Guo, L., & He, Q. (2011). Freshwater flocculation of suspended sediments in the Yangtze River, China. *Ocean Dynamics*, *61*, 371–386.

Han, A., Dai, M., Kao, S.-J., Gan, J., Li, Q., Wang, L., Zhai, W., & Wang, L. (2012). Nutrient dynamics and biological consumption in a large continental shelf system under the influence of both a river plume and coastal upwelling. *Limnology and Oceanography*, *57*(2), 486–502.

Han, H., Xiao, R., Gao, G., Yin, B., Liang, S., et al. (2023). Influence of a heavy rainfall event on nutrients and phytoplankton dynamics in a well-mixed semi-enclosed bay. *Journal of Hydrology*, *617*, 128932.

Harris, C. K., & Wiberg, P. (2002). Across-shelf sediment transport: Interactions between suspended sediment and bed sediment. *Journal of Geophysical Research: Oceans*, *107*(C1), 8–1.

Hatcher, A., Hill, P., & Grant, J. (2001). Optical backscatter of marine flocs. *Journal of Sea Research*, *46*(1), 1–12.

Hay, W. W. (1998). Detrital sediment fluxes from continents to oceans. *Chemical Geology*, *145*(3-4), 287–323.

Heezen, B. C., Menzies, R., Schneider, R., Ewing, W., & Granelli, N. (1964). Congo submarine canyon. *AAPG bulletin*, *48*(7), 1126–1149.

Hermand, R., Salen-Picard, C., Alliot, E., & Degiovanni, C. (2008). Macrofaunal density, biomass and composition of estuarine sediments and their relationship to the river plume of the Rhone River (NW Mediterranean). *Estuarine, Coastal and Shelf Science*, *79*(3), 367–376.

Hicks, D. M., Hill, J., & Shankar, U. (1996). Variation of suspended sediment yields around New Zealand: the relative importance of rainfall and geology. *IAHS publication*, 149–156.

Hieronymi, M., Müller, D., & Doerffer, R. (2017). The OLCI neural network swarm (ONNS): A bio-geo-optical algorithm for open ocean and coastal waters. *Frontiers in Marine Science*, *4*, 140.

Hovis, W. A., Clark, D. K., Anderson, F., Austin, R. W., Wilson, W. H., Baker, E. T., Ball, D., Gordon, H. R., Mueller, J. L., El-Sayed, S. Z., et al. (1980). Nimbus-7 Coastal Zone Color Scanner: system description and initial imagery. *Science*, *210*(4465), 60–63.

Hovius, N. (1998). Controls on sediment supply by large rivers.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M., & Claustre, H. (2007). Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean. *Biogeosciences*, *4*(5), 853–868.

Hutchings, L., Pitcher, G. C., Probyn, T. A., Bailey, G. W., et al. (1995). The chemical and biological consequences of coastal upwelling. *Environmental Sciences Research Report Es*, *18*, 65–82.

Hutchings, L., Van der Lingen, C. D., Shannon, L. J., Crawford, R. J. M., Verheye, H. M. S., Bartholomae, C. H., Van der Plas, A. K., Louw, D., Kreiner, A., Ostrowski, M., et al. (2009). The Benguela Current: an ecosystem of four components. *Progress in Oceanography*, *83*(1-4), 15–32.

Iken, K., Brey, T., Wand, U., Voigt, J., & Junghans, P. (2001). Food web structure of the benthic community at the Porcupine Abyssal Plain (NE Atlantic): a stable isotope analysis. *Progress in Oceanography*, *50*(1-4), 383–405.

Imbeau, E., Vincent, W. F., Wauthy, M., Cusson, M., & Rautio, M. (2021). Hidden stores of organic matter in northern lake ice: Selective retention of terrestrial particles, phytoplankton and labile carbon. *Journal of Geophysical Research: Biogeosciences*, *126*(8), e2020JG006233.

Iqbal, M. F., Javed, M. F., Rauf, M., Azim, I., Ashraf, M., Yang, J., & Liu, Q.-f. (2021). Sustainable utilization of foundry waste: Forecasting mechanical properties of foundry sand based concrete using multi-expression programming. *Science of the Total Environment*, *780*, 146524.

Ittekkot, V. (1988). Global trends in the nature of organic matter in river suspensions. *Nature*, *332*(6163), 436–438.

Jacobs, Z., Roberts, M., Jebri, F., Srokosz, M., Kelly, S., Sauer, W., Bruggeman, J., & Popova, E. (2022). Drivers of productivity on the Agulhas Bank and the importance for marine ecosystems. *Deep Sea Research Part II: Topical Studies in Oceanography*, *199*, 105080.

Jassby, A., Cloern, J., Mueller-Solger, A., et al. (2003). Phytoplankton fuels delta food web. *California Agriculture*, *57*(4), 104–109.

Jiang, D., Scholze, J., Liu, X., Simis, S. G., Stelzer, K., Müller, D., Hunter, P., Tyler, A., & Spyrakos, E. (2023). A data-driven approach to flag land-affected signals in satellite derived water quality from small lakes. *International Journal of Applied Earth Observation and Geoinformation*, *117*, 103188.

Joint, I., & Groom, S. B. (2000). Estimation of phytoplankton production from space: Current status and future potential of satellite remote sensing. *Journal of Experimental Marine Biology and Ecology*, *250*(1-2), 233–255.

Jones, R. I. (1998). Phytoplankton, primary production and nutrient cycling. In *Aquatic humic substances: Ecology and biogeochemistry* (pp. 145–175). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kankaanpää, H. T., Holliday, J., Schröder, H., Goddard, T. J., von Fister, R., & Carmichael, W. W. (2005). Cyanobacteria and prawn farming in northern New South Wales, Australia—a case study on cyanobacteria diversity and hepatotoxin bioaccumulation. *Toxicology and Applied Pharmacology*, *203*(3), 243–256.

Keiner, L. E. (1999). Estimating oceanic chlorophyll concentrations with neural networks. *International Journal of Remote Sensing*, *20*(1), 189–194.

Kemp, P., Sear, D., Collins, A., Naden, P., & Jones, I. (2011). The impacts of fine sediment on riverine fish. *Hydrological Processes*, *25*(11), 1800–1821.

Kench, P. S. (1999). Geomorphology of Australian estuaries: Review and prospect. *Australian Journal of Ecology*, *24*(4), 367–380.

Kiørboe, T. (1993). Turbulence, phytoplankton cell size, and the structure of pelagic food webs. In *Advances in marine biology* (pp. 1–72). Elsevier.

Kiørboe, T. (2008). Watershed management strategies to prevent and control cyanobacterial harmful algal blooms. In *Cyanobacterial harmful algal blooms: State of the science and research needs* (pp. 259–273). Springer New York.

Kranck, K. (1980). Experiments on the significance of flocculation in the settling of fine-grained sediment in still water. *Canadian Journal of Earth Sciences*, *17*(11), 1517–1526.

Kratzer, S., & Plowey, M. (2021). Integrating mooring and ship-based data for improved validation of OLCI chlorophyll-a products in the Baltic Sea. *International Journal of Applied Earth Observation and Geoinformation*, *94*, 102212.

Kratzer, S., & Vinterhav, C. (2010). Improvement of MERIS level 2 products in Baltic Sea coastal areas by applying the Improved Contrast between Ocean and Land processor (ICOL)-data analysis and validation. *Oceanologia*, *52*(2), 211–236.

Kravitz, J., Matthews, M., Lain, L., Fawcett, S., & Bernard, S. (2021). Potential for high fidelity global mapping of common inland water quality products at high spatial and temporal resolutions based on a synthetic data and machine learning approach. *Frontiers in Environmental Science*, *9*, 587660.

Kruger, F., & Bigalke, R. (1984). Fire in fynbos. *Ecological effects of fire in South African ecosystems*, 67–114.

Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., & Casal, G. (2016). Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. *Remote Sensing*, *8*(6), 497.

Kyryliuk, D., & Kratzer, S. (2019). Evaluation of Sentinel-3A OLCI products derived using the Case-2 Regional CoastColour processor over the baltic sea. *Sensors*, *19*(16), 3609.

Lain, L. R., Bernard, S., & Evers-King, H. (2014). Biophysical modelling of phytoplankton communities from first principles using two-layered spheres: Equivalent Algal Populations (EAP) model. *Optics express*, *22*(14), 16745–16758.

Lamberth, S. J., & Turpie, J. K. (2003). The role of estuaries in South African fisheries: Economic importance and management implications. *African Journal of Marine Science*, *25*, 131–157.

Lamberth, S. J., Van Niekerk, L., & Hutchings, K. (2008). Comparison of, and the effects of altered freshwater inflow on, fish assemblages of two contrasting South African estuaries: The cool-temperate olifants and the warm-temperate breede. *African Journal of Marine Science*, *30*(2), 311–336.

Landsberg, J. H. (2002). The effects of harmful algal blooms on aquatic organisms. *Reviews in Fisheries Science*, *10*(2), 113–390.

Langbein, W. B., & Schumm, S. A. (1958). Yield of sediment in relation to mean annual precipitation. *Eos, Transactions American Geophysical Union*, *39*(6), 1076–1084.

Largier, J. L., Chapman, P., Peterson, W. T., & Swart, V. P. (1992). The western Agulhas Bank: circulation, stratification and ecology. *South African Journal of Marine Science*, *12*(1), 319–339.

Lathrop, R. G., & Lillesand, T. M. (1989). Monitoring water quality and river plume transport in Green Bay, Lake Michigan with SPOT-1 imagery. *Photogrammetric Engineering & Remote Sensing*, *55*(3), 349–354.

Lee, H. J., & Chu, Y. S. (2001). Origin of inner-shelf mud deposit in the southeastern Yellow Sea: Huksan Mud Belt. *Journal of Sedimentary Research*, *71*(1), 144–154.

Lemley, D. A., Adams, J. B., Taljaard, S., & Strydom, N. A. (2015). Towards the classification of eutrophic condition in estuaries. *Estuarine, Coastal and Shelf Science*, *164*, 221–232.

Levin, L. A., Boesch, D. F., Covich, A., Dahm, C., Erséus, C., Ewel, K. C., Kneib, R. T., Moldenke, A., Palmer, M. A., Snelgrove, P., et al. (2001). The function of marine critical transition zones and the importance of sediment biodiversity. *Ecosystems*, *4*, 430–451.

Li, H.-M., Tang, H.-J., Shi, X.-Y., Zhang, C.-S., & Wang, X.-L. (2014). Increased nutrient loads from the Changjiang (Yangtze) River have led to increased harmful algal blooms. *Harmful Algae*, *39*, 92–101.

Li, S., Zhang, Z., Zhou, M., Wang, C., Wu, H., & Zhong, Y. (2022). The role of fronts in horizontal transports of the Changjiang River plume in summer and the implications for phytoplankton blooms. *Journal of Geophysical Research: Oceans*, *127*(8), e2022JC018541.

Liu, H., & Dagg, M. (2003). Interactions between nutrients, phytoplankton growth, and micro-and mesozooplankton grazing in the plume of the Mississippi River. *Marine Ecology Progress Series*, *258*, 31–42.

Liu, H., He, X., Li, Q., Hu, X., Ishizaka, J., Kratzer, S., Yang, C., Shi, T., Hu, S., Zhou, Q., et al. (2021). Evaluation of ocean color atmospheric correction methods for Sentinel-3 OLCI using global automatic in situ observations. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–19.

Lohrenz, S. E., Fahnenstiel, G. L., Redalje, D. G., Lang, G. A., Dagg, M. J., Whitledge, T. E., & Dortch, Q. (1999). Nutrients, irradiance, and mixing as factors regulating primary production in coastal waters impacted by the Mississippi River plume. *Continental Shelf Research*, *19*(9), 1113–1141.

Lohrenz, S. E., Redalje, D. G., Cai, W., Acker, J., & Dagg, M. (2008). A retrospective analysis of nutrients and phytoplankton productivity in the Mississippi River plume. *Continental Shelf Research*, *28*(12), 1466–1475.

Loisel, H., & Morel, A. (1998). Light scattering and chlorophyll concentration in case 1 waters: A reexamination. *Limnology and Oceanography*, *43*(5), 847–858.

Lowe, M. L., Morrison, M. A., & Taylor, R. B. (2015). Harmful effects of sediment-induced turbidity on juvenile fish in estuaries. *Marine Ecology Progress Series*, *539*, 241–254.

Lunven, M., Guillaud, J. F., Youénou, A., Crassous, M. P., Berric, R., Le Gall, E., Kérouel, R., Labry, C., & Aminot, A. (2005). Nutrient and phytoplankton distribution in the Loire River plume (Bay of Biscay, France) resolved by a new Fine Scale Sampler. *Estuarine, Coastal and Shelf Science*, *65*(1-2), 94–108.

Lutjeharms, J. R. E., Meyer, A. A., Ansorge, I. J., Eagle, G. A., & Orren, M. J. (1996). The nutrient characteristics of the Agulhas Bank. *South African Journal of Marine Science*, *17*(1), 253–274.

Mäkelä, A., Witte, U., & Archambault, P. (2017). Ice algae versus phytoplankton: Resource utilization by Arctic deep sea macroinfauna revealed through isotope labelling experiments. *Marine Ecology Progress Series*, *572*, 1–18.

Mann, K. H. (1988). Production and use of detritus in various freshwater, estuarine, and coastal marine ecosystems. *Limnology and Oceanography*, *33*(4part2), 910–930.

Marañón, E., Holligan, P. M., Barciela, R., González, N., Mouriño, B., Pazó, M. J., & Varela, M. (2001). Patterns of phytoplankton size structure and productivity in contrasting open-ocean environments. *Marine Ecology Progress Series*, *216*, 43–56.

Matthews, M. W., & Bernard, S. (2013). Using a two-layered sphere model to investigate the impact of gas vacuoles on the inherent optical properties of Microcystis aeruginosa. *Biogeosciences*, *10*(12), 8139–8157.

McCabe, R. M., MacCready, P., & Hickey, B. M. (2009). Ebb-tide dynamics and spreading of a large river plume. *Journal of Physical Oceanography*, *39*(11), 2839–2856.

Mehdizadeh Allaf, M., & Peerhossaini, H. (2022). Cyanobacteria: Model microorganisms and beyond. *Microorganisms*, *10*(4), 696.

Mei, Z.-P., Finkel, Z. V., & Irwin, A. J. (2009). Light and nutrient availability affect the size-scaling of growth in phytoplankton. *Journal of Theoretical Biology*, *259*(3), 582–588.

Mertes, L. A. K., & Dunne, T. (2007). *Effects of tectonism, climate change, and sea-level change on the form and behaviour of the modern Amazon River and its floodplain*. Wiley Chichester.

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., et al. (2006). Inverse estimates of anthropogenic CO2 uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, *20*(2).

Milligan, T. G., Hill, P. S., & Law, B. A. (2007). Flocculation and the loss of sediment from the Po River plume. *Continental Shelf Research*, *27*(3-4), 309–321.

Milliman, J. D., & Farnsworth, K. L. (2013). *River discharge to the coastal ocean: A global synthesis*. Cambridge University Press.

Milliman, J. D., & Syvitski, J. P. M. (1992). Geomorphic/tectonic control of sediment discharge to the ocean: The importance of small mountainous rivers. *The Journal of Geology*, *100*(5), 525–544.

Mitchell-Innes, B. A., Richardson, A. J., & Painting, S. J. (1999). Seasonal changes in phytoplankton biomass on the western Agulhas Bank, South Africa. *African Journal of Marine Science*, *21*.

Mobley, C. D. (2001). Radiative transfer in the ocean. *Encyclopedia of Ocean Sciences*, *4*, 2321–2330.

Moran, M. A., Ferrer-González, F. X., Fu, H., Nowinski, B., Olofsson, M., Powers, M. A., Schreier, J. E., Schroer, W. F., Smith, C. B., & Uchimiya, M. (2022). The Ocean's labile DOC supply chain. *Limnology and Oceanography*, *67*(5), 1007–1021.

Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color 1. *Limnology and Oceanography*, *22*(4), 709–722.

Morel, A. (1974). Optical properties of pure water and pure seawater. *Optical Aspects of Oceanography*.

Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., & Tieche, F. (2007). Optical properties of the "clearest" natural waters. *Limnology and oceanography*, *52*(1), 217–229.

Mulholland, P. J. (2003). Large-scale patterns in dissolved organic carbon concentration, flux, and sources. In *Aquatic ecosystems* (pp. 139–159). Elsevier.

Murray, T. S., Cowley, P. D., Bennett, R. H., & Childs, A.-R. (2018). Fish on the move: Connectivity of an estuary-dependent fishery species evaluated using a large-scale acoustic telemetry array. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(11), 2038–2052.

Neil, D. T., Orpin, A. R., Ridd, P. V., & Yu, B. (2002). Sediment yield and impacts from river catchments to the Great Barrier Reef lagoon: a review. *Marine and Freshwater Research*, *53*(4), 733–752.

Newell, R. I. E., Marshall, N., Sasekumar, A., & Chong, V. C. (1995). Relative importance of benthic microalgae, phytoplankton, and mangroves as sources of nutrition for penaeid prawns and other coastal invertebrates from Malaysia. *Marine Biology*, *123*, 595–606.

Ogawa, H., Amagai, Y., Koike, I., Kaiser, K., & Benner, R. (2001). Production of refractory dissolved organic matter by bacteria. *Science*, *292*(5518), 917–920.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S. A., De Marvao, A., Dawes, T., O'Regan, D. P., et al. (2017). Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, *37*(2), 384–395.

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., & McClain, C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans*, *103*(C11), 24937–24953.

Osadchiev, A., & Korshenko, E. (2017). Small river plumes off the northeastern coast of the Black Sea under average climatic and flooding discharge conditions. *Ocean Science*, *13*(3), 465–482.

Osburn, C. L., & Morris, D. P. (2003). Photochemistry of chromophoric dissolved organic matter in natural waters. *UV Effects in Aquatic Organisms and Ecosystems*, *1*, 185–217.

Ostendorf, B. (2011). Overview: Spatial information and indicators for sustainable management of natural resources. *Ecological Indicators*, *11*(1), 97–102.

Pace, G., Meloni, D., & Di Sarra, A. (2005). Forest fire aerosol over the Mediterranean basin during summer 2003. *Journal of Geophysical Research: Atmospheres*, *110*(D21).

Paerl, H. W., Hall, N. S., Hounshell, A. G., Rossignol, K. L., Barnard, M. A., Luettich, R. A., Rudolph, J. C., Osburn, C. L., Bales, J., & Harding, L. W. (2020). Recent increases of rainfall and flooding from tropical cyclones (TCs) in North Carolina (USA): implications for organic matter and nutrient cycling in coastal watersheds. *Biogeochemistry*, *150*, 197–216.

Palmer, S. C. J., Kutser, T., & Hunter, P. D. (2015). Remote sensing of inland waters: Challenges, progress and future directions.

Paola, C., Twilley, R. R., Edmonds, D. A., Kim, W., Mohrig, D., Parker, G., Viparelli, E., & Voller, V. R. (2011). Natural processes in delta restoration: Application to the Mississippi Delta. *Annual Review of Marine Science*, *3*, 67–91.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,

M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Penven, P., Lutjeharms, J. R. E., Marchesiello, P., Roy, C., & Weeks, S. J. (2001). Generation of cyclonic eddies by the agulhas current in the lee of the Agulhas Bank. *Geophysical Research Letters*, *28*(6), 1055–1058.

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

Pitarch, J., Falcini, F., Nardin, W., Brando, V., Di Cicco, A., & Marullo, S. (2019). Linking flow-stream variability to grain size distribution of suspended sediment from a satellite-based analysis of the tiber river plume (tyrrhenian sea). *Scientific reports*, *9*(1), 19729.

Poulton, A. J., Mazwane, S. L., Godfrey, B., Carvalho, F., Mawji, E., Wihsgott, J. U., & Noyon, M. (2022). Primary production dynamics on the Agulhas Bank in autumn. *Deep Sea Research Part II: Topical Studies in Oceanography*, *203*, 105153.

Rabalais, N. N., Turner, R. E., & Wiseman Jr, W. J. (2002). Gulf of Mexico hypoxia, aka "the dead zone". *Annual Review of Ecology and Systematics*, *33*(1), 235–263.

Rabotyagov, S. S., Kling, C. L., Gassman, P. W., Rabalais, N. N., & Turner, R. E. (2014). The economics of dead zones: Causes, impacts, policy challenges, and a model of the Gulf of Mexico hypoxic zone. *Review of Environmental Economics and Policy*.

Ritchie, J. C., Cooper, C. M., & Schiebe, F. R. (1990). The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon Lake, Mississippi. *Remote Sensing of Environment*, *33*(2), 137–148.

Rudorff, N., Rudorff, C. M., Kampel, M., & Ortiz, G. (2018). Remote sensing monitoring of the impact of a major mining wastewater disaster on the turbidity of the Doce River plume off the eastern Brazilian coast. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 349–361.

Sambrotto, R. N., Niebauer, H. J., Goering, J. J., & Iverson, R. L. (1986). Relationships among vertical mixing, nitrate uptake, and phytoplankton growth during the spring bloom in the southeast Bering Sea middle shelf. *Continental Shelf Research*, *5*(1-2), 161–198.

Sano, I., Mukai, S., Nakata, M., Holben, B. N., & Kikuchi, N. (2011). Optical properties of biomass burning aerosols during Russian forest fire events in 2010. *Remote Sensing of Clouds and the Atmosphere XVI*, *8177*, 118–125.

Schalles, J. F. (2006). Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal. In *Remote sensing of aquatic coastal ecosystem processes* (pp. 27–79). Springer.

Schmidt, N., Lipp, E. K., Rose, J. B., & Luther, M. E. (2001). ENSO influences on seasonal rainfall and river discharge in Florida. *Journal of Climate*, *14*(4), 615–628.

Schulz, R. (2001). Rainfall-induced sediment and pesticide input from orchards into the Lourens River, Western Cape, South Africa: importance of a single event. *Water Research*, *35*(8), 1869–1876.

Schumann, E. H., Perrins, L. A., & Hunter, I. T. (1982). Upwelling along the south coast of the Cape Province, South Africa. *South African Journal of Science*, *78*(6), 238–242.

Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., & Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: An ocean color case study. *Optics Express*, *26*(6), 7404–7422.

Shannon, L. V., Hutchings, L., Bailey, G. W., & Shelton, P. A. (1984). Spatial and temporal distribution of chlorophyll in southern African waters as deduced from ship and satellite measurements and their implications for pelagic fisheries. *South African Journal of Marine Science*, *2*(1), 109–130.

Shao, T., & Wang, T. (2020). Effects of land use on the characteristics and composition of fluvial chromophoric dissolved organic matter(CDOM) in the Yiluo River watershed, China. *Ecological Indicators*, *114*, 106332.

Shi, J. Z. (2010). Tidal resuspension and transport processes of fine sediment within the river plume in the partially-mixed Changjiang River estuary, China: a personal perspective. *Geomorphology*, *121*(3-4), 133–151.

Shi, W., & Wang, M. (2009). Satellite observations of flood-driven Mississippi River plume in the spring of 2008. *Geophysical Research Letters*, *36*(7).

Shumway, S. E. (1990). A review of the effects of algal blooms on shellfish and aquaculture. *Journal of the World Aquaculture Society*, *21*(2), 65–104.

Smith, M. E., Bernard, S., & O'Donoghue, S. (2013). The assessment of optimal MERIS ocean colour products in the shelf waters of the KwaZulu-Natal Bight, South Africa. *Remote Sensing of Environment*, *137*, 124–138.

Soriano-González, J., Urrego, E. P., Sòria-Perpinyà, X., Angelats, E., Alcaraz, C., Delegido, J., Ruız-Verdú, A., Tenjo, C., Vicente, E., & Moreno, J. (2022). Towards the combination of C2RCC processors for improving water quality retrieval in inland and coastal areas. *Remote Sensing*, *14*(5), 1124.

South african national biodiversity institute (sanbi), (2018). beta marine ecosystem map [vector] 2018. available from the biodiversity gis website, downloaded on 11 august 2023. (n.d.).

Stagl, J. C., & Hattermann, F. F. (2016). Impacts of climate change on riverine ecosystems: Alterations of ecologically relevant flow dynamics in the Danube River and its major tributaries. *Water*, *8*(12), 566.

Steinmetz, F., Deschamps, P.-Y., & Ramon, D. (2011). Atmospheric correction in presence of sun glint: application to MERIS. *Optics Express*, *19*(10), 9783–9800.

Steinmetz, F., & Ramon, D. (2018). Sentinel-2 MSI and Sentinel-3 OLCI consistent ocean colour products using POLYMER. *Remote sensing of the open and coastal ocean and inland waters*, *10778*, 46–55.

Sterckx, S., Knaeps, S., Kratzer, S., & Ruddick, K. (2015). SIMilarity Environment Correction (SIMEC) applied to MERIS data over inland and coastal waters. *Remote Sensing of Environment*, *157*, 96–110.

Sterckx, S., Knaeps, E., & Ruddick, K. (2011). Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: The use of the near infrared similarity spectrum. *International Journal of Remote Sensing*, *32*(21), 6479–6505.

Stumpf, R. P., & Pennock, J. R. (1989). Calibration of a general optical equation for remote sensing of suspended sediments in a moderately turbid estuary. *Journal of Geophysical Research: Oceans*, *94*(C10), 14363–14371.

Summerhayes, C. P., Sestini, G., Misdorp, R., & Marks, N. (1978). Nile Delta: Nature and evolution of continental shelf sediments. *Marine Geology*, *27*(1-2), 43–65.

Swart, V. P., & Largier, J. L. (1987). Thermal structure of Agulhas Bank water. *South African Journal of Marine Science*, *5*(1), 243–252.

Thomas, J., Joseph, S., & Thrivikramji, K. P. (2018). Estimation of soil erosion in a rain shadow river basin in the southern Western Ghats, India using RUSLE and transport limited sediment delivery function. *International Soil and Water Conservation Research*, *6*(2), 111–122.

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lake water quality parameters with sentinel-2 msi imagery. *Remote Sensing*, *8*(8), 640.

Urquhart, E. A., & Schaeffer, B. A. (2020). Envisat MERIS and Sentinel-3 OLCI satellite lake biophysical water quality flag dataset for the contiguous united states. *Data in Brief*, *28*, 104826.

Uudeberg, K., Ansko, I., Põru, G., Ansper, A., & Reinart, A. (2019). Using optical water types to monitor changes in optically complex inland and coastal waters. *Remote Sensing*, *11*(19), 2297.

Van der Lingen, C. D., Hutchings, L., Lamont, T., & Pitcher, G. C. (2016). Climate change, dinoflagellate blooms and sardine in the southern Benguela Current Large Marine Ecosystem. *Environmental Development*, *17*, 230–243.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. *The carbon cycle and atmospheric CO2: Natural variations Archean to present*, *32*, 99–110.

Vundavilli, H., Mullarney, J. C., MacDonald, I. T., & Bryan, K. R. (2021). The interaction of buoyant coastal river plumes with mangrove vegetation and consequences for sediment deposition and erosion in a tidal environment. *Continental Shelf Research*, *222*, 104417.

Walling, D. E. (2006). Human impact on land–ocean sediment transfer by the world's rivers. *Geomorphology*, *79*(3-4), 192–216.

Walling, D. E., & Moorehead, P. W. (1989). The particle size characteristics of fluvial suspended sediment: An overview. *Sediment/Water Interactions: Proceedings of the Fourth International Symposium*, 125–149.

Wang, B. (2006). Cultural eutrophication in the Changjiang (Yangtze River) plume: History and perspective. *Estuarine, Coastal and Shelf Science*, *69*(3-4), 471–477.

Wang, J., Lee, Z., Wei, J., & Du, K. (2020). Atmospheric correction in coastal region using same-day observations of different sun-sensor geometries with a revised POLYMER model. *Optics Express*, *28*(18), 26953–26976.

Wang, M. (2010). Atmospheric correction for remotely-sensed ocean-colour products. *Reports and Monographs of the International Ocean-Colour Coordinating Group (IOCCG)*.

Wang, Y., Wu, H., Lin, J., Zhu, J., Zhang, W., & Li, C. (2019). Phytoplankton blooms off a high turbidity estuary: A case study in the Changjiang River Estuary. *Journal of Geophysical Research: Oceans*, *124*(11), 8036–8059.

Waquet, F., Peers, F., Ducos, F., Goloub, P., Platnick, S., Riedi, J., Tanré, D., & Thieuleux, F. (2013). Global analysis of aerosol properties above clouds. *Geophysical Research Letters*, *40*(21), 5809–5814.

Weithoff, G., & Beisner, B. E. (2019). Measures and approaches in trait-based phytoplankton community ecology–from freshwater to marine ecosystems. *Frontiers in Marine Science*, *6*, 40.

Werdell, P. J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, *98*(1), 122–140.

Whitfield, A. K. (1989). The benthic invertebrate community of a southern cape estuary: Structure and possible food sources. *Transactions of the Royal Society of South Africa*, *47*(2), 159–179.

Wihsgott, J. U., Sharples, J., Hopkins, J. E., Woodward, E. M. S., Hull, T., Greenwood, N., & Sivyer, D. B. (2019). Observations of vertical mixing in autumn and its effect on the autumn phytoplankton bloom. *Progress in Oceanography*, *177*, 102059.

Wolanski, E. (1995). Transport of sediment in mangrove swamps. *Hydrobiologia*, *295*(1-3), 31–42.

Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, *6*(5), e1373.

Yang, S. L. (1998). The role of ScirpusMarsh in Attenuation of Hydrodynamics and Retention of Fine Sediment in the Yangtze Estuary. *Estuarine, Coastal and Shelf Science*, *47*(2), 227–233.

Yin, K., Qian, P.-Y., Chen, J. C., Hsieh, D. P., & Harrison, P. J. (2000). Dynamics of nutrients and phytoplankton biomass in the Pearl River estuary and adjacent waters of Hong Kong during summer: preliminary evidence for phosphorus and silicon limitation. *Marine Ecology Progress Series*, *194*, 295–305.

Zhang, Z., Zhu, M., Wang, Z., & Wang, J. (2006). Monitoring and managing pollution load in Bohai Sea, PR China. *Ocean & Coastal Management*, *49*(9-10), 706–716.

Zhou, M.-j., Shen, Z.-l., & Yu, R.-c. (2008). Responses of a coastal phytoplankton community to increased nutrient input from the Changjiang (Yangtze) River. *Continental Shelf Research*, *28*(12), 1483–1489.

Zhou, Y., Shen, Q., Peng, H., Yao, Y., Wang, R., Ma, W., Shi, J., Ren, Y., Zhao, J., Xu, W., et al. (2023). Suitability of Sentinel-3 chlorophyll a products based on optical water types. *International Journal of Remote Sensing*, *44*(4), 1369–1389.

Zhu, W., Yu, Q., Tian, Y. Q., Becker, B. L., Zheng, T., & Carrick, H. J. (2014). An assessment of remote sensing algorithms for coloured dissolved organic matter in complex freshwater environments. *Remote Sensing of Environment*, *140*, 766–778.

Zhu, W., Yu, Q., Tian, Y. Q., Chen, R. F., & Gardner, G. B. (2011). Estimation of chromophoric dissolved organic matter in the Mississippi and Atchafalaya river plume regions using above-surface hyperspectral remote sensing. *Journal of Geophysical Research: Oceans*, *116*(C2).

Zwolsman, J. J. G., Berger, G. W., & Van Eck, G. T. M. (1993). Sediment accumulation rates, historical input, postdepositional mobility and retention of major elements and trace metals in salt marsh sediments of the scheldt estuary, SW Netherlands. *Marine Chemistry*, *44*(1), 73–94.

Appendix A

Appendix

A.1 *In situ* [Chl-a] and TSM concentration maps

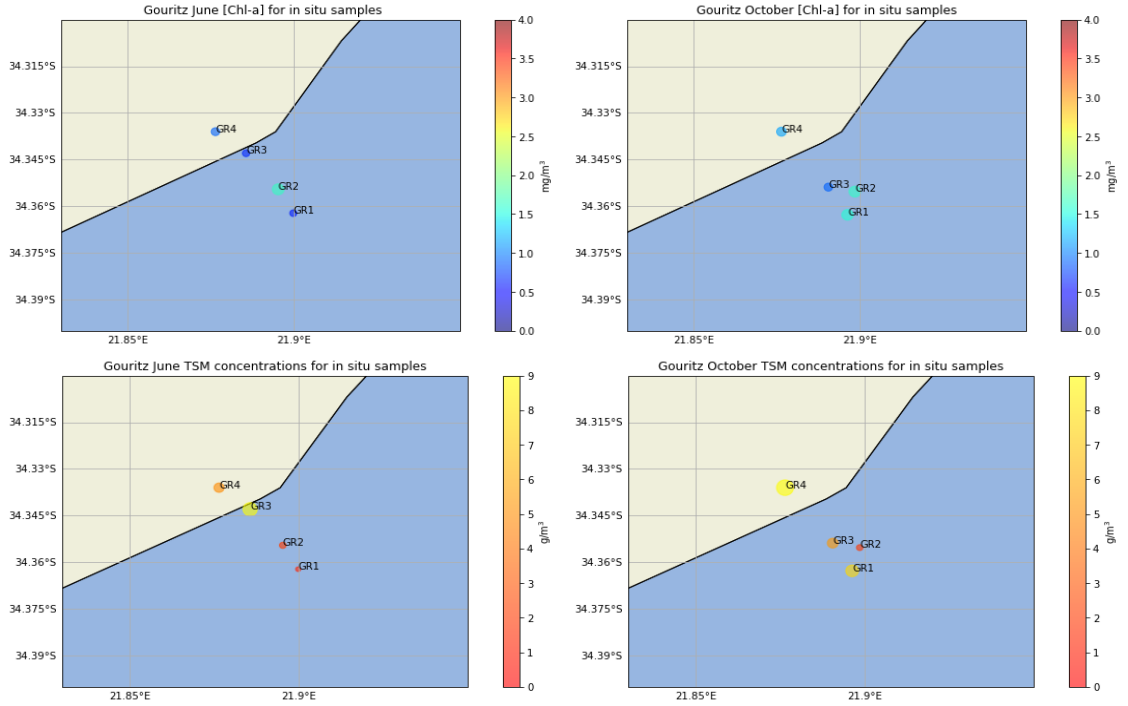

Figure A.1: Maps of the *in situ* Chl-a (top; mg/m^3) and TSM (bottom; g/m^3) concentrations measured during the June (left) and October (right) 2022 sampling trips near the Gouritz River. The diameter of the circles and their respective colours represent the intensity of the concentration of the samples collected and analysed.

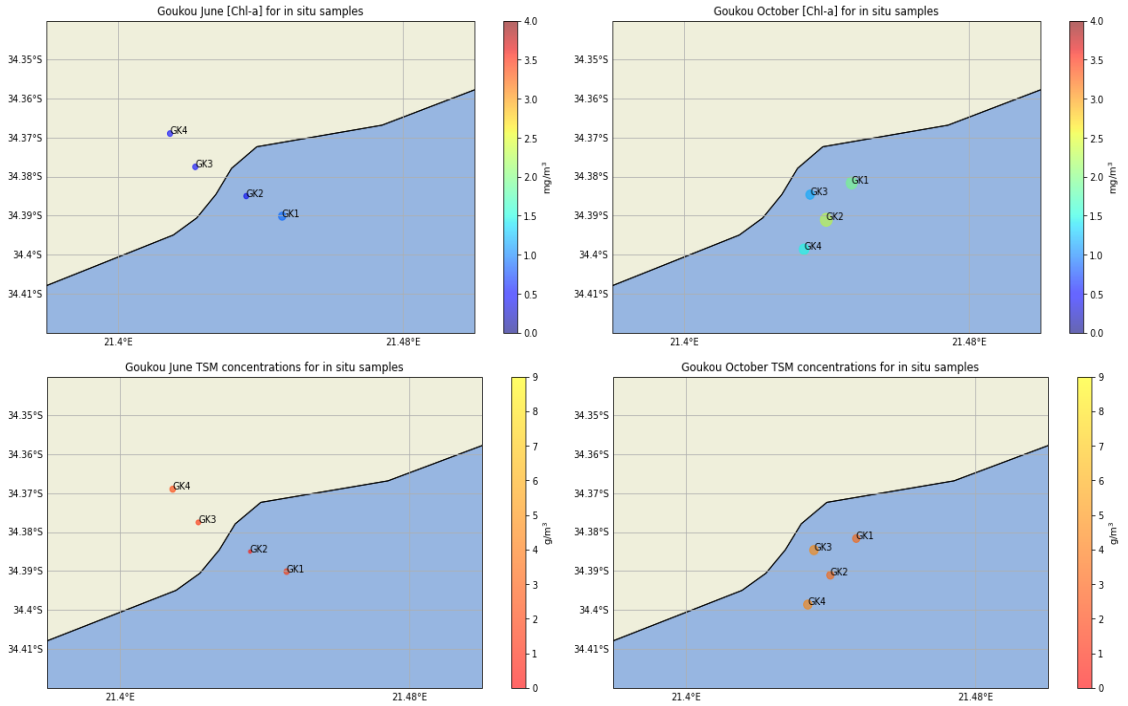

Figure A.2: Maps of the *in situ* Chl-a (top; mg/m^3) and TSM (bottom; g/m^3) concentrations measured during the June (left) and October (right) 2022 sampling trips near the Goukou River. The diameter of the circles and their respective colours represent the intensity of the concentration of the samples collected and analysed.

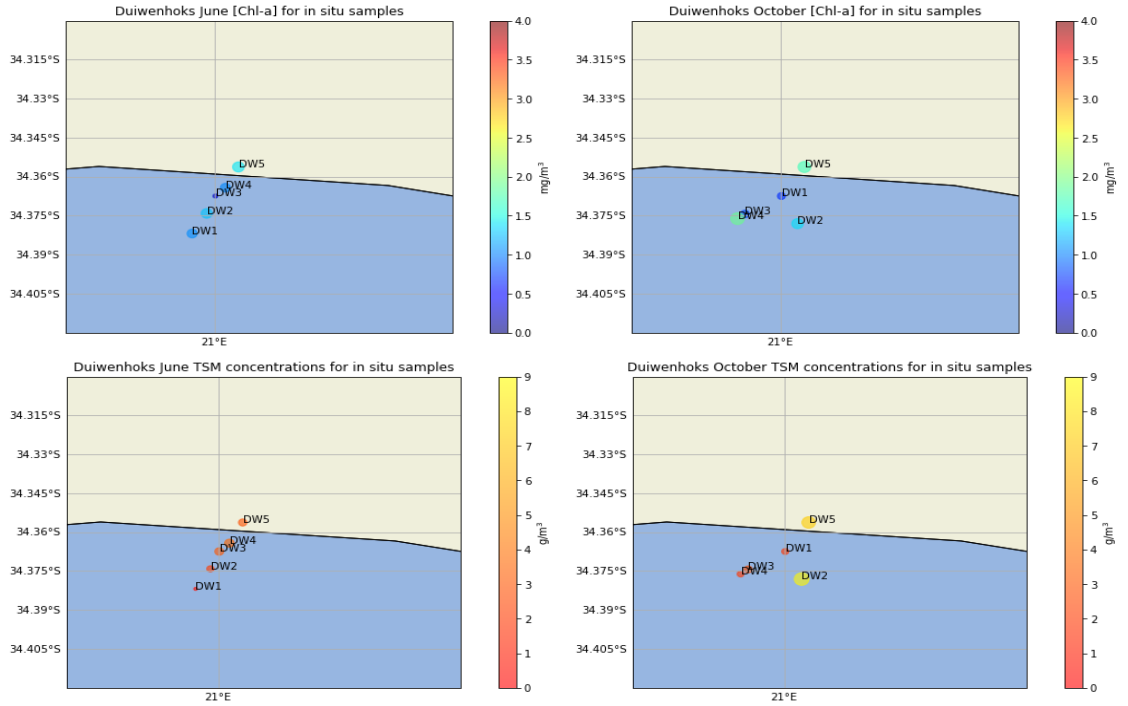

Figure A.3: Maps of the *in situ* Chl-a (top; mg/m^3) and TSM (bottom; g/m^3) concentrations measured during the June (left) and October (right) 2022 sampling trips near the Duiwenhoks River. The diameter of the circles and their respective colours represent the intensity of the concentration of the samples collected and analysed.

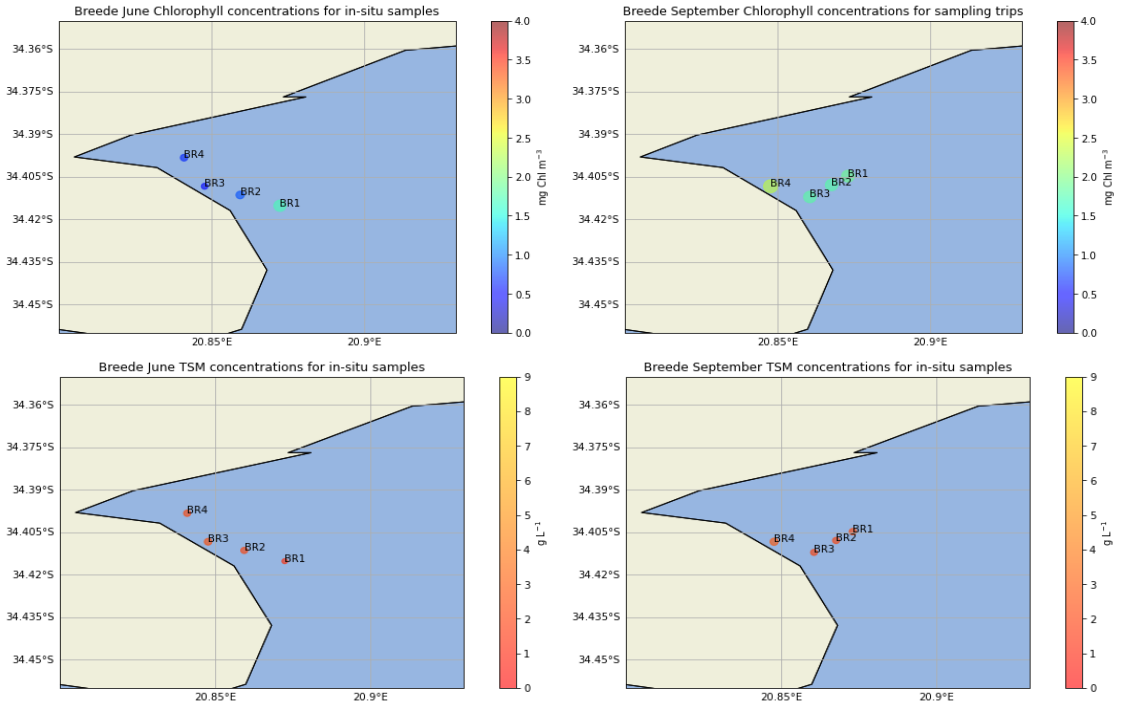

Figure A.4: Maps of the *in situ* Chl-a (top; mg/m^3) and TSM (bottom; g/m^3) concentrations measured during the June (left) and October (right) 2022 sampling trips near the Breede River. The diameter of the circles and their respective colours represent the intensity of the concentration of the samples collected and analysed.

A.2 R_{rs} retrievals

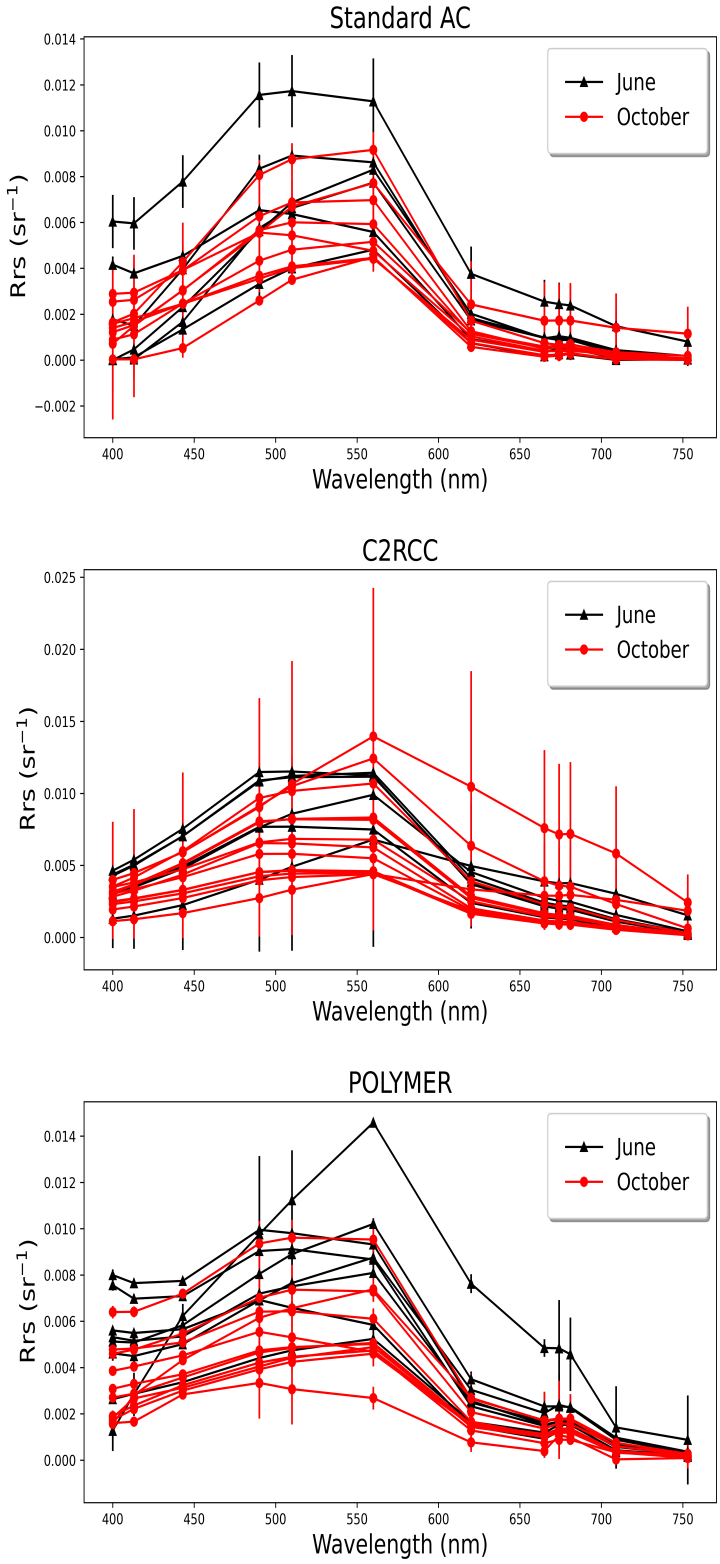

Figure A.5: The R_{rs} and standard deviations of retrievals for the standard L2 (top), C2RCC (middle) and POLYMER (bottom) ACs for the June (black) and October (red) match-up station.

A.3 Scatter plots for model match-ups

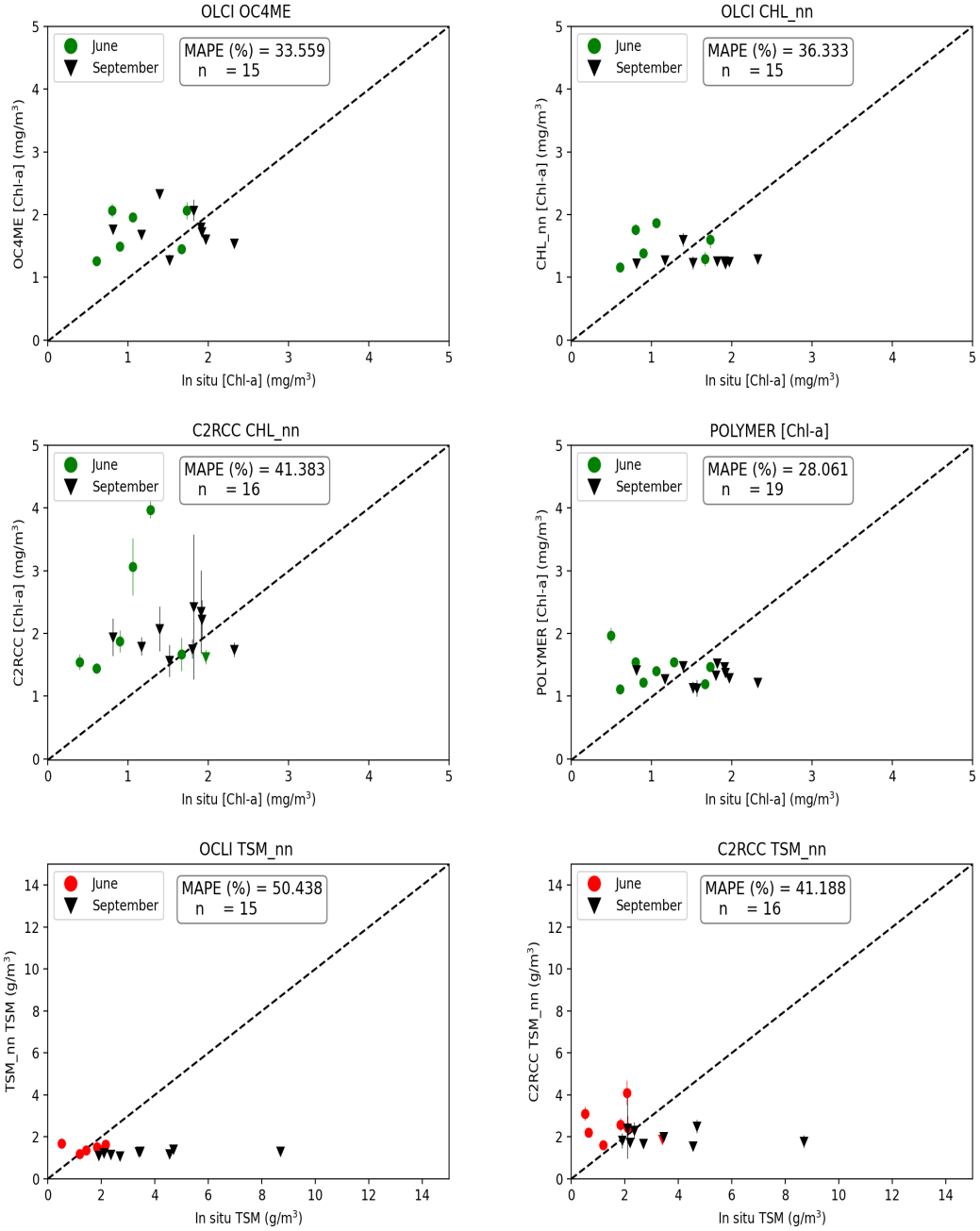

Figure A.6: The measured *in situ* [Chl-a] (green/black) and TSM (red/black) versus the available ocean colour satellite products evaluated in this study.

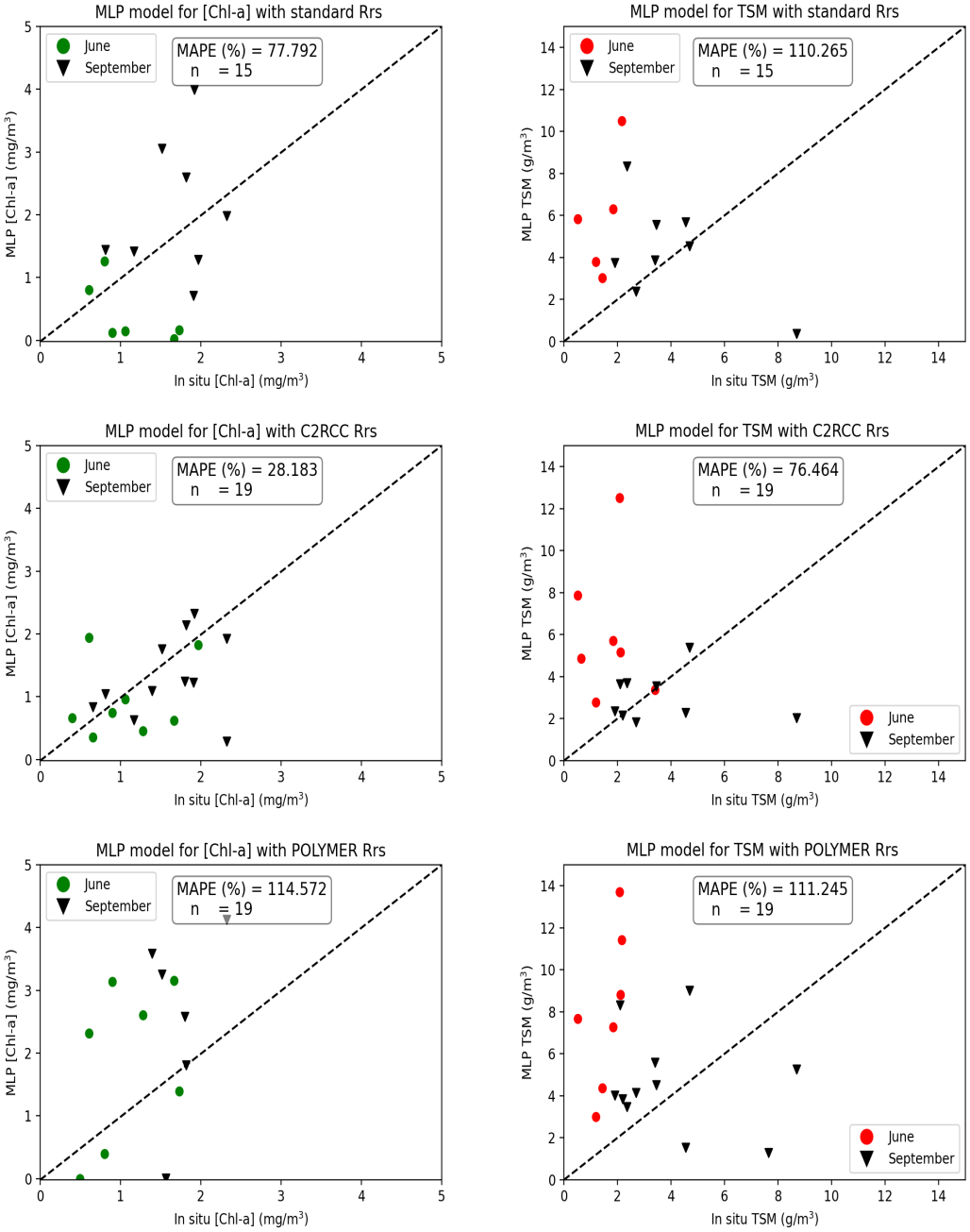

Figure A.7: The measured *in situ* [Chl-a] (green/black) and TSM (red/black) versus the full MLP model predictions.

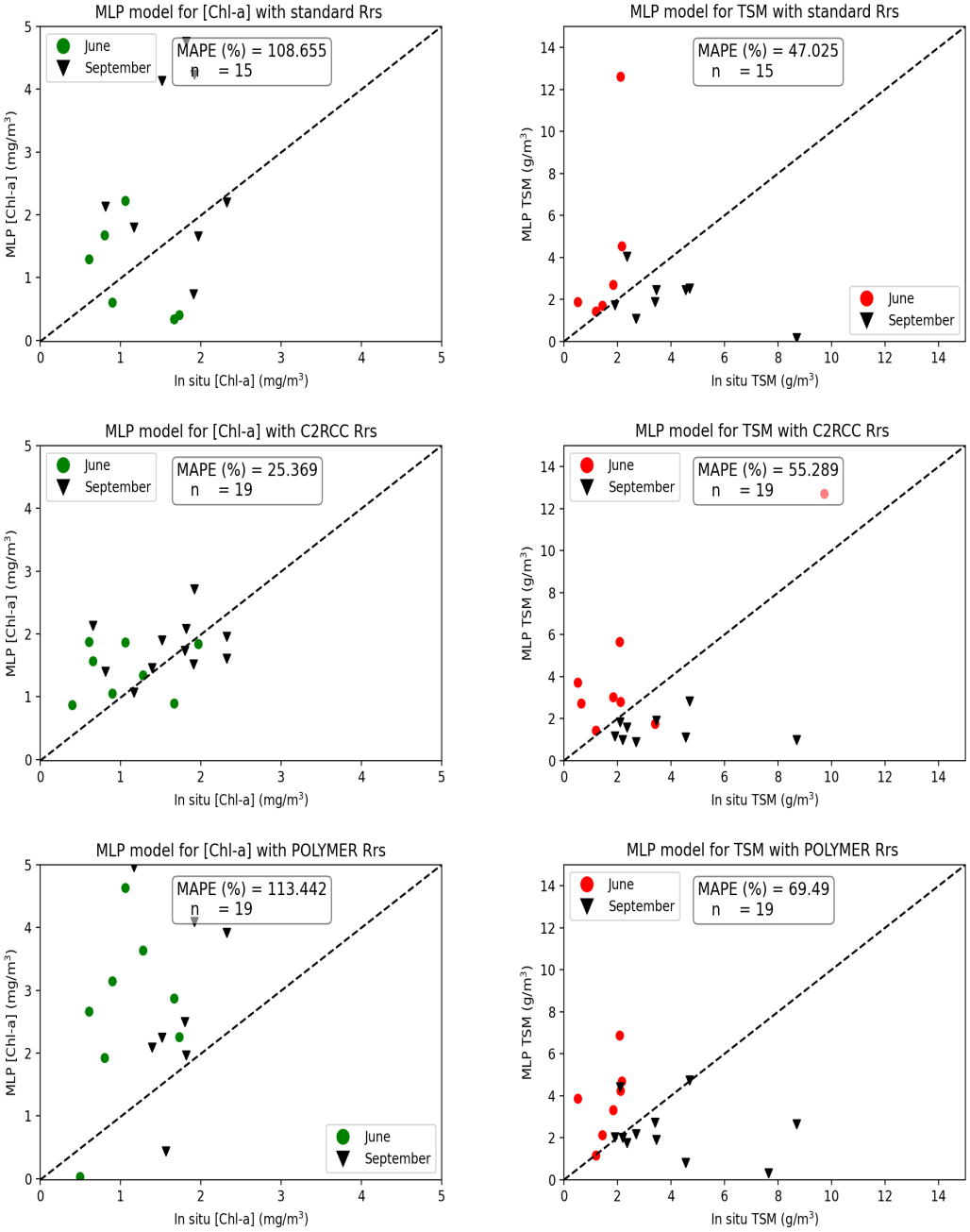

Figure A.8: The measured *in situ* [Chl-a] (green/black) and TSM (red/black) versus the constrained MLP model predictions.

A.4 Other TSM MLP model regional maps

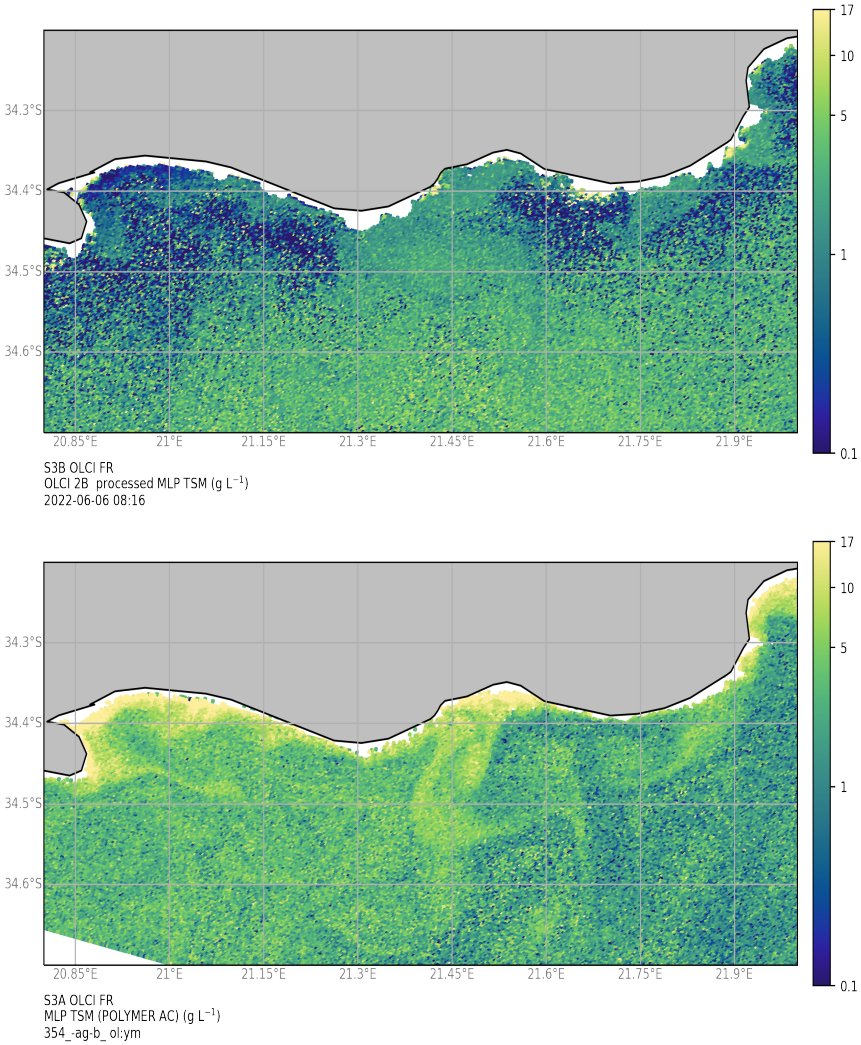

Figure A.9: Maps of the constrained MLP TSM model outputs when applied to standard (top) and POLYMER (bottom) ACs respectively, on the 6^{th} of June 2022.