Department of Physics

University of Strathclyde

Neural networks applied to ocean colour remote sensing for environmental monitoring

Madjid Hadjal

2022

A thesis submitted in accordance with the requirements for the degree

Doctor of Philosophy

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

Abstract

Autotrophic algal organisms that perform photosynthesis are the basis of the marine food web and absorb up to 30% of the anthropogenic emitted CO_2 (Gruber et al., 2007). The algal concentration can be measured by retrieving their chlorophyll a concentration from remote sensing. The light signal measured from satellites needs to be corrected for various atmospheric and water surface effects. After atmospheric correction, the water leaving signal can be isolated and different algorithms exist to retrieve chlorophyll a. In open waters, blue-green ratios perform well (O'Reilly et al., 1998). In coastal waters, other water constituents (dissolved matter and sediments) make both the atmospheric correction process and the chlorophyll a estimates harder as they alter the light signal. Current chlorophyll algorithms therefore tend to perform poorly for turbid coastal waters. To develop a better algorithm, a northwest European shelf seas matchup dataset is built by collecting *in situ* chlorophyll a and MODIS Aqua sensor data. Different neural network algorithms are developed to make the best possible estimates of chlorophyll a, using either the bottom of atmosphere reflectance Rrs, commonly used by other algorithms, or top of atmosphere reflectances, when no atmospheric correction was applied to the light signal.

It is found that the uncorrected top of atmosphere signal produces better and more reliable estimates over the entire dataset which contains mainly nearshore samples. Small neural network architectures containing 3 hidden layers of 15 neurons show good performances. The randomness involved in what a single neural network produces is tackled by using an ensemble approach of ten networks. The use of the whole light spectrum from 412 nm to 2130 nm produced the best estimates and should lead future dataset creations that currently do not include all available spectral information. The neural network algorithm developed here for chlorophyll a works well for turbid coastal waters where other algorithms either fail or mask out the data after applying quality control flags. The greatest impact is likely to be for nearshore waters where turbidity tends to be greatest. The approach developed here for NW European shelf seas has potential to be extended to global scale operation if a suitable training data set can be collected in future. The process is repeated for a modelled dataset and shows almost perfect estimates of the three different water constituents that alter the light signal. Temperature can also be estimated with better performance using the same approach which allows creation of a single temperature algorithm that works during both day and night, for any month and position on Earth. The benefit of using an enhanced chlorophyll a algorithm is later evaluated from the eutrophication assessment point of view. Remote sensing techniques produce several orders of magnitude more data than the current *in situ* approach to assess the eutrophication status and should help policymakers in producing coherent and improved assessment of the environment.

Acknowledgements

The first person I thank is my main supervisor David McKee, who supported me during the whole thesis and who is an incredible enthusiastic human being. Thanks for the coffees, the laughs and the advice provided throughout the past 3 years and half. David is probably the best definition of a “super chouette” supervisor!

I would like to say thank you to the other supervisors who shared their thoughts with me and contributed to the work in general, Alejandro Gallego for the more practical information about how useful my work may be, Jinchang Ren for the machine learning knowledge, and Encarni Medina for her expertise in general and very useful help. I also thank Jacqueline Tweedle for her support during the first year and half of the thesis.

I am thankful to the girls from the office, Ina who showed me around at the beginning, Marilisa who helped me improve my Italian, Stacey and Cait for their good company and funny moments shared together. I also thanks the students from Encarni’s group that I had the chance to meet a few times, especially James that helped me getting on with neural networks. I thank Jinchang’s student that helped me at the beginning of the thesis too.

I thank the people from John Anderson building in general, John and Gordon for the administrative help; the IT team, especially Mr Briggs for the ultra-reactivity and efficiency; Brian and Oliver for the friendly discussions; Naresh for never letting me win a badminton game, and the students from the fluorescence team for the wee strolls around.

Thank you to Marine Scotland Science for allowing me to attend the December 2019 cruise and the Stonehaven coastal station. Thanks to Berit for her kindness, Matt for the good company and Kerry for showing me the chemistry laboratory.

I thank my friend Nils for laughing at my stupid jokes 9 times out of ten and hosting me in his lovely flat when I needed it, or cooking some surprisingly delicious German food!

I am thankful to all my friends in France for the help and support during the thesis and the good times spent together online or on a bike climbing the mountains.

Table of Contents

Publications

Hadjal, Madjid; Medina-Lopez, Encarni; Ren, Jinchang; Gallego, Alejandro; McKee, David. An Artificial Neural Network Algorithm to Retrieve Chlorophyll a for Northwest European Shelf Seas from Top of Atmosphere Ocean Colour Reflectance. *Remote Sensing.* **2022**, **14**, **3353**. https://www.mdpi.com/2072-4292/14/14/3353.

Hadjal, M., Paterson, R. and McKee, D., **2023**. Neural networks to retrieve in water constituents applied to radiative transfer models simulating coastal water conditions. *Frontiers in Remote* Sensing Multi-and Hyper-Spectral Imaging, 4.https://www.frontiersin.org/articles/10.3389/frsen.2023.973944/abstract.

List of abbreviations

Abbreviation	Definition
a*Ph	Specific absorption from Phytoplankton
a*MSS	Specific absorption from MSS
a*BD	Specific absorption from Biogenic Detritus
a*CDOM	Specific absorption from CDOM
b*Ph	Specific scattering from Phytoplankton
b*MSS	Specific scattering from MSS
BOA	Bottom of Atmosphere
BT	Brightness Temperature
Chl	Chlorophyll a
CDOM	Coloured Dissolved Organic Matter
CMEMS	Copernicus Marine Environment Monitoring Service
CZCS	Coastal Zone Color Scanner
GSM	Garver-Siegel-Maritorena algorithm
HPLC	High-performance liquid chromatography
ICES	International Council for the Exploration of the Sea
IOP	Inherent Optical Properties
JMP–EUNOSAT	Joint Monitoring Programme of the Eutrophication of the North Sea with Satellite data
LN	Log Normal
LF	Log Flat
NASA	National Aeronautics and Space Administration
NN(s)	Neural Network(s)
MAE / MAD	Mean absolute error / difference
MSS	Mineral Suspended Sediments
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MTL	Multi-Task Learning
NOMAD	NASA bio-Optical Marine Algorithm Dataset
OCX	Ocean Colour X (X = number of bands used)
OCRS	Ocean Colour Remote Sensing
OLCI	Ocean and Land Colour Instrument
P90	90^{th} percentile of a data distribution
PACE	Plankton, Aerosol, Cloud, ocean Ecosystem
PML	Plymouth Marine Laboratory
R	Pearson correlation coefficient
RMSE / RMSD	Root mean square error / difference
ReLU	Rectified Linear Unit
Rhos	Rayleigh corrected reflectance

Rhot	Top of atmosphere reflectance
Rrs	Remote Sensing Reflectance
SeaBASS	SeaWiFS Bio-optical Archive and Storage System
SeaDAS	SeaWiFS Data Analysis System
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor
SENZ	Sensor Zenith Angle
SOLZ	Solar Zenith Angle
SST	Sea Surface Temperature
TOA	Top of Atmosphere
WAPD	Weighted absolute percentage difference

1. Introduction

Since the development of industrialisation in the 18th century, the impact of humanity on the environment has increased rapidly through a variety of processes, but very notably by releasing various materials into the air, land and water. The focus of this thesis is on the development of improved Earth observation techniques to support monitoring of anthropogenic impacts on the environment. The domain of interest is the ocean, particularly coastal waters where anthropogenic impacts and sensitivity to their effects is greatest. The process to be addressed is monitoring to identify signs of eutrophication, the process whereby increased nutrient loading in natural waters caused by human activity on land, e.g. agriculture, leads to disruption of ecological balance in downstream waters. This is represented in Rockstrum et al.'s Doughnut model for economic sustainability as the nitrogen and phosphorus cycles, and is one of 9 'planetary boundaries' that Rockstrum et al. (2009) identified as environmental factors with potential to cause global planetary disruption if pushed beyond tipping points. The nitrogen cycle, biodiversity, land use change and climate change (outer boundaries) already reached the harmful level (Rockstrum et al., 2009) that cause "unacceptable" changes to the environment that degrade the inner boundaries (water access, education, etc.).

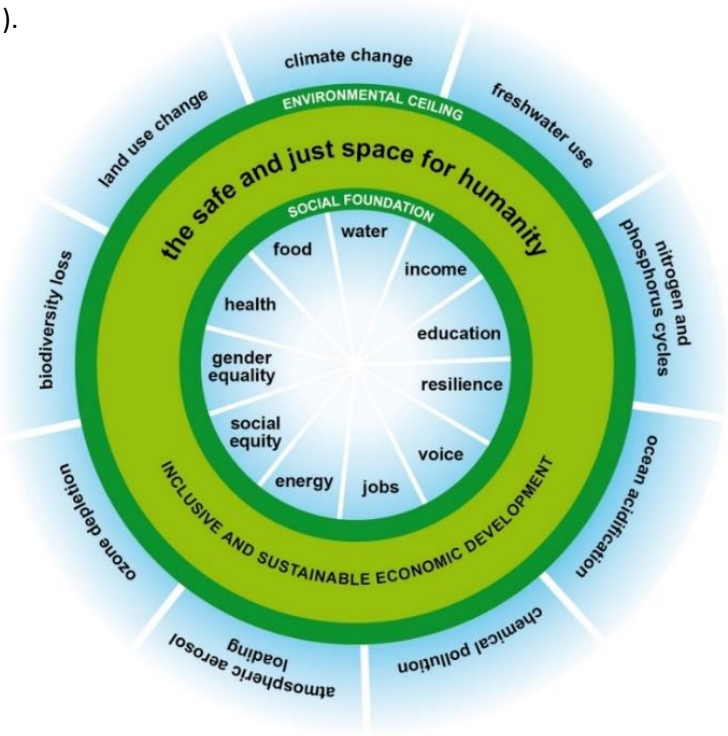

Figure 1.1: The "planetary boundaries" doughnut, adapted from Rockstrum et al., 2009. Source: Wikimedia commons, public domain.

Eutrophication manifests as raised levels of algal biomass when compared against historical background levels and can be identified through raised concentrations of the plant pigment chlorophyll a (Chl). Chlorophyll a allows organisms to absorb light energy and to convert it into chemical energy through the photosynthesis. Chlorophyll a is the main indicator of algal biomass, as all organisms performing the oxygenic photosynthesis contain Chl. It makes the Chl measurement a key indicator that can be later converted into a biomass. The photosynthesis process is also potentially impacted by changes in water temperature which helps to regulate algal growth (Raven and Geider, 1988). Satellite ocean colour remote sensing (OCRS) provides global daily coverage of the Earth's oceans (subject to cloud cover) and is capable of providing information on both Chl concentration and sea surface temperature (SST). There is growing interest in incorporating satellite observations of Chl into eutrophication monitoring schemes. However, uncertain and variable quality in satellite data products is an inhibiting factor in the uptake of these data sets by environmental monitoring agencies. In this thesis the performance of existing satellite Chl products is examined with particular focus on the optically complex waters of northwest European shelf seas. This is a particularly challenging area for interpretation of ocean colour signals due to the confounding influence of non-covarying concentrations of mineral suspended sediments (MSS) and coloured dissolved organic materials (CDOM). The aim of the thesis is to establish if application of machine learning (ML) techniques can result in development of improved algorithms that provide more reliable data with well understood performance attributes. Additionally, the aim is to maximise data availability, particularly in coastal waters where existing Chl algorithms are known to perform poorly (e.g. Siegel et al., 2000; McKee et al. 2007) and where current state of the art processing approaches use quality control flagging to eliminate poorly estimated data. Many marine regions are covered by international legislation: the European Water Framework Directive (WFD); the Oslo-Paris commission (OSPAR) in European western waters or the Helsinki Commission (HELCOM) for the Baltic; which bind national signatories to make efforts to both protect and monitor these regions. Adoption of ocean colour data has potential to massively increase data availability for this purpose, but data quality also has to be assured in order to provide confidence in the outcome of assessments. This thesis has been developed in close partnership with Marine Scotland Science (MSS), the agency of the Scottish Government tasked with monitoring Scottish marine waters and reporting against the relevant legislation. The intention is to shape the analysis to take into account how OCRS data is likely to be used by monitoring agencies such as MSS in addition to the general interest of producing optimal algorithm performance. Understanding the requirements of end users and adapting product development to suit is an essential element of the work presented in this thesis.

1.1. Eutrophication in coastal and shelf seas – the need for wide scale monitoring

The initial goal to observe how the environment evolves is to understand the different processes involved and to give us the capacity to predict what will happen in the future, to be able to react and adapt. In the early 20th century, the Haber process enabled conversion of dinitrogen (N_2, commonly called nitrogen) from the atmosphere into ammonia (NH_3), nutrients used to increase the development of crops (Haber and Rossignol, 1913). This induced an intensification of agriculture in general. Land use has drastically changed during the 20th century (Ramankutty and Foley, 1999) and human population growth and land use are correlated (Meyer and Turner, 1992) thanks to increased quantities of food newly available. Lands are connected to the oceans through rivers and atmospheric dust. With the increased concentration of nutrients on lands, more nutrients are released to the ocean. Organisms living in the ocean perform the photosynthesis to grow. They consume light as a primary source of energy, inorganic carbon and different nutrients during the process. The direct availability of the nutrients affect the phytoplankton communities' distribution (Barcelos e Ramos et al., 2017). Therefore, additional sources of nutrients released in the ocean by anthropogenic activities can directly increases the concentration of phytoplankton when nutrients are a limiting factor. The effect is even more pronounced in a climate changing with increased amount of CO_2 released by anthropogenic activities which is known to increase phytoplankton (Schippers et al., 2004). This process of increasing the algal biomass concentration due to human activities releasing nutrients in the environment is called eutrophication. Following Nixon's definition (Nixon, 1995), eutrophication is "*an increase in the rate of supply of organic matter to an ecosystem*".

Increasing the amount of phytoplankton potentially has both benefits and issues associated. The main benefit is that with more plants, more carbon dioxide is consumed and can be exported down to the ocean floor (the biological pump, Hulse et al., 2017), which directly counters the anthropogenic impact on atmospheric CO_2 levels (Basu and Mackey, 2018). With more food at the bottom of the food web, species relying on it (grazers that eat phytoplankton) have access to more food which potentially increases their development (Hansen et al., 1997) which can induces reduced phytoplankton biomasses (Watras et al., 1985). However, when phytoplankton develop too much and create massive blooms, it can become harmful for the environment. When they die, phytoplankton cells lose their buoyancy and sink through the water column. Bacteria then consume the degraded organic matter produced by phytoplankton, turn it into inorganic matter and make nutrients available again for the remaining phytoplankton cells. This processes is called the remineralisation from the microbial loop (Azam et al., 1983). During this process,

aerobic bacteria consume dioxygen (O_2, commonly called oxygen). Massive blooms induce massive remineralisation and oxygen consumption (Devol, 1978). An area with low amounts of oxygen is hypoxic and is called an oxygen minimum zone (OMZ, Wyrtki, 1962). There is no meaningful threshold to define an OMZ as it depends on the organisms present (Seibel 2011). Every species from small grazers to whales need oxygen to survive, hence a massive algal bloom can become harmful to the rest of the food chain and induce death. Moreover, a rise in temperature reduces the capacity of the waters to ventilate owing to stratification and reduces oxygen solubility (Keeling et al., 2010; Helm et al., 2011). Complex interactions exist between the different trophic levels which involve physics, biology and chemistry.

Eutrophication has been linked to the riverine nutrient concentration in the North Sea in the 1970s (Gieskes and Kraay, 1975). Nitrogen and phosphorus enriched rivers in the Netherlands were linked to an increase in the biomass of phytoplankton during the growing season. Decisions were made to tackle the problem in the 1980s by trying to reduce the nutrients released to the ocean by 2-3% each year (De Jong, 2006). Reducing the amount of nutrients released by rivers is called de-eutrophication (Terseleer et al., 2010). The OSLO-PARIS (OSPAR) commission is responsible for assessing the state of eutrophication of different areas in Europe and comes from an agreement between 35 European countries. Eutrophication is currently based on *in situ* observations of chlorophyll a, which is the green pigment present in chloroplasts contained inside phytoplanktons. This pigment allows phytoplanktons to absorb light and convert radiation energy into chemical energy (Falkowski and Kiefer, 1985). Chlorophyll a is a proxy for primary production and has been measured indirectly since 1890 (Boyce et al., 2012). Chlorophyll a is considered to be the best indicator of phytoplankton carbon biomass (Huot et al. 2007). Secchi disk measurements were linked to the biological activity of the surface of the ocean (Falkowski and Wilson 1992) and compared to chlorophyll a measurements in the open ocean (Lewis et al. 1988), and used to get a proxy of chlorophyll a concentrations. Direct measurement methods (based on photometry and fluorescence) currently in use to measure chlorophyll were developed in the 1960s and are detailed in Chapter 2, Section 2.1. Different areas of the North Sea region are now routinely sampled by different environmental agencies. For example, Marine Scotland Science (MSS) collects chlorophyll in the northern North Sea multiple times a year, and weekly at the Stonehaven and Loch Ewe coastal monitoring stations (Bresnan et al., 2015). The current eutrophication assessment made by OSPAR (called the Common Procedure, more details in Chapter 7) divides European waters into different regions which are hydrologically and ecologically similar (OSPAR report, 2017). All *in situ* measurements made for these specific regions are merged for a growing season, along with oxygen and nutrient concentrations, and eventually a eutrophication index is created to evaluate if a region is facing eutrophication or not over a

six years period. An example of a beach in Brittany, France that faces eutrophication of macroalgae, *Ulva armoricana*. These “green tides” frequently occur in Brittany since the 1970s, supposedly from increased nutrient release from the rivers (Charlier et al., 2007; 2008).

Figure 1.2: *Ulva Armoricana* algae during a green tide in northern Finistère (Dossen beach, Sieck Island, Santec, France). Source: Wikimedia commons, public domain.

A region can be considered in a eutrophic state when an increased phytoplankton biomass is observed during the growing season over several years compared to a reference level. In the northern hemisphere and for European waters, the growing season is defined as March to October (or September, redefinition currently discussed). The evolution of the phytoplankton biomass does not only depend on an increased supply of nutrients from the rivers or atmosphere, but complex relationships exist between multiple factors depending on the area observed (Sathyendranath et al., 2009; Berge et al., 2010; Chen and Liu, 2010; Wirtz et al., 2011). For the North Sea region which is where most data available in this thesis come from, nutrients and temperature are the two major variables responsible of the eutrophication status (Desmit et al., 2019). The North Sea surface temperature rose by 1.6°C between 1974 and 2014 (Desmit et al., 2019). The development of different phytoplankton species is temperature dependent (Edwards et al., 2016). Increasing the temperature induces a decrease in phytoplankton species size (Sommer et al., 2017), a loss in biodiversity (Rasconi et al., 2017) and can lead to a switch in species dominance (Graham and Vinebrooke, 2009). Growth rate or nutrient uptake can also be expected to increase with warming of the ocean (Litchman et al., 2007). Understanding the phenomenon fully requires consistent observations of chlorophyll and temperature as a minimum. But addition of further information such as macro and micronutrients, salinity, currents, wind, partial pressure of carbon dioxide, and

taxonomic composition of the algal population would help understanding how the system is evolving, due to the complexity of the interactions.

Since the development of passive remote sensing techniques in the later 1970s (Section 1.2), temperature and chlorophyll a can be observed from space with a relatively good precision in open waters when compared with *in situ* retrievals. Satellite ocean colour remote sensing has provided global daily cover (every 1-2 days) since 1997 and most of the larger platforms can measure both chlorophyll and temperature at the same time. Passive visible sensors are primarily limited by the presence of clouds. The current eutrophication assessment relies on *in situ* sampling efforts, which is expensive due to the lack of automation, and is therefore limited by data availability. The performance of ocean colour chlorophyll algorithms is affected by the composition of materials found in seawater and by the quality of radiometric data produced after atmospheric correction (Siegel et al., 2000). Both of these factors are known to be particularly problematic in turbid coastal waters (discussed in more detail later). Uncertain data quality assurance for remotely sensed chlorophyll has been a limiting factor in adoption of this potentially rich data source. They were therefore not used by the different policymakers to report the eutrophication status. However, with recent improvements in our understanding of chlorophyll product quality, there has been renewed consideration of the potential merit of inclusion of remote sensing chlorophyll in eutrophication assessment. A comparison between remote sensing and *in situ* efforts regarding data availability and quality is performed in Chapter 7. Increased data coverage provides the opportunity to redefine the current assessment methods to make benefit of access to more information. These environment agencies require that the quality of remote sensing product is of similar quality than the information retrieved from *in situ* measurements. **Being able to provide these different groups in charge of the assessment with an improved quality chlorophyll remote sensing product is the main objective of this thesis**. An algorithm trying to make improved estimates in the regions concerned by the OSPAR commission is developed in Chapter 3.

1.2. Monitoring chlorophyll from space

Ocean colour remote sensing consists of making an observation of the visible light spectrum leaving the ocean that can be later converted into an environmental variable of the surface of the ocean. Most of the visible light received on Earth is created by the Sun. The Sun is composed of Hydrogen atoms $_1H$ that can fuse with each other under specific conditions of temperature and pressure and emit photons (light) of different wavelength during the process. The radiative energy carried by a photon depends on its wavelength ($E = h \cdot c/\lambda$), with E the energy in Joules; the Planck constant $h = 6.62\ 10^{-34}$ Js; the speed of

light constant in the vacuum c = 299,792,458 ms^{-1} and λ the wavelength expressed in meters m. The photons reach the Earth, interact with the electrons of the different particle composing Earth (gases in the atmosphere, water, lands). A proportion of these photons is reflected back to space after the different interactions. Spectroradiometers onboard satellites at approximately 700 km above the surface filter and count photons that reach them after being reflected by the Earth. Knowing the number of photons gives access to the total energy per unit solid angle received for this specific part of the spectrum, known as the radiance, L, with units W m^{-2} nm^{-1} sr^{-1} (more details in Chapter 2, Section 2.2). The total radiance L_t measured at the satellite sensor's altitude is affected by different factors due to the interactions of photons with the medium prior reaching the sensor (more details provided in Chapter 2, Figure 2.4). It can be simplified as a sum of different radiances from each contributor, mostly the atmosphere, the ocean and the surface of the ocean. The common procedure prior to using a remote sensing signal consists of correcting the total radiance from these different contributors to isolate the oceanic signal. This approach is called the atmospheric correction (AC).

The reason why such spectroradiometer sensors are available nowadays come from previous relationship observed between the light signal and the water constituent concentrations. In 1970, Clarke et al. measured the water leaving light signal in the Sargasso Sea from an aircraft along with the chlorophyll concentration of the same water body at the same time from a ship. They observed that the light signal in the visible was evolving with the concentration. Higher concentrations of chlorophyll were correlated to lower signals in the blue (<490 nm), and higher in the green (>490 nm). These observations marked the birth of ocean colour as a field of research. Ocean colour consists of using the light signal leaving the ocean to indirectly measure the concentration of its constituents, with chlorophyll being the most commonly measured. In 1976, Jerlov defined five different water types based on their light spectrum, depending on where the water was observed, near the coast or offshore and the relationship with the optical depth. This number was reduced to two a year later by Morel and Prieur (1977). The signal leaving the water was simplified as coming from Case 1 or Case 2 waters. In Case 1 waters, the signal is driven by the chlorophyll concentration, similarly to what Clarke et al. observed in 1970, and a simple relationship can be found using these colours based on the blue and green signal. Case 1 waters represent approximately 90% of the total surface of the ocean. In Case 2 waters, the signal is also impacted by other constituents present in the waters, mostly in coastal conditions. The mineral suspended sediments (MSS) encompass minerals that can be brought to the water column by currents and their interactions with the seabed. They interact with light, with absorption increasing approximately exponentially towards the blue and scattering across the visible spectrum. MSS make the water look brown when its concentration is

high. The coloured dissolved organic matter (CDOM) is formed by the breakdown of plant material (terrestrial or marine) and is typically found in higher concentrations in coastal waters and lakes (Kuster et al., 2005). CDOM also increases absorption almost exponentially into the blue. (Bricaud et al., 1981). Their original name is “gelbstoff”, which means yellow substance in German, due to their tendency to absorb the blue and some green parts of the light signal and to leave the water with a yellow colour. Figure 1.3 below shows an example of a red green blue (RGB) image created with MODIS sensor. We can see different shades of colour in the ocean, each associated with different water constituent concentrations. High concentrations of sediments are expected in the Irish Sea, near the Thames plume or the Bristol Channel and would be characterised as Case 2 waters.

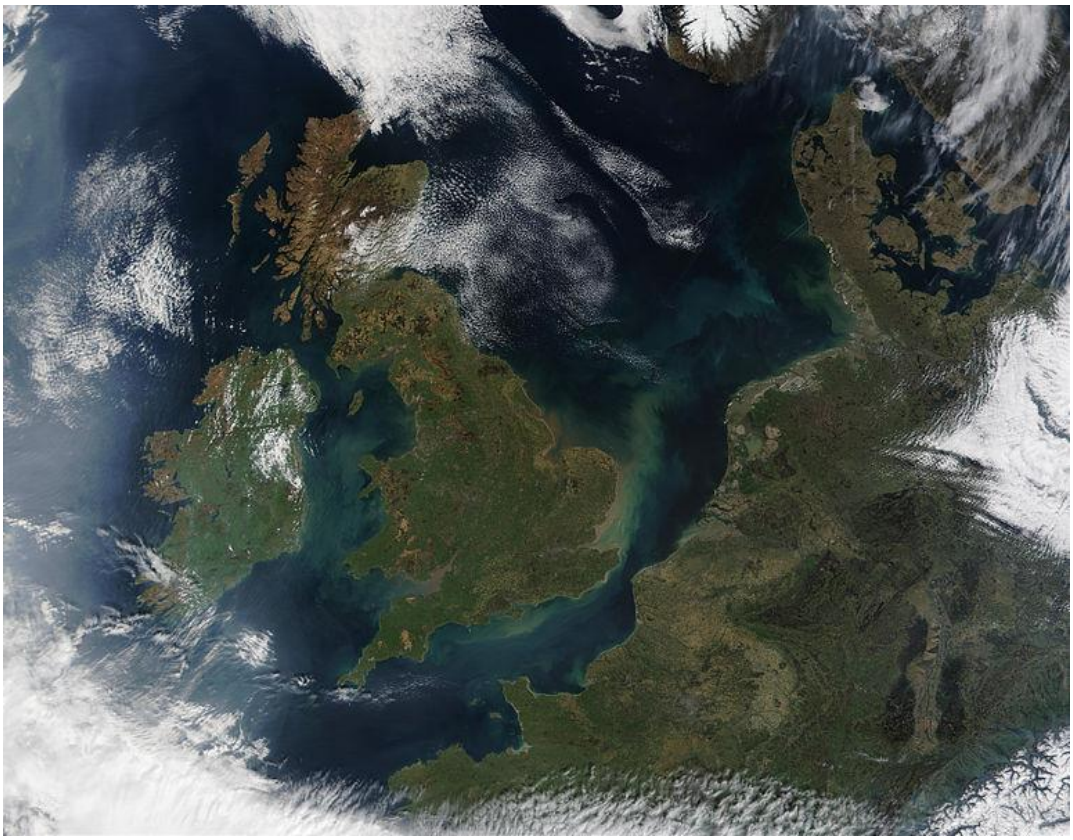

Figure 1.3: Quasi true colour image of Northwestern Europe, captured by MODIS, 6 April 2002 at 11:10:00 (produced by Jacques Descloitres, MODIS land rapid response team). Source: Wikimedia commons, public domain.

Algorithms using the behaviour observed by Clarke et al. (1970) were developed after the launch of the first dedicated ocean colour sensor, the Coastal Zone Color Scanner (CZCS), in 1978. This satellite sensor could observe ocean colour at six different wavebands in the visible. The ratio between the blue and green bands were used to quantify the chlorophyll concentration (Gordon et al., 1980; Smith and

Baker, 1982). The satellite was not equipped with infrared bands and could perform only partial atmospheric correction (Gordon, 1981). What is now called the "standard atmospheric correction" was initially developed by Gordon and Wand (1994a, 1994b) and Wang and Gordon (1994) for the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) sensor. The main hypothesis assumes that the light signal measured in the red and near infrared (750 to 870 nm) does not come from the water but comes mainly from aerosols. Aerosols are particles in suspension in the atmosphere that light can interact with. A measure of the light signal coupled to pre-computed models gives access to an estimation of the aerosols in the atmosphere. Their contribution to the total signal in the visible is corrected. With estimations of the different contributors to the total light signal, it becomes possible to isolate the water leaving signal L_w (more details in Chapter 2, Section 2.2). This water leaving signal is then transformed into a reflectance, which is the ratio of light leaving the ocean in a particular angle against the light entering the ocean from any directions. This reflectance is called the remote sensing reflectance (Rrs) and is the main optical property used to build algorithms in ocean colour. While the assumption that the infrared signal is driven by the aerosol content of the atmosphere in Case 1 waters conditions, the assumption breaks down in Case 2 coastal waters (Siegel et al., 2000). The reasons of the failure comes from the presence of minerals or the proximity of the seabed in shallow waters which also return a signal in the infrared for the bands commonly used in the atmospheric correction procedure. Thereby, the contribution from the aerosols is overestimated, and the total signal overcorrected. This leads to negative reflectances in coastal waters, which is not physically realistic (see Chapter 3, Figure 3.5 for an example of negative Rrs). The available Rrs signal is linked to the constituents present at the surface of the ocean. The common definition of the "surface of the ocean" from a remote sensing perspective is the euphotic depth (depth where 1% of the below surface photosynthetic available radiation light is measured, Gordon and McCluney, 1975; Lee et al., 2007). This depth can vary from a meter in very turbid waters (Majozi et al., 2014) to several dozens in open water conditions. An example of Rrs spectra is shown below (Figure 1.4). The light spectrum evolves based on the content of the water (Chl, CDOM and MSS). Only the Chl concentrations were changed for this example. If we compare low Chl (blue line, 0.001 mg m^{-3}) and high Chl (red line, 50 mg m^{-3}) concentrations, we observed that higher Chl values lead to higher signals backscattered in the green and red parts of the spectrum and lower signals absorbed in the blue.

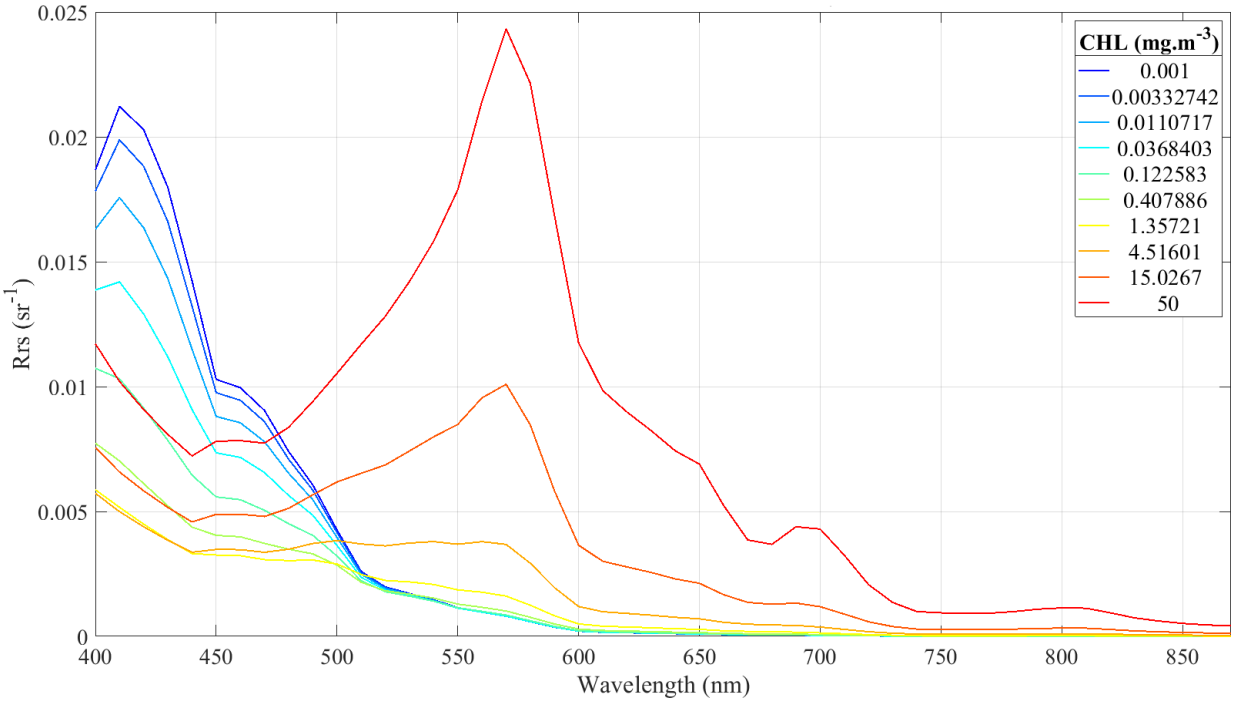

Figure 1.4: Spectra examples for different Chlorophyll a concentrations (see legend). The data were produced using Hydrolight 5.2 and a generic bio optical model simulating case 1 waters conditions with low amounts of coloured dissolved organic matter (0.01 m^{-1}) and mineral suspended sediments (0.1 g m^{-3}). For colour blind people, the dark blue curve has a value of 0.001 and is the top curve on the left, while the red curve has a value of 50 and is the top curve on the right that peaks near 570 nm.

Multiple approaches have been developed to tackle the standard atmospheric overcorrection problem in coastal waters. Ruddick et al. (2000) developed the Management Unit of the North Sea Mathematical Models (MUMM) approach to overcome the issue. It is based on a homogeneity assumption for the aerosols, and measures made over clear pixels can be extended to the coastal ones. While this shows interesting benefits at a small scale, it fails when applied to a larger scale as it requires locally fine-tuned parameters used by the method to be applied to a whole image. Wang and Shi (2007) proposed to switch from the red and infrared bands to the short-wave infrared bands (SWIR). The SWIR bands are less impacted by the content of the water and more linked to the aerosol content. However, while this atmospheric correction scheme produces coherent signal in very coastal waters, it fails at producing good quality reflectances in open waters. SWIR bands have access to a lower quality signal with lower signal to noise ratios than visible bands and are not usually used due to the high-quality requirements. One possibility could be applying a classification method to switch between the standard

and the SWIR atmospheric correction depending on the water type (Liu et al., 2019). Other atmospheric correction methods tried to fix other issues, such as the sun induced glint that deteriorates the light signal (Steinmetz et al., 2011). The research for an improved atmospheric correction method in turbid waters conditions is still active (Ilori et al., 2019; Renosh et al., 2020; Vanhellemont and Ruddick, 2021). Each sensor usually comes with its own atmospheric correction method, especially if it lacks infrared bands or was designed for land observation. To cite a few examples of atmospheric correction, the method iCOR (Image correction for atmospheric effects) is available for Sentinel–3 sensors (Wolters et al., 2021); ACOLITE (atmospheric correction for OLI "lite" version) for Sentinel–2 sensors (Vanhellemont and Ruddick, 2015; Maciel and Pedocchi, 2022), ARSI (the Atmospheric and Radiometric Correction of Satellite Imagery) for Landsat sensors (Clewley et al., 2014), SEN2COR the atmospheric correction processor (Main-Knorn et al., 2017) for sentinel sensors which can also be applied to Landsat sensors (Marujo et al., 2021), C2RCC (Doerffer and Schiller, 2007) or C2X (Brockmann et al., 2016) used for Sentinel or Landsat sensors. The main objective of each method tries to figure out the content of aerosols present in the atmosphere to correct the signal measured by a satellite sensor, and going into details for each method is not the objective of this thesis. Except identification of the different contributors that affect a TOA signal, there is no agreed method on how to correct a top of atmosphere signal for any satellite, and a vast amount of method exist. Some of the available Chl products are now using different atmospheric correction methods to assess the different conditions that can be met. For example, the OC-CCI (Ocean Colour Climate Change Initiative) reflectance product, which merges light signal in the visible over the European waters, uses the standard atmospheric correction (Gordon and Wang, 1994) along with the Polymer method (Steinmetz et al., 2011) to correct areas impacted by glint. Recently, Shroeder et al proposed a merged AC method using ensemble of neural networks (Shroeder et al., 2022). Correct classification of a pixel to apply the most appropriate atmospheric correction seems to be one of the future general method used to perform the atmospheric correction in ocean colour.

Once the signal is corrected and the bottom of atmosphere (BOA) reflectance Rrs available, different algorithms can be applied to retrieve the desired variable. For chlorophyll, algorithms use an empirical relationship based on the ratio of one blue band (400–510 nm) and one green band, near 550 nm. These algorithms are called the ocean colour algorithms (OCX). The first one developed by Gordon (1980) or Smith and Baker (1982) were using a single band in the blue and the green, therefore the algorithm was called OC2. With the band availability increasing with more sophisticated sensors, the algorithms used to retrieve chlorophyll evolved. The OC3 algorithm (O'Reilly et al., 1998) added one additional blue band near 490 nm that helped the estimation over lower concentrated waters. Algorithms

kept being improved, with OC4 (O'Reilly et al., 1998; 2000) adding the 510 nm band available for some sensors such as SeaWiFS. The algorithms using blue and green information were found to fail in case 2 waters, similarly to the atmospheric correction (Siegel et al., 2000; McKee et al., 2007). The reason comes from the presence of other constituents (MSS and CDOM) or other light contributors that alter the light signal (land adjacency effect (Zibordi et al., 2009); shallow waters (Cannizzaro et al., 2006), etc. The presence of other constituents alters both the atmospheric correction and the Chl algorithm quality. Gohin et al. (2002) proposed an improved algorithm (OC5, ocean colour algorithm using 5 bands) to provide improved estimations over waters impacted by MSS and CDOM. Using the 412 nm band, they managed to evaluate the CDOM presence linked to its higher absorption of light in the blue. They also added the 550 nm band to independently evaluate the concentration of MSS in the water, which increases the light backscattered at green wavelengths. This OC5 algorithm could therefore make more accurate estimations in the English Channel and northwest European coastal waters. The OC5 algorithm is based on look-up-tables that precomputes the chlorophyll based on the light signal, and is still limited when high loads of MSS or CDOM are present. Similarly, other algorithms have been developed to overcome the presence of CDO M (Carder et al., 1991; Gholizadeh et al., 2016) and MSS in coastal waters (Gons et al., 2002). Other failures of the open water algorithm were highlighted in oligotrophic waters (when the Chl concentration is low, <0.2 mg m^{-3}), where OCX algorithm struggle to reach low values. The Color Index (CI) algorithm (Hu et al., 2012) was developed to overcome the issue and is now commonly used in such conditions.

At the time of writing, two remote sensing laboratories propose an operational chlorophyll product to the public, PML and ACRI-ST in European waters. Their chlorophyll products rely on the OC5 and CI algorithms and are referred to as OC5-CI algorithms. These algorithms will be called OC5-PML and OC5-ACRI in the rest of the manuscript. The OC5-PML algorithm uses a merged satellite product, the Ocean-Colour Climate Change Initiative (OC-CCI). This satellite product merges reflectance signals available from different sensors and PML applies its own version of the OC5 algorithm to provide a daily chlorophyll map of the European waters since 1997. ACRI provides a similar product, named GLOBCOLOUR (https://www.globcolour.info/, accessed on 11 July 2022). For the ACRI product, the reflectances coming from different sensors are not merged. The OC5 algorithm is applied separately, and chlorophyll values are then merged for each day (Garnesson et al., 2021). Both products rely on classification of the pixels prior to the application of an algorithm. In remote sensing, the different types of products get a level attributed to identify what type of information they provide. A raw image of digital counts of photons for each wavelength is level 0. When it is converted into an energy and the satellite sensors calibration are

applied, it becomes level 1. When atmospheric correction is applied to the image and different variables such as the reflectance at a specific wavelength or chlorophyll is available, it becomes level 2. If the image is reprojected, it becomes level 3. The PML product is available daily and merges any available images of a day, and it is a level 3 product. If a temporal merge is realised, the product becomes level 4. The OC5-ACRI chlorophyll data are level 4. Daily products are merged to remove the missing data mostly induced by clouds or quality control applied by different agencies. Standard level 4 products are usually available on an 8-day or monthly basis, to help visualising the seasonal signal. The OC5-ACRI products applies 30 days sliding window to fill the clouds when direct observations are not available. Other data for some sensors are available from NASA website, mostly at level 2, but can be processed from level 0 to level 4 using the SeaDAS software (https://seadas.gsfc.nasa.gov/, accessed on 11 July 2022). The OC5-PML and OC5-ACRI products are distributed by different parties such as NEODAAS (https://www.neodaas.ac.uk/), CMEMS (https://marine.copernicus.eu/) or GLOBCOLOUR.

Environmental agencies need to be assured that the data quality obtained from remote sensing products is comparable to the quality retrieved from *in situ* measurements for reporting purposes. Currently available algorithms from PML, ACRI or NASA may not meet the requirements to be applied in all water types which possibly explains why they are not used by the reporting agencies. Overestimations still affect coastal waters when high concentrations of MSS or CDOM are present. A recent method merging different algorithms has been developed to bypass the overestimations returned by the OC5-PML algorithm from high concentrations of sediments. The method relies on classification and merges the OC5-PML product with the Gons algorithm (Gons et al., 2002) when turbid waters are detected. This product is called the JMP-EUNOSAT (Joint Monitoring Programme of the Eutrophication of the North Sea with Satellite data, Van Der Zande et al., 2019). The Gons algorithm uses band information in the red, available for the European sensors, such as MERIS equipped on Envisat–2 satellite or OLCI equipped on Sentinel-3. The Gons algorithm cannot be applied for other sensors due to the lack of available bands between 600 and 800 nm required. This product is available since 2016 continuously with the availability of Sentinel-3 satellites, or between 2002 and 2012 when the MERIS sensor was available. I spent some time reproducing this algorithm over the Scottish waters during the first year of the thesis in 2019. But with the official method not entirely done and the requirement to use European satellites only to access the algorithm fully, other directions have been followed after that attempt that will be developed in Chapter 3. An example of the type of product that can be developed using remote sensing for Earth monitoring is shown below in Figure 1.5. It represents the chlorophyll a content in the ocean for a whole

month, June 2002, observed by SeaWiFS sensor (1997-2010). The dark areas are clouds (especially near the equator) or masks applied (for the southern pole).

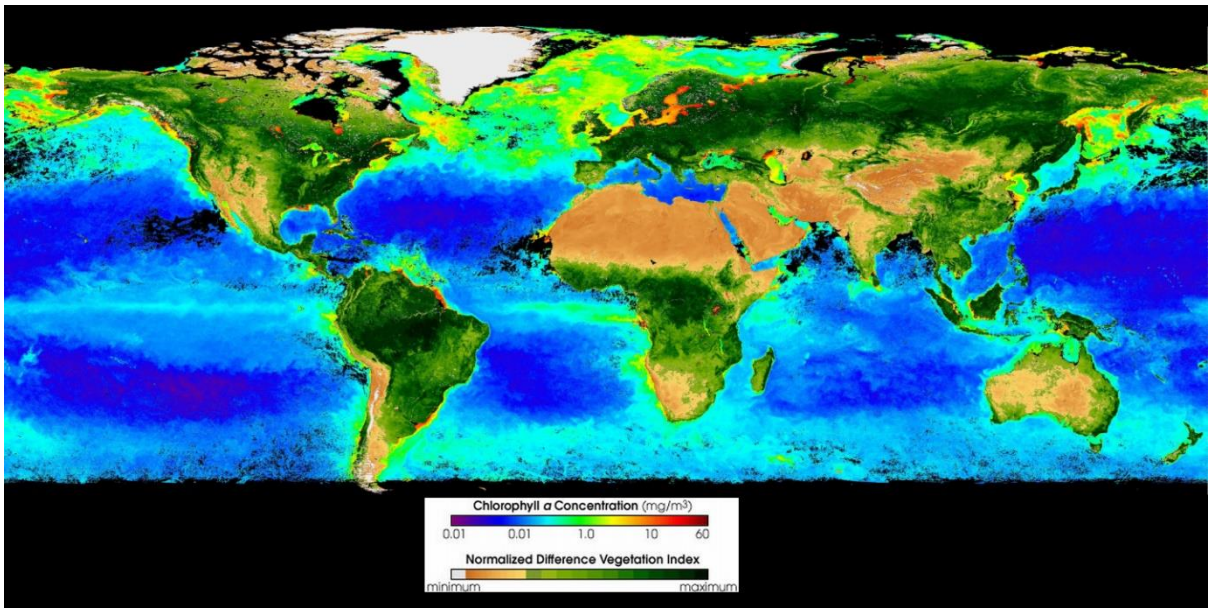

Figure 1.5: Map example of the global biosphere, with chlorophyll a content in the oceans (mg m^{-3}, OC4 algorithm) and indication of the density of plant growth on lands (NDVI algorithm). Averaged observations from SeaWiFS sensor during June 2002. Source: Wikimedia commons, public domain.

1.3. Monitoring Sea Surface Temperature from space

The surface temperature on Earth is commonly monitored from remote sensing sensors. It is one of the most important variable to retrieve in Earth observation due to the current state of climate change, and the temperature evolution is a direct effect of anthropogenic activities that emit CO_2. Earth is an imperfect black body whose surface emits photons in the infrared (from 700 nm to 1 mm) and microwave (1 mm to 1 m) parts of the light spectrum, based on its temperature. Observation of SST from space started in 1964 with the first American sensor Nimbus–1 equipped with infrared radiometers (Nordberg, 1965). Remote sensing techniques allow retrieval of Earth's surface temperature using either the infrared (4 or 11 µm) or microwave wavelengths (30 cm to 3cm). Many ocean colour sensors nowadays measure simultaneously both the visible and infrared spectrum (MODIS, the Advanced Very High Resolution Radiometer (AVHHR) or the Visible Infrared Imaging Radiometer Suite, VIIRS). The visible and infrared

sensors are usually equipped on sun synchronous satellites and come with a resolution at nadir of approximately 1 km, up to 100m for recent sensors (Landsat 8 & 9 with a reduced swath). While the light spectrum is continuous, the spectral information near 4 and 11 µm were selected because they are less impacted by the absorption of the atmosphere due to a higher transmissivity of the signal (0.970 to 0.985 respectively, Minett et al., 2019, example on Figure 1.5 below). For temperature retrievals, the main issues concerning infrared sensors are there sensitivity to cloud cover which masks the signal leaving the ocean, and the impact by sun glint during daylight for the 4000 nm bands which contribute to the signal and artificially increases the temperature estimated. The cloud cover is one of the main problem in Earth observation to access surface information: on average, 66% of Earth is covered by clouds, 53% for the lands and 71% for the oceans (Mao et al., 2019, based on 2003-2012 observations). The gases and water concentration present in the atmosphere are the main source of signal absorption (Figure 1.6).

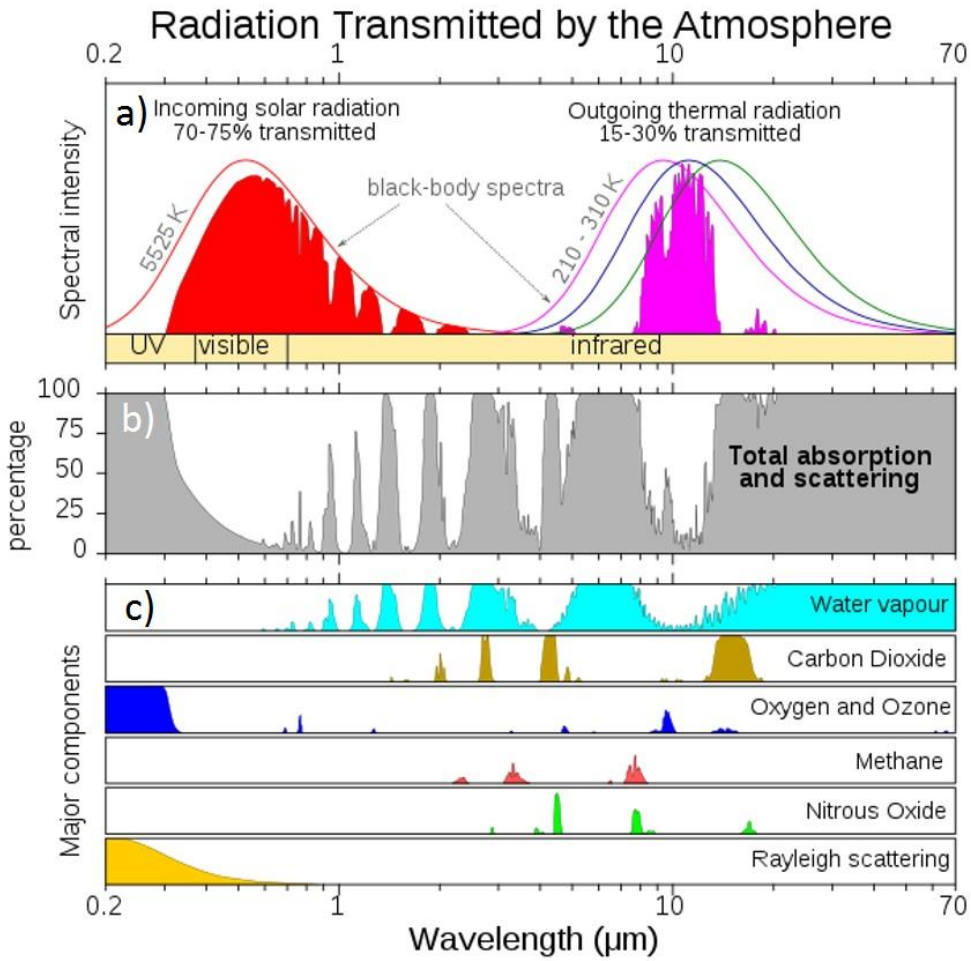

Figure 1.6: Example of the visible and infrared radiances observed on Earth and the different contributors to the signal absorption. In panel a), the red curve is the solar energy received on Earth at TOA; the pink, blue and green curves are the thermal radiative bands for a black body of 210, 260, and 310 K. The filled red curve represents the energy available at BOA; in violet the energy emitted. Panel b) presents the total percentage of available signal. Panel c) presents the spectral magnitude of the major contributors to the signal absorption. Source: Wikimedia commons, public domain.

Inversely, microwave bands are almost unaffected by clouds, and several microwave sensors were launched in the past (list in Table 1 in Gentmann et al., 2010). The main strength of microwave sensors is their access to an increased visibility which solves the cloud problem. However, microwave sensors come with a reduced pixel resolution of approximately 25 km² for the best sensors due to the lower signal in the microwave part of the spectrum. Therefore, they cannot make accurate predictions in coastal waters. Temperature retrievals are sensitive to waves produced by wind (Hanafin and Minett, 2005). Both techniques come with their own limits and infrared sensors are generally used for coastal observation

while microwave sensors can be used to get a global image of the Earth's surface temperature daily not impacted by clouds. Because the same visible and infrared sensors are available for Chl, they will be used in this thesis to retrieve the sea surface temperature, microwave sensor will not be used.

Algorithms developed to retrieve SST in the infrared initially relied on combinations of two close pairs of bands near 4 or 11 µm (Kilpatrick et al., 2015). In this thesis, we will call them SST4 if they use the pair of bands near 4 µm and SST11 if they use the pair of bands near 11 µm. The most recent SST algorithms for MODIS sensors include the use of additional information, such as the sensor zenith angle, the latitude, a temporal information (month of year) and which of MODIS mirror is used (Kilpatrick et al., 2014; Luo et al., 2019). Each of these additional information improve the quality of the retrieval. The latitude is linked to the thickness of the atmosphere, being thinner at the poles. The sensor zenith angle plays a similar role, and very high sensor zenith angle comes with a degraded signal. The inclusion of which MODIS mirror is used tries to eliminate minor corrections associated to instrumentation. An algorithm using a triple combination of bands at 4, 10 and 11 µm is available for the VIIRS sensors (Minett et al., 2020) and the field is still actively looking for improved retrievals (Zhang et al., 2017; Pastor et al., 2017; Huang et al., 2020). Algorithms relying on the 4 µm bands, SST4, are used during night time due to their higher sensitivity to sun glint (smaller signal magnitude compared to the 11 µm bands but varying faster than it). In contrast, SST11 algorithms using the 11 µm bands are used for daylight estimations of SST due to their higher signal magnitude and greater resistance to variations of the signal induced by sun glint. One important key point of SST observations is that the brightness temperature observed by sensors comes from the very top surface of the ocean, i.e. the first 10 micrometers for infrared sensors, the first mm for microwave sensors (Minett et al., 2019). Therefore, different definitions have been attributed depending on where the signal comes from. Infrared sensors measure the skin sea surface temperature (SSTskin) while the microwave sensor measure the sub-skin temperature (SSTsubskin). However, it is almost impossible to measure the temperature for these precise layers of the ocean, and most *in situ* data coming from drifting buoys usually sample the temperature at approximately 20 cm below sea level. Hence, SST estimations depend on the windspeed (Harris et al., 1994). Recently, visible bands of the 10 m resolution satellite Sentinel-2 have shown capacity to estimate SST using neural networks (Medina Lopez et al, 2019) with relatively lower performances compared to dedicated infrared sensors, but with a much higher resolution. The operational algorithms developed by NASA rely on multiple parameters (see above) and rely on previous temperature measurement to make an estimate. There is an opportunity to develop updated algorithms thanks to the vast amount of data available in the field, and test hypothesis that would not be available from a regional Chl development. This analysis will be conducted in Chapter 6.

1.4. Can neural networks provide better quality remote sensing products?

Most algorithms used in remote sensing to retrieve a parameter are based on statistics that let the machine "learn" how to make an estimation of a variable based on different inputs carefully selected. OCx algorithms used in ocean colour are empirical relationship using linear regressions to fit an estimated chlorophyll against an *in situ* value. However, the expression "machine learning" in now commonly associated to more advanced technique than regressions, such as artificial neural networks (NNs) which nowadays encompass the deep learning field, support vector machine (SVM) or random forest to cite only the most frequently used. All three main methods have been tested during a collaboration (Yan et al., 2021) and artificial neural networks outperformed other methods and were therefore selected during the thesis development. Artificial neural networks have been proposed to simulate biological neurons (McCulloch & Pitts, 1943, Hebb, 1949). Single artificial neurons have been trained to learn how to estimate a target relying on different input information, they are called perceptrons (Rosenblatt, 1958). Perceptrons consist of an input signal transformed into an output using an activation function with weights associated to each connection. Connections between multiple neurons and the definition of backpropagation of the error were developed later (Rumelhart et al., 1985, McClelland and Rumelhart, 1986). As a result of increased computation power availability, modern neural networks, especially deep learning networks, can contain up to billions of parameters and handle complex problems such as natural language processing (Brown et al., 2020). An example of how a small multi-layer perceptron works mathematically is shown in Chapter 2, Section 2.3.

NNs have initially been used in ocean colour for water classification (Benediktsson et al., 1993). The idea of using NNs to invert the light signal for Chl estimation emerged in 1994 (Doerffer and Schiller, 1994). Buckton et al. (1999) applied NNs to modelled data and discussed the possibility of including non-light information in network training. NNs have been applied for Chl retrieval for Case 1 waters using either above surface measurements (Keiner and Brown, 1999) or simulated data (Gross et al., 1999) or a mixture (Tanaka et al., 2004), with Keiner and Brown showing that NNs outperformed state-of-the-art algorithms at that time. Over optically complex waters, Schiller and Doerffer (1999) used NNs with simulated Rayleigh-corrected reflectances, while D'Alimonte and Zibordi (2003) applied the technique to a real coastal data set. In both cases the NNs returned promising results and / or better performance than state-of-the-art algorithms. NNs have been applied as operational products for case 2 waters constituents' retrieval for the Medium Resolution Imaging Spectrometer (MERIS, Doerffer and Schiller, 2007) and the Ocean and Land Colour Instrument (OLCI, Brockmann et al. 2016) radiometer sensors. Hieronymi et al.

(2017) have proposed a network trained on modelled data using the method developed in Schiller and Doerffer (2006) applied to real satellite images with neural networks developed for classified water types being the key feature. Pahlevan et al. (2020) trained NNs for lakes with Sentinel 2 and 3 satellite data using above surface measurements showing good performance, while Cao et al. (2020) have shown that NNs can outperform current state-of-the-art algorithms for Chl predictions in Chinese lakes. NNs have also been used to retrieve other parameters, such as photosynthetically available radiation (Schiller, 2006), other pigments (Bricaud et al., 2007), inherent optical properties (IOPs, Ioannou et al., 2011, Chen et al., 2014) and the spectral diffuse attenuation Kd (Jamet et al., 2012). Recently, NNs have been applied to retrieve surface temperature and salinity using TOA visible bands from the high resolution satellite Sentinel-2 (Medina-Lopez and Urena-Fueñtes, 2019; Medina-Lopez, 2020). Top of atmosphere signals have seldom been directly used by the OCRS community with only a few publications describing techniques relying on it (Niang et al., 2003; Vepsäläinen et al., 2005), largely resulting from the fact that <10% of the TOA signal in the blue comes from within the ocean for case 1 waters. It is clear that NNs have significant potential to improve retrieval of Chl and other important water quality and light field parameters from ocean colour signals in optically complex coastal waters. A review of the use of deep learning methods (i.e. more than one hidden layer) developed for Earth observation can be found in Yuan et al. (2020) with a section dedicated to ocean colour.

1.5. Objectives of the thesis

Established, reliable and comprehensive data are essential for reporting against national and international water quality standards. Limited performance of existing ocean colour Chl algorithms in optically complex coastal waters is a major inhibiting factor in take-up of the technology by national environmental monitoring agencies. Many of the algorithms for coastal Chl are restricted in scope either geographically; through optical water type classification; cannot be applied to all sensors due to additional spectral information required; by restricting application through extensive use of flags to eliminate the most challenging conditions (such as very turbid or shallow waters), many of which are regularly found in northwest European shelf seas. This problem has persisted for over twenty years and there is little scope to believe that further development of blue-green reflectance ratio algorithms will significantly advance the issue (O'Reilly and Werdell, 2019). However, satellite data have the potential to provide a degree of spatial and temporal coverage of Chl concentrations that is highly challenging or most likely impossible with alternative present-day *in situ* observational technology, particularly in open sea or offshore areas (Van der Zande et al., 2019). The recent advancement of machine learning techniques suggests that it is

time to develop a new framework for exploiting their strengths. Whilst neural networks have been discussed in the ocean colour literature since as far back as 1997 (Buckton et al., 1999), they remain a new and unfamiliar territory for many researchers operating in ocean colour or remote sensing in general, because empirical relationships are easier to build when only a few spectral information are used. Whilst there is a small but growing body of literature in this area (see above), understanding of the mathematical techniques involved and how to properly implement them remains confined to a relatively small element of the ocean colour community.

Machine learning algorithms are data-based approaches, and therefore primarily limited by the amount and variety of data available. They are often considered regional and not used as operational algorithms due to their tendency to overfit when data availability is limited. Overfitting occurs when an algorithm fails to make coherent estimations for specific conditions because it did not learn the expected behaviour. The reasons can be multiple. For neural networks, overfitting usually come from the complexity of the algorithm used or the lack of data available to represent all conditions. The whole process of developing a useful neural network algorithm consists of applying different specific steps to avoid overfitting. Satellite sensors have been sampling the Earth continuously since 1997 and growing datasets are now available for both chlorophyll and temperature. Due to the complexity of the light signal and the capacity of neural networks to deal with complex tasks, there is an opportunity to retrieve chlorophyll or temperature using NNs without following the physical approach by using only specific spectral information. The current state-of-the-art physical approach to retrieve chlorophyll from space relies massively on classification and application of an appropriate algorithm for each satellite pixel. For temperature retrieval, the current approach used by NASA divides the Earth into small latitude boxes and on a month-of-year basis. It means that the algorithms have to be fine-tuned for each box and for each month of year. . Moreover, current SST algorithms also require knowing the temperature based on other sensors or previous measurements prior to make a temperature estimation.

Neural networks are considered as black boxes due to the difficulty to visualise them, especially if massive architectures are used with hundreds of neurons. There is a tendency to avoid using algorithms that cannot be manually tuned like the empirical ones, easier to visualise. However, current algorithms for both chlorophyll and temperature estimation are not trivial and as they become increasingly complex are impacted by the same issue of being hard to visualise and understand *in fine*. This is particularly true for the OC5 algorithm that relies on look-up-tables fine-tuned by the different groups, which are not publicly available. Neural networks have the capacity to propose a unique algorithm that can be applied

anywhere, if the dataset is sufficiently representative. They are the main method used in this thesis and different experiments are conducted in the results chapters. Due to its capacity to observe simultaneously the visible and infrared spectrum coupled to a 20-year availability (long enough to generate a sufficiently large ground truth data set), the MODIS sensor onboard the Aqua satellite is the main source of remote sensing information used in this thesis.

The structure of the thesis is as follows:

- First, the methods and tools used common to the different chapters are described in Chapter 2. Details are provided about how *in situ* samples are collected, how remote sensing measurements are corrected, and a brief description of neural networks and statistical metrics to evaluate the algorithms performance is conducted.
- The development of a neural network algorithm to retrieve chlorophyll in European shelf seas is conducted in Chapter 3. Top and bottom of atmosphere reflectances are tested for retrieving chlorophyll a. The different steps at identifying optimal architectures or how to avoid the randomness produced by neural networks are described. Comparisons with state-of-the-art algorithms are assessed, and the advantages and disadvantages of each approach are highlighted. Independent coastal data are used to further evaluate performances reached. Evaluation of noise impacting performance is attempted, and inclusion of non-light signal to improve performance is assessed.
- Current operational state-of-the-art algorithms use five bands in the visible spectrum. It is known that other bands in the red or infrared can provide useful information about the water and atmospheric content. The impact of which information is available from the remote sensing perspective when a neural network algorithm is developed in Chapter 4. An automatic clustering machine learning method is applied and compared to a manual selection. Neural networks are also applied to other available ocean colour datasets to see if they can retrieve chlorophyll with improved performance even in case 1 conditions, and can be generalised.
- The potential consequence of data availability on performances is carried out in Chapter 5 with the help of a modelled dataset providing access to the light signal and the different water constituents. Several tests are included, first, addition of noise to turn the perfect modelled dataset into a more realistic measurement. Then, multi-task learning which retrieves different variable (Chl, CDOM and MSS) at the same time. Eventually, due to the

nature of the modelled dataset created, the benefit from having access to a hyperspectral signal compared to a multi-spectral one is evaluated.

- To evaluate if a general approach on how to develop a neural network exists for remote sensing or if the observations made in Chapters 3 to 5 were chlorophyll specific, the same method is applied to SST retrieval in Chapter 6. Combinations of different remote sensing signals are tested to evaluate the potential benefit of having access to more information for both day and night retrievals. Noisy or non-noisy remote sensing data are evaluated separately. A regional dataset is also available with access to visible and infrared bands simultaneously. Using it, assessment of temperature and salinity retrieval are performed. A further experiment consists of training an algorithm over a specific period (2002-2003) and apply it to the rest of the available data to evaluate if the current month-of-year basis algorithm is required.
- The current eutrophication assessment relies on use of *in situ* chlorophyll a samples to evaluate the status of a region. The spatial and temporal information available from both remote sensing and *in situ* sampling are compared in Chapter 7. Benefits from using a remote sensing chlorophyll a algorithm for the eutrophication report is highlighted, and potential changes in the methodology are discussed.

2. Materials and methods

This chapter provides more details about the materials and methods used in the thesis and should help the reader understand the decision making process that shaped the project. The data come from two main different sources: *in situ* samples collected at the surface of the ocean detailed in Section 2.1 and satellite images collected from passive light measurements detailed in section 2.2. Section 2.3 presents the main tool used to create algorithms used during this project, multi-layer perceptron neural networks. The different metrics used to evaluate performances are presented in Section 2.4. Finally, a brief summary of the processing of data is presented in Section 2.5.

2.1. *In situ* data

2.1.1. Chlorophyll a

Chlorophyll a (Chl) is contained inside phytoplankton and plants cells, inside what is called the chloroplast (Figure 2.1a). Chl is a molecule, a green pigment, which interacts with visible light (Figure 2.1b). It absorbs light near the blue (peak near 440-460 nm) and red (peak 630-670 nm) bands (Figure 2.2). The radiative energy received from the photons absorbed by the Chl is converted into chemical energy (adenosine triphosphate, ATP) inside the chloroplast. This energy is required by the organism to drive photosynthesis, which consumes water, nutrients and inorganic carbon. Oxygenic photosynthesis creates O_2 and organic matter (sugars, which can be consumed later by the organism) during the process. More details about the different steps involved in the photosynthesis can be found in Falkowski and Raven (2013) or Sarmiento (2013). The idea of this section is to provide some information about how Chl is measured by the different available techniques and to provide an indication of their relative accuracy. Why is the focus on Chl and not the other pigments, or all of them? Other chlorophyll pigments exist (Chl b, c, d, etc.) and other pigments in general exists (*pheophytin*, carotenoids, etc.). However, Chl found in the photosystems that drive photosynthesis and is the only one present in all organisms that conduct oxygenic photosynthesis (creation of O_2, Bjorn et al., 2009). Chl represents a part of phytoplankton biomass, of the order of 2% (Desortova, 1981) and this fraction varies with environmental conditions. Organisms may grow more chloroplasts to absorb more light when light is a limiting factor, which is called photoacclimation (MacIntre et al., 2002). Being able to measure Chl from plants is a key tool to evaluate the amount of phytoplankton and indirectly provide information about the primary production and the consumption of carbon dioxide by the organisms.

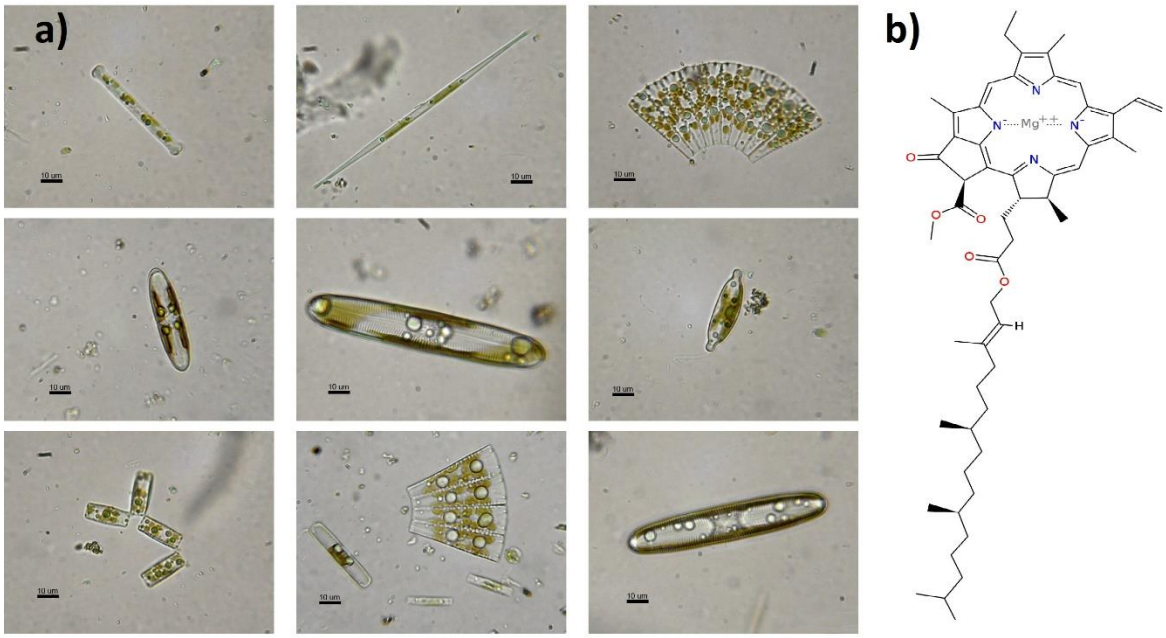

Figure 2.1: a) Image of a different species of diatoms. The green parts contain chloroplast, inside which the chlorophyll a pigments are (each species can contain other pigments). b) Skeletal formula of the chlorophyll a molecule. Source: Wikimedia commons, public domain.

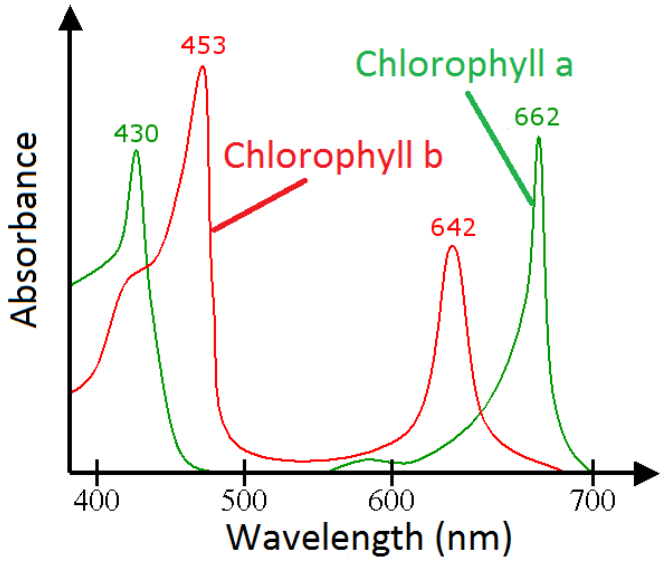

Figure 2.2: Diagram of the absorption of Chlorophyll a and b molecules in the visible part of the spectrum with peak near <430-450nm and near 630-670nm. Source: Wikimedia commons, public domain.

There are a variety of ways to measure the chlorophyll concentration of a volume of water. Differences include whether to remove a sample from the water column in order to extract Chl from a filtered sample (*in vitro*) or to use a submersible fluorometer to make an *in vivo* measurement. *In vitro*

samples can be analysed using high performance liquid chromatography (HPLC), spectrophotometry or fluorometry with various options for sample processing including use of different extraction agents. There are even differences in how water samples are collected in the first place. The most common method uses rosette of bottles connected to a CTD (conductivity, temperature and depth), that get plunged inside the water and close themselves when they reach a specific depth (Figure 2.3a). At the Stonehaven coastal station (57°N, 2°W) sampled by Marine Scotland Science, a 10 m Lund tube is plunged inside the water and closed when fully immersed, containing a volume of water from the first 10 meters.

The Chl is contained inside the chloroplasts, which are themselves located within the phytoplankton cells. *In vitro* sampling techniques require extraction of the Chl by chemical and mechanical processing resulting in the production of a solution of Chl within the chemical used for extraction, often 90% acetone. This is a destructive technique which removes the Chl from any influence of the original plant structure. In vivo fluorescence measurements, in contrast, do not disrupt the cell structure so the observed fluorescence signals influenced by the physiological state of the cells, meaning that signals can be affected by e.g. solar quenching effects (Berry et al., 2018). *In situ* fluorometers are typically calibrated prior to deployment using standard solutions of extracted Chl.

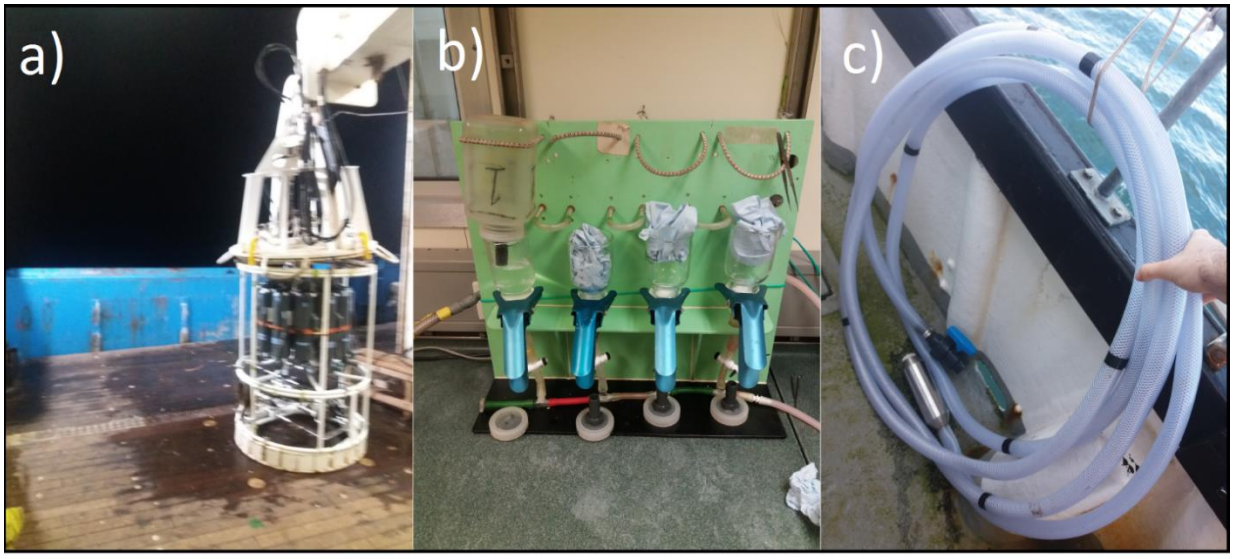

Figure 2.3: a) Photo of a rosette equipped with Niskin bottles on the MRV Scotia, during the December 2019 cruise campaign conducted annually by Marine Scotland Science. b) Filtration of approximately 1 L of water onto a glass fibre filter. c) Photo of the Lund tube used at the Stonehaven coastal monitoring station used to collect water from the first ten meters. Photos taken by myself during the December 2019 cruise campaign and during my visit at the Stonehaven coastal station in March 2020.

2.1.2. Preparation of the solution for *in vitro* sampling

For *in vitro* techniques, phytoplankton cells present inside the volume of water pulled out from the sea need to be isolated. A typical volume of 1L is filtered onto a 47 mm GF/F fibre glass filter. A higher volume is used if the water is clear to ensure a measurable amount of material is collected, and a smaller volume is used if the water looks highly concentrated to reduce the duration of filtration. GF/F filters typically retain material larger than ~0.7 μm. When the operation is conducted from a ship, if the further steps that involve chemical measurements cannot be done directly, the filter is generally frozen after filtration, kept at a temperature of –40 to –10°C and protected from light with aluminium. This process is operated to reduce the Chl degradation which will change the Chl concentration with time. When the measurement can be done, Chl cells need to be extracted from the algal cells. The phytoplankton cells present onto the filter paper are often broken using mechanical forces (grinder) and a solvent composed of 90% acetone and 10% of water is added to dissolve the cells and release the Chl inside the solution. The cell parts and materials heavier than Chl present in the solution are separated by centrifugation. A specific volume of the Chl and acetone solution can then be analysed by one of several different methods.

2.1.3. Spectrophotometry

Chl absorbs light across the full visible spectrum (Figure 2.2). The spectrophotometry technique consists of a measure of absorbance which represents the quantity of light absorbed by a solution. A volume of solution obtained after steps described in 2.1.2 is put in a cuvette of a spectrophotometer with known dimensions of path length *L*. The amount of light transmitted through a Chl sample is referenced against the amount of light transmitted through a blank consisting of the chemical used for extraction (usually 90% acetone). This measure is called the absorbance. Less light can cross the more the solution is concentrated. Using Beer-Lambert law (*A = kLC*, Luther and Nikolopulos, 1913), the absorbance *A* is linked to the concentration *C*, the extinction coefficient *k* and the path length *L*. The standard method (Jeffrey and Humphrey, 1975) requires absorbance measurements at different wavelength (630, 645, 663, and 750 nm) to correct for different effects, such as turbidity from the 750 nm measure. Standard equations (e.g. Parsons and Strickland, 1963) are usually used to convert the absorbance signal into a Chl concentration.

2.1.4. Chlorophyll Fluorescence

Chl absorbs light with peak at shorter wavelengths (<450 nm) where more energy from the sun is available (see Chapter 6, Figure 6.1). When Chl molecules absorb photons of short wavelength, different interactions occur. The absorption contributes to an increase of temperature of the system and can change the state of excitation of available electrons. When electrons return to their initial excitation state, the process emits photons of lower energy with a peak near 682 nm (Krause and Weis, 1984). The amount of light emitted by fluorescence is small in comparison of the total light absorbed (0.5%-10%; Latimer et al. 1956; Brody and Rabinowitch 1957; Barber et al. 1989; Porcar-Castell et al. 2014). However, the fraction of energy used for photosynthesis is inversely proportional to the fluorescence intensity (Duysens and Sweers, 1963, Kalaji et al., 2017). Therefore, a measure of the fluorescence is a proxy for photosynthetic activity. The emission process is called chlorophyll fluorescence. Therefore, measuring the light in the red directly using a suitably calibrated sensor gives the Chl concentration. The fluorescence field is vast, and researchers are still arguing about what exact processes that lead to fluorescence are involved (see question 21 in the review paper by Kalaji et al., 2014). A fluorescence measurement can be conducted onto an extracted volume of Chl, or a water volume directly. The main advantages of a fluorescence measurement are the speed of the measurements (within seconds to minutes). Fluorometers typically use a blue LED to stimulate red fluorescence which is then collected by a suitably filtered detector. This process is highly sensitive and can be easily adapted for *in situ* observations as well as for *in vitro* samples.

Note however, that *in vivo* fluorescence signals are subject to physiological influences while extracted samples are not. Therefore, although an *in situ* fluorometer will have most likely been calibrated using known concentrations of extracted Chl, there will be variations in output signal that are not necessarily concentration related.

2.1.5. High Performance Liquid Chromatography (HPLC)

Chromatography separates the different chemical components of a mixture inside a solution. High Performance Liquid Chromatography (HPLC) is an enhanced version that adds pressure and pumps the mobile phase (containing the mixture) through a column (the stationary phase) to speed up the process of normal chromatography. The mobile and stationary phases have different hydrophobic affinities, which will grip the different molecules based on their hydrophobic status. Each component of the mixture has a specific affinity, and while being pumped through the column, they move at a different speed that separates them. Once separated, an absorbance measurement can be conducted at different wavelengths depending on the component analysed, usually in the ultraviolet and visible spectrum. The chromatograph represents the absorbance of the isolated molecule, which usually consist of a curve of a particular area at a specific wavelength, which is then converted into a concentration. A fluorescence measurement can also be used. The light detector used inside an HPLC sensor needs to be calibrated using standards for each pigment of interest. The key advantage of HPLC is its ability to physically separate each pigment allowing individual analysis.

2.1.6 Advantages and disadvantages of the measurements

Each technique has its own advantage and disadvantage. In terms of cost and deployment facility, the fluorescence conducted *in vivo* is the cheapest and can produce automated high frequency measurements. Therefore, fluorimeters are attached to drifting buoys such as BGC-AGRO floats. There is no automated measurement method of extracted Chl due to the requirement of multiple chemical products and steps involved. Methods that require water samples to be filtered still require a human presence. A fluorometer can produce a very high frequency Chl estimate (one value a minute for fluorimeters attached to Ferries for example). In terms of data quality and precision of the measurement, an automated fluorescence measurement produces the poorest estimate of Chl. A water sample taken at sea can contain multiple phytoplankton species, of different shape and size, which can each contain different pigments. Fluorescence is also sensitive to the ambient light when the measure is conducted. Finally, cells can hide each other and only a part of them will receive the incoming light (quenching effect).

Other pigments may not emit light at the exact same wavelength than Chl but can contribute to the signal measured, such as *phaeophytin*, a degraded product of Chl. All these effects lead to poorer quality of Chl estimates.

The *in situ* Chl dataset assembled for Chapters 3 and 4 contains a mixture of measurements produced using all of the different methods briefly described above. In 2004, Claustre et al (2004) conducted a cruise campaign that collected water samples and measured triplicates of pigments presents inside the same water sample using HPLC. They calculated that despite using the same technique and water sample, absolute percentage difference errors of approximately 21.5% occurred for the pigments in general, and to 7% for Chl. They managed to reduce the error down to 13.9 and 5.5% using a round robin exercise for the other pigments and Chl, respectively. Round robin exercises have also demonstrated uncertainties in HPLC concentrations up to 40% (Sørensen et al. 2007). The data set collected in this thesis is much more diverse, including in vivo and *in vitro* samples, and therefore could easily present errors of 50% or more depending on the measurement conditions (Jaccard et al, 2018). Tilstone et al. (2012) go even further by returning estimates of errors of 80%. The important point to note is that as a source of ground truth for remote sensing, great care has to be taken to consider the uncertainties in the Chl data collected at sea.

Spectrophotometry and spectrofluorometry using Chl solutions can reach similar estimates and an example is shown from the SeaBASS dataset in Chapter 4 (Figure 4.14). Estimates from HPLC are the most accurate that can be achieved now. This is because HPLC can isolate the different pigments precisely, while other methods are impacted by presence of these other pigments that contribute to the light signal measured. Spectrophotometric or spectrofluorometric measurements can be comparable to total Chl measured by HPLC (Stramska et al., 2006) when carefully measured. However, is a very precise Chl estimate meaningful for ocean colour? Chl is obviously present in most organisms that the community is interested in and a Chl values provide information about the total biomass of the organism. Nevertheless, the sun light interacts with the whole phytoplankton cell, present inside the ocean, not the Chl pigment only. The different effects that can affect the accuracy of a Chl estimate also influence the light scattered back to the surface of the ocean that will potentially reach a sensor at the end. In ocean colour, the volume of water from which the Chl concentration is measured is much lower than the volume of water that contributed to a satellite signal (1 L versus 1 km² of various depth). The depth distribution of Chl is highly variable (Smith, 1981). The Chl concentration inside a cell can also evolve based on the ambient light to overcome potential light deficiency under specific conditions (see recent work done by Cornec et al.,

2021). Deep Chl maxima can reach 50 meters in some conditions. Light can interact and reach the surface from depths down to at least 80 m under some conditions (Gieskes et al., 1987). Each of these details is a consideration for how accurate a Chl *in situ* estimate should be to become useful when creating satellite algorithms. Moreover, there is not a single sampling group performing all the *in situ* measurements on Earth, and triplicates of HPLC data also returned discrepancies up to 20% when made by the same research team (Claustre et al., 2004). The data used in this thesis are public and come from multiple sources or were directly provided by the different producers. Therefore, limiting the development to a single Chl extraction method does not make sense, especially when we consider that data availability is one of the main limiting factors when developing neural networks. Therefore, despite each of the method having its own limits, accuracy and detection range, data from multiple organisations and any technique briefly described above are used in this thesis

2.2. Satellite data

2.2.1. Measurement of the signal

Photons emitted during the fusion process that occur inside the Sun reach the Earth surface approximately 8 minutes and 20 seconds after leaving the Sun's surface. Photons interact with the matter present on Earth, the atmosphere, lands, and the oceans with two main mechanisms: absorption and scattering. A photon can be absorbed by a material's electron and contribute to increase its temperature and vibrational state. Absorption removes this photon from the environment. A photon can be scattered by a material's electron which will lead to a change of its direction. Scattering can be elastic (conservation of energy) or inelastic (loss of energy, transferred to the electron(s) it interacted with). The nature of the absorption and scattering depends on the material, especially its size. Most of the photons (90 – 98%) that reach the sea surface enter the ocean and interact with water constituents (Ocean optics book, https://www.oceanopticsbook.info/, accessed on 11 July 2022, Mobley, 1994; Kirk, 1994). A proportion of these photons interact with the water constituents, is scattered back to the surface and leaves the ocean. Some of that proportion reaches the satellite sensor and can be measured. The proportion of the backscattered photons is linked to the properties of the medium and its constituents. Ocean colour satellites are equipped with spectroradiometer sensors to count the number of photons received at a specific wavelength. When photons enter the sensor from the aperture, they are reflected by a mirror and directed towards a detector. Prior to reaching the detector, they cross a spectral interference filter (Barnes et al., 1998) which only lets photons of a specific wavelength width pass, called a waveband. Photons that pass reach the radiometer sensor and are absorbed by the material present. With the

appropriate material used, the absorption of photons leads to the emission of electrons (Gordon, 2019, Section 2.2). It is considered that the elastic mechanical-like “collision” between the photon and the electron causes the latter to be ejected from the material due to the transfer of energy. When ejected, the electrons also called *photoelectrons*. Electrons are negatively charged and can contribute to the flow of electricity when moving. This is called the *photoelectric effect* (Einstein, 1905). Since Einstein's paper, photons are considered to carry quanta of energy linked to their wavelength with Planck’s law (Planck, 1901). The emitted electron can interact with its environment, and a precise measure of the temperature, electric current or conductivity changes of a material depends of the number of photons that were absorbed by the material. Photons that reach the sensor come from different contributors linked to the atmospheric content, water or surface interactions. Figure 2.4 summarises the different interactions that can occur for visible and near infrared photons before they reach a satellite sensor.

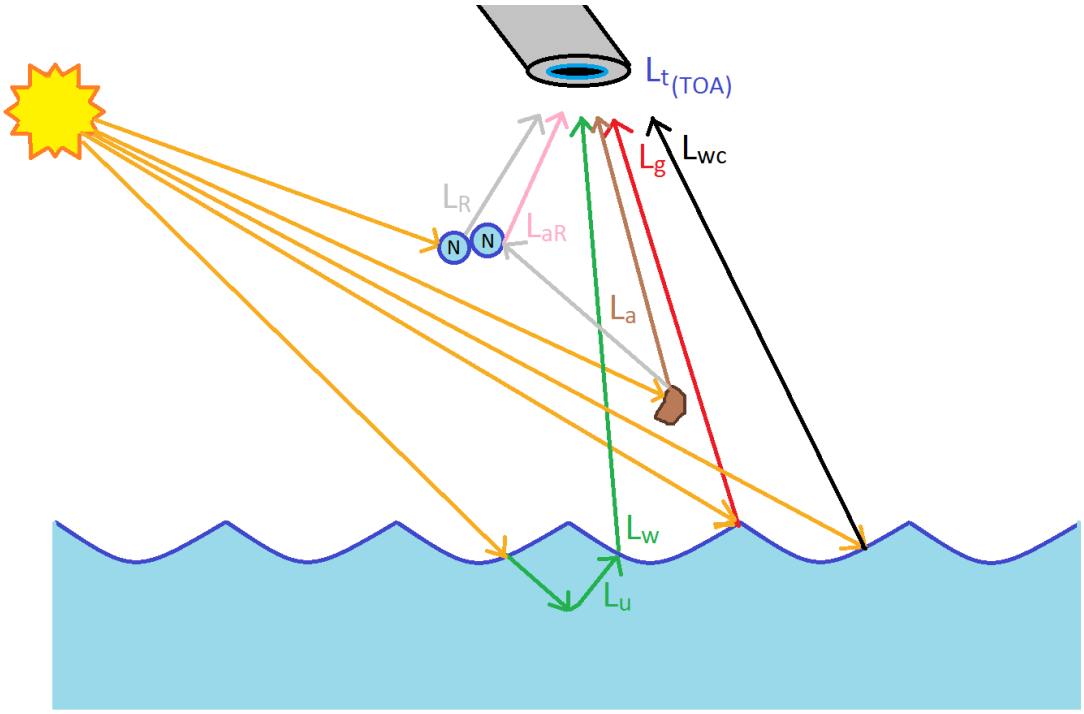

Figure 2.4: Diagram of the different processes that can contribute to the top of atmosphere radiance measured by spectroradiometer on-board a satellite. Incoming sunrays (yellow) are in reality parallel between each other but this representation eases the understanding of interactions. Multiple interactions can lead to a photon reaching the sensor. The different radiances contribution to the top of atmosphere

signal (L_{TOA}) are, at TOA: L_R (Rayleigh scattering), L_{aR} (Rayleigh-Aerosols interaction), L_a (Aerosols), L_w (water leaving signal), L_g (solar glint) and L_{wc} (whitecaps).

The measured signal is considered to be at the "top of atmosphere" (TOA), because it is the signal received at an altitude of 700-800 km, where polar sun synchronous orbit satellites usually are. L_t is the TOA total radiance measured by the sensor. Defining where the Earth atmosphere stops is not trivial and depends on the task. It is considered that more than 99.99% of the atmosphere molecules are below the Karman line (Valentine, 2016), defined at approximately 100 km (McDowell, 2018). Therefore, photons have a chance to interact with most of the atmosphere before they reach the sensor. The raw signal measured by the sensor consists of digital counts. This raw signal then needs to be calibrated to account for sensor degradation since the launch. Sensors attached to satellites are initially calibrated in the laboratory when they are built. Once they leave the laboratory, they must be recalibrated frequently to keep making accurate measurements. A vicarious calibration team oversees the maintenance of the satellite sensor's quality, for example by observing the Moon's surface or the sun by using a solar diffuser stability monitor for the later (Xiong et al., 2003). Terra and Aqua spacecrafts are equipped with a MODIS sensor each (MODIS Terra or MODIS Aqua) and were launched in 1999 and 2002 respectively. MODIS Terra faced sensor degradation early in its life, which impacted some algorithms relying on it, especially for blue bands (Wang et al., 2012). The ocean and atmosphere are mostly blue, therefore MODIS Aqua was selected rather than Terra due to calibration issues for its blue bands (Xiong et al., 2003a; 2003b; 2006). It is important to consider that after 20 years in space, the quality of MODIS Aqua decreased over time but the calibration team managed to maintain the quality of the measurements to a high level. MODIS Aqua is expected to be available until 2025 at best, when it will have consumed all of its fuel required to correct its orbit. It was originally designed to live 7 years.

L_t represents the radiance (W m^{-2} sr^{-1}) after photons have been counted and converted into an energy, for a specific angle, for a specific wavelength, integrated over time and after the different calibrations have been applied by the calibration team. At the top of atmosphere, the total radiance is the sum of each contributor that reached the sensor shown in equation below (Eq. 2.1). All radiances **L** (Luminance in French, which gave the letter L) are wavelength and viewing angle dependent. In Eq. 2.1, they are displayed at the top of atmosphere for easy readability. Some (L_w and L_g) have been altered by the atmosphere transmissivity.

$$\mathbf{L_t} = \mathbf{L_R} + \mathbf{L_a} + \mathbf{L_{aR}} + \mathbf{L_g} + \mathbf{L_{wc}} + \mathbf{L_w} \quad (2.1)$$

2.2.2. Atmospheric correction

Historically, to evaluate the composition of oceanic constituents, a measure of the water leaving radiance L_w has been required at the bottom of atmosphere (BOA). The different contributors to L_t need to be estimated to isolate L_w, to be able to apply the algorithms later on. In ocean colour, the process of isolating the water leaving signal is called the atmospheric correction. Atmospheric corrections have been initially proposed by Gordon (1978), to account for the atmospheric signal for the Landsat-1 satellite sensor. A light measurement made by a satellite at TOA is affected by other contributions, mostly the glint on the ocean surface and interactions with atmosphere constituents. In ocean colour, it is often mentioned that 90% of the signal comes from the other contributors and the ocean represents only 10% of the total measurement. Hence, the overall measurement made by the sensor needs to be of high precision and requires high signal-to-noise ratios. A rule of ten is generally mentioned when estimating errors and their impact onto an indicator estimation due to only 10% of the signal coming from the ocean. It also explains why the ocean colour field spent a large amount of time to carefully develop atmospheric correction methods. However, this is an incomplete vision of the problem, as the 90%-10% rule is only valid for some blue bands in open waters conditions. When the conditions are coastal or over a massive phytoplankton bloom, the green and red photons mostly come from the ocean and not the atmosphere. Gordon developed the method that would be improved and applied later to the CZCS sensor on the Nimbus-7 satellite. The different contributors are represented in Figure 2.4. With air particles of size S much smaller than the visible light wavelength ($S<<\lambda$), the light is scattered following the Rayleigh scattering theory (Rayleigh, 1871). Rayleigh scattering is the dominant scattering interaction that occurs in the atmosphere when visible photons of the 400–800 nm encounter water and air molecules, which are smaller by three orders of magnitude. L_R is the radiance from the Rayleigh scattering induced by air molecules that compose the atmosphere. L_a is the radiance coming from the scattering induced by aerosols. Aerosols are molecules that are injected into the atmospheric column from interactions of wind with land particles, ocean particles or are emitted by human activities. L_{aR} is the radiance induced by the interaction between aerosols and Rayleigh scattering of atmosphere molecules. L_g is the radiance that is induced by the glint of the sun of the ocean's surface. L_g can include multiple terms, such as the direct specular reflection of the sun by the sea surface, often called L_{gs}, and the reflection of the irradiance of the sky by the surface, L_{sky}. Whitecaps and foam at the surface of the sea are created by wind interactions

with the ocean and can contribute to the light signal due to their higher reflectivity (they are white), called $\mathbf{L_{wc}}$. If the signal is from an interaction with the ocean, a direct atmospheric transmissivity $\boldsymbol{t}$, wavelength dependent is applied to convert the signal from BOA to TOA. In Equation 2.1, all contributors are considered at the TOA to ease readability. MODIS Aqua faces uncertainty measurements of approximately 2% for the total radiance at most for its visible bands when observing Earth at nadir (Xiong et al, 2019). It is important to consider that the measure of the uncertainty from a satellite sensor is conducted when comparing BOA signals with light measured just above water. The latter process is conducted at AERONET stations (Holben et al., 1998) originally developed to quantify the aerosol composition of the atmosphere.

There are multiple methods to estimate the different contributors and retrieve the water leaving radiance L_w (see the four methods available for MODIS Aqua compared in Goyens et al., 2013, or Ilori et al., 2019). Here, only the NASA official algorithm initially developed by Gordon and Wang (1994) called "standard atmospheric correction" is briefly described (Stumpf et al., 2003; Bailey et al., 2010). From the multitude of gases present in the atmosphere, L_R is impacted by four gases that heavily absorb photons when at high concentration, O_2, O_3, NO_2, and H_2O. With careful band selection during the sensor's design, O_2 and H_2O gases can be considered as not interacting with photons (Mobley et al, 2016). For high concentrations of O_3 and NO_2, a contribution to the gas transmittance of 4-6% is expected and simulated using radiative transfer models (Ahmad et al., 2007). Non absorbing gases interactions are described in Wang et al. (2002) and are impacted by windspeed, surface reflectance effects and pressure of the atmosphere. The glint L_g is considered wind and viewing angle dependent (Wang and Bailey, 2001). Some sensors are tilted towards the west to avoid direct sun glint, but MODIS Aqua is not. Whitecaps will affect the $\mathbf{L_{wc}}$ radiance depending on their own surface reflectance and the fraction of surface area they cover (Gordon and Wang, 1994). Aerosols $\mathbf{L_a}$ and $\mathbf{L_{aR}}$ estimations are described in Ahmad et al. (2010).

Following observations from Morel and Prieur (1977), the water leaving signal is clustered into Case-1 phytoplankton dominated waters or Case-2 complex waters. The main hypothesis to process the aerosol retrieval considers that the ocean absorbs all light in the near infrared (750 – 1000 nm) and can therefore be considered black. This is called the "*black pixel assumption*" (Gordon and Wang, 1994). Hence, a light measurement at these wavelengths is solely linked to the atmospheric content and provides direct information of the light scattered by aerosols and air molecules. Relationships between light signals near 750 and 870 nm are usually used to correct the full visible spectrum from aerosols. Different concentrations of aerosols and relative humidity combinations are simulated into 80 different models for MODIS Aqua (Ahmad et al., 2010). The signal measured by the sensor is compared to the 80 models and

the closest model is selected to evaluate the aerosols and water vapour concentrations. If the black pixel assumption is wrong due to other constituents present in the water that produce a signal in the near infrared, an iteration process is conducted until aerosols can be estimated, otherwise the process is stopped and the pixel masked (Bailey et al., 2010). Whilst initially, 7 clusters have been proposed by Jerlov (1976), the community used the Case-1 and Case-2 clustering method for long and it is still appropriate to use nowadays in terms of physics. The open waters have a certain amount of Chl which lead the signal leaving the ocean, and the coastal waters have other constituents that alter it. More cluster have been proposed through the years (Jackson et al., 2017), up to 13 clusters recently (Spyrakos et al., 2018).

Exact details of the standard approach used for the MODIS Aqua sensor can be found in the NASA technical document written by Mobley et al. (2016). The atmospheric correction process is a general approach applied to correct a signal measured by a sensor. When it fails for various reasons such as measuring a cloud or being affected by too much glint, pixels are commonly masked. The most used atmospheric correction is the NASA one. More models have been proposed, and try to improve different parts of the process, such as making more accurate estimates under glint conditions (Steinmetz et al., 2011) or in coastal areas (Hu et al., 2000; Ruddick et al., 2002; Wang and Shi, 2007; Brockmann et al., 2016; Fan et al., 2017; Pahlevan et al., 2017; Liu et al., 2019; Pahlevan et al., 2021). While the Nasa standard atmospheric correction is well accepted as a general method in case 1 waters, there is no clear agreement on how to perform the atmospheric correction in case 2 waters. Turbid coastal waters with high sediment concentrations are known to be particularly problematic due to strong backscattering from mineral particles generating significant water leaving radiance in the red and near infrared.

Much of the research conducted by the ocean colour community during the past 25 years has focused on two main issues. First, finding a way to make an appropriate atmospheric correction, especially in case 2 conditions. Second, using the atmospherically corrected light signal to estimate the desired water constituent, especially chlorophyll. This is called the inverse problem and more details are provided below in Section 2.2.3. The work presented in this thesis challenges this paradigm. It will be shown in Chapter 3 that as soon as the TOA L_t signal is available, good quality estimates can be performed and atmospherically corrected signals are not required, at least in coastal waters. However, the whole process is still required in order to mask pixels where the signal does not seem to come from the ocean (such as clouds).

2.2.3. Inverse problem

The initial objective of ocean colour is to retrieve water constituents using a light signal. This is called the inverse problem because it consists of inverting the light signal into a constituent concentration. The common procedure requires access to the water leaving radiance obtained after the atmospheric correction process in remote sensing (Gordon et al., 1978; 1980; Smith and Baker, 1982). The water leaving radiance L_w is directly linked to the interactions of light and the constituents of the ocean. The ocean colour community developed the bottom of atmosphere remote sensing reflectance Rrs (O'Reilly et al., 1998), to generate estimates of constituents. By definition, the remote sensing reflectance is the ratio of the water leaving radiance L_w and the downwelling radiance E_d measured just above the sea level (Eq. 2.2). Rrs represents how much irradiance that reached the sea level went into the water, interacted with the constituents and was scattered back to the surface into a small solid angle ΔΩ. Rrs is a ratio of radiance and irradiance. The information carried depends on the solid angle and is therefore expressed in per steradian (sr^{-1}). See Figure 2.6 for details. The angle Θ is the angle between the centre of the solid angle and the zenith. The angle ϕ is the solar azimuth angle. Rrs is called an apparent optical property, as for a given viewing observation, its value will evolve with the constituents inside the water and will not be sensitive to sky conditions. Therefore, it is a good candidate to be able to evaluate the water constituents. Other reflectance signals exist and will be used in Chapter 3 and 4, they are all ratios of light and expressed in per steradians. The Rayleigh corrected reflectance Rhos is the ratio of light at the top of atmosphere when the Rayleigh signal L_r has been corrected (Eq. 2.3). The TOA reflectance Rhot is the ratio of light at the top of atmosphere, prior applying the atmospheric correction process (Eq. 2.4).

$$\mathrm{Rrs} = \frac{L_w}{E_d} \tag{2.2}$$

$$\mathrm{Rhos} = \frac{\pi.\left(\left(\frac{L_t}{t_{gsen} * t_{gsol}}\right) - L_r\right)}{F_o.\mu_o.t_{sen}.t_{sol}} \tag{2.3}$$

$$\mathrm{Rhot} = \frac{\pi.L_t}{F_o.\mu_o} \tag{2.4}$$

with F_0 the extra-terrestrial irradiance, μ_0 the cosine of the solar zenith angle, t_{gsen} and t_{gsol} the solar to sensor and surface to sensor gaseous transmittances, t_{sen} and t_{sol} the solar to sensor and surface to sensor diffuse transmittances and E_d the downwelling radiance at the sea surface. While Rhot or Rhos can be used for quasi true colour image generation, Rrs is the apparent optical property used for most of the ocean colour algorithms.

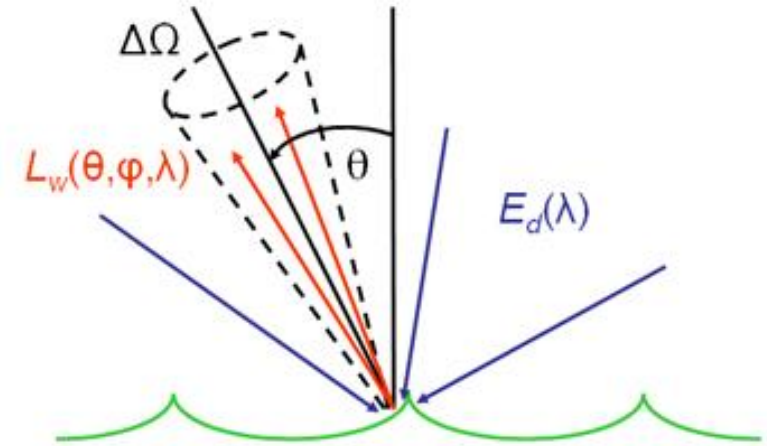

Figure 2.5: Diagram that show the downwelling light that reaches the ocean (E_d) from any direction above the surface and the upwelling radiance that leaves the ocean after interacting with its content (L_w) into a solid angle ΔΩ. Image from oceanopticsbook website, unchanged: https://www.oceanopticsbook.info/.

Different algorithms exist to convert the remote sensing reflectance into a desired water constituent concentration. For chlorophyll, the method requires a dataset of simultaneous retrievals of chlorophyll concentrations and light, called a matchup dataset. Polar satellites usually sample a region once a day at subtropical latitudes. For MODIS Aqua, the pixel resolution at nadir (the closest point on Earth) is approximately 1km². Public matchup datasets usually consist of a valid satellite Rrs signal and chlorophyll a concentrations made between the surface and 10 meters, in a +/- 3 hour interval, over a 3x3 pixels area and require a coefficient of variation to be lower than 0.15 for several specific wavebands (Bailey and Werdell, 2006). One important point concerning interaction of photons with the medium is that the absorption of light by the water is wavelength dependent, with a minimum absorption near 490 nm, and a maximum absorption in the red-infrared (red visible from Figure 2.2). The information carried by the 490 nm band led to the creation of the optical depth Kd_{490} which informs on the depth of the signal provenance (Preisendorfer, 1986). A common value for Kd_{490} found in coastal waters studies is 0.1 m^{-1}, which means that the light signal comes from 1/0.1 = 10 m. This 10 m value explains why most of the datasets, especially designed for coastal algorithms, use Chl data measured between the surface and a 10 meters depth. The light signal that goes deeper is not considered to be measured in coastal conditions, as it is mostly absorbed. The Kd_{490} value is highly dependent on atmospheric correction process, and performs poorly in coastal waters (McKee et al., 2007) and will not be used in this thesis. The signal

measured by a sensor is available at different wavebands, each containing information from interactions with the ocean down to a certain depth that varies with wavelength.

Once a matchup dataset is available, different algorithms can be applied. Historically, the ratio between the blue and green light has been used to retrieve chlorophyll a for CZCS (Gordon et al., 1980; Smith and Baker, 1982). Green and red ratios have been used for Chl measures in coastal waters (Hovis et al., 1977). In 1998, a massive study followed the launch of the SeaWiFS sensor in September 1997 by NASA. SeaWiFS was the first sensor equipped with red and near infrared bands which allowed development of a more advanced atmospheric correction. Efforts were made to collect data to build the main SeaWiFS algorithm used for open waters was developed early (O'Reilly et al., 1998; O'Reilly et al., 2000). The algorithm is called OC4, because it is an Ocean Colour algorithm using 4 bands. . OCX (x stands for the number of bands used) algorithms in general are also called blue green algorithms (Eq. 2.5). They use the ratio of single blue bands over a single green band, and empirical algorithms find the best fit possible to make an estimate of chlorophyll. When more than a single blue band is available, the maximum value between all blue bands available is used. The algorithm is of the form:

$$\log_{10}(\text{chlor}_a) = a_0 + \sum_{i=1}^{4} a_i \left(log_{10} \left(\frac{\text{Rrs}(\lambda_{blue})}{\text{Rrs}(\lambda_{green})} \right) \right)^i \quad (2.5)$$

with a_0 to a_4 = coefficients tuned for each sensor and where Rrs λ_{blue} is the maximum Rrs value between from the different blue bands used (443 to 510 nm), and Rrs λ_{green} is the Rrs near 550 nm. More details are provided by NASA at https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/ (accessed on 11 July 2022). The general idea is that the backscattering in the green part of the spectrum and the absorption in the blue part is directly linked to the amount of green pigment, chlorophyll a, and can therefore be estimated with measures at these two wavelengths.

MODIS sensors only contain two (443 and 488 nm) of the three (443, 490 and 510 nm) blue bands used initially for SeaWiFS. Including the use of the green band near 555 nm, the MODIS Aqua algorithm is therefore called OC3. Algorithms for more recent sensors use additional bands, up to six (OC_6 algorithm, O'Reilly and Werdell, 2019). Using more blue bands helps making better estimations for different concentration ranges of chlorophyll, with bands closer to 400 nm more accurate for estimations over oligotrophic waters (very low concentrations of Chl, <0.1 mg m^{-3}). The OC3 algorithm will be evaluated in Chapter 3 when applied to coastal waters. A NASA dataset called SeaBASS (for SeaWiFS Bio-optical Archive and Storage System) is freely available and used by NASA to build the different ocean colour algorithms

OCX. Neural networks (section 2.3) will be applied to this dataset in Chapter 4 and compared with standard algorithms. The SeaBASS dataset was used to tune the OC3 algorithm coefficients a_0 to a_4 mentioned above.

Similarly to the atmospheric correction problem with case 1 and case 2 waters returning different quality of light signals, the development of Chl algorithms is impacted by the presence of non-algal constituents as they alter the blue and green signals. OCX algorithms have been in use in case 1 waters since the past 40 years and made good estimates for almost 90% of the ocean where the light signal is dominated by the phytoplankton concentration. However, problems arise when the chlorophyll a is retrieved in coastal waters using OCX algorithms (Siegel et al., 2000). They tend to overestimate by several orders of magnitude the chlorophyll concentration due to presence of CDOM, sediments, proximity of the seabed and non-oceanic signals. In coastal waters, different approaches have been attempted to make more realistic estimations of chlorophyll a. Algorithms that use light information in the red part of the spectrum were developed to make more accurate estimations in coastal waters. Different algorithms exist (Gons et al., 2002; Dogliotti et al., 2015), mostly developed for the MERIS and OLCI sensors carried by Envisat and Sentinel-3 satellites. MODIS Aqua does not contain the appropriate red bands required to develop such algorithms, and red-green algorithms are usually preferred but developed at a very local scale (see the recent different algorithms developed for the Chesapeake Bay, Wynne et al., 2022). In very oligotrophic waters, it has been observed that OCX algorithms did not produce good performance which led to the development of the Colour Index algorithm (Eq. 2.6, Hu et al., 2012). It uses three bands and compare the reflectance measured in the green (555 nm) and the difference between a green linearly interpolated bands based on a blue (443) and red (670) band.

$$\mathrm{CI} = \mathrm{Rrs}(555) - [\mathrm{Rrs}(443) + \frac{555-443}{670-443} * (\mathrm{Rrs}(670) - \mathrm{Rrs}(443)] \quad (2.6)$$

The ocean colour community produced multiple algorithms that solve different water types, from the open ocean, oligotrophic waters or coastal waters. The blue-green algorithms produce good estimates in open waters; the CI algorithm produces good estimates when Chl concentrations are below 0.25 mg m^{-3}; and red algorithms produce good estimates in coastal waters, when available. The community hence developed classification techniques to decide when to apply a specific algorithm. Classification of the light is the main method used in ocean colour, as it originally came from the Jerlov observation of different light spectra (Jerlov, 1976). There are a numerous number of studies that tried to cluster the light signal prior to applying an appropriate algorithm to each (Jerlov, 1977; Morel and Prieur, 1977; Moore et al., 2009; 2014; Spyrakos et al., 2011; Vanterpotte et al., 2012; Melin and Vanterpotte, 2015; Neil et al., 2019).

When the signal is too complex to develop a physical algorithm, selecting appropriate specific algorithms is a possible solution. By constraining the range of conditions they will be applied over, it is easier to develop very specific algorithms that use only two or three bands when the light conditions are very similar. Developing a general algorithm which would use multiple bands at the same time is much harder and has not been achieved yet using empirical relationships. Classification is a strong tool and can also be applied to the atmospheric correction process (Liu et al., 2019; Shroeder et al., 2022).

The main algorithm used nowadays in the European shelf seas region was developed in 2002 to produce good estimates of chlorophyll in coastal waters using blue-green algorithms (Gohin et al., 2002). The algorithm is called OC5 and relies on both Rrs and L_w bands from 412 to 555 nm. For MODIS Aqua, it is basically the OC3 algorithm with inclusion of the 412 nm band to account for the absorption by CDOM in coastal waters and the inclusion of L_w 555nm to stand for the presence of sediments that do alter the green signal. The algorithm was trained with both open waters data (for the OC3 part) and coastal data from the English Channel. OC5 is based on pre-computed look-up-tables. The light signal gets a position attributed to the look-up-table which contains Chl concentrations between 0.01 and 65 mg m^{-3}. It can deal with relatively low CDOM and MSS concentrations (Gohin et al., 2002). When CDOM concentration are too high, it has been observed that the OC5 algorithm tends to overestimate Chl (Pittarch et al., 2016). Two version of OC5-like algorithms are available from different institutions, the Plymouth Marine Laboratory (PML) and ACRI-ST (in France). Both OC5 algorithms apply the CI algorithm described above (Eq. 2.6) when the pixel is classified as coming from oligotrophic waters. Each group provides its own version of the OC5 algorithm with its own look-up-tables. Important to note, both groups only provide the Chl concentration after application of the algorithm, and not the algorithm coefficients or look up tables used to inverse the Rrs into Chl. Only NASA algorithms come with simple formulae and coefficients and are therefore easily reproduced. Moreover, each procedure may apply extra quality controls and remove estimates in certain conditions that come with a complex final method, and transforming a light signal into an estimate requires the full knowledge of the method.

The original title and idea of this thesis was “Automate classification of ocean colour …", to improve current algorithm selection by selecting more appropriate algorithms using a machine learning classification algorithm. This was in line with general community thinking that progress would be best achieved by partitioning images into classified regions and applying region-specific algorithms. This is exemplified by Van Der Zande et al. (2019), where they merge OC5-PML with a red algorithm (Gons et al., 2002) using careful classification of each pixel. I initially reproduced their approach during the first year

of the project as a sanity check and because they did not share the final chlorophyll product as it was not available officially at that time. The main concern I had about their approach was the use of the red algorithm which requires bands that MODIS Aqua does not have access to (near 700 nm), and only MERIS (2002-2012) and OLCI (2016+) sensors are equipped with. The fact that other groups were already doing the same thing helped me to switch from reproducing what other ocean colour researchers were doing to become more "independent" and apply my own ideas to the problem, with the use of advanced machine learning techniques. In the end, what end users need is a chlorophyll (or other constituent) estimation, and the method used to achieve that is not particularly relevant so long as data quality can be assured. The ocean colour field seems trapped in a loop. Algorithms developed using *in situ*, BOA or in water radiometry struggle to perform when applied to satellite data that is poorly corrected for atmospheric effects. There seems to be little prospect of further substantial progress using this traditional approach. Here I present a new approach that aims to overcome both the atmospheric correction and water type-specific problems. With the use of neural networks (Section 2.3), it will be shown in Chapter 3 and 4 that it is possible to overcome these issues and directly invert the top of atmosphere signal into a chlorophyll a concentration.

In Chapter 6, the method developed in Chapters 3 to 5 is applied to the retrieval of the sea surface temperature (SST) from remote sensing. Temperature algorithms use long wave infrared bands, also called "brightness temperatures". They are emission of light from the Earth's surface based on its temperature, and only partially impacted by reflection of photons coming from the sun. As the main focus of this thesis is on Chl estimation, SST details will be provided in Chapter 6 only.

2.3. Multi-layer perceptrons

Neural networks are an increasingly commonly used tool due to their capacity to learn complex relationships and make accurate estimations or predictions. The neural networks used in this thesis are the simplest version invented originally to simulate the behaviour of real neurons, perceptrons. In his most famous publication in 1958, Rosenblatt invented the concept of perceptrons. It was an attempt at using mathematics to understand "*the capability of higher organisms for perceptual recognition, generalization, recall, and thinking*" (Rosenblatt, 1958). It ended with the creation of the perceptron, which connects different inputs to an artificial neuron with weights. The neuron is then capable of making an estimation of a desired variable. This concept was enhanced later in the 1980s by Rumelhart et al. (1985, 1986) and multi-layer perceptrons were created. They consist of layers of perceptrons connected

to each other, similarly to what occurs in the brain with neurons connected to the surrounding ones. The final target of a multi-layer perceptron neural network is to make an estimation of a specific variable, called the target, using different inputs. For that, an error is calculated between the output of the neural network and the target (a difference), and corrections of the weights are made to find the best fit possible. It is important to note that neural networks operate by minimising the global error (sum of errors for all of the training samples). In ocean colour, it has been observed that the colour of the ocean is linked to the concentration of its constituents. Typical inputs would be the different light bands available from a spectroradiometer, and the output would be the constituent altering the light signal, such as chlorophyll.

An example of a multi-layer perceptron is shown below in Figure 2.6, using three inputs (x_1 to x_3), two neurons (y_1 and y_2) on a single hidden layer that estimates a single output z_1. This output produces an estimate of the desired indicator, for example the Chl concentration. The inputs are different measurements available to estimate a variable. In this case they consist of measurements of light at different wavelengths, for example Rrs at 412 nm, 443 nm, etc. A training dataset consists of simultaneous measurements of inputs x and associated targets t_1. For this thesis, the main target t_1 is the Chl concentration, but other water constituents, salinity or temperature are also tested in different chapters. The links between the inputs and the neurons, or the neurons and the output (or other layers of neurons when more hidden layers are used) are called weights (w_{ij}). Initially, random values between specific ranges are given to each weight. The general concept of neural networks consists of finding the best weights combination to minimize the error between the estimation z_1 and the target t_1. To do so, two main steps are used to update the weights. The training and backpropagation processes. The training (feedforward) consists of making an estimation of the desired target from the inputs to the target. The backpropagation consists of correcting all the weights of the neural network based on the error observed between the estimation and the target. When the error correction stops improving, the training stops and the neural network can be applied to independent test data. For the example shown in Figure 2.6, three inputs x_1, x_2 and x_3 and the bias x_0 are connected to our two neurons y_1 and y_2 independently. The bias term value is always 1. The weights between the inputs and neurons are called w_{ij}. The subscript i gives the information to where the weights is connected for the next layer; 1 is a connection to the first neuron y_1. The subscript j give the information from where neuron is connected related to the input layer; j = 1 when it is related to the first input. The subscript 0 is always used for the bias unit of each layer. Neurons consists of application of an activation function g to the sum of the multiplication of inputs and their specific weights of the precedent layer. This sum is called net_1 for the first neuron y_1 for example. A single output is returned for each neuron (details in equations below). The selection of an activation function is

one of the decisions that has to be made, and one of the most used historically is the sigmoid or logistic function (Eq. 2.7).

$$S = \frac{1}{1+e^{-x}} \tag{2.7}$$

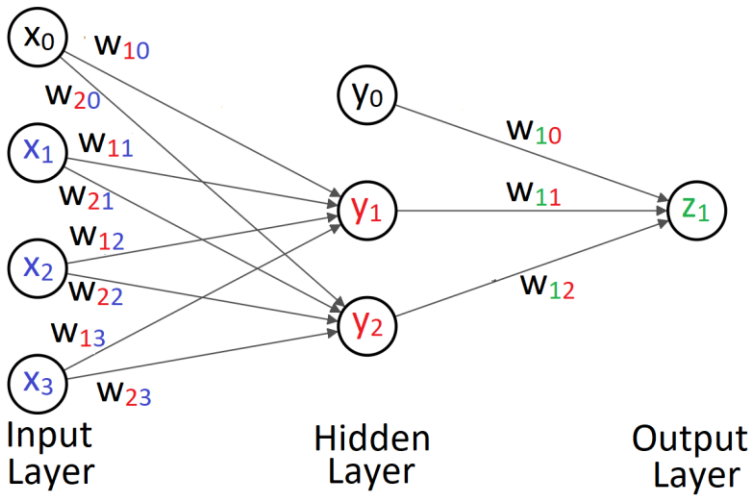

Figure 2.6: Example of a multilayer perceptron neural network using three inputs, two neurons on a single hidden layer, and one output. The bias term of the input and hidden layers are represented with subscript of zero.

2.3.1. First step: Feedforward

After an activation function g has been selected, the initial step consists of creation of random weights w_{ij}. The signal received by the top neuron y_1 is the application of the activation function g to the sum of the inputs coming from the previous layer multiplied by their specific weights, including the bias. The multiplication of the inputs $\vec{x}$ and associated weights $\overrightarrow{w_i^T}$ written as matrices is a dot product.

$$y_1 = g(x_0 w_{10} + x_1 w_{11} + x_2 w_{12} + x_3 w_{13}) \quad (2.08)$$

$$y_1 = g(net_1) \quad (2.09)$$

$$y_1 = g\left(\overrightarrow{w_1^T}\vec{x}\right) \quad (2.10)$$

$$y_2 = g(x_0 w_{20} + x_1 w_{21} + x_2 w_{22} + x_3 w_{23}) \quad (2.11)$$

$$y_2 = g(net_2) \quad (2.12)$$

$$y_2 = g(\overrightarrow{w_2^T}\vec{x}) \quad (2.13)$$

The same method is applied to the second neuron y_2. Once all neurons of a layer have been calculated, the next layer applies the same procedure. The neurons outputs y_1 and y_2 simply become the inputs of the layer that comes after, and the same operations are applied. For our example, the final output estimated by the network is:

$$z_1 = g(y_0 w_{10} + y_1 w_{11} + y_2 w_{12}) \quad (2.14)$$

$$z_1 = g(net_1) \quad (2.15)$$

$$z_1 = g(\overrightarrow{w_1^T}\vec{y}) \quad (2.16)$$

The final output of the neural network can be compared to the desired target and the performance evaluated. A cost function J is defined to measure the sum of the error between the estimation and the target for all training points available. The mean squared error function is usually used (Equation 2.17). For our example, we have only one output, hence k=1, but the general equation is:

$$J = \sum_1^k \frac{1}{2}(t_k - z_k)^2 \quad (2.17)$$

2.3.2. Second step: Backpropagation of the error

The objective is now to minimize the cost function *J* to make the best estimations z_1 of the targets t_1. For that, each weight has to be updated. The method used to reach the global minimum of the cost function (where errors between the target and output are minimum) is called the gradient descent (Eq. 2.18). The value for a specific weight w_{pq} of layer *p* and position *q* has to be updated (:= is not an equal sign, but a "update" sign). η is the learning rate (a constant) that needs to be defined.

$$w_{pq} := w_{pq} - \eta \cdot \frac{\partial J}{\partial w_{pq}} \tag{2.18}$$

For example, consider the weights connecting the output layer and the hidden layer, w_{1j}. We want to know how slightly changing each of the weights will affect the total error, $\frac{\partial J}{\partial w_{ij}}$. We need to calculate the impact of each of them to the total error estimation, $\frac{\partial J}{\partial w_{10}}$, $\frac{\partial J}{\partial w_{11}}$, $\frac{\partial J}{\partial w_{12}}$. For that, we use the chain rule of derivatives.

$$\frac{\partial J}{\partial w_{10}} = \frac{\partial J}{\partial net_1} \cdot \frac{\partial net_1}{\partial w_{10}} \tag{2.19}$$

$$\frac{\partial J}{\partial w_{11}} = \frac{\partial J}{\partial net_1} \cdot \frac{\partial net_1}{\partial w_{11}} \tag{2.20}$$

$$\frac{\partial J}{\partial w_{12}} = \frac{\partial J}{\partial net_1} \cdot \frac{\partial net_1}{\partial w_{12}} \tag{2.21}$$

The different terms need to be calculated. The first term is called the sensitivity of the output unit k, k=1 for our example because we have only one output. It is not trivial and the chain rule can be applied again to ease the calculation:

$$\frac{\partial J}{\partial net_1} = -\delta_1 \tag{2.22}$$

$$\delta_1 = -\frac{\partial \frac{1}{2}(t_1 - z_1)^2}{\partial net_1} \tag{2.23}$$

$$\delta_1 = -\frac{\partial \frac{1}{2}(t_1 - z_1)^2}{\partial z_1} \cdot \frac{\partial z_1}{\partial net_1} \tag{2.24}$$

$$\delta_1 = -\frac{2}{2}(t_1 - z_1) \cdot (-1) \cdot \frac{\partial z_1}{\partial net_1} \tag{2.25}$$

$$\delta_1 = (t_1 - z_1) \cdot \frac{\partial z_1}{\partial net_1} \tag{2.26}$$

The last term $\frac{\partial z_1}{\partial net_1}$ depends on the activation function used when designing the network. It can be generalised as:

$$\delta_1 = (t_1 - z_1)\frac{\partial g(net_1)}{\partial net_1} = (t_1 - z_1) \cdot g'(net_1) \tag{2.27}$$

With *g'* the primitive function of the activation function *g*. For the sigmoid function (Eq. 2.7), *g'* is:

$$g'(net_1) = g(net_1) \cdot (1 - g(net_1)) \tag{2.28}$$

Therefore, δ_1 becomes:

$$\delta_1 = (t_1 - z_1) \cdot (g(net_1) \cdot (1 - g(net_1)) \tag{2.29}$$

For the second terms of Eq. 2.19–2.21, the partial derivative of net_1 (the sum of the inputs and their weights associated) to each weight can be calculated:

$$\frac{\partial net_1}{\partial w_{10}} = \frac{\partial (y_0 w_{10} + y_1 w_{11} + y_2 w_{12})}{\partial w_{10}} = 1 + 0 + 0 \tag{2.30}$$

$$\frac{\partial net_1}{\partial w_{11}} = \frac{\partial (y_0 w_{10} + y_1 w_{11} + y_2 w_{12})}{\partial w_{10}} = 0 + y_1 + 0 \tag{2.31}$$

$$\frac{\partial net_1}{\partial w_{12}} = \frac{\partial (y_0 w_{10} + y_1 w_{11} + y_2 w_{12})}{\partial w_{12}} = 0 + 0 + y_2 \tag{2.32}$$

The terms in Eq. 2.19–2.21 become:

$$\frac{\partial J}{\partial w_{10}} = -(t_1 - z_1)\,(g(net_1)\big(1 - g(net_1)\big) \cdot 1 \tag{2.33}$$

$$\frac{\partial J}{\partial w_{11}} = -(t_1 - z_1)\,(g(net_1)\big(1 - g(net_1)\big) \cdot y_1 \tag{2.34}$$

$$\frac{\partial J}{\partial w_{12}} = -(t_1 - z_1)\,(g(net_1)\big(1 - g(net_1)\big) \cdot y_2 \tag{2.35}$$

Or, as a general form:

$$\frac{\partial J}{\partial w_{kj}} = -(t_k - z_k)\,(g(net_k)\big(1 - g(net_k)\big) \cdot y_j \tag{2.36}$$

Which gives access to the weight correction from the gradient descent function (Eq. 2.18):

$$w_{1j} := w_{1j} + \eta \cdot (t_1 - z_1) \cdot g'(net_1) \cdot \frac{\partial net_1}{\partial w_{1j}} \tag{2.37}$$

It can be simplified for any output neurons as:

$$w_{kj} := w_{kj} + \eta \cdot \delta_k \cdot y_j \tag{2.38}$$

and for hidden neurons, as:

$$w_{ji} := w_{ji} + \eta \cdot \delta_j \cdot x_i \tag{2.39}$$

These two main steps define how neural networks are trained, and obviously require a few constants and specific functions to be selected prior the training, such as the activation function or the learning rate. The training usually stops when the cost function value between two corrections does not evolve anymore and becomes smaller than a specific threshold several times in row. When it happens, the network reached convergence by "learning" the features that links the inputs and the target. The same neural network can be applied to test data for an independent test and should be able to make good estimations. When the network fails at converging, it usually produces uncorrelated estimations of the target. The main issue arises when the network overfits the input signals, by remembering the training data and performing poorly on the test data. In practice, the amount of data is usually limited and overfitting is the main problem in neural network development. Multiple solution attempts exist to try to counter overfitting, such as using different activation functions (Karlic and Olgac, 2011), removing some connections between neurons and their inputs (called the dropout method, Hinton et al., 2012), or using different architectures (Benardos and Vosniakos, 2007) etc. The field is very active and the best practices to follow to avoid overfitting continue to evolve.

While any dataset can be used as soon as it contains inputs and a target, good practices help the networks to produce more consistent and better quality estimates. These practices are of two kinds: pre-processing the data so it can be understood more easily by the network, and fine-tuning the neural network parameters, by using a specific activation function for example, or number of neurons and layers, etc. After conducting this thesis, the preparation of the data prior to the training is clearly more impactful than fine-tuning the neural network parameters. For example, using a specific activation function because it works best, or using one more hidden layer only improves the performance by a few percent. With the vast amount of parameters that more advanced neural network techniques can reach, fine-tuning all parameters could help reach better estimates, and there is probably some specific decisions that produce consistently more accurate results. However, the main improvement in this thesis was reached with a few pre-processing steps and by not limiting the dataset with preconcepts, such as not using information that has not been used before because nobody observed a relationship with the signal. Most of the previous work conducted in ocean colour using artificial neural networks limited itself by not using the full light spectrum for example. Another example of useful pre-processing of data, normalising the data in a manner that each input has the same "spread" ranging from 0 to 1 or log transforming the target when it

is log distributed clearly had more impact than decisions about network structure. These specific steps can also be fine-tuned, with different normalisation methods producing different results, with no exact method to follow (Cao et al., 2017). There is always potential for further improvement.

Neural networks are considered to be black boxes because visualising the different weights when high numbers of neurons and hidden layers are used is not comprehensible for humans. Even an oversimplified example as shown in this section is not easy to visualise even though it contains only 11 weights. The use of more and more neurons through time comes from observations that increasing the number of layers and neurons per layer improves performance achieved by a neural network for different tasks, which opened the development of what is now called the deep learning field (deep stands for neural networks that contain more than one hidden layer). Google recently released one of the most advanced neural networks for natural language processing which can contain up to 500 billion parameters (Chowdhery et al., 2022). Since the invention of the transistor (Bardeen and Brattain, 1948; Shockley, 1949), computational power continuously increased in the past 70 years, along with data availability, two main requirements for improved neural networks algorithms. The power of neural networks is such that it outperforms most humans or existing methods relying on physics on specific tasks for various fields, from recognition of tumours (Heckler et al., 2019) to creation of long protein structures (Jumper et al., 2021). Tasks from self-driving cars to prompt-to-draw algorithms are already available and artificial intelligence is slowly taking more and more space in our everyday life. In this thesis, the main task consists of a simple inverse problem of converting a light signal into an ocean surface indicator (e.g. Chl) using a minimally complex method. It is hoped that by providing detailed information on the construction of the neural networks and by keeping complexity of the approach to a minimum, this work might be helpful in persuading more of the community to accept the potential benefits of this statistical approach where the goal is purely to minimise a sum of errors.

2.3.3. Example of how to create neural networks in Matlab

Using Matlab, it is very easy to create neural networks as the default settings already contain most of the main methods used to return good performances, such as the min max normalisation or early stops of the learning process when the error does not evolve above a threshold (10^{-6}) six times in a row. In Chapters 3 to 6, the neural networks were built using the following method using the Matlab script shown in Figure 2.7.

```
input1N -> % Inputs (light signal notmalised)
targetN -> % Target (Chla, Temp) normalised if necessary
nbNeurons = 15; % Number of neurons per hidden layer
nbLayers = 3; % Number of layers
net = fitnet([repmat(nbNeurons,nbLayers,1)'], 'trainlm'); % Architecture creation
net.performFcn = 'mae'; % Error function to minimise
for l=1:nbLayers % Layer activation function
    net.layers{l}.transferFcn = 'poslin'; % poslin = ReLU
end
net.trainParam.showWindow = false; % false = not visible, true = visible
[net1,tr1] = train(net, input1N', targetN'); % Train the neural network
save([num2str(nbLayers),'L',num2str(nbNeurons),'NN_1','.mat'],'net1','tr1','PS1');
load([num2str(nbLayers),'L',num2str(nbNeurons),'NN_1','.mat']);
y1 = net1(input1N'); % Apply NN weights to normalised inputs
trInd = tr1.trainInd; % Train dataset (70%)
testInd = tr1.testInd; % Test dataset (15%)
% For figures to compare X and Y:
X = targetN;
Y = y1;
```

Figure 2.7. Example of how to train a neural network in Matlab.

- The number of neurons per hidden layer is set up using nbNeurons.
- The number of hidden layers is selected using nbLayers.
- The “fitnet” function creates the neural network architecture (using the specific number of layers and neurons per layer) with random weights. This is also where the Levenberg-Marquart algorithm is selected (‘trainlm’) which is used to update the weights during the training. The Levenberg-Marquart algorithm is an optimisation method widely used to solve nonlinear least squares minimisation problems, and more details can be found in literature (Ranganathan, 2004; Gavin, 2019).
- The performance function selected is the mean absolute error (default settings).
- The activation function used in hidden layers is the ReLu, called poslin in Matlab, because the value is always positive and linear (Y = X when X > 0, else Y = 0).
- The gui is hidden by setting showWindow to false.
- The training process to update the weights, applying the forward and backward processes mentioned in Sections 2.3.1. and 2.3.2. above is performed using the “train” function. The architecture used is specified under “net”; the inputs under “inputN1” and the target under “targetN”.
- Eventually, the trained network is saved in “net1” and contains the weights of each link between neurons. The index of the samples is saved under “tr1” and returns the sample group

(if it belongs to the training, validation or test set). The configuration used to normalise the data is also saved as it is required to preprocess the inputs and is saved under the "PS1" variable.

- Once a neural network is available, it can be applied to collocated data of inputs and targets (matchups). For this process, a previously neural network trained is loaded in Matlab using load function. The network weights are then applied to the inputs to make an estimate.

2.4. Performance metrics

When an estimate of an indicator has been done using an indirect method, using empirical models or neural networks for example, performance needs to be evaluated. Determining which algorithm returns the best performance requires the selection of one or more performance metrics, which has previously been shown to be non-trivial for ocean colour applications (Seegers et al, 2018). The Pearson Correlation Coefficient metric R (Equation 2.40) is a common statistical descriptor for assessing algorithm performance but is known to be impacted by density fluctuations in the distribution of the dataset. Seegers et al. (2018) recommended use of the Mean Absolute Error (MAE - Eq. 2.41) as being robust over several orders of magnitude, and as an absolute metric avoids being overly influenced by higher values. Even with a dedicated publication to statistical tools used to evaluate the different algorithms, there is no particular mandatory equation to be used among the various available. It is important to note that an "error" is by definition the difference between a measurement and what is expected theoretically. The different techniques to measure *in situ* samples are not theoretical, therefore the word "difference" is preferred over the word error in the metrics used for most of this thesis, and Eq. 2.41 therefore represents the Mean Average Difference (MAD). In Chapter 5, when a modelled dataset will be compared to estimates made by a neural network, the term "error" will be used. R and WAPD are unit less; MAD can be expressed in percentage or with the same units as the observed variable depending on the formula, RMSD and the BIAS are expressed with the same concentration as the observed variable,

The different statistical metrics to evaluate performance used even differ for each community of different fields, such as the sea surface temperature field. The problem is the same, comparing the estimates made between algorithms using a satellite light measurement, with a value measured by a method inside the ocean. For the temperature, the remote sensing community often use the root mean square deviation (RMSD, Eq. 2.42), the weighted absolute percentage difference (WAPD, Eq. 2.43) and the BIAS (Eq. 2.44). R is known to be sensitive to outliers in Chl algorithms, and the same behaviour is expected here, although it is still useful to observe it. Similarly, the RMSD showed good correlations in the

past with lower values produced when the algorithm is visually improved, but is expected to produce inconsistent results when outliers are present and should be considered with reservation. The MAD generally behaves well and is the main tool used in this thesis for Chlorophyll, CDOM, MSS and temperature estimations. It is robust enough to avoid being led by a few points of any range. An MAD of 1.8 means that there is a relative measurement error of 80%. A smaller MAE highlights a better prediction from an algorithm, and the best value possible is 1. The MAPD (mean absolute percentage difference) cannot be used for any indicator for which some observed data can reach zero (e.g. for salinity or temperature in Celsius) and would need to be removed. The weighted absolute percentage difference WAPD was developed to overcome the issue. It is the MAD divided by the mean, expressed as a percentage.

$$\mathrm{R} = \frac{\sum(Mi-\bar{M})(Oi-\bar{O})}{\sqrt{\sum(Mi-\bar{M})^2 \sum(Oi-\bar{O})^2}} \tag{2.40}$$

$$\mathrm{MAD} = \frac{\sum_{i=1}^{N}|Mi-Oi|}{N} \tag{2.41}$$

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(Mi-Oi)^2}{N}} \tag{2.42}$$

$$\mathrm{WAPD} = \frac{\sum_{i=1}^{N}|Mi-Oi|}{\sum_{i=1}^{N} O_i} \tag{2.43}$$

$$\mathrm{BIAS} = \bar{M}_i - \bar{O}_i \tag{2.44}$$

where *M* is the modelled value, *O* the observed value, *i* the i^{th} value, N the total number of values.

For this thesis, *M* is always the modelled measurement obtained from the algorithm and *O* the *in situ* or model output, used as the target for the neural network development. Chl can span over several order of magnitudes (<0.01 mg m^{-3} in the subtropical pacific gyres to >1000 mg m^{-3} in some estuaries). Following Seegers et al. (2018), both *M* and *O* are log transformed (Chl -> $\log_{10}$ (Chl)) prior to application of the different metrics.

2.5 Processing of the data for the thesis

To develop ocean colour algorithms, collocated data from a remote sensor and *in situ* samples are required. *In situ* data for Chl were collected from online sources (CMEMS, ICES websites), or directly provided by the different environmental agencies around the area, between 48°N and 65°N; 25°W and 13°E. The reason for selection of this area is to support comparisons with other COPERNICUS products

available in European waters from the CMEMS website. As mentioned earlier, Chl data are concentration of Chl from any method available, available between 0 and 10 meters, 08:00 and 16:00 (local), and between 2002 and 2020. The 0 to 10 meters depth limit come from the fact that most data are coastal (a few kilometres from the coast). The 08:00 to 16:00 time limit comes from the requirement for comparison against other Chl algorithm products available. The OC5-PML and OC5-ACRI products are daily concentrations of Chl, and do not provide temporal information. Therefore, limiting data available during daylight avoids getting matchups that are more accurate for the developed algorithm and poorer quality (due to a potentially increased temporal window) for daily products. The 2002 to 2020 limits comes from the MODIS Aqua availability required to create the matchup dataset.

The Aqua spacecraft equipped with a MODIS spectroradiometer was launched in May 2002 and the images are available since the 7th of July 2002 for the area. It produces approximately six images per day over the area, during daylight. MODIS Aqua data are freely accessible from NASA ocean colour website (https://oceancolor.gsfc.nasa.gov/, accessed on 11 July 2022). The list of all files available was downloaded from NASA ocean colour web browser. Level 0 data (counts of photons) measured by the sensor are transformed into Level1A radiances by NASA and the Level1A files are available on their server. A batch script in shell (.sh) was written to process the following steps. It reads each file available from the list of all files available, downloads it from NASA servers at Level 1A and process it using the SeaDAS software. SeaDAS is provided by the OBPG team (Ocean Biology Processing Group), in charge of developing remote sensing algorithms at NASA and can process data from Level 1 to Level 4 (temporal interpolations). The Level1A data are top of atmosphere radiances. Using SeaDAS, the "MODIS L1B" processor transforms Level1A data into Level 1B calibrated radiances by applying the different calibrations of each band. Level1B calibrated radiances are converted into the different reflectances products (Rhot, Rhos, Rrs) using the 'l2gen" process in SeaDAS. This process applies the standard atmospheric correction method which transform radiances into reflectances and brightness temperature products from 412 nm to 12,000 nm (therefore, a Level 2 product). To allow simple matchup creations, the Level 2 reflectances are transformed into Level 3 interpolated reflectances (reprojected onto a consistent grid) using subsequently the "L2BIN" and then "L3MAPGEN" processors. Four types of data are available after the process, Rrs, Rhos and Rhot reflectances, along with brightness temperature emissions in the infrared. In total, 15 bands in the visible and 4 bands in the long wave infrared were processed. Each image contains 51 layers of information (49 light bands, two coordinates) and a single file size is approximately 200 MB. Slightly less than 50,000 images were processed from July 2002 to January 2020. A single image takes approximately 15 minutes to be transformed from Level 1 to Level 3 by SeaDAS. The satellite pixels are

later matched with *in situ* Chl, temperature and salinity and used in Chapter 3, 4 and 7. The processing of the images was parallelised to help produce data faster, and took approximately 2 months on a dedicated Linux server.

For Chapter 5, the different datasets are built using a radiative transfer model called Hydrolight (Sequoia Scientific) and more details will be provided in the Chapter. Rrs are available along with Chl, CDOM and MSS for Chapter 5. For Chapter 6, data come from both the method described above and the SeaBASS dataset for MODIS Aqua sensor, available from the SeaBASS website (https://seabass.gsfc.nasa.gov/search/sst, accessed on 11 July 2022). It is a matchup dataset for sea surface temperature algorithm development and more details are provided in Chapter 6. MATLAB software was used for all processing, neural network creation and production of figures.

Chapter 3: Development of a neural network for chlorophyll a retrieval in European shelf seas using ocean colour

3.1. Introduction

European shelf sees contain both case 1 and case 2 waters, which limits performance of case-specific chlorophyll algorithms. State-of-the-art algorithms for this region or other similarly complex areas rely on water classification to select the right algorithm for each pixel (Liu et al., 2019), if available. This method involves addressing three component parts: 1) having an operational classification method, 2) having access to successful water type-specific algorithms, and 3) having similar performance of atmospheric correction over all kind of water classes. The main limit of the classification process happens when a pixel is classified between two or more clusters, which puts pressure on the thresholds that are used for the classification. Most methods tend to average values returned by different classes when such conditions are met (Hu et al., 2012, Zhang et al., 2019). The requirement for adequate atmospheric correction over the range of water types selected is also non-trivial. In many cases, apparent improvements in algorithm performance are achieved by elimination of problematic pixels through application of additional flags, as well as by altering the substance of specific algorithms. Whilst these procedures can improve performance metrics, it is done through reduction in data availability which limits spatio-temporal coverage.

Here I present an alternative approach. This chapter will focus on the development of a neural network to retrieve chlorophyll a from the MODIS ocean colour sensor for northwest European shelf seas, without any attempt to use classification to address the range of optical complexity found in these waters. Moreover, one of the major features of the NN approach highlighted in this chapter will concern the very permissive approach used for satellite data processing (Section 3.2.2.) involving minimal use of flagging to ensure maximum data availability as well as improvement in data quality. The second important characteristic of the NN approach will be the possibility to use uncorrected TOA signal, Rhot, to retrieve chlorophyll a (Sections 3.3. to 3.5.) therefore overriding the need for atmospheric correction. Finally, because chlorophyll concentrations are a spatio-temporal variable, the last section in the chapter will evaluate the impact of non-light signals on NN estimations. Performance limits of the NN approach will be discussed in Section 3.7. The majority of work presented in this chapter has been published in a very similar way in Remote Sensing journal (MDPI, https://www.mdpi.com/2072-4292/14/14/3353). Additional details that are briefly discussed in the publication are provided in this chapter.

3.2. Materials and methods

Building an ocean colour satellite chlorophyll algorithm requires co-located *in situ* chlorophyll a and satellite spectral information obtained with clear sky conditions that match within specified spatial and temporal windows. The common procedure (Bailey and Werdell, 2006) proposes to use a ±3 h window and a 3x3 satellite pixel area centred onto the matching pixel (the pixel's coordinate that matched with an *in situ* sample). It also requires a coefficient of variation between specific bands to be lower than 15% to valid a homogeneity test for the 9 pixels (with at least 5 of them available). In this project, the areas are very coastal, which invalidate a lot of matchups when 4 pixels only are available due to land presence. The homogeneity test can fail due to the gradients in the light signal from the shore to the open ocean, with higher turbidity near the coast in general. Moreover, the approach has to be compared with other products to be validated. These products are daily Chl values. Therefore, a different approach have been selected and is described below.

3.2.1. *In situ* chlorophyll a samples

The study area for this work is shown in Figure 3.1 and extends from 25°W to 13°E and 48°N to 65°N, including samples from the North Sea, Irish Sea, English Channel, and the western Baltic Sea. These are predominantly optically complex Case 2 shallow shelf seas with many areas presenting high sediment and or CDOM loads (Babin et al., 2003) that influence OCRS signals either persistently or seasonally / episodically. There are also stations, e.g. in the NE Atlantic sector or from the northern North Sea, Arctic and Norwegian area, that are deeper and further from land which would satisfy the Case 1 classification but are clearly under-represented with respect to our data set (Figure 3.1).

A Chl matchup data set covering the years 2002 - 2020 has been assembled from different sources of *in situ* samples: CMEMS (https://marine.copernicus.eu, accessed 10 March 2022); International Council for the Exploration of the Sea (ICES, www.ices.dk, accessed 10 March 2022); and data from countries included in the area that were directly provided by different institutions. For Danish marine waters, chlorophyll data were derived from the ODA database (DCE, 2021) and provided by the Department of Ecoscience, Aarhus University (Denmark). Data from the Norwegian and Barents Seas were provided by the Plankton Research Laboratory at the Institute of Marine Research (Bergen, Norway). For the Scottish waters and Stonehaven station, data were provided by Marine Scotland Science, Data from the waters of England and Wales were provided by the Centre for Environment, Fisheries and Aquaculture Science (CEFAS, https://www.cefas.co.uk/) and the Plymouth Marine Laboratory (PML,

https://www.westernchannelobservatory.org.uk/, last accessed 10th of March, 2022). The data set contains a mixture of Chl measurements produced using different methods (Jaccard et al, 2018), including: High Performance Liquid Chromatography (Claustre et al. 2004), fluorescence (Yentsch and Menzel, 1963), and spectrophotometry (Parsons and Stickland, 1963). In all cases samples with volumes typically ~1L will have been collected and filtered onto 25 mm GF/F glass fibre filters and frozen. The Chl pigments would generally have been extracted with 90% acetone and one of the methods specified above applied to measure their concentration using methods described in Section 2.1. Additionally, this data set includes data from *in situ* fluorometry. Such a diverse data set naturally suffers from a range of complicating factors including differences between *in vivo* and extracted Chl concentration estimates due to factors such as solar quenching and also due to practical constraints such as pigment extraction efficiency. Round robin exercises have previously demonstrated uncertainties in HPLC concentrations up to 40% (Sørensen et al. 2007) and this more diverse data set could easily demonstrate errors of 50% or more depending on the measurement conditions (Jaccard et al, 2018). Note that the CMEMS data set includes a large volume of data from ferrybox systems operating along the Norwegian coast. Unfortunately, this data set had to be eliminated from our analysis due to unresolved data quality issues. Other than these data from the Norwegian coast (approx. 60-65°N), no data were removed from the three data sets identified above. Approximately one million *in situ* Chl samples were available initially, but this number includes duplicates between different datasets, samples at different depths and Norwegian ferrybox data that were removed prior to averaging. The focus of this work is an attempt to establish satellite Chl products that provide equivalent quality data to that currently used by organisations like CMEMS and ICES. As such, we note that this validation data set is subject to unquantified and potentially significant uncertainty and that this should be considered in our analysis of satellite algorithm performance. Figure 3.1 represents the map of all *in situ* data available from different agencies, between 2002, July and 2020, January, the first 30 meters of the ocean, and prior any data cleaning (removal of poor-quality data or duplicates, temporal window applied or averaging).

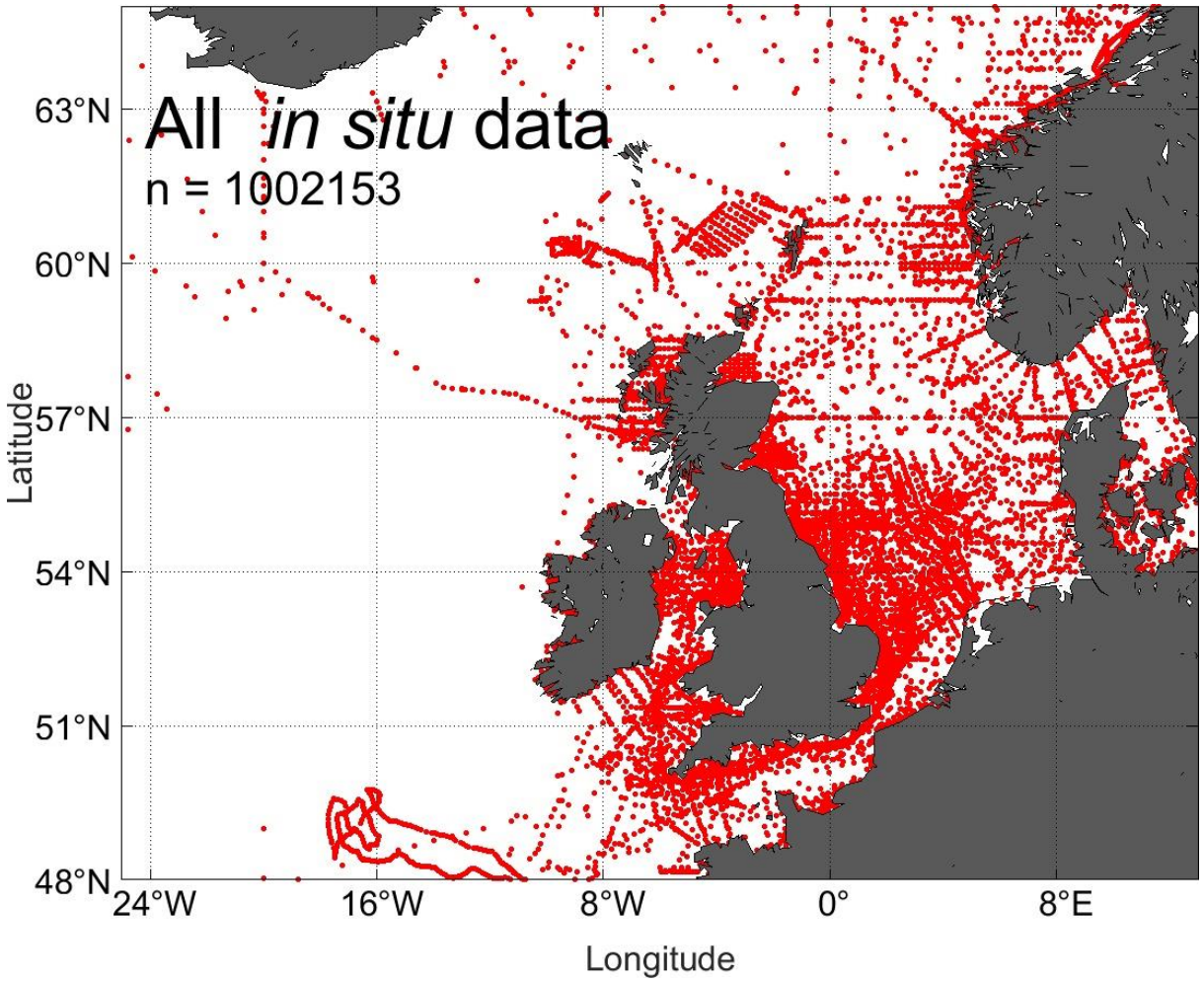

Figure 3.1: Map of all *in situ* samples available between 0 and 30m before data cleaning (removal of duplicates, averaging over the same stations, removal of Norwegian FerryBoxes).

3.2.2. Satellite data

The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Aqua spacecraft has produced images since early July 2002. For this study, all MODIS-Aqua images available between 48°N and 65°N, 25°W and 13°E during daylight from July 2002 to January 2020 were downloaded as L1A products from the National Aeronautics and Space Administration (NASA) ocean colour servers (https://oceancolor.gsfc.nasa.gov/cgi/browse.pl, accessed 10 March 2022), using the R2018 calibration. The Aqua satellite has an ascending node orbit crossing the equator at 13:30. The MODIS sensor, with a swath of 2330 km and a pixel resolution of ~1 km at nadir observes approximately 80% of the specified area each day. In order to maximise the information content for the NNs to operate on, the following bands were saved for this study: 412, 443, 469, 488, 531, 547, 555, 645, 667, 678, 745, 859, 869, 1240 and 2130nm. Their characteristics can be accessed from the NASA website

(https://modis.gsfc.nasa.gov/about/specifications.php, 10 March 2022). Bands 17-19 at 905nm, 936nm and 940nm were not processed due to their high correlation to cloud cover. Band 6 (1640 nm) has malfunctioned since 2009 (Wang et al., 2009) and therefore was not used for this study which leaves 15 bands. The inclusion of Bands 5 and 7 (1240 and 2130nm) follows the study of Wang and Shi (2007) who used these SWIR bands to perform enhanced atmospheric correction in coastal waters with MODIS Aqua. L1A files were downloaded, processed using l2gen and converted into L3 mapped files with a plate carrée projection using SeaDAS 7.5.1 following implementation of the NASA standard atmospheric correction using only the 'ATM FAIL' flag from the l2gen, with the "fudge option" set to 3. Images available from the same day were not merged in order to enable access to the temporal information and to provide optimal matchup conditions. This permissive approach, whereby flags that are usually applied by other data producers are not applied in this study, is intended to produce as broad a data set as possible in order to provide a test bed for assessing the potential for NNs to accommodate the most challenging optically complex waters. This has the added benefit of maximising the number of potential matchups which is the main limiting factor in NN development. While the method has already been described in Section 2.2, the version published in the publication (Hadjal et al., 2022) are kept below.

The total radiance measured at TOA by satellite sensors can be described as the sum of contributions from multiple physical effects:

$$L_t(\lambda) = L_R(\lambda) + L_a(\lambda) + L_{aR}(\lambda) + L_g(\lambda) + L_{wc}(\lambda) + L_w(\lambda) \quad (3.1)$$

with L_t the total radiance at TOA measured by the sensor. The terms on the right hand side of Eq. 1 are TOA radiances due to: L_R total Rayleigh scattering by air molecules, L_a scattering by aerosols only, L_{aR} aerosol-Rayleigh scattering, L_{wc} whitecaps and foam, L_g sun glint, and L_w the water leaving radiance. Three different reflectances were obtained as outputs of l2gen process. The TOA reflectance, Rhot:

$$\mathrm{Rhot} = \frac{\pi . L_t}{F_o . \mu_o} \quad (3.2)$$

The Rayleigh corrected reflectance, Rhos:

$$\mathrm{Rhos} = \frac{\pi . \left(\left(\frac{L_t}{t_{gsen} * t_{gsol}} \right) - L_r \right)}{F_o . \mu_o . t_{sen} . t_{sol}} \quad (3.3)$$

The BOA remote-sensing reflectance, Rrs:

$$\mathrm{R}_{rs} = \frac{\mathrm{L}_w}{\mathrm{E}_d} \tag{3.4}$$

with F_0 the extra-terrestrial irradiance, μ_0 the cosine of the solar zenith angle, t_{gsen} and t_{gsol} the solar to sensor and surface to sensor gaseous transmittances, t_{sen} and t_{sol} the solar to sensor and surface to sensor diffuse transmittances and E_d the downwelling radiance at the sea surface. While Rhot or Rhos can be used for quasi true colour image generation, Rrs is the apparent optical property used for most of the ocean colour algorithms.

The fourth-order polynomial ocean colour algorithm designed for case 1 waters for MODIS Aqua sensor (OC3M, O'Reilly et al., 1998) was applied to available Rrs from MODIS Aqua matchups following equations 3.5:

$$\log_{10}(\mathrm{chlor}_a) = a_0 + \sum_{i=1}^{4} a_i \left(log_{10} \left(\frac{R_{rs}(\lambda_{blue})}{R_{rs}(\lambda_{green})} \right) \right)^i \tag{3.5}$$

with $a_0 = 0.2424$; $a_1 = -2.7423$; $a_2 = 1.8017$; $a_3 = 0.0015$ and $a_4 = -1.2280$ and where Rrs blue is the maximum Rrs value between 443 and 488 nm, and Rrs green is the Rrs at 547nm.

3.2.3. Copernicus remote sensing chlorophyll a products used for comparison

A European Space Agency satellite product merging MODIS Aqua, SeaWIFS, MERIS, the Visible Infrared Imaging Radiometer Suite (VIIRS) and OLCI sensors created by the OC-CCI group is available for download on the CMEMS website. Rrs spectra of each sensor are realigned to MERIS wavebands at 412, 443, 490, 510, 560 and 665 nm and provides a daily merged and bias-corrected product for European shelf seas waters. It is available from 1998 to the present period. The NASA standard and polymer atmospheric corrections (respectively Gordon and Wang, 1994a; Steinmetz et al., 2011) are applied depending on the sensor and area of study. One surface Chl product is available for download on the CMEMS website based on this merged daily Rrs product, named OC5-CCI, and uses the OC5 algorithm (Gohin et al., 2002) in coastal waters. This algorithm was developed by IFREMER in collaboration with PML. It is available as a daily observed product and applies extra masking in certain conditions to avoid failure of the algorithm in coastal waters (see the latest Quality Information Document for this product, Pardo et al., 2021). This algorithm will be referred to here as OC5-PML.

Another product is available from CMEMS using a similar approach (OC5 algorithm), provided by ACRI-ST, known as the European Space Agency GlobColour project, and is available as a daily interpolated product, with a +/- 30 days sliding window to create "cloud free" surface Chl maps (Garnesson et al.,

2021). This version will be referred to as OC5-ACRI. Rather than using OC-CCI merged Rrs data, it averages chlorophyll from each sensor separately, and creates an averaged version from multiple sensors afterwards.

Both of these algorithms (OC5-PML and OC5-ACRI) use a similar method, applying the OC5 and OCI algorithms in optically complex and clear waters respectively, therefore relying on classification. The OC5 algorithm was initially designed with 5 wavebands at 412, 443, 490, 510 and 550 nm and was developed to work for complex Case 2 waters impacted by constituents other than Chl, such as Coloured Dissolved Organic Matter (CDOM) or Total Suspended Sediments (TSS, Gohin et al., 2002). The 412 nm band is used to take into consideration the CDOM absorption and the 550 nm band alone as a normalised water leaving signal to highlight the impact of sediments in water. It was initially trained with English Channel data. In open ocean waters, OC5 returns values close to the blue-green OC3 or OC4 algorithms (O'Reilly 1998). The Colour Index (CI) algorithm, developed by Hu et al., (2012), is a three band algorithm (443, 555 and 670 nm) used for oligotrophic waters with very low values of Chl (<0.2 $mg.m^{-3}$), where blue-green algorithms can lead to failure because of different effects such as the glint, stray light etc. It was trained with oligotrophic to mesotrophic matchups. CI is used if the value returned by OC5 is below a threshold, usually around 0.15 or 0.20 $mg.m^{-3}$. If the output falls between 0.15 and 0.2 $mg.m^{-3}$, an average between both OC5 and CI results is returned. OC5 products used by both algorithms were developed using their own specific LUTs and available wavebands. The OC5 algorithm cannot return values above 65 $mg.m^{-3}$ due to its design being based on LUTs not including values greater than 65 mg m^{-3}. OC5-PML (OCEANCOLOUR_ATL_CHL_L3_REP_OBSERVATIONS_009_067) and OC5-ACRI (OCEANCOLOUR_ATL_CHL_L4_REP_OBSERVATIONS_009_098) were downloaded from the CMEMS website in Aug 2020 and Feb 2021 respectively.

3.3. Matchups creation

The formation of matchup data sets between *in situ* Chl concentrations and satellite data must take into consideration both spatial and temporal components. For comparison convenience, the same conditions were used for all three products. The temporal window considered all *in situ* data sampled from the first 10 meters, between 8:00 and 16:00 GMT+0 as a candidate. If multiple *in situ* samples were available for a given pixel at different depths, the median value was saved as a unique pixel matchup. The OC5-PML and OC5-ACRI products were reprojected using the MODIS Aqua grid as a standard to allow comparison over the exact same pixels. Below are figures of available matchups for all three products and histograms associated.

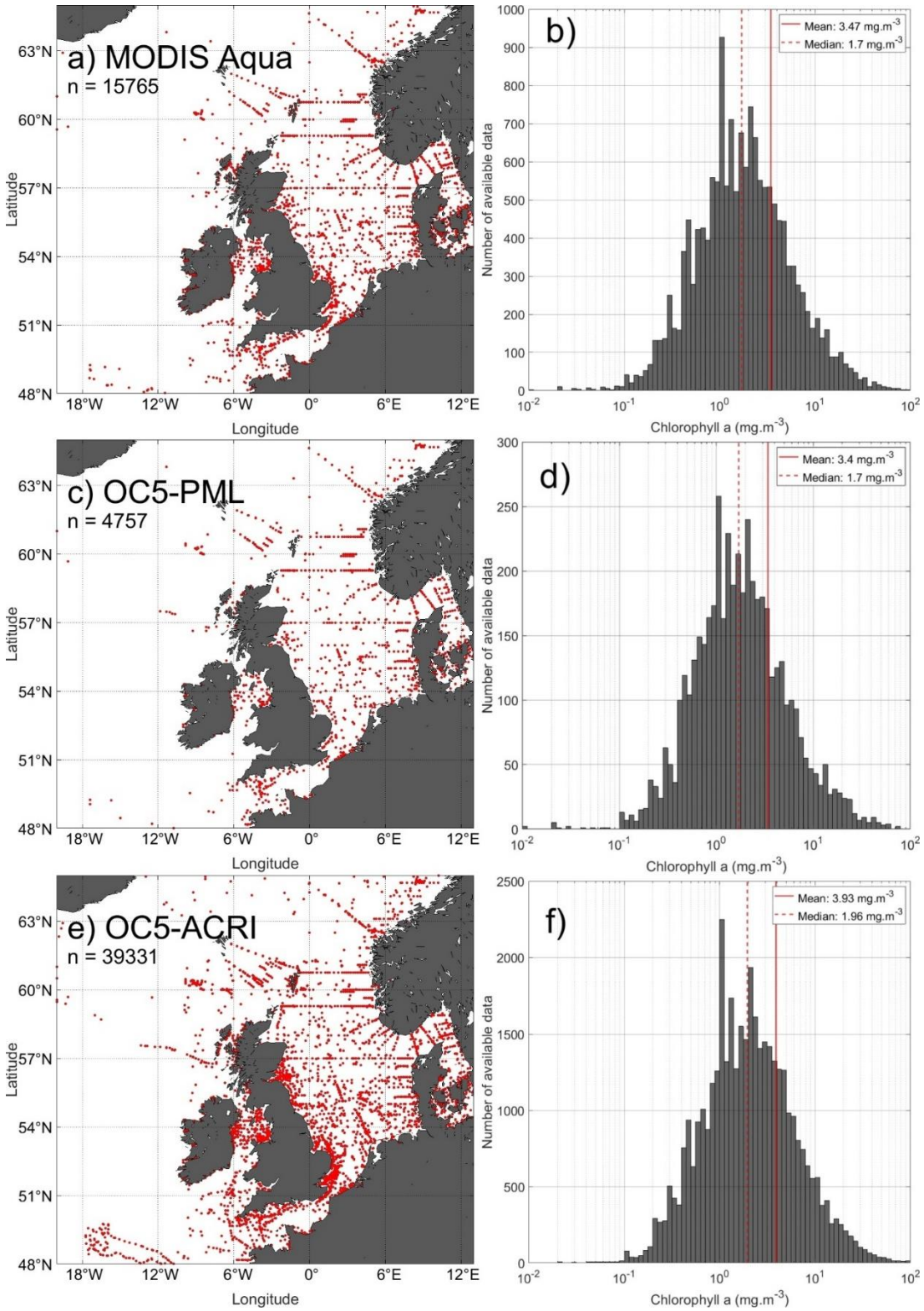

Figure 3.2: Map of all matchup points available for a) MODIS-Aqua sensor, c) PML product and e) ACRI product. July 2002 to January 2020, 0-10m, 08:00-16:00, exact pixel. b), d) and f) Histogram of the *in situ* samples associated to each product.

The final matchup matrix consists of geo coordinates of the satellite pixel's centre, the three different reflectances (Rhot, Rhos, Rrs) at the 15 wavebands, and the median of the 0-10m Chl averaged between 08:00 and 16:00. Data duplicates from different sources (ICES, MSS and CMEMS) were removed before calculation of median Chl values. The process has been repeated for the OC3, OC5-PML and OC5-ACRI algorithms, with different number of matchups available for each product (Figure 3.2). Final matchup numbers are determined by the details of processing for each algorithm tested later, but vary from 4757 for OC5-PML to 39331 for OC5-ACRI, with a total of 15765 matchups being available for the NN approach developed here using MODIS Aqua data.

The MODIS Aqua matchup data set is dominated by coastal waters, with the vast majority of the data sampled close to the coast. The final distribution is close to a normal distribution, with spikes at exactly 1 and 2 $mg.m^{-3}$ and a median of 1.7 $mg.m^{-3}$, which could come from sensor or human rounding. Over the 15765 matchups available for the MODIS Aqua sensor, 13246 are unique observations for different time and locations while 2519 matched at least two different MODIS Aqua images usually within a 1-hour interval due to the temporal window used and areas being seen twice by the satellite during the same day at these latitudes. These data were not merged and kept as unique matchups in order to add noise to the NNs as this has been shown to help NNs generalize (Jim et al. 1993, Buckton et al. 1997, Gross et al. 1999). There are approximately 1300 matchups per year between 2003 and 2006, 800 between 2007 and 2016 and less than 600 for each year afterwards. Seasonal coverage for the data set is not even, with approximately 1200-1400 matchups for months between February and September, less than 1000 for November and January, and less than 500 during December, mainly due to the increase in cloud cover during winter.

3.4. Neural network design

When designing a neural network, numerous decisions can impact time, complexity or reliability of the neural network. The main idea of this Chapter was not to fine tune NN for this specific task, but taking a general approach to the problem, using physics to make decisions and avoiding human impact. In NN development it is important to obtain an optimal architecture that produces good quality output data on both train and test sets and there is often a trade-off between the network complexity and the prediction accuracy. Selecting a small network structure may be computationally efficient, although this can lead to under-fitting, where there are too few connections to adequately resolve a complex signal. On the other hand, an overly big network is likely to be computationally inefficient and may introduce overfitting, whereby the network uses a complex curve to predict a simple signal. In this case, the network

may produce excellent results with the training set because it will remember the data set rather than learn from it, but it will give poor results with independent data. This is guarded against by testing the trained network against an independent test set. For this study, the whole process was conducted using Matlab R2020b's "fitnet" function from the deep learning toolbox to create the network and the "train" function to train the network, which is later applied to either the matchups or an image. We used the same number of neurons per layer each time (example in Figure 3.3), the Rectified Linear Units activation function for every node (y = x if x>0, else y = 0), and the Levenberg-Marquart function to minimise the error based on the Mean Squared Error. Scripts used in the production of data presented in this chapter are available from the link in the reference section, and figures of neural network is available from Section 2.3, Chapter 2.

One of the more significant challenges of constructing a successful NN for this application is the need to be able to operate over a wide range of Chl concentrations. Our data set extends over ~3 orders of magnitude. In order to spread weights more evenly across the data set, the target (Chl) was log-transformed (Chl becomes log10(Chl)) prior to training and the inputs (15 reflectances) were normalised between 0 and 1 using the min-max method (Eq.5). Log transforming the target both improves network performance and prevents the network from returning negative values that would be unphysical. Normalizing the inputs prevents the NN from relying too much on a dominant signal. Application of NNs to remote sensing images requires knowledge of the normalisation parameters used (the min and max values used for the normalization prior to training) and output values need to be back-transformed from log10 to obtain Chl concentrations.

3.4.1. Neural network architecture size

Various strategies can be adopted for finding optimal network architectures. It is reasonably common practice to take at least the same number of neurons as inputs and to evaluate the impact of adding more hidden layers. Because we do not have an overly large dataset and since even reasonably priced modern computers have good performance characteristics, we opted to systematically explore the impact of selecting different network architectures. MAD scores were obtained for NNs operating on Rrs, Rhos and Rhot inputs. In each case, we tested NN architectures varying between 1 to 10 layers and 1 to 20 neurons per layer. We repeated the application of each architecture 10 times. The median MAD value of the 10 runs applied to the test sets (15% of the total data set) is displayed in Figure 3.3. Similar performances are obtained with the training sets. There is a general tendency to obtain higher MAD scores for architectures using between 6 to 20 neurons per layer, and less than 5 layers. However, identifying a

truly optimal 'winner' for each input data type (Rrs, Rhos and Rhot) is probably not meaningful. Rather there are regions in this space where performance is broadly equivalent, and will be slightly different each time due to the effect of the random initialisation of the weights. In this case prediction results are similar once we have at least 6 neurons per layer, which could highlight that there may be elements of redundancy over the 15 bands available. For this study, we used networks composed of 3 layers of 15 neurons for the Rrs, Rhos and Rhot reflectances as they produce nearly optimal results without becoming overly computationally intensive. Choosing a higher architecture, say 8 layers of 20 neurons, was found to give better performance on the training sets (higher metrics) but poorer performance on the test sets. Differences in performance between test and training data sets can be a sign of overfitting, i.e. failure of the network to generalise. It is clear from Figure 3.3 that having too many hidden layers without enough neurons per layer generally degrades performance.

The heat maps in Figure 3.3 also reveal differences in the level of performance between NNs operating on different input sources. Interestingly, NNs operating on uncorrected TOA Rhot input data perform best, slightly better than Rhos, with Rrs showing the poorest performances. NNs operating on fully atmosphere corrected Rrs values produce higher MADs, though the differences do not appear to be very large and any of the three reflectance could still be used effectively. At first glance, it may seem surprising that uncorrected TOA reflectance inputs, Rhot, produce such apparently stronger results despite the atmospheric reflectance signal being present within the input. This will be examined in more detail later, but it should be realised from the outset that this effectively means that in this case the NN is having to account for the impact of atmospheric scattering and having to derive Chl for a wide range of coastal water types. On the other hand, Rrs and Rhos NNs are operating on reflectance data that is imperfectly corrected for atmospheric effects. As shall be highlighted later (Figure 3.5), both the full atmospheric correction and even the Rayleigh correction have potential to generate unphysical (negative) reflectance data, especially for turbid coastal waters. In these cases, the NN is effectively having to compensate for these AC errors and then derive Chl for optically complex waters. Taking this into consideration, it is perhaps less surprising that the TOA Rhot NN performs slightly better than the other two due to the loss on information during the AC process.

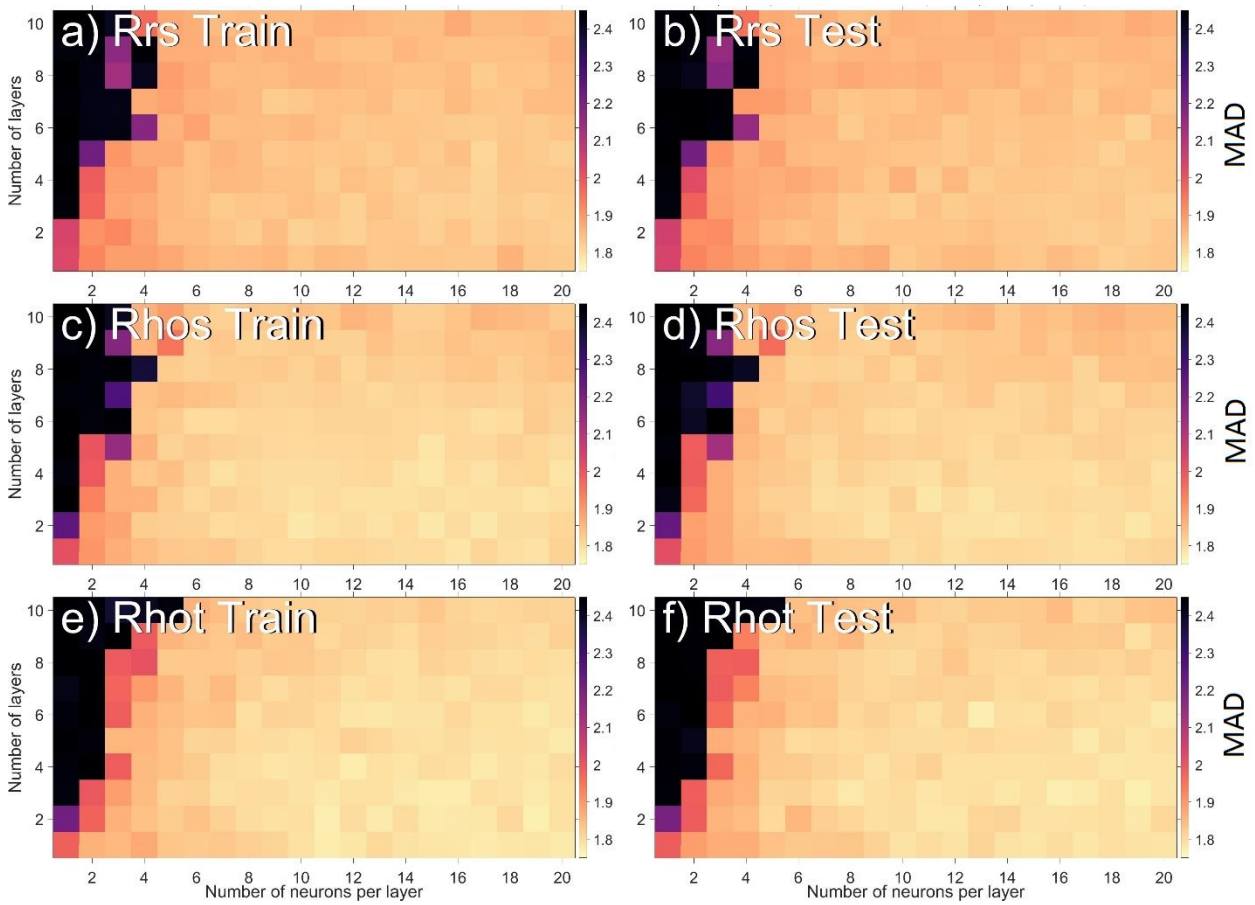

Figure 3.3: Mean Absolute Error (MAD) heat maps of each architecture possible for the train (left) and test (right) data set, operating on all three reflectances.

3.4.2. Selection of reflectance

Figure 3.4 shows the performances of the NNs for both the training (70%) and test set (15%) applied to the 15,765 matchups available, the last 15% (1,459 points) being used as the validation set (not shown). All three reflectances show similar performances. There is a slight tendency to overestimate low values (<1 $mg.m^{-3}$) and to underestimate high values (>10 $mg.m^{-3}$), possibly reflecting limits of representation in the training data set (not enough training data available for these ranges). Approximately 70% of the points fall between the 1:2 / 2:1 dashed line (a ratio of 2 between *in situ* samples and satellite estimation), close to the *in situ* error measurements for such a data set. The gap between training and test set performances is small, with Rrs showing higher differences than Rhot. Rhot achieves the best performance for the test set and at this stage is the candidate of choice for use in the rest of this

chapter. However, any of the reflectances could be used, as only some spectra were over corrected by the AC, and thanks to normalisation of data prior to training all show relatively good performance for a coastal data set.

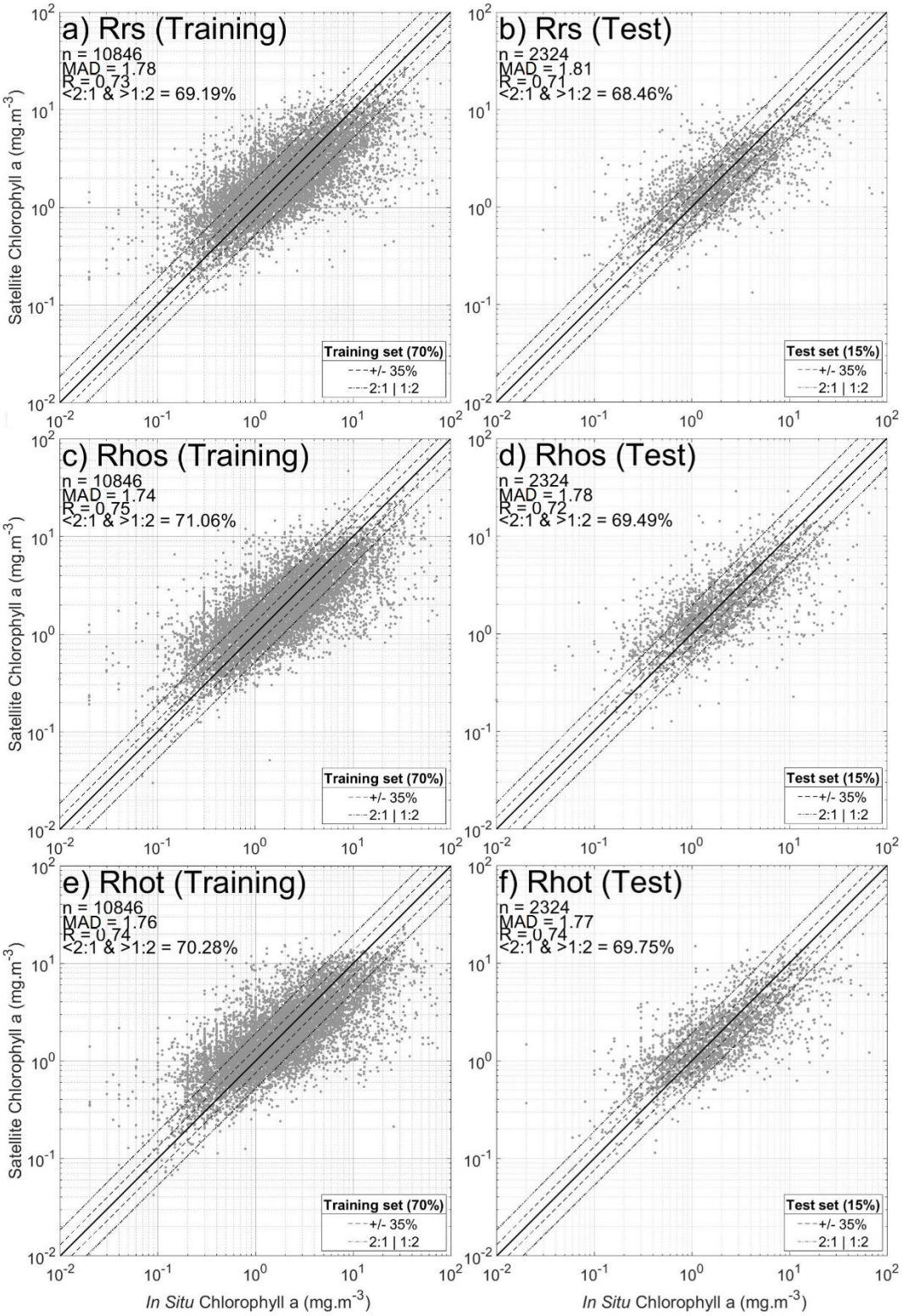

Figure 3.4: Neural network performances for an architecture based on 3 layers of 15 neurons each, using Rrs (a and b), Rhos (c and d), and Rhot (e and f) for the training set (70% of the total matchups) and the test set (15% of the total matchups) respectively.

To further support the choice of Rhot, Figures 3.5a, c and e show one example of a winter map that highlights the impact of atmospheric over-correction on NN performances for both Rrs and Rhos. It is visible as both dark patches some distance off the coast and as high values associated with image striping in the western North Sea. Features like these commonly occur in areas where turbidity is known to be high when using NN based on Rrs or Rhos, but are generally absent using Rhot. This can be illustrated by examining a subset of the available matchup spectra (Figure 3.5b, d and f). It is well known that the standard, black pixel atmospheric correction causes over-correction of Rrs spectra for turbid coastal waters, seen here most obviously as the occurrence of negative Rrs values (Figure 3.5b), but potentially being true even for non-negative data. Intriguingly, negative values are also found for a smaller number of Rhos data (Figure 3.5d), implying that even taking the preliminary step of applying the Rayleigh correction can sometimes be sufficient to produce unphysical data. None of these negative Rrs or Rhos values are realistic, and we suspect this is the main reason why the Rrs and Rhos NNs show slightly poorer performances, with the NN having to overcome this type of over-correction at the same time as determining the Chl signal. In contrast, Figure 3.5f shows TOA reflectance (Rhot) signals which are always physically plausible, even if they are still obviously impacted by the contribution of atmospheric scattering. MODIS Aqua striping effects seem to be more visible in winter compared to other seasons, which may highlight a link with high solar zenith angles and a need for removal prior to application of any algorithm (Bouali and Ladjal, 2011). Our results suggest that it is marginally easier for the NN to handle the uncorrected atmospheric scattering signal in Rhot than it is to undo imperfect atmospheric correction of Rhos and Rrs signals. The combination of higher statistical metric performances, the training and test set performances being closer, and observation of unrealistic oceanic coastal features possibly due to the failure of atmospheric correction all lead to the choice of Rhot as the preferred data source for NN development.

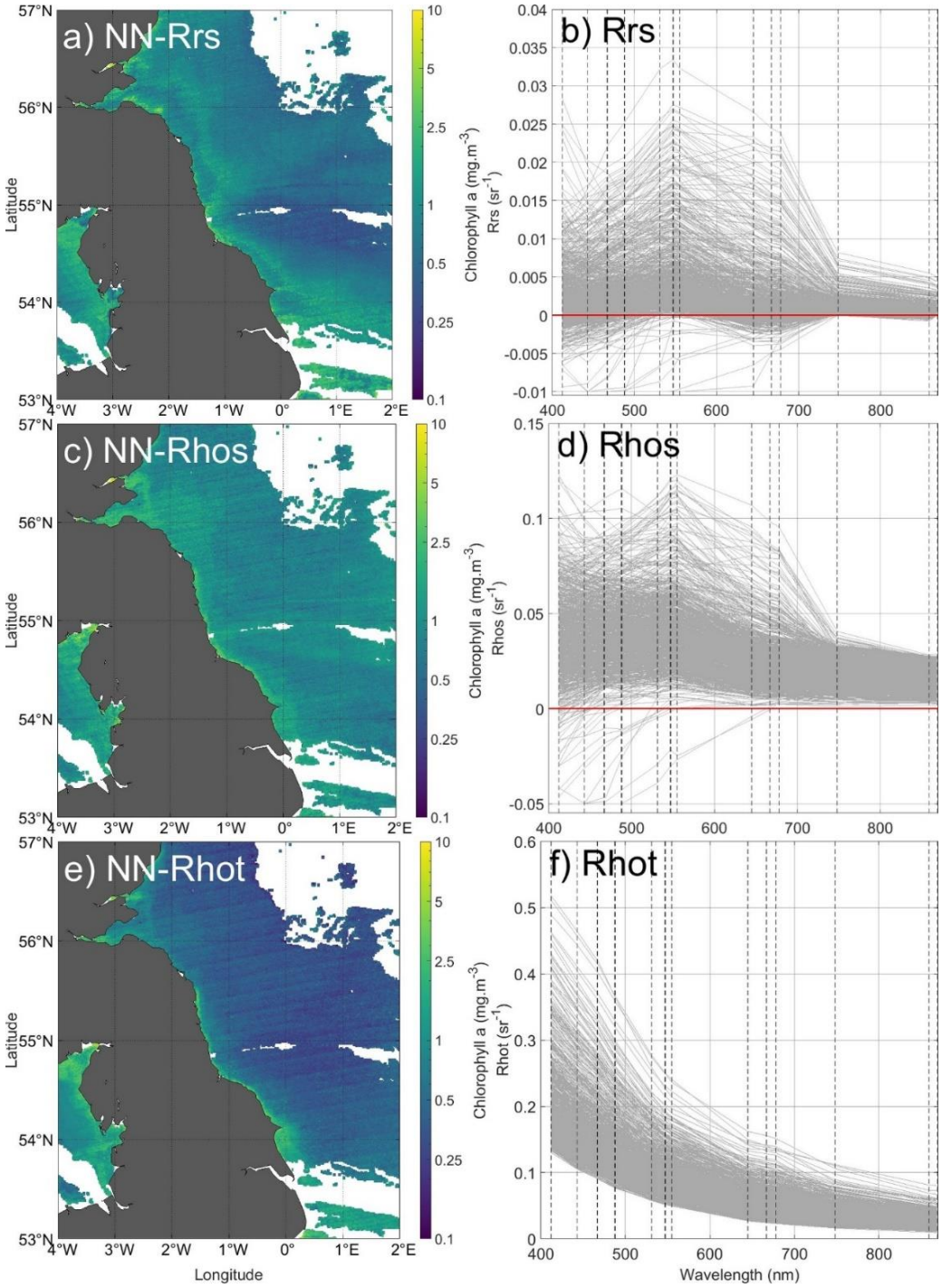

Figure 3.5: Winter image of the 31th of December, 2019, 13:05 from MODIS-Aqua, highlighting the atmospheric correction noise present near the Firth of Forth, western North Sea for the Rrs a) and Rhos c). Absent from the Rhot product e). 1000 spectra examples from the matchup data set, for Rrs b), Rhos d) and Rhot f). Only the first 13 of the 15 bands of MODIS Aqua available for this study are displayed with dashed lines.

3.4.3 Ensemble approach

The nature of NNs is such that, each time a network is trained, it will produce a network that is specific to the training data set employed. Randomly re-sampling the available training data generates subsequent NNs with properties that are not exactly the same each time. Figures 3.6a and 6c display the results of two NNs using Rhot and that return similar yet slightly different performances. These differences are even more apparent from their respective images (Figure 3.6b and 3.6d) for the 23rd of July, 2019 at 12:25 from MODIS Aqua, where a coccolithophore bloom near 57°N and 0°W is retrieved differently by the two NNs (see Figure 3.7 for example). This discrepancy is largely due to the fact that coccolithophore blooms are underrepresented in our matchup data set and the performance of the resulting NNs is therefore heavily dependent on how many such stations are included in the associated training data sets. Elsewhere the two NNs produce images that are visually very similar. The obvious solution to this problem is clearly to attempt to expand the training data set. However, this is impractical in the short term. Instead, we must look to develop an approach that is more robust for any given training data set.

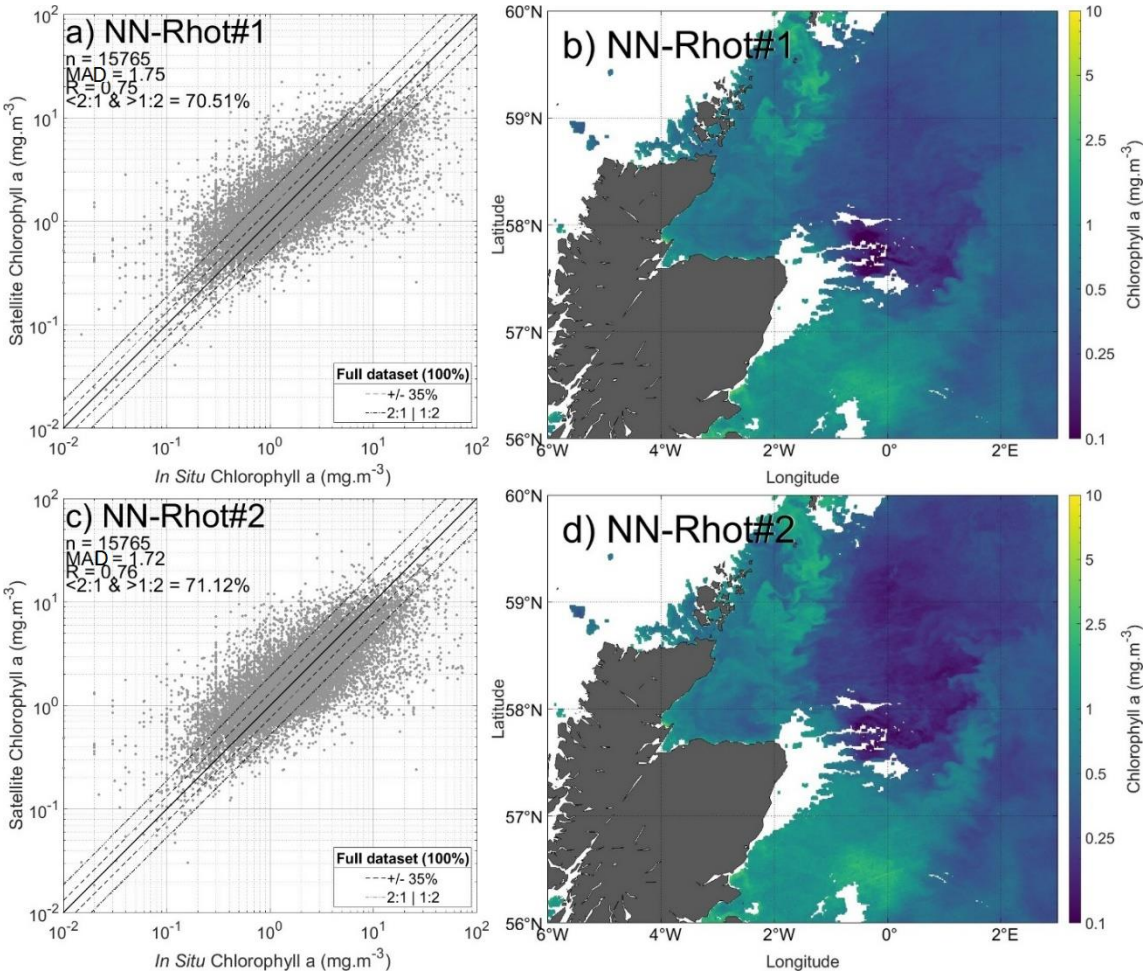

Figure 3.6: Two examples of the same architecture of a Rhot neural network using 3 layers of 15 neurons showing slight differences. a) and c) Neural network performances for the whole matchup data set. b) and d) Image of the 23rd of July 2019, 12:25 from MODIS-Aqua (same as image 8). Notice the difference around 0°E and 57.5°N coming from a coccolithophore bloom returning very low Chl values.

Figure 3.7: Brightness enhanced quasi true colour image showing the coccolithophore bloom for the 23rd of July 2019 captured by Sentinel-3A OLCI sensor at 10:06 am.

To this end and to minimise the impact of under-sampled features in our training data set such as coccolithophores blooms, we decided to use a standard NN architecture (3 layers of 15 neurons) but resample the training data set multiple times, generating multiple NNs that could subsequently be analysed to produce a single, hopefully convergent, median data product. This approach has the further merit of being able to provide a measure of product uncertainty through the standard deviations of the resulting distributions of Chl values for each pixel. Conversely, the computational effort involved needs to be considered if the approach is to be used in an operational sense for image processing, so establishing an optimal number of NN iterations is essential. Figure 3.8 displays median values for the same image of the 23rd July 2019 using 10 (Figure 3.8a) and 100 (Figure 3.8b) networks. Visual inspection of these panels (and others – not shown - representing different numbers of iterations) suggest that an ensemble approach using the median of 10 NNs is sufficient to achieve convergence with a version merging 100 NNs. Figures 3.8c and 3.8d show corresponding relative standard deviations from the 10 and 100 iteration networks respectively, expressed as percentages relative to the median (rather than the mean). Again, there is broad consistency between these two images suggesting that 10 iterations is sufficient to capture

the performance of the NN approach. The standard deviation of NN outputs varies across the image, reflecting variable confidence in NN output for each pixel. Figure 3.8e shows that this percentage uncertainty varies from ~7% to more than 100% with the vast majority of data falling between 10 and 30%. Looking at a worst case scenario, Figure 3.8f shows the impact of this uncertainty for a randomly selected single point inside the coccolithophore bloom near 57°N and 0.2°W where NN performance is worst, revealing up to an order of magnitude uncertainty (up to 100% rSTD errors).

This approach has the further merit of being able to provide a measure of product uncertainty through the standard deviations of the resulting distributions of Chl values for each pixel. Conversely, the computational effort involved needs to be considered if the approach is to be used in an operational sense for image processing, so establishing an optimal number of NN iterations is essential. Figure 3.8 displays median values for the same image of the 23rd of July 2019 using 10 (Figure 3.8a) and 100 (Figure 3.8b) networks. Visual inspection of these panels (and others – not shown - representing different numbers of iterations) suggest that an ensemble approach using the median of 10 NNs is sufficient to achieve convergence with a version merging 100 NNs. Figures 3.8c and 3.8d show corresponding relative standard deviations from the 10 and 100 iteration networks respectively, expressed as percentages relative to the median (rather than the mean). Again, there is broad consistency between these two images suggesting that 10 iterations is sufficient to capture the performance of the NN approach. The standard deviation of NN outputs varies across the image, reflecting variable confidence in NN output for each pixel. Figure 3.8e shows that this percentage uncertainty varies from ~7% to more than 100% with the vast majority of data falling between 10 and 30%. Looking at a worst case scenario, Figure 3.8f shows the impact of this uncertainty for a randomly selected single point inside the coccolithophore bloom near 57°N and 0.2°W where NN performance is worst, revealing up to an order of magnitude uncertainty.

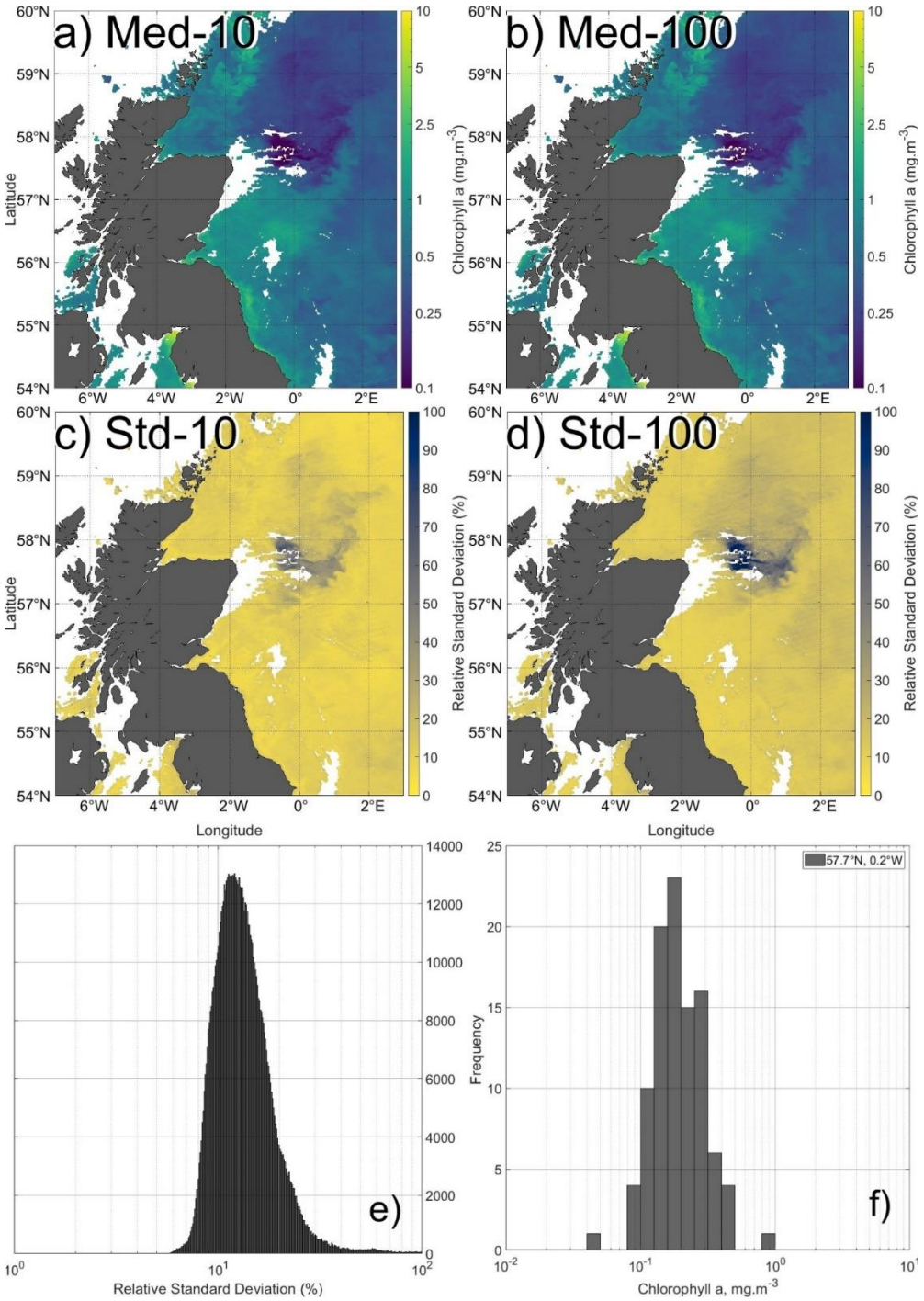

Figure 3.8: Chlorophyll estimation from the neural networks using a) 10 networks and b) 100 networks averaged (median). Relative standard deviation (relative to the median) for the same image expressed in percentage using c) the same 10 or d) 100 networks. e) Histogram of the relative standard deviation of the panel d) using 100 networks expressed in percentage. f) Histogram of a random point inside the coccolithophore bloom.

3.5. Neural network algorithm performance evaluation

3.5.1. Comparison with other algorithms

In order to establish a baseline for performance evaluation with current well established Chl algorithms, Figure 3.9 shows results for all the available matchups for a) OC3, b) OC5-PML, c) OC5-ACRI and d) MODIS Aqua data processed using the median of 10 NNs with the same 3 layers of 15 neurons architecture, using Rhot. The OC3 algorithm was applied to MODIS Aqua matchups Rrs obtained using the permissive approach and therefore shows a massive spread between measured and retrieved Chl, with a strong bias towards over-estimation. The OC3 algorithm was not designed for optically complex coastal waters, and initial NASA development of this algorithm relied on application of masks to eliminate pixels with obvious data quality issues; see https://oceancolor.gsfc.nasa.gov/atbd/ocl2flags/ to see when specific flags are applied at level 2 or 3 or the data. While some data follow the 1:1 line, most of the matchups are overestimated by the algorithm by varying amounts but up to by several orders of magnitude. This is likely due to the influence of turbid waters on both atmospheric correction performance and OC3 algorithm performance, which usually return very high Chl values over sediment plumes on images. This is not surprising as the algorithm was developed for clear open ocean waters and is not expected to perform well in turbid coastal waters, where a good proportion of the available matchups come from. Of course, it is also worth noting that this algorithm (and others with similar structure and performance in these waters) remains in common use by unwary end users who perhaps have less familiarity with the field, and who would potentially benefit from more robust guidance by data providers.

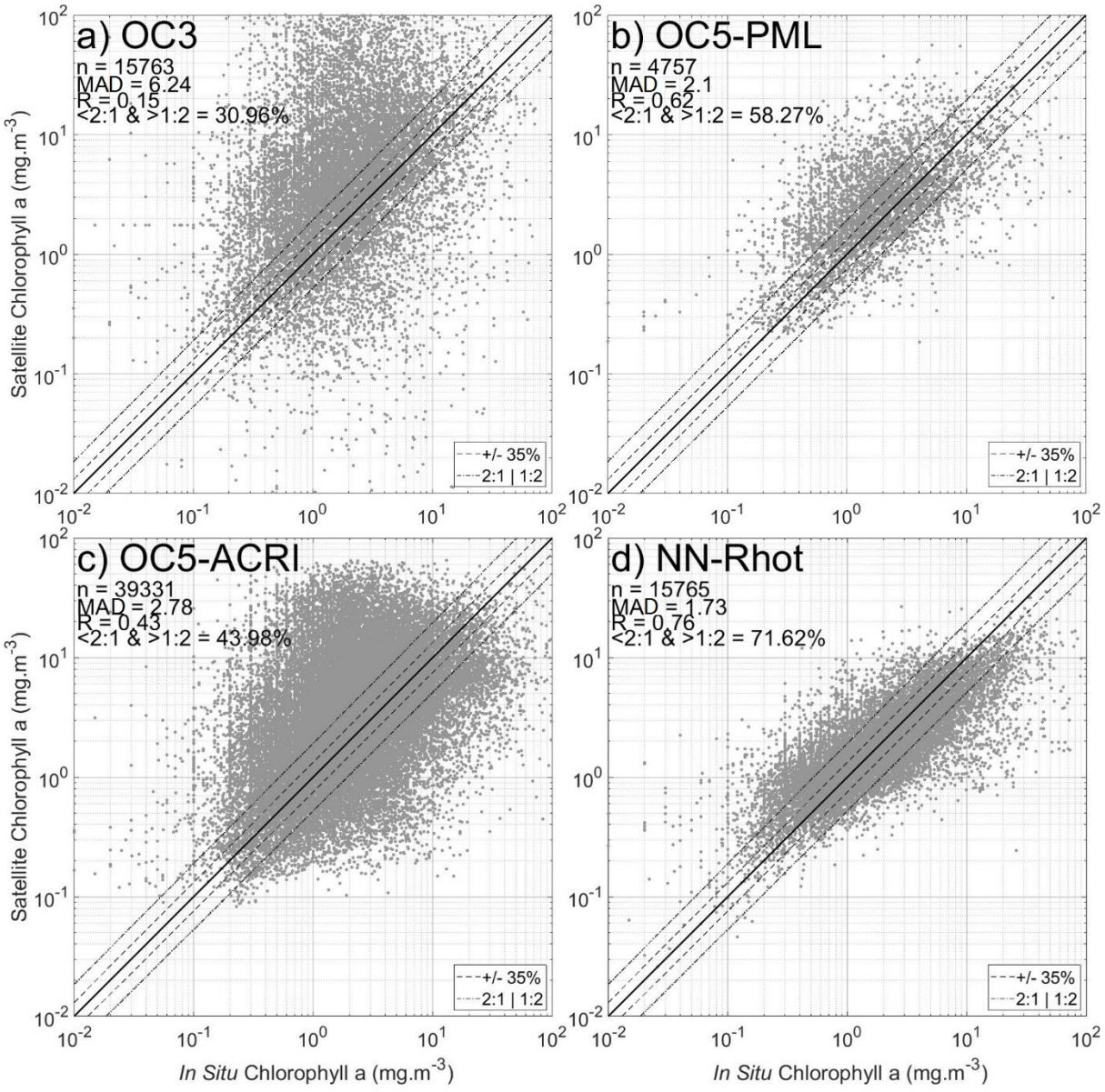

Figure 3.9: Performances over the matchup dataset for a) OC3, b) OC5-PML, c) OC5-ACRI and d) NN-Rhot, the median of 10 random neural networks using Rhot with MODIS Aqua data.

The OC5-PML algorithm shows a clear improvement compared to OC3, with overestimation restricted to a maximum of ~1 order of magnitude for Chl between 1 and 10 $mg.m^{-3}$ and possibly a tendency towards underestimation at high concentrations. Importantly, only 4757 matchups were available for this product despite being based on the accumulation of data from between 2 and 5 satellites at any given time. This significant reduction in data availability comes from: 1) the OC5 algorithm itself which can only be applied in certain conditions (based on the signals at 412 and 550 nm), which removes a significant number of potential matchups; and 2) additional quality control flags which exclude more

problematic waters such as coccolithophore blooms, very coastal or shallow waters, glint impacted areas, low sun angles, etc. It should also be noted that the maximum value allowed by the OC5 algorithm is 65 $mg.m^{-3}$, which can be problematic for coastal waters where higher concentrations are possible.

The OC5-ACRI product (Figure 3.9c), despite using a broadly similar algorithm to OC5-PML, tells a different story, with many more matchups being available, but with significantly greater ranges of over and under-estimations. The increase in data availability is directly due to use of a +/- 30 days sliding window average rather than single direct observations. It is likely that the apparently stronger performance of the PML variant is achieved through use of additional flagging to remove poor quality data rather than actual improvement in algorithm performance *per se*. Conversely, while the OC5-ACRI product has a clear advantage of having almost 100% coverage for the area except in winter, increasing by ~3 times the number of available matchups, this appears to be achieved at the price of data quality. There is clear potential merit in using this type of merging approach at global scales that are dominated by case 1 waters where the algorithm may perform well, but these results suggest that there may be significant issues in coastal regions.

The proposed ensemble NN-Rhot product (Figure 3.9d), offers several advantages. The distribution is clearly better constrained towards the 1:1 line than either OC3 or the OC5-ACRI product (Figures 8a and 8c), and is somewhat tighter than the OC5-PML product (Figure 3.9b) with R of >0.75 vs 0.61 and MAD of <1.8 vs 2.1. The NN approach produces more than twice the matchups available for OC5-PML (13246 “daily” matchups for the permissive MODIS Aqua approach against 4757), and it does not require application of further flags to eliminate optically complex waters or outliers to reach similar performances. Very importantly, the NN performance is achieved without requiring the application of any atmospheric correction, even though this contains a wide variety of optically complex water conditions. The NN product is far from perfect and there is evidence of a tendency to over-estimate at low concentrations and vice versa, with the range of error remaining at approximately one order of magnitude on a per pixel basis. However, two thirds of the NN data lie between the 1:2 and 2:1 lines, broadly reflecting the level of performance that can be attained for *in situ* measurement of Chl using the diverse methods used to generate the training data set. It is worth noting that both the PML and ACRI algorithms could potentially return higher performances if retrained using this specific matchup dataset. That said, the NN approach appears to offer a useful combination of high quality performance and maximal data availability.

3.5.2. Restricting the comparison to the exact same matchups

To avoid data availability bias between the different products used above as comparators, we have repeated the analysis but this time have restricted the comparison data set to matchups that are available for all four algorithms. As the most restrictive algorithm examined here, this second data set is largely constrained by the flagging procedures adopted by the OC5-PML algorithm. However the process of establishing clean matchups for each algorithm means that there are fewer common matchups than were originally available even for OC5-PML. Therefore, only samples commonly available for OC5-PML, OC5-ACRI and the permissive MODIS Aqua dataset are shown in Figure 3.10. The vast majority of OC5-ACRI matchups following this approach should come from direct observations due to the requirement of data being available from the MODIS Aqua and OC5-PML daily observations. 3896 points met the requirements of being available from all products at the same time. Performances of OC3 and OC5-ACRI products are improved (with a very similar distribution). While the OC5-PML algorithm returns similar metrics (MAD of 2.09), the NN-Rhot approach benefitted from flagging data following OC5-PML approach with an improved MAD below 1.7. For reference, the MAD obtained for case 1 waters algorithms (GSM, OCI, OC3) using the NOMAD dataset reach ~1.6 (Seegers et. al., 2018), with best and worst performances reached over oligotrophic and eutrophic waters respectively from the GSM algorithm (MAD of 1.47 and 2.05). It is notable that the NN approach developed in this study appears to achieve performance metrics for optically complex coastal waters that are comparable with standard algorithm performance in Case 1 waters.

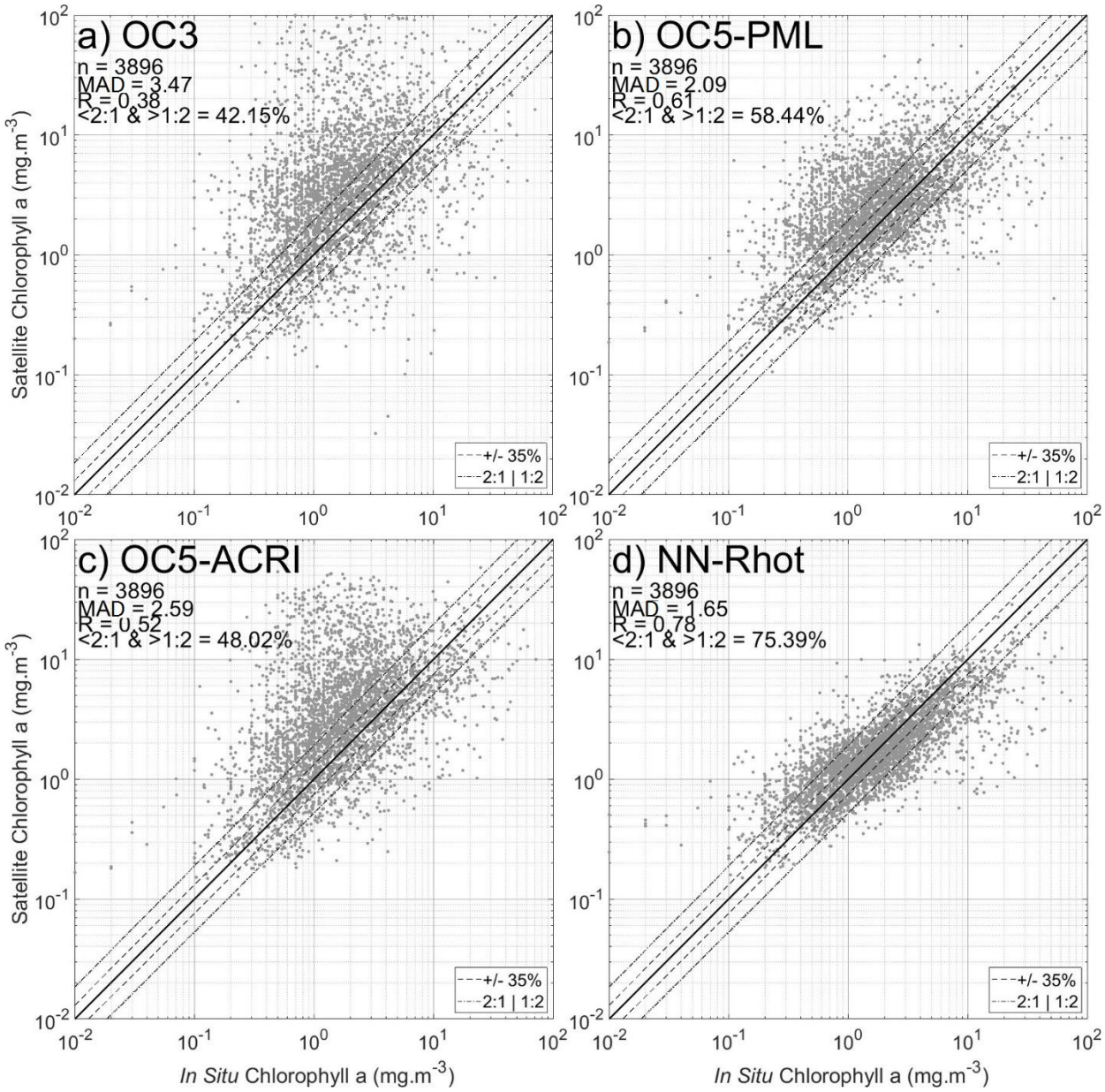

Figure 3.10: Performances over the exact same match-up dataset for a) OC3, b) OC5-PML, c) OC5-ACRI and d) NN-Rhot - the median of 10 random neural networks using Rhot with MODIS Aqua data.

3.5.3. Neural networks generalisation capacity

Further comparative evaluation of algorithm performance is achieved through analysis of images from spring (Figure 3.11) and winter (Figure 3.12). Figure 3.11 shows an example of a spring day at the start of the spring bloom season (20th of April 2005). In general, OC3 produces higher maximum values, well beyond the top end of the colour scale used in the plots, with a maximum of ~4300 mg.m^{-3}. In comparison, the OC5-ACRI, OC5-PML and NN-Rhot products reach maximum values around 62-65 mg.m^{-3}

due to OC5 not allowing any value above 65 $mg.m^{-3}$ while no threshold was defined for the NN. OC3 and OC5-ACRI display broadly similar results across the scene including a patch of high Chl values off the east coast of Scotland which could potentially be an artefact of image merging. This observation is consistent with the previous section and Figure 3.10 where both products returned similar distributions. OC5-PML and the NN are in broad agreement with relatively small differences between them in the North Sea area of this image, the main difference being in the Baltic Sea where very coastal waters are returned as high Chl values by the PML product with no particular feature seen from the NN. It has been previously observed that the PML product overestimates Chl in the Baltic part of the image due to presence of CDOM (Pitarch et. al., 2016). Another area where we can spot differences, this time between the NN and the blue-green algorithms, is in the NE Atlantic where the NN produces higher values than any of the other products which generally agree with each other by returning values below 0.25 $mg.m^{-3}$, largely due to the fact that the OCx, GSM and CI algorithms are applied for this area and there is generally only limited difference in performance for these relatively clear waters. In this case, it is likely that the OCx products are performing well and the NN would benefit from inclusion of additional training data from oligotrophic waters. This difference over an under-sampled area like the NE Atlantic highlights a need for the matchup dataset to contain more open case 1 waters and re-emphasises that this is a data driven approach. NNs do not have the capacity to make realistic estimations for under-sampled scenarios (only ~3% of the current dataset comes from the NE Atlantic).

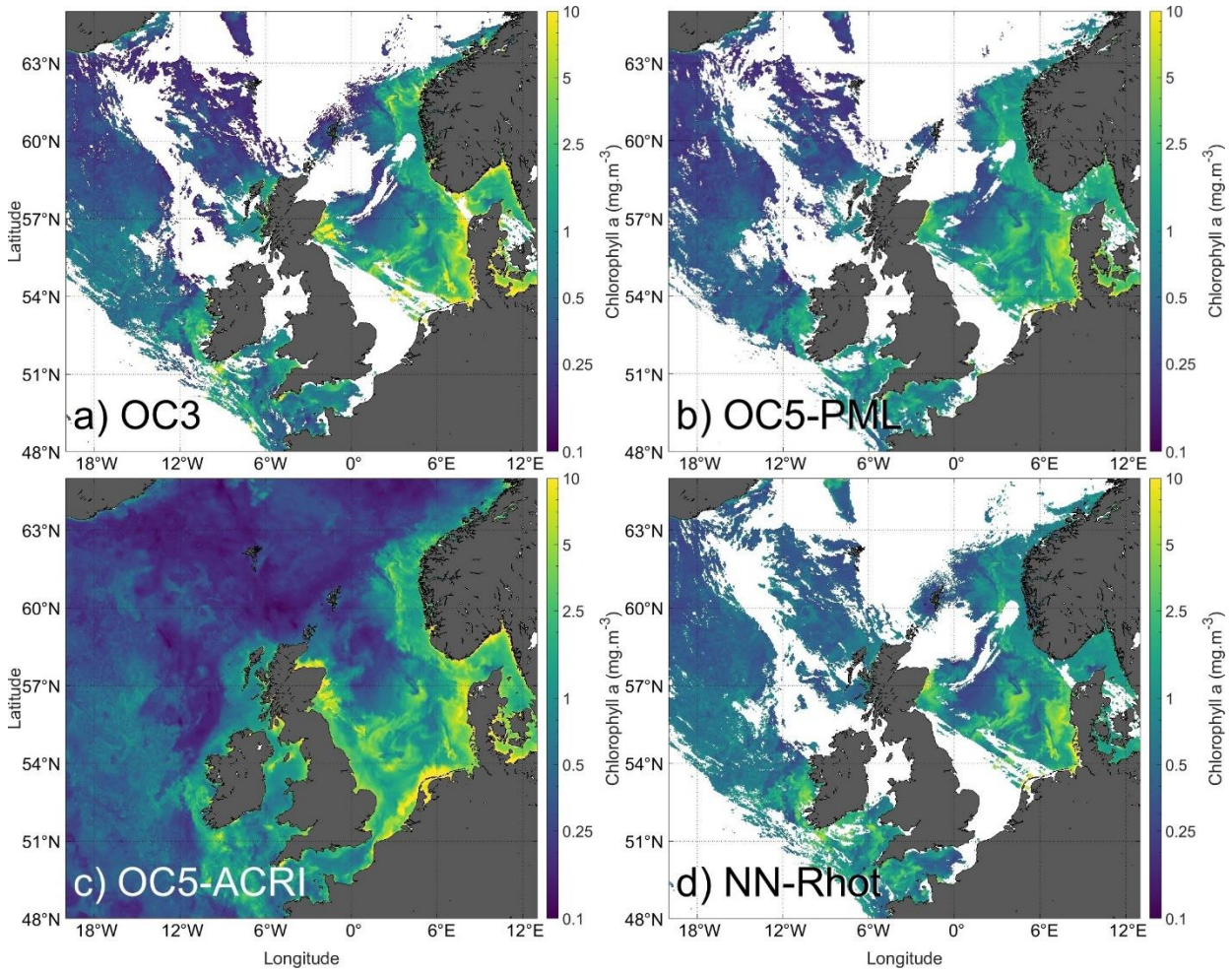

Figure 3.11: Daily Chlorophyll a surface concentration for the 20th of April, 2005 from a) OC3, b) OC5-PML, c) OC5-ACRI and d) the median of 10 neural networks using Rhot applied to MODIS Aqua. For a) and d), MODIS Aqua images at 12:15, 13:50 and 13:55 were merged. For c), the image was interpolated using +/- 30 days by ACRI-ST to get a cloud-free product.

Figure 3.12 shows the daily Chl image from 31st December 2019 (same as Figure 3.5). Given the latitude and time of year, this image represents an example of algorithm performance for high solar zenith angle. Both OC5-PML and OC5-ACRI products apply additional solar zenith angle flags which limit data availability at higher latitudes in winter, with the PML product in this case offering no data availability, while ACRI is using a visible solar zenith angle threshold. In contrast, both OC3 and the NN-Rhot provide data across the scene having been produced from a more permissive dataset using only the ATM FAIL mask and not applying any solar or sensor zenith angle threshold. The OC3 and OC5-ACRI algorithms return significantly higher values than NN in many areas in the southern North Sea and various other coastal waters that are known to present higher sediment loads at this time of year due to winter mixing

in shallow waters, or which are generally tidally mixed. The NN produces lower values which are more consistent with previously measured distributions in this region e.g. usually below 0.5 $mg.m^{-3}$ at Stonehaven (57°N, 2°E, Bresnan et al., 2015) or in the English Channel (Lacroix et al., 2007). Overall, it seems likely that the NN product is both outperforming and is more available in winter than the other direct observation algorithms tested here (OC5-ACRI provides more data points through the wide time frame used, but is therefore not an entirely direct observational algorithm). Whilst the general performance of the NN-Rhot algorithm is reasonably well documented above (e.g. Figure 3.9d and 3.10d), we note the occurrence of relatively high Chl values (between 1 and 5 $mg.m^{-3}$) in a number of coastal areas including the Solway Firth, Morecambe Bay and the Wash. These are regions of known high turbidity and also extensive mudflats at low tide. Algorithm performance under these conditions remains uncertain. Moreover, winter and low values are underrepresented in this dataset, the main limit for machine learning algorithm development. Indeed, further direct validation effort is required for the most optically complex waters and other challenging situations such as areas affected by cloud shadows or immediately adjacent to clouds and land. Masking these areas (based on the rSTD) may be the best option at the moment and would require further development.

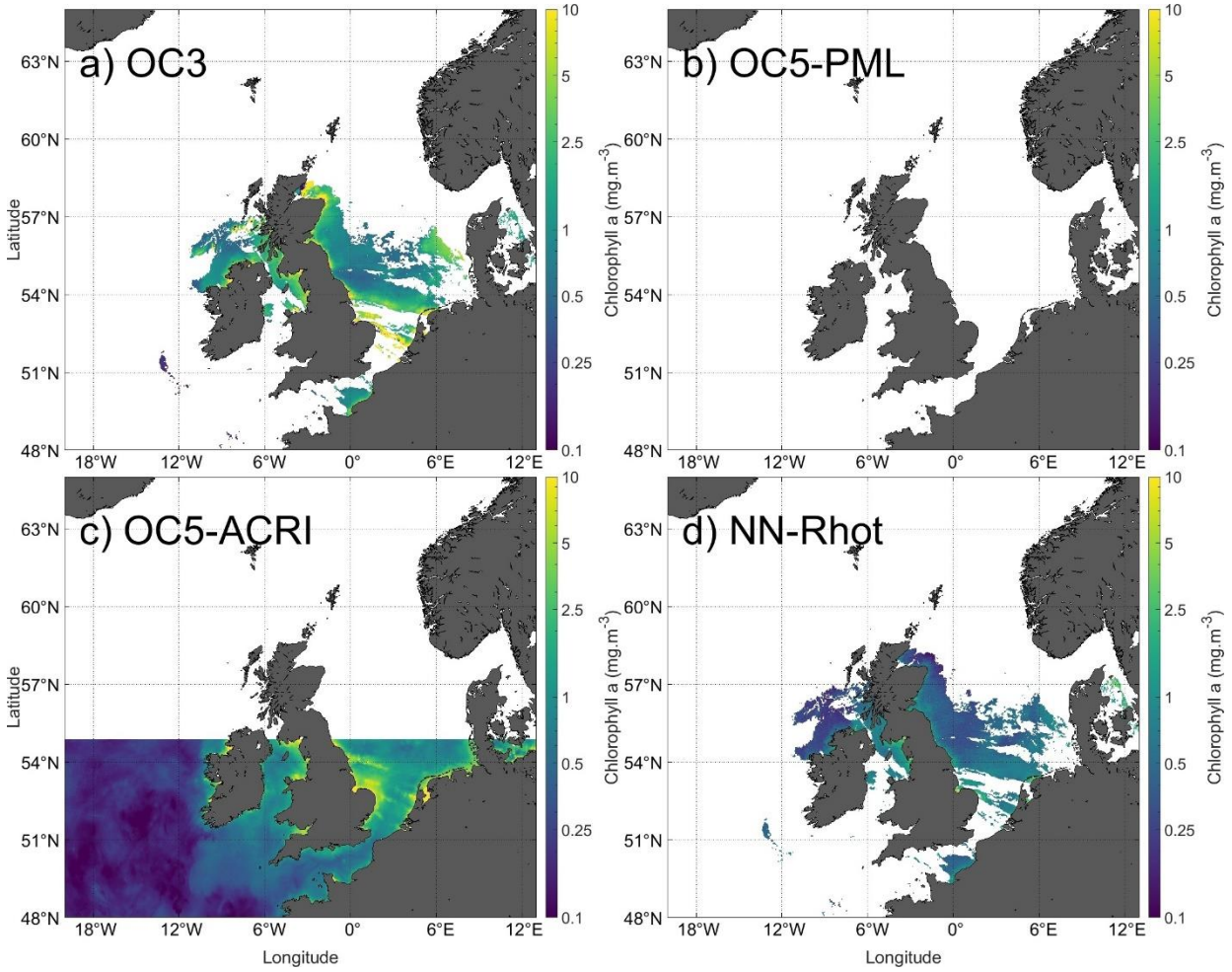

Figure 3.12: Chlorophyll a surface concentration for the 31st of December, 2019 from a) OC3, b) OC5-PML, c) OC5-ACRI and d) the median of 10 random neural networks using 3 layers of 15 neurons for the Rhot reflectance. Single MODIS-Aqua image at 13:05 for a) and d). Averaged daily image for b). Daily image interpolated (+/- 30 days) for c).

3.5.3. Evaluation of NN performance onto independent coastal time series: Example of Stonehaven.

One possibility to independently evaluate performance of the NN algorithm for estimating Chl in coastal waters is the use of a coastal time series. Weekly Chl samples have been collected by Marine Scotland Science since 1997 from the top 10 meters of the ocean at the Stonehaven station (east coast of Scotland). These samples have been co-located with satellite products. Stonehaven matchups were not used for the training, hence their estimation is totally independent. Performances for this specific coastal time series (Figure 3.13) are slightly worse than the global dataset for both PML and NN products, but statistical metrics on such low amounts of data may not be fully representative. The OC5-ACRI product is not shown as it did not show seasonal correlation at any time and produced significant over-estimates

most of the time. 81 matchups are available for the PML product, and slightly more than 2.5 times more for the permissive MODIS Aqua product with 214 daily averaged samples. Data availability comparison between both products is similar to previously observed values for the full dataset (2.8 times more data for MODIS Aqua permissive approach). Compared to the NN, winter data are underrepresented in the PML product due to application of solar zenith angle flags, with solar zenith angles commonly above 70° during winter at this latitude (57°N). Matchups for OC5-PML are available from the middle of February to the middle of November; while they are available at any time using the NN-Rhot algorithm. Low Chl values (<1 $mg.m^{-3}$), usually sampled between October and March, tend to be overestimated by both algorithms, but the discrepancy tends to be much lower for the NN-Rhot algorithm. NN-Rhot produces consistent estimations with independent *in situ* observations (Figure 3.13d). General performance metrics for this independent data set are broadly comparable with the original training data set (Figures 3.9d and 10d). Low values are still systematically overestimated, as a result of under representation of such data from the training dataset. This observation is supported by Figure 3.13e and 3.13f that respectively display the standard and relative standard deviation for NN-Rhot estimates of Chl using the ensemble approach. The relative standard deviation is typically greatest in winter and with values reaching ~50%, whereas values drop to ~20% at other times of the year. Again, this reflects scarcity of training data from winter months due to increased cloud cover, reduced daylight hours and potentially reduced sampling effort at this time of year. Note: the observation shown here concerns the Stonehaven time series only, but similar observations not shown here have been observed for other time series available in the region, such as SOMLIT (Gohin, 2011); stations E4 or L1 from PML (Smyth et al., 2010) or the CEFAS times series (Greenwood et al., 2006).

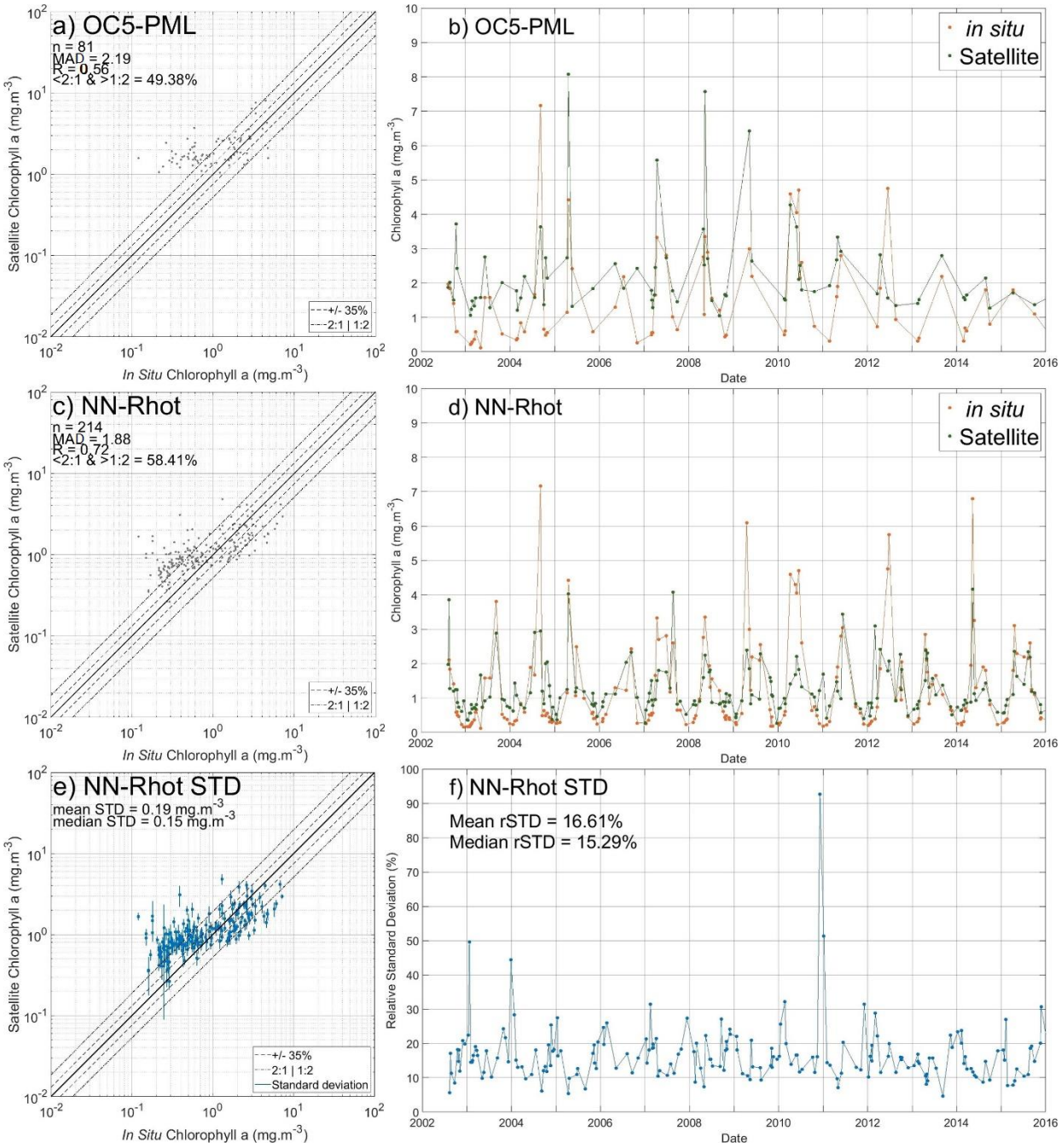

Figure 3.13: Chlorophyll estimations for the Stonehaven time series, for a) and b) OC5-PML algorithm; and c) and d) NN-Rhot. Standard deviation and relative standard deviation associated to the ensemble NN-Rhot approach (e and f).

3.5.4. Flag impact on performance

Following comparison in Section 3.5.2 of same matchups only, one limit comes from the fact that PML or ACRI products do not come with a list of flags applied for each pixel, therefore it is hard to evaluate

which types of noise are impacting algorithm performance, and to consider which (if any) of the available flags might improve performance of the NN approach. To figure out which flags impact performances, another analysis has been conducted. The match-up analysis of this chapter has been conducted on level 3 satellite images reflectances. Level 2 data were not saved during the process due to hard drive space requirements (several Tb) and processing time required (~15 min to process a single image, with ~ 50 000 images available). Moreover, level 2 flags "l2flags" information (list of flags applied to each pixel) are not available at Level 3 from SeaDAS software due to the information not being a variable that can be interpolated, but a binary number, using 32 bits to include as many flags as possible. Only values of 1, 2, 4, etc. can be obtained. Creation of a Level 3 reprojected pixel does not allow use of any flags with value of 1.5 for example, if the l2flags of the Level 2 pixels come with values of 1 and 2. Because it would take several months to reprocess all the data at Level 2 for all 3 reflectances and create a new Level 2 match-up dataset with it, data at Level 2 and Level 3 data will be compared in this subsection. Data at Level 2 have been processed saving NASA's standard atmospheric correction l2flags product. The main limit from this comparison arises from the different data projections used. The Level 2 uses the raw projection captured by the sensor, the Level 3 was reprojected on a constant grid to compare all images evenly. Similar to an algorithm or atmospheric correction choice, there is currently no best method to select a specific projection due to the Earth's curvature at these latitudes, and the choice will come with both positive and negative impacts. Because most data were sampled in coastal areas, the atmospheric correction associated to a pixel in such conditions can mask Level 2 pixels due to atmospheric correction failure due to the presence of lands. In some conditions, reprojected data can provide valid pixels at higher resolution if a neighbouring pixel is available.

Regarding the 15,765 matchups available at level 3, information from a close and similar pixel at level 2 was available for 13,906 of them. Over the 25 flags available from l2flags (see https://oceancolor.gsfc.nasa.gov/atbd/ocl2flags/ for details about the flags, accessed on 11 July 2022), 11 occurred more than 1000 times (Figure 3.14) and impact on NN performance after removal of each of them separately is evaluated (Figure 3.15). The flag occurring the most, atmospheric correction failure, is masked at level 3 and comes from interpolated pixels close to clouds or land at level 2. Product failure (PRODFAIL) is highlighting that one of the products proposed by the process (absorption, Kd490 or a Chlorophyll algorithm for example) is not available. Contamination from sun glint (STRAYLIGHT), and coastal shallow water (COASTZ) also occurs often (>5000 times) due to the fact that most match-up data come from very coastal waters. Other flags present come from product quality warning (PROD WARN); possible cloud contamination, haze presence most of the time (CLDICE); high degree of polarization

(HIPOL); turbid water presence detected in the Red/infrared (TURBIDW); high solar zenith angle, due to inclusion of winter data (HISOLZEN), and the last two (MAXAERITER and ATMWARN respectively) come from atmospheric correction process that used more than 10 iterations to converge or suspicious output signal after AC (negative reflectance for example). One last observation that supports results observed in Section 3.4.3 where coccolithophores were suspected to be underrepresented from the dataset, is that flag 10 (coccolithophore presence) only occurs a few dozen times. A pixel can be impacted by several flags simultaneously.

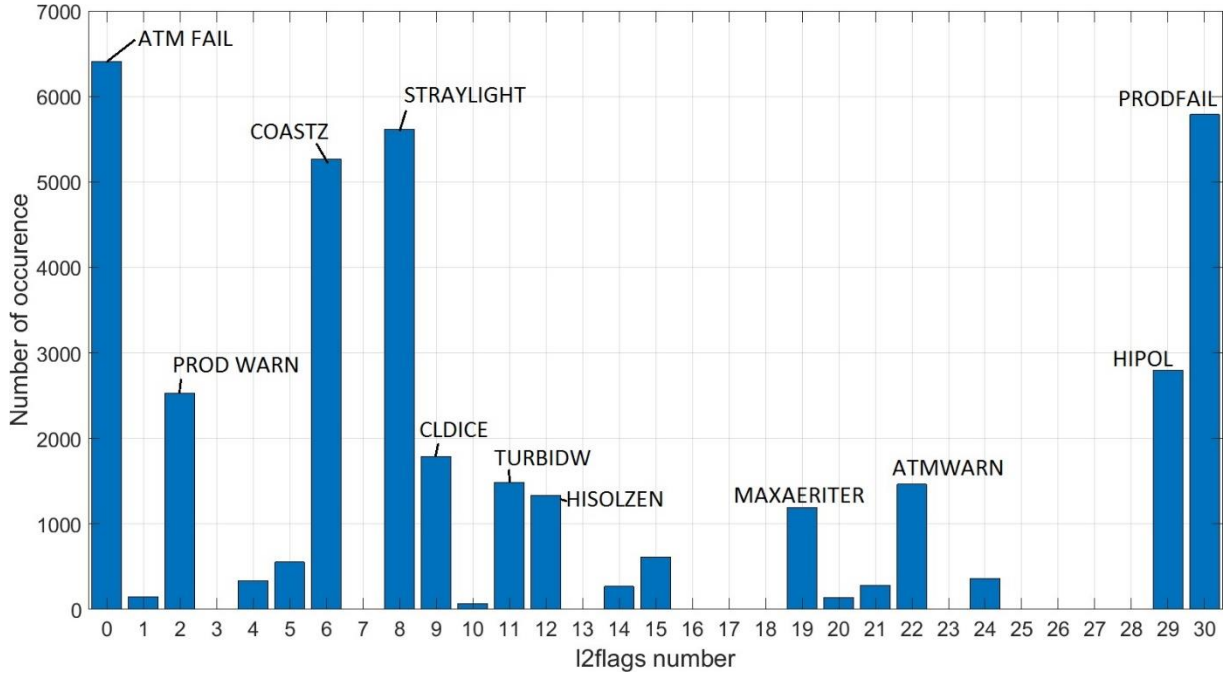

Figure 3.14: Frequency of each flag for the MODIS Aqua match-up dataset at level 2.

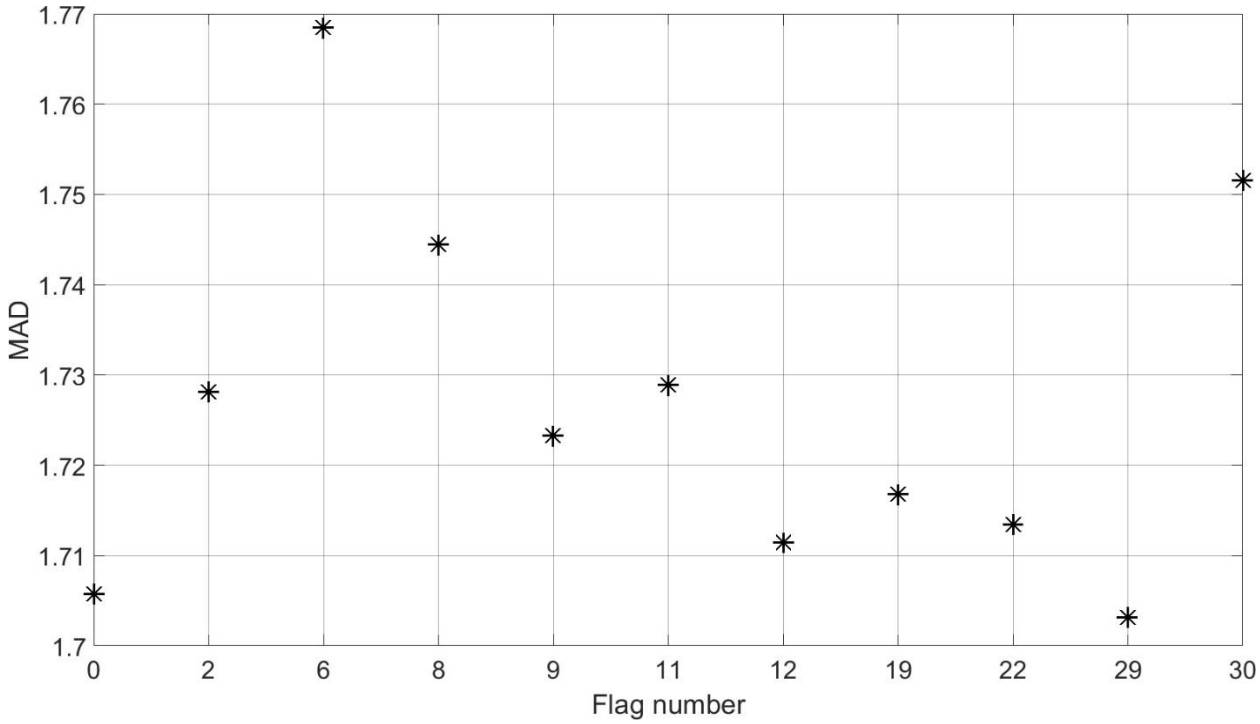

Figure 3.15: Ensemble Mean Absolute Difference (MAD) performance, obtained for the full dataset with each neural network ensemble developed using a subset of the data. The different flag number (x axis) were removed separately for each test (data that are positive to the flag boolean). Please refer to the previous Figure 3.14 to read flags' names.

Data containing one of the specific 11 high-occurrence flags mentioned previously were removed and NNs were trained using the remaining data each time, to simulate performance that could be obtained if a non-permissive approach would have been selected earlier during the satellite data creation. Three flags (number 6, 8 and 30, Figure 3.15) did not seem to improve performance when removed from the analysis compared to results obtained for the general dataset (MAD of 1.76 on average). On the contrary, flag number 0 (the atmospheric correction failure, which would be an obvious choice if data would be available at level 2 for this study) and 29 (high polarisation) show the highest improvement when removed from available data. The rest of the flags showed some improvement and could be considered for applications where achieving highest performance overrides data availability.

The last test concerning flag application/removal consists of the development of a neural network based on a less and less noisy dataset after application of more and more flags each time. Flags have been classified based on the highest performance they can reach when removed from Figure 3.15, with flag 29, high polarisation being the most impactful (MAD of 1.70+) and flag 6, shallow coastal waters, being the least impactful (same errors as the general dataset, which means no improvement made when this flag is

applied).. Figure 3.16 shows performances reached for this dataset based on the number of flags applied to the data. For example, when 2 flags have been applied, it means that flag 29 and 0 (HIPOL and ATM FAIL) have been applied and removed from the dataset. When no flags were applied (using the raw level 3 dataset), MAD <1.8 can be reached with 100% of data available. However, as soon as more flags are applied, MAD improves on the training set, and is slightly followed by the test set, down to MAD = 1.55. However, as more flags are applied, less data become available, and percentage number drops drastically below 40% after only the first 2 flags have been applied. When less than 20% of original data is available (~2600 pts remaining), overfitting becomes a real issue with train and test sets showing almost opposite improvement direction, and difference of MAD > 0.1. This effect highlights the main limiting factor from neural network development and may explain why the ocean colour remote sensing community do not use approaches like this yet. NNs need more data than usual empirical algorithms to outperform them and be robust (resistant to overfitting). Here, NNs seem to be able to reach lower MAD than ever (<1.6) over the train set, with less than 1000 points available, but poor performances on the test set, which is much smaller (150 samples, MAD of <1.7). The dataset is so small that it becomes impossible to reach generalisation compared to when the dataset was 10 times bigger, without applying any of the flag. It is clear that a less noisy dataset allows more accurate estimations, and main performance limitations come from data availability.

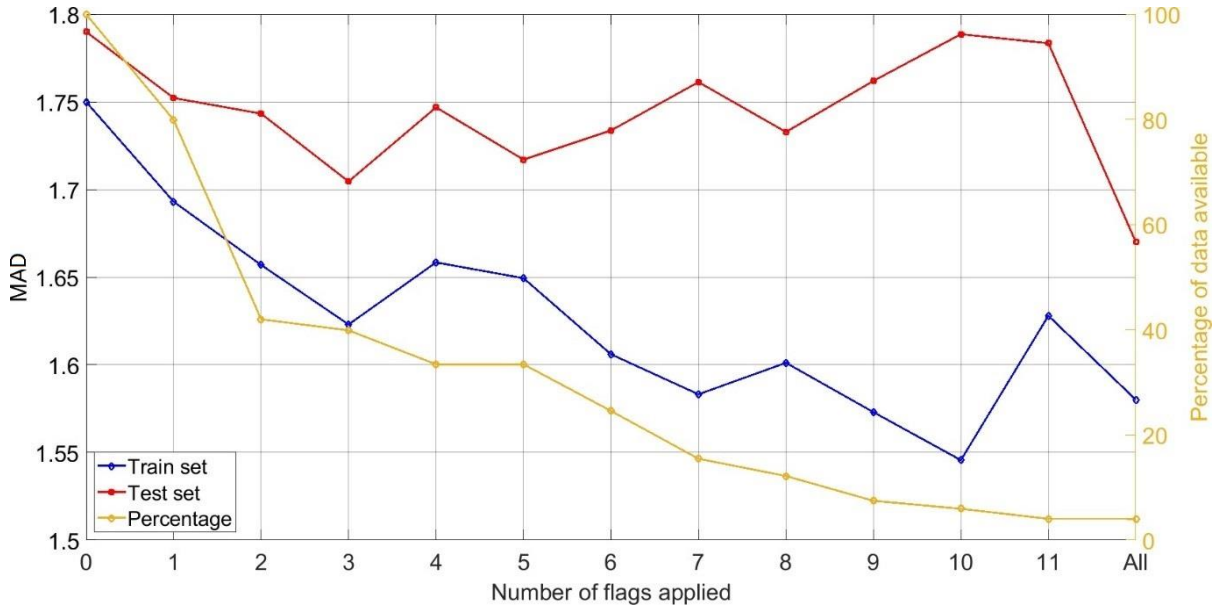

Figure 3.16: MAD performances obtained for a neural network developed including all flags previously applied. Train set performance (blue), test set performance (red) and percentage of data available (yellow).

3.6. Incorporation of non-optical information to improve NN performance.

The ability of NNs and other machine learning approaches to derive statistically meaningful relationships for seemingly poorly or uncorrelated data is one of the major attractions of the approach. However, there are potential pitfalls that one must also be aware of. An obvious source of potentially useful additional information would be inclusion of geo-spatial and temporal information in the training data set to facilitate recognition of regional and seasonal / inter-annual variations that chlorophyll concentration is sensitive to, mainly due to complex interactions between advection, sunlight, predators and nutrients. Inclusion of spatial coordinates, day of year or seasonality signal in addition to TOA reflectance signals was attempted and found to produce significantly improved results. Latitude, longitude, day of year and season information were normalised using *mapminmax* function shown in Section 3.4, with the same 0 – 1 limits, later rescaled automatically by MATLAB between -1 and 1. Day of year is the date of the year (month and day) expressed in days since the start of the year, with January the 1st being 1 and December 31st 365. Season is a number between 1 and 4 corresponding to Jan-Mar for 1, Apr-Jun for 2, Jul-Sep for 3 and Oct-Dec for 4. It has a ~10 days shift compared to the real calendar seasons, which should not negatively impact how meaningful this “seasonal” definition is. A last test was conducted including all 4 variables mentioned together with the TOA signal. The same 3 layers of 15 neurons architecture was used. Figure 3.17 below shows performance obtained from these different tests.

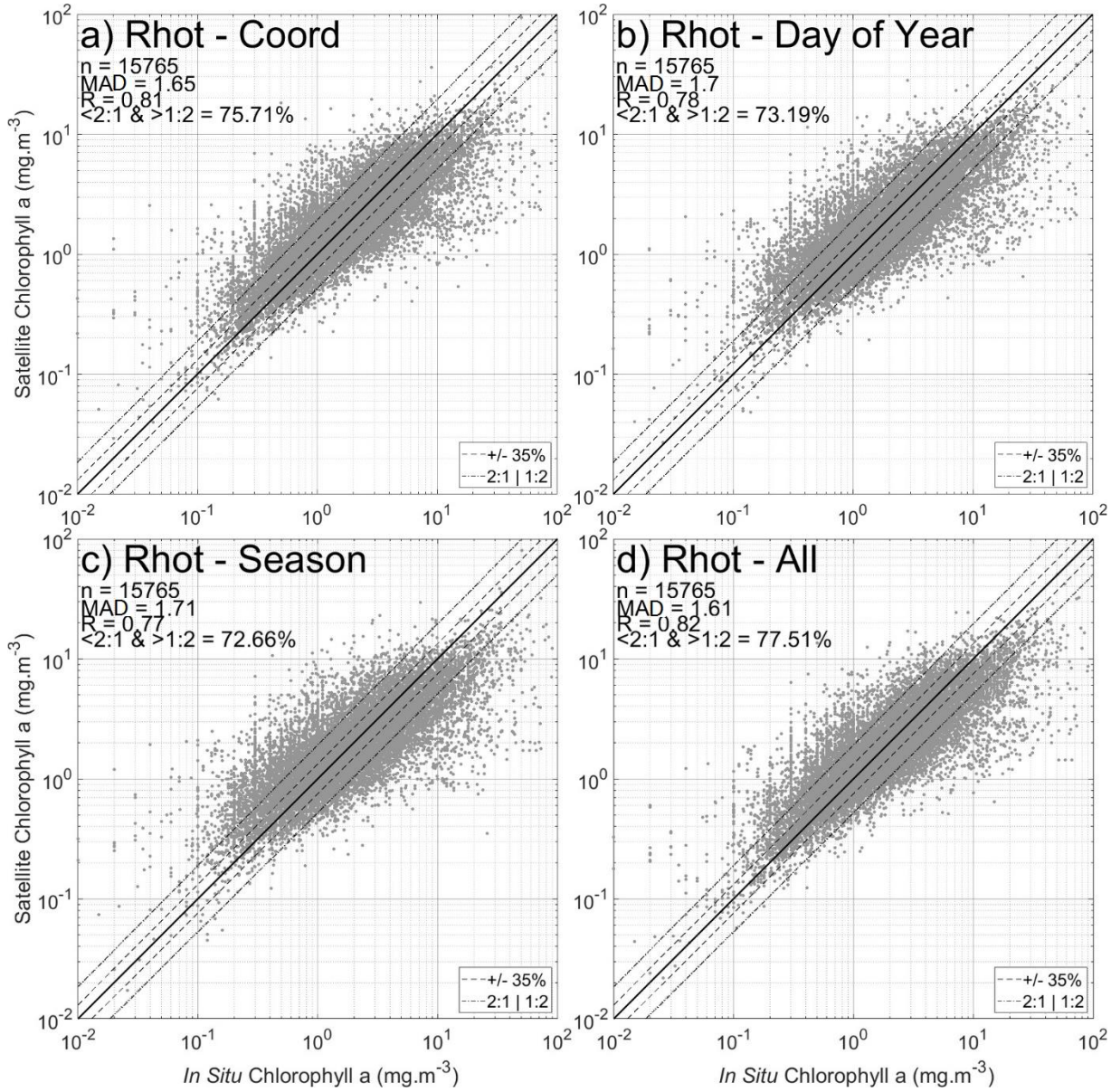

Figure 3.17: Performances of neural network developed using the Rhot light signal and a non-light input. Separate inclusion of a) latitude and longitude, b) day of year, c) season, or d) all information mentioned for panels a) to c).

Whilst the associated statistical metrics are improved, the resulting NNs can sometimes create poorer images in terms of generalisation, showing fewer spatial details than would be expected, for example predicting smooth features over large areas hiding mesoscale features associated with surface Chl. Figure 3.18 shows a chlorophyll map example for a single day of a summer representing MODIS Aqua image of the 28th of July 2019, for each NN developed above. Figure 3.19 represents the same image for OC3 algorithm known to perform well in case 1 conditions.

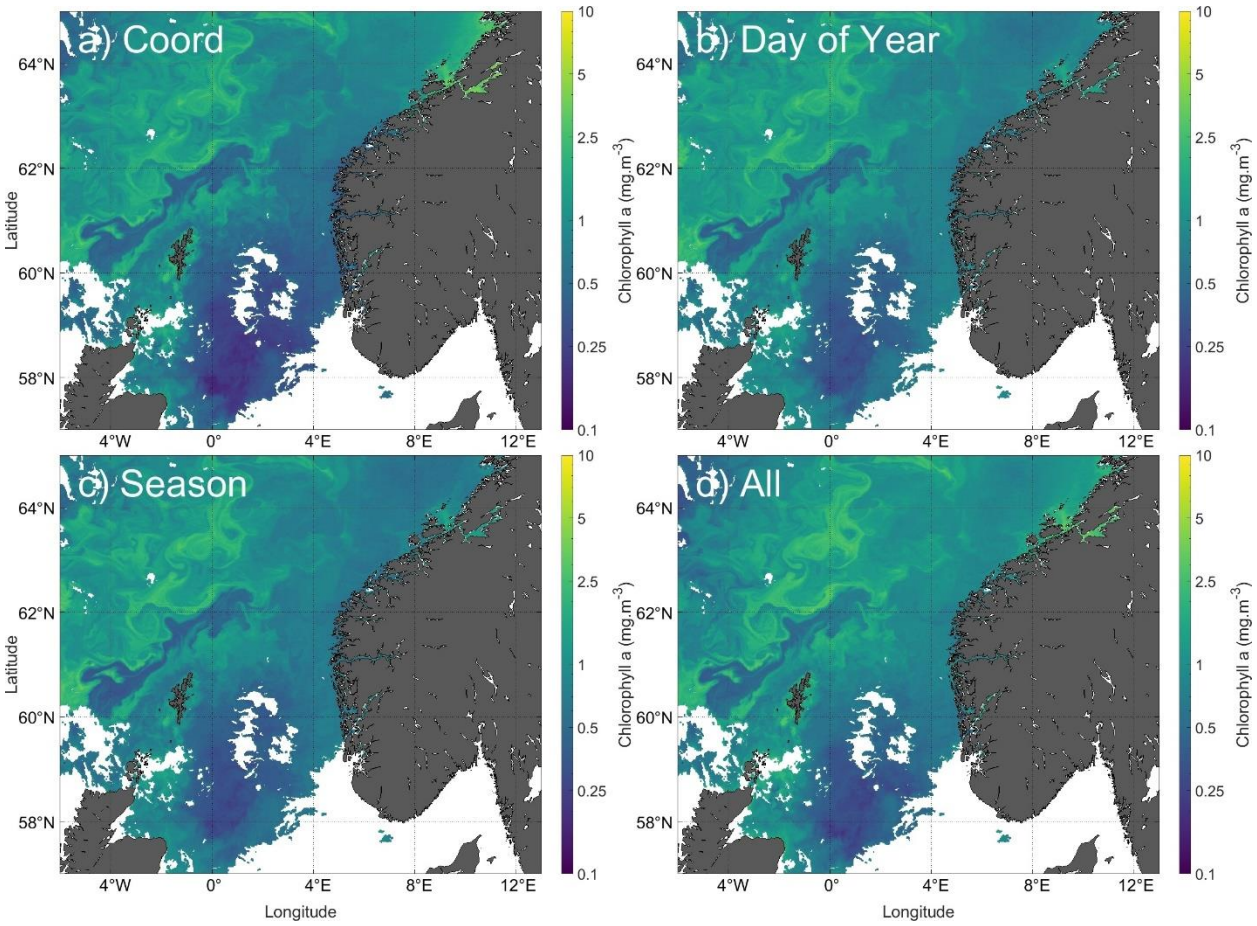

Figure 3.18: Chlorophyll a surface concentration for the 27th of September 2010, 12:10 pm, from a neural network developed including Rhot light signal and a) latitude and longitude, b) day of year, c) season and d) all 4 inputs.

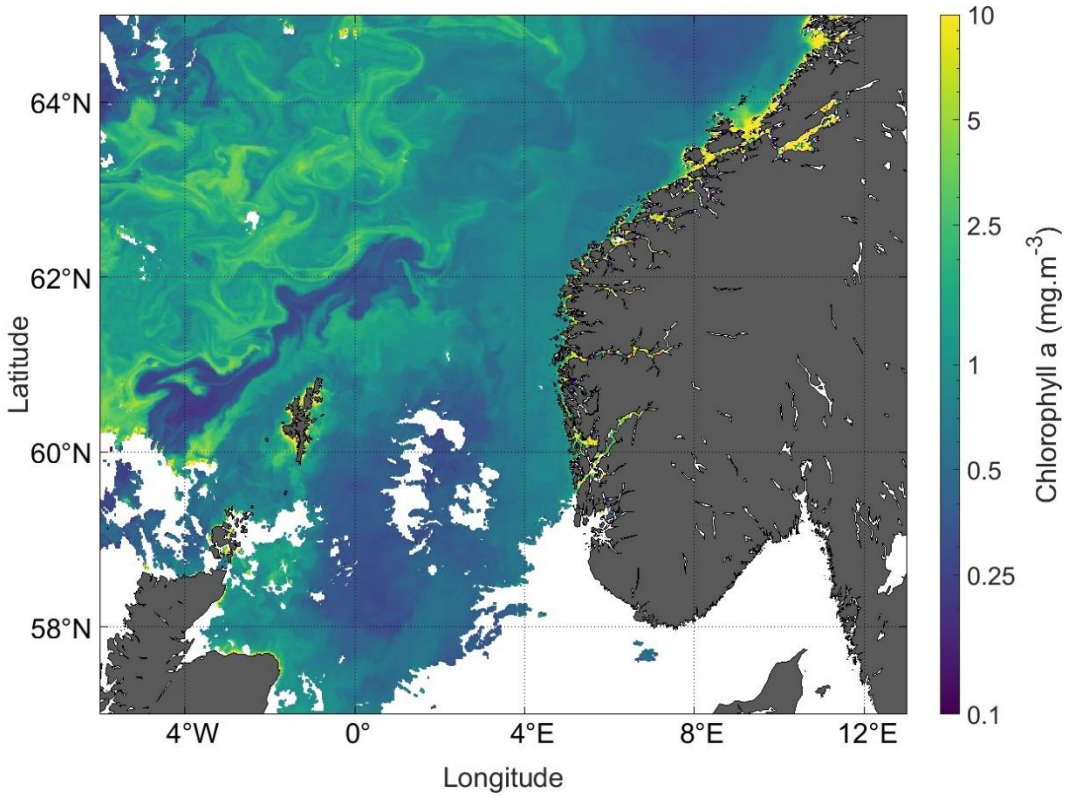

Figure 3.19: Chlorophyll a surface concentration for the 27th of September 2010, 12:10 pm, using OC3 algorithm.

All images are similar to OC3 outputs for the open waters, with main differences coming from coastal areas where OC3 return very high values (>100 mg.m-3) whilst NNs return continuous values from inshore to open seas. Chlorophyll mesoscale features from different NN look realistic compared to OC3, but they may have been rubbed out for some areas in the Atlantic part of the image, where thin mesoscale features are missing. It seems likely that inclusion of geospatial and temporal data as inputs to the training data set allows NN to identify key features of the data set, but reduces the weighting put onto the directly observed light data. Inclusion of geospatial data, in particular, is likely to overstate regional attributes and to seriously impinge on performance away from areas of dense *in situ* sampling, especially for the Atlantic part where only a few data were available from this study. Previous work (not shown here) involving NN estimations that included large volumes of Norwegian ferry box data in the training (40,000 matchups from the Norwegian coast only, measured by uncalibrated fluorometers onboard ferries, removed since due to suspicious quality) clearly showed significant spatial averaging rather than direct observation due to the massive density of matchups available from the ferry box system (40 000 points available for Norwegian coast, against 15 765 everywhere else).

The NN presented in this Chapter is a regional algorithm as a consequence of the nature of the training data set used to develop it. However, the methodology presented has scope to be scaled up to global scale by accumulating a sufficiently extensive and robust training data set. When this is attempted, it will be interesting to test the effect of including geo-spatial information as an additional input (potentially attractive if the training data is truly global). Based on my experience to date, I expect that such an approach would also result in a loss of spatial resolution in resulting images, and that future NNs should focus on use of all-optical inputs to maximise generalisation capability (i.e. smooth the signal over a region and remove fine scale details such as eddies).

Inclusion of solar and sensor zenith angle or log transformation of inputs (for Rhot only) prior to normalisation were also tested but only returned slightly improved performances (MAD is 2% better, 1.73 compared to 1.76) and were not used for this analysis but could potentially be more impactful at a global scale. However, overfitting due to addition of more inputs could become an issue each time a non-light signal is added, and further analysis will be required to validate use of any non-light signal as inputs. This specific test will be conducted in Chapter 6, but with sea surface temperature estimations due to the massive data availability at global scale thanks to drifting buoys. NNs including other variables such as wind speed will be evaluated to see the impact on estimations as such scale.

3.7. Discussion

It has been shown that the performance of several blue-green reflectance ratio algorithms, including the latest versions of the most up to date variants, have variable performance in the optically complex waters of the northwest European shelf. At the same time, there is growing interest in the potential use of OCRS to contribute to monitoring the environmental status of territorial waters, particularly with respect to the impacts of eutrophication. There is therefore a profound need to understand and quantify the performance characteristics of satellite Chl products and, if possible, to improve upon the existing array of algorithms. There is growing appreciation of the potential for machine learning approaches to be the key to improved data quality. Here, we have shown that an artificial neural network is able to return favourable performance compared to existing algorithms in direct matchup analyses and can produce realistic images of Chl distributions across the region. It seems likely that efforts of this nature will continue to be developed in coming years and that there will indeed be significant advances in Chl algorithm performance as a result.

In developing a new algorithm like this, it is important to appreciate where previous algorithms have struggled (generally, not specific to the dataset available) and to understand the limitations of new techniques. It is also important to understand the quality of data being used for validation. Here we are using a set of *in situ* Chl observations including filtered water samples (analysed using HPLC and fluorometric analyses) and *in situ* fluorometry. Of these, HPLC is probably the most accurate source of data. Claustre et al. (2004) demonstrated that with suitably stringent quality control measures in place, it was possible to achieve average uncertainties below 10%. However, this represented a round-robin exercise by leading laboratories rather than general community capability, and the metric that really matters in this analysis is the range of uncertainty, which was ~±20% for the Claustre et al. (2004) study. More recently Sørensen et al. (2007) and Tilstone et al. (2012) have both reported Chl results that suggest greater levels of uncertainty for HPLC data sourced from across the community, with uncertainty ranges up to ~±40%. When considering the performance of satellite Chl products, particularly with respect to the potential for using satellite data to complement *in situ* observations for reporting against legislative requirements, it is very important to consider that the quality of the *in situ* data is likely to be of the order of ±40% *at best*, and that in many cases it may be considerably more uncertain than that. Results from the NN-Rhot algorithm may well be reaching levels that are consistent with the quality of data being used to train the networks, at least for scenarios that have been sampled relatively frequently. That said, there is undoubtedly a need to target future sampling efforts towards scenarios that are currently under-sampled e.g. low Chl in winter, open Case 1 waters and specific events such as coccolithophore blooms.

3.7.1. Chlorophyll algorithm and atmospheric correction failure in northwest European shelf seas

Northwest European shelf seas are optically diverse, with optical properties ranging from Case 1 conditions at the margins and in certain summer stratified shelf waters to highly tidal regions characterised by high sediment loads and areas of freshwater influence where high concentrations of CDOM impact strongly on reflectance signals (Babin et al., 2003). Many areas show strong seasonal variations, with shallow regions often exhibiting higher sediment loads in winter associated with increased wind-driven mixing associated with winter storms (Mitchell and Cunningham, 2015). The occurrence of sediment dominated waters degrades the performance of standard atmospheric correction algorithms based on the black pixel approximation. At its extreme, this leads to negative reflectance values in AC-corrected Rrs values, as seen in Figure 3.5b. However, it is important to realise that over correction is not restricted to spectra with negative values and that many of the non-negative spectra shown in that figure will also be poor representatives of the true remote sensing reflectance signal at sea level (due to the

overcorrection. A further measure of the true difficulty of atmospheric correction over these waters is revealed in Figure 3.5d, where application of just the first stage of atmospheric correction, the Rayleigh correction, is sufficient to drive a number of Rhos spectra into negative values. It is clear that, for an optically complex area like northwest European shelf seas, atmospheric correction is a potentially limiting step. Variable atmospheric correction performance will almost certainly have deleterious consequences for subsequent Chl algorithm performance. Several atmospheric correction methods exist (Gordon and Wang, 1994; Ruddick et. al., 2002; Wang and Shi, 2006; Steinmetz et. al., 2011; Brookman et. al., 2016). However there is no generally agreed, optimal choice of atmospheric correction that can be successfully applied over such a complex area. More recently, similar observations of higher performances has been made for either temperature estimations from the visible spectra (Medina-Lopez, 2020) or land observations (Lin et al., 2015), promoting the use of uncorrected top of atmosphere signal until more performant atmospheric correction methods emerge. The fact that the NN-Rhot algorithm operates successfully from TOA reflectances without requiring selection of an atmospheric correction algorithm is therefore an attractive feature of the approach.

Moving beyond atmospheric correction issues, it has previously been established, using *in situ* radiometry (therefore unaffected by AC issues) that the presence of independent and variable concentrations of sediment and / or CDOM has potential to disrupt the performance of OCx blue-green reflectance ratio algorithms. As one of many examples in the literature, McKee et al. (2007) demonstrated that the OC4v4 algorithm performed poorly across the Irish Sea (part of this study area) with notably worse performance in highly turbid, sediment-rich waters. Thus the second stage of Chl estimation from radiometry, the Chl algorithm itself, is further challenged by the level of optical complexity found in northwest European shelf seas. The extremely variable performance of OC3 (Figure 3.9a) is probably most associated with the performance of the empirical Chl algorithm rather than the AC (similar overestimations with the more restricted version, Figure 3.10a). The more sophisticated OC5 algorithm version from PML reduces error to within an order of magnitude, although some of this is achieved by masking out identifiably poor quality reflectance data rather than producing inherently better estimates. Lavigne et al. (2021) have shown limitations of this specific OC5-PML algorithm when used for high Chl coastal waters. In both cases, the failure of state of the art algorithms is not only due to atmospheric correction, but is intrinsic to the performance of the algorithms for optically complex waters.

There is considerable interest and optimism in the field that machine learning techniques can be used to develop a new generation of ocean colour algorithms that will perform more robustly in optically

complex shelf seas. In this study, we have attempted to develop a baseline approach where we take into account the issues affecting both atmospheric correction and Chl algorithm performance, and where we seek to establish performance characteristics for one of the simplest forms of machine learning techniques. Here we have attempted to systematically explore the various decisions that go into constructing a NN. Notably this has included the option of using any of fully corrected BOA Rrs, partially corrected Rhos and completely uncorrected TOA Rhot as inputs. Remarkably, we have shown that similar levels of performance can be obtained with any of these input data sets, with uncorrected TOA Rhot data providing marginally superior results than the other two. This otherwise surprising result can be explained by the limited performance of AC for these conditions. It is true for both NNs and other, more traditional algorithms, that poorly AC-corrected Rrs data is a hurdle to be overcome. In this case it would appear that eliminating AC altogether and operating directly on Rhot facilitates the job of deriving Chl for the NN.

Advanced algorithms such as OC5 undoubtedly do a more robust job of directly dealing with optical complexity found in coastal waters than the OCx algorithms designed for Case 1 waters. However it is clear from results presented here that a significant portion of the apparent improvement in performance is derived from the relatively stringent flagging used to eliminate the most problematic scenarios. The OC5 –PML product is a good example of a trade-off between improved data quality vs reduced data availability. Conversely, the NN-Rhot approach has been designed to maximise both data quality and availability simultaneously. The computational flexibility offered by the NN allows us to operate directly on TOA reflectances directly and to accommodate the optical complexity of north-western European shelf seas. In doing so the NN-Rhot approach is able to improve both data quality and quantity, and through the ensemble approach it can also provide a measure of data uncertainty. At this point in time the availability and comprehensiveness of the training data set appears to be the limiting factor for the NN-Rhot approach. Further extension of the training data set is perfectly feasible through data mining existing historical data and targeted future sampling.

Another important feature of the NN approach presented here is the use of all relevant spectral bands. Rather than attempting to find an optimum set of wavebands or trying to ascribe physical significance to any particular band, our approach has been able to provide the NN with all available wavebands in the Vis-NIR-SWIR range in order to allow it to resolve the combined problem of dealing with AC and Chl retrieval. The NN approach developed here effectively ingests all of the available spectral information and the neural network is free to determine statistically robust relationships free from human intervention or bias. For example, reducing the number of inputs to the 3 RGB bands (490, 550, 670 nm)

used by D'Alimonte and Zibordi (2003) produced significantly poorer quality results (not shown). It is likely that our approach does in fact carry elements of redundancy e.g. using the 859 and 868 nm or the 547 and 555 nm bands simultaneously may not be meaningful as the input information content is presumably almost identical within each pair. This may point to further simplification that could improve computational efficiency in the future, but is unlikely to improve product quality. We note that El-Habashi et al. (2019), Gilerson et al. (2021) and Pahlevan et al. (2021) identify failure of the atmospheric correction and resulting impact on blue BOA Rrs wavebands as major limiting factors in their incorporation in NN to retrieve Chl from satellite data. This is consistent with our observations in Figure 3.5 and our interpretation of why the NN operating on TOA Rhot outperforms the NN operating on BOA Rrs. The inclusion of Red - NIR – SWIR bands, elsewhere used for AC, has an unresolved, but potentially crucial role for successful exploitation of TOA Rhot as input as they can be directly linked to in water sediment concentrations. Recent work supports the need for NIR bands for high Chl content that blue-green reflectance ratio algorithms have problems with (Lavigne et al., 2021). The neural network approach developed in this study uses a combination of visible and infrared bands and there is potential for sensitivity to fluctuations in atmospheric signals such as impacts of volcanic eruptions (Hervo et. al., 2012). We have not observed such impacts in this study, but caution that there remains scope for this to occur under specific circumstances.

3.7.2. Benefits and limits of neural networks

NN approaches are generally assumed to be computationally intensive. However, modern computing power is such that all of the work presented here was easily achieved on a relatively modest computing platform. Testing of all the architectures presented in Figure 3.3 took approximately 1 day using a desktop PC with 16Gb RAM and Matlab R2020a. On the same computer, applying the 10 iteration NN to the image shown in Figure 3.8a took ~3 minutes. Whilst the computational requirements should not be underestimated, it seems quite feasible that the NN approach developed here is practically implementable due to their low dimension.

In this study we have used the NN as a black box, deliberately avoiding introducing user bias into the production of the NN, though noting that there are inevitable elements of user choice in the design of the NN e.g. normalisation method, choice of activation function, choice of error metrics used to assess performance that do in fact impact on eventual NN performance. Unfortunately, the resulting statistical relationships that emerge from NN development are not amenable to physical interpretation. The 3 layers of 15 neurons architecture adopted in this study, despite being relatively small, still represents

approximately 690 weights connecting the neurons and is therefore essentially impossible to interpret physically. It should be noted that there is tremendous potential for further refinement of the NN structures, for example with inclusion of dropout layers (Srivastava et. al., 2012), use of more complex activation functions such as leaky ReLU (Maas et. al., 2013) or different neural network architectures (e.g. long-short term memory networks, Hochreiter and Schmidhuber, 1997). The simple feed forward networks used here provided good performances and whilst further complexity is possible, there is perhaps merit in minimising the complexity of NN structures used and addition of further complexity should be based on demonstrated merit only.

Whilst the NN approach presented here operates on TOA reflectance data with minimal flagging, masks are still applied for clouds, ice, glint, saturation and stray light. It is interesting to note that neural networks have already demonstrated good potential to identify these areas (Mohajerani et al., *2019)* and could replace current threshold methods in the near future. The TOA NN does not require additional ancillary data products and is therefore independent of availability of other data sources. However, there is an opportunity to include these ancillary data that impact the light signal of the atmosphere, and could lead to further improvement.

A novel feature of our approach is the development of a bootstrap-like, iterative approach to produce distributions of Chl estimates for each pixel rather than a single value. The resulting descriptive statistics are potentially useful for providing end users with estimates of confidence in each pixel and for identifying water quality scenarios that are under-represented in training data sets. This can be used to direct future *in situ* sampling efforts to maximise impact on development of future versions of NN algorithms.

NN performance is ultimately determined by how representative the training data set is. For example, in this study, there are only a few hundred samples available for the NE Atlantic and NN performance is currently questionable for that region and for open ocean waters more generally. The focus of this work was to develop an algorithm that worked well in optically complex shelf seas rather than open waters where standard algorithms such as OC3 and CI are expected to work reasonably close to the mission target of +/-35% differences between samples and estimates (e.g. Hooker et al. 1992), with MADs of 1.4 for oligotrophic to 1.6 for general case 1 waters usually reached by these algorithms (Seegers et. al., 2018). More generally, the training data set assembled for this study has relatively small numbers of data outside the 1-10 $mg.m^{-3}$ range (Figure 3.2) with potential implications for NN performance towards both extremes of the data range (Figure 3.9d). Further expansion of the training data set is imperative,

particularly if the NN will be used for open ocean waters, even inadvertently as we have done in this chapter. Recent work by Hu et al. (2020) highlights the potential for machine learning approaches to simplify Chl retrieval in open ocean Case 1 and oligotrophic waters. However, it is less clear if the NN will be able to recover very low values in more turbid waters, a scenario that our current data set does not properly encompass. It is possible that the contribution to reflectance of a low concentration of phytoplankton amongst a high concentration of sediment is sufficiently small that it is not identifiable, even using a NN (e.g. Doerffer and Schiller 1994). Moreover, adjacency effects from land is significant in coastal waters and may increase the backscattered measured signal (Bulgarelli and Zibordi, 2018). Development of machine learning enhanced OCRS algorithms is likely to increase rather than decrease demand for *in situ* validation data, with particular emphasis on directing effort towards novel and rare features that are underrepresented in existing training data sets.

As it currently stands, the NN developed here is unashamedly a regional algorithm, with our focus being on establishing a methodology that can in future be scaled up to global levels. Testing NNs performances on independent datasets is a major limit as matchup datasets created in the past for other studies typically did not use Rhot nor the full set of bands used for this study. Here we have tested the NN-Rhot approach on an independent coastal time series, with the NN returning encouraging results across the annual cycle. This study highlights the need for agreed matchup datasets to be shared and used by the community for algorithm development, including top, middle and bottom of atmosphere reflectances and the full set of flags. OC5-PML and OC5-ACRI products would benefit from having access to the data gathered for this study to further refine their LUTs, and could potentially return higher performances. However, there would still be limits associated with the nature of the OC5 algorithm and reliance on flagging to eliminate the more difficult scenarios.

Restricting NN inputs to optical signals only is potentially key to ensuring translation of the NN approach beyond current geographic confines. However, there is also a limitation on applicability to a particular satellite sensor, in this case MODIS Aqua. This is partly due to the availability of specific bands for each sensor but also reflects specifics of sensor calibration. Directly translating the current NN to another sensor is unlikely to be easy and is likely to require collation of a suitable matchup data set for that instrument followed by repetition of the methodology outlined above generating another instrument-specific NN. Development of a long term, consistent TOA time series, incorporating data from multiple satellites along the lines of the OC-CCI project, is essential and in this case may be key to developing a global data set for exploiting the capabilities offered by machine learning data analysis

techniques in this field. Introduction of hyperspectral OCRS data in the future, e.g. the forthcoming PACE mission, has potential to support development of improved NNs that may be able to exploit enhanced spectral resolution to improve accuracy of Chl retrieval.

The NN developed here proceeds straight from TOA Rhot to estimates of Chl, effectively bypassing the need for production of atmospherically corrected, BOA Rrs values. Whilst this is efficient, it precludes the possibility of applying the NN to reflectance signals measured *in situ*. Ironically, the Rrs NN discussed in Figure 3.4 might not work well with *in situ* Rrs data as a result of having been trained on poorly AC corrected satellite Rrs values. Whilst there is clearly merit in avoiding the need for atmospheric correction, there is undoubtedly interest in generating accurate BOA Rrs values, not least because it is a Global Climate Observing System established Essential Climate Variable, but also because it facilitates functional links between satellite and ground truth optical observations. The NN methodology development proposed here is translatable to deriving surface Rrs values instead of Chl, but requires provision of an adequate training data set. Extensive efforts to produce global sets of *in situ* optical and biogeochemical data have been made by the community (e.g. NOMAD, Werdell and Bailey (2005); MERMAID; Valente et al. (2016)) and it is likely that future AC algorithms will be developed using neural networks techniques such as the NN approach discussed here (e.g. Brockmann et al. 2016, Fan et al., 2017). Again, there will be increasing value to be had from future *in situ* optical sampling, with increasing focus on the use of sensors deployed on moorings and other autonomous platforms providing an efficient means of generating necessary matchups with satellite data.

3.8. Conclusion

This chapter presented the main objective of this PhD project, the creation of a remote sensing chlorophyll algorithm that returns closest possible values compared to what is currently reported by European sampling agencies, and maximises data availability in the process. Major differences compared to state-of-the-art OC5 algorithms currently provided by different group in the same region were shown and discussed.

First, while most currently available chlorophyll products still use empirical relationships, there is widespread understanding that machine learning techniques may offer useful alternatives and here the development of a neural network approach was detailed. The architecture size of a neural network impacts its performance and the selection of an optimal size has been discussed to avoid overfitting and underfitting problems. Log transformation of the target was a crucial step as NNs struggle to predict non-

linear signals in general, even with the use of non-linear activation functions. Because neural networks are based on error minimisation, returning negative values close to 0 when a very low value is expected is acceptable in terms of error measurement from a NN point of view. However, similar to negative light signals returned by atmospheric correction, negative concentrations are meaningless in terms of physics, and log transformation of the target eliminates this possibility. The other important design step presented was the use of an ensemble approach to counter the randomness involved in neural networks, which also allows users to see how confident neural networks are for each estimated value and can help highlight an under-sampled feature. Other fine-tuning possibilities (for example different normalisation algorithms, activation or error function) were not discussed as none of them seemed to produce drastically different estimations.

The second major difference compared to what is currently done elsewhere by the OCRS community is the use of TOA signals as input, making this approach insensitive to atmospheric correction and the optical water type. Because there is no single specific atmospheric correction method that performs similarly well over both case 1 and 2 waters, it becomes difficult to successfully estimate chlorophyll for all available BOA pixels, and common methods consist of flagging out inferior quality but available data. Application of atmospheric corrections known to be harder to perform in coastal conditions has been shown here to result in unphysical Rhos and Rrs data that either has to be removed from the training data set, or if included in the training, forms an additional hurdle for the NN to overcome. Based on these observations, it makes sense that NNs operating on uncorrected TOA Rhot perform better than BOA and TOA-RC versions. The two main differences with best available algorithms were the estimation quality, with MAD <1.8 compared to 2.1; and quantity, with access to more than 2.5 times data in coastal areas usually flagged for quality requirements on the signal. By using a very permissive approach during the match-up creation, neural networks showed the capacity to make coherent estimations over a noisy dataset. Winter data at these latitudes are flagged by other RS agencies because the solar zenith angle is high at this time of the year (>70°), but results close to ground truth have been obtained for the Stonehaven coastal station using the NN approach, with slightly overestimations possibly due to the lack of data at such low concentrations and time of the year from the dataset. Application of different flags improved performances but reduced the data availability significantly and is therefore not recommended. Some areas may still need to be flagged, for example coccolithophores, until enough training data have been collected.

Finally, the capacity of neural networks to include any signal as an input was assessed with inclusion of geo-spatial and temporal information along with light signals. While good improvement in match-up statistics was possible by including meaningful information related to natural distribution of a variable at the surface of the ocean like chlorophyll, the need for neural networks to be good at generalisation to form meaningful images was highlighted and as a result I would not recommend such application for an open waters product with scattered data availability. Inclusion of such information would be helpful if matchups data were gridded (i.e. available at regular times and places) which is not possible for observation but is for models outputs.

By not correcting the atmospheric signal, any TOA algorithm becomes sensitive to interaction with the atmosphere, and end users of such products should be aware of potential discrepancies that could happen if the TOA reflectance signal is impacted by unusual features such as volcanic eruptions, desert aerosols, human gaseous emission, etc. This algorithm was designed using 15 bands from the visible to the SWIR, and is MODIS Aqua specific. A direct transferability of the algorithm developed in this study to other ocean colour sensors is not possible yet, but the same methodology can be repeated for other sensors. The major limit is the number of matchups available for a younger sensor than MODIS Aqua and how representative of the region complexity they are. It is also important to note that there is some level of equivalence in the spectral coverage of candidate sensors. For example, it is known that some specific MODIS wavebands are highly related to atmospheric signals (469nm, 748nm and 869nm), and this was the main reason of their inclusion in the current NN approach. The next chapter will test several options for NN algorithm creation by selecting different wavebands among the 15 available and see if similar performances can be obtained with less spectral reflectance information.

Take home message: Chapter 3

The main objective of the thesis is to improve estimates of chlorophyll a in coastal waters to help monitoring agencies to characterise eutrophication, where current methods overestimate by several orders of magnitude the chlorophyll due to the complexity of the signal.

- An algorithm using the top of atmosphere signal is trained with 15 MODIS Aqua bands using neural networks.
- It outperforms other methods (more data available and closer estimates to the ground truth measurements) that rely on bottom of atmosphere signals, clustering and band ratios
- Various methods to overcome the limitations involved by the use of a neural network are developed, such as merging multiple networks together or looking for optimal architectures.
- By not applying any masks and gathering a large dataset (15,000 matchups), the algorithm seems to perform well on usually flagged out areas.
- The main limits involve under sampled areas (coccolithophores, extremely low values)
- Applications of flags improve performance but drastically reduce data availability.
- With a more carefully selected environment, by application of flags when the signal may be altered, results can be improved by 20% (MAE 1.6 vs 1.8).
- Access to winter at high latitudes seems possible with this approach and produces realistic estimates. Estimates could be improved by including more matchups from winter.

One of the main question that remains after Chapter 3 concerns the use of 15 bands to develop the algorithm. Are all 15 bands really required? This is the main subject of Chapter 4.

4. Assessment of waveband availability for neural networks development using multi-spectral ocean colour sensors

4.1. Introduction

The previous chapter developed a neural network using matchups between *in situ* chlorophyll (Chl) collected in European northwest shelf seas and 15 reflectance wavebands at top of atmosphere collected by MODIS Aqua sensor. No other algorithms ever used the 15 MODIS Aqua bands to evaluate Chl in coastal waters nor go full spectrum with a multispectral sensor, which may partly explain the success of the approach, especially regarding the use of noisy satellite data, impacted by other contributors such as glint or haze, which would usually be flagged by other algorithms. One potential explanation is that rather than applying an atmospheric correction process in coastal waters, including more light information related to different noise candidates such as band 469 nm related to aerosol optical thickness (Glantz et al., 2012) or the common bands in the infrared used to perform the atmospheric correction helped the network. The most comparable work to what has been presented in Chapter 3 comes from Hieronymi (2017), when 11 bands were used to cluster the light signal after normalisation and water type specific dedicated NN algorithms, based on modelled data, were applied. Otherwise, when applied to real matchups, most operational algorithms usually still rely on blue and green bands in open waters (O'Reilly et al., 1998, Gohin et al., 2002), even though other algorithms exist in coastal waters using red (Gons et al., 2002) or NIR bands (Moses et al., 2012).

In data prediction and classification of hyperspectral images, when 100 to 400 wavebands are available in the visible and/or near-infrared spectrum, most recent NN techniques do not rely on the full spectrum, but usually apply data reduction methods. The main reasons why such methods are required are mainly due to the highly redundant information in the spectral domain, resulting in little variability in neighbouring bands, which may lead to the curse of dimensionality, i.e. less samples than the number of spectral bands as variables. This easily causes overfitting issues and requires much increased computational complexity, yet produces degraded performance (Zabalza et al., 2014a, 2014b). The goal of data reduction here aims to reduce the data redundancy whist preserving the most important information for ease of the training process and improved data modelling. Typical methods here include those using band selection and feature extraction. The former selects a specific subset of the bands, based on band clustering or band ranking (Ren et al., 2014; Sun et al., 2021; Sun et al., 2022), and the latter

reduces the data redundancy by some feature transforms such as Principal component analysis and its variations (Zabalza et al., 2014a, 2014b; Zabalza et al., 2015).

The algorithm developed in Chapter 3 used all 15 bands from the Vis-SWIR that were saved during the data processing. This reflects MODIS design and our understanding of the utility of each waveband, for example not processing long wave infrared bands because they are related to the temperature signal and their addition could be harmful due to the potential seasonality links between chlorophyll and temperature signals. Among the 15 bands used, some are very close to each other, such as bands at 547 and 555 nm, or 859 and 869 nm. The main differences between different bands apart from their wavelengths are their width (from 10nm for the ocean colour specific bands to 50 nm in the infrared), the spatial resolution (250m, 500m or 1000m at Nadir for MODIS), and the signal-to-noise-ratio (SNR) associated, which is usually linked to the previous two. For example, bands at 859 or 555 nm are dedicated to land observation which requires lower SNR due to increased backscattering signal by lands (usually 250 (unitless) compared to more than 500 required for ocean colour), and in exchange leads to higher saturation threshold limits. To the extent of the our knowledge, this is the first attempt to include 15 bands and this thesis considers how appropriate it might be to use 'land' or 'atmospheric correction' bands for a Chl algorithm.

The purpose of this chapter is to test the impact caused by removing spectral information and establishing what performance would have been reached if the initial band selection had followed previous practice in the field closely. This chapter is based on the exact same dataset used in Chapter 3 and most experiments concern addition or removal of particular wavebands, or combinations of wavebands. Two methods are selected for band selection. First, this chapter relies on the ocean colour literature by either replicating wavebands used in existing algorithms or testing hypothesis of specific waveband impacts (Section 4.3.). One of the main reason for such tests comes from the potential capacity to build robust algorithm that use less bands, and therefore can be developed for smaller and more numerous satellite sensors, or shared directly with other sensors. Next, a data reduction method developed in Sun et al. (2020) is applied to classify and select from 1 to 15 bands to create an algorithm with each combination (Section 4.4.) and thereby establish a hierarchy of the most useful waveband contributors. Finally, the addition of more bands for Chl algorithm design compared to the blue green bands used in open water algorithms is conducted using independent datasets (Section 4.5.). Two independent datasets are used. First, the SeaBASS dataset that provides MODIS Aqua remote sensing matchups over different places around the world. Data from the SeaBASS dataset were mostly sampled

for case 1 waters conditions (because the OC3 algorithm works well to make estimates there, which means that there is no impact by the presence of sediments). However, some of the samples may come from case 2 conditions near coastal areas. This dataset is therefore representative of different optical conditions. This dataset will allow seeing if the technique of this chapter promoting the use of as many bands as possible can be applied to another independent dataset. The second dataset used is called the NOMAD dataset, which comprise *in situ* radiometry matchups. By definition, this dataset cannot be affected by the atmosphere which allows an evaluation of the technique can when applied to any kind of inversion problem in ocean colour. If the same observations are made with this dataset, the method is general and can be extended to any remote sensing inversion problem and considered as the main method to follow in future.

4.2. Material and methods

The matchups used in this study are those described previously in Chapter 3. Figure 3.2a from Chapter 3 shows where *in situ* data matched MODIS Aqua light signal, with only the atmospheric failure flag being used to further select data. To summarise, *in situ* data were collected by different institutions and methods in northwest European shelf seas from the first 10 meters of the ocean using any sampling technique available. MODIS Aqua Rhot signal (at top of atmosphere) was processed between 2002 and 2020 for any image available in the area (48°N to 65°N, 25°W to 13°E), and matchups were created using a 08:00 to 16:00 temporal window to be able to compare it with daily algorithms available in the region. Any *in situ* Chl samples available inside a pixel meeting the depth and time requirement were averaged (median) and a matchup was created between the *in situ* value and the 15 available bands. 15 765 matchups are available.

4.2.1. Band combination used

Exact wavelengths and SNRs of the 15 bands saved during the process are given in Table 4.1 along with the different band combinations used by each experiment. Different selections were made based on prior physical knowledge or to replicate waveband availability for other ocean colour sensors (Section 4.2.1.1. and 4.3.) or as determined by machine learning results (Section 4.2.1.2. and 4.4.). Details on each of these data selection approaches are given below.

4.2.1.1. Manual selection

Different wavebands were manually selected for this experiment, mostly based on literature. Experiment **A** is the direct application of the Chapter 3 algorithm, using all 15 bands available. First, SWIR

bands have not been used in any algorithm development in the past and were therefore removed for experiment **B**. Redundant signal present in **B** from neighbouring bands at 555, 645 and 859 nm with lower SNRs was removed in **C.** Band at 469nm impacted by the atmosphere was further removed for **D**. **E** uses only ocean colour dedicated wavebands. **F** switched the use of IR bands from 869 to 748 nm to evaluate their respective impact. **G** proposes total removal of blue information based on observation made by El Habashi et al., (2019) and Gilerson et al. (2021). **H**, **I**, **J** and **K** experiments try to simulate others sensors by restricting the band availability to what is available within MODIS Aqua. **L** and **M** are restricted to data commonly used by European agencies (OC-CCI and OC5). **N**, **O** and **P** are testing performance for RGB algorithms as done previously (D'Alimonte et al., 2003) with **Q** including a fourth band at NIR in the process. **R** and **S** use knowledge learned from the machine learning clustering results from Section 4.4. and use the best performers. **U** and **T** add the simulation of a 510 nm band missing from MODIS Aqua by taking the median of 488 and 531 nm bands, turning it into a pseudo 510nm band, to evaluate if this specific band available for any other sensors would improve performance.

Table 4.1: 15 bands available for the MODIS Aqua sensor and the combination used for each experiment.

	Waveband (nm)	412	443	469	488	531	547	555	645	667	678	748	859	869	1240	2130
	SNR	880	838	243	802	754	750	228	128	910	1087	586	201	516	74	110
Exp	Spectrum	VISIBLE											NIR		SWIR	
A	All (15, from Hadjal et al. 2022)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
B	SWIR Removed (13)	x	x	x	x	x	x	x	x	x	x	x	x	x		
C	No repetitive signal (10)	x	x	x	x	x	x			x	x	x		x		

D	469 band impact (9)	x	x		x	x	x			x	x	x		x		
E	Ocean colour bands (8)	x	x		x	x	x			x	x			x		
F	MODIS Visible (8)	x	x		x	x	x			x	x	x				
G	OC minus BLUE (7)				x	x	x			x	x	x		x		
H	SeaHawk-like (8)	x	x		x	x	x			x		x		x		
I	SeaWiFS-like (7)	x	x		x		x			x		x		x		
J	MERIS-like (8)	x	x		x		x			x	x	x		x		
K	CZCS-like (5)		x			x	x			x		x				
L	OC-CCI (6)	x	x		x	x	x			x						
M	OC5 (5)	x	x		x	x	x									
N	RGB 1 (3)	x					x			x						
O	RGB 2 (3)		x				x			x						
P	RGB-NIR (4)		x				x			x		x				
Q	3 Best (ADBH)		x			x				x						
R	6 bands #1 (ADBH)	x	x			x	x		x				x			

S	6 bands #2 (ADBH)	x	x			x	x		x					x		
T	SeaWiFS-like and 510 (8)	x	x		x*	x*	x		x		x					
U	All and 510	x	x	x	x*	x*	x	x	x	x	x	x	x	x	x	x

The symbol * used for T and U experiments means that the 510 nm band was created using the average of 488 and 531 nm bands. Experiment U differs with experiment A by having 16 bands (the interpolated 510 nm is included along with all 15 bands).

4.2.1.2. Classification methods from Sun et al. (2020)

The 15 bands available can be clustered to reduce their number if a repetitive signal is present between two or more of them. The Enhanced Fast Density-Peak-based Clustering algorithm (E-FDPC, Jia et al., 2015) was used to cluster the bands into subsets, followed by ranking the importance of each band and band subset. It uses the different band characteristics and their local density and compares them to the nearest high-density bands. The bands are normalised before the operation (log10).

When grouping the bands, each of the noisy bands tend to form a separate group due to their les similarity to the neighbouring bands. Therefore, an adaptive distance-based band hierarchy (ADBH) was introduced to tackle this particular problem (Sun et al., 2020). To this end, the data redundancy between different bands has been effectively reduced by grouping neighbouring bands into a subset, followed by selecting the representative bands based on ranking. Eventually, the reduced bands can be taken as features to train the machine learning algorithms for data modelling and prediction, including regression analysis and classification. The machine learning methods may include support vector machine (SVM), artificial neural networks (NNs), random forest, etc. The data set available here contains only 15 bands, however, some bands are close to each other, or may contain a signal already largely present from nearby bands, therefore band selection methods could help understand the impact of each specific wavelength available.

4.2.2. Independent matchup datasets: SeaBASS and NOMAD

Two external datasets were used to test the hypothesis of "using more spectral information leads to better performances". They were downloaded from online sources (https://seabass.gsfc.nasa.gov/ and https://seabass.gsfc.nasa.gov/wiki/NOMAD, accessed on 11 July 2022). NNs have been developed using information they contain to evaluate the impact of inclusion of more light information for performance improvements.

SeaBASS (SeaWiFS Bio-optical Archive and Storage System) dataset contains matchups between *in situ* Chl samples within the first 10 meters and MODIS Aqua level 2 Rrs. Satellite data were atmospherically corrected using the standard Nasa algorithm (Gordon and Wang, 1994). Matchups consists of a 3x3 pixel box around the *in situ* sample within +/- 3 hours with at least 4 pixels available among all pixels from the box (not counting land pixels), and a coefficient of variation below 0.15 (unitless) between remaining available pixels for specific wavebands (details provided in Bailey and Werdell, 2006). The first 10 MODIS Aqua bands from 412 nm to 678 nm are available from SeaBASS dataset (see Table 4.1 for more information). Figure 4.1 shows the location and numerical distribution of *in situ* Chl samples in the dataset, with the vast majority coming from Case 1 waters and with a peak in the distribution at 10 $mg.m^{-3}$ mostly coming from the Chesapeake Bay, therefore meeting Case 2 conditions.

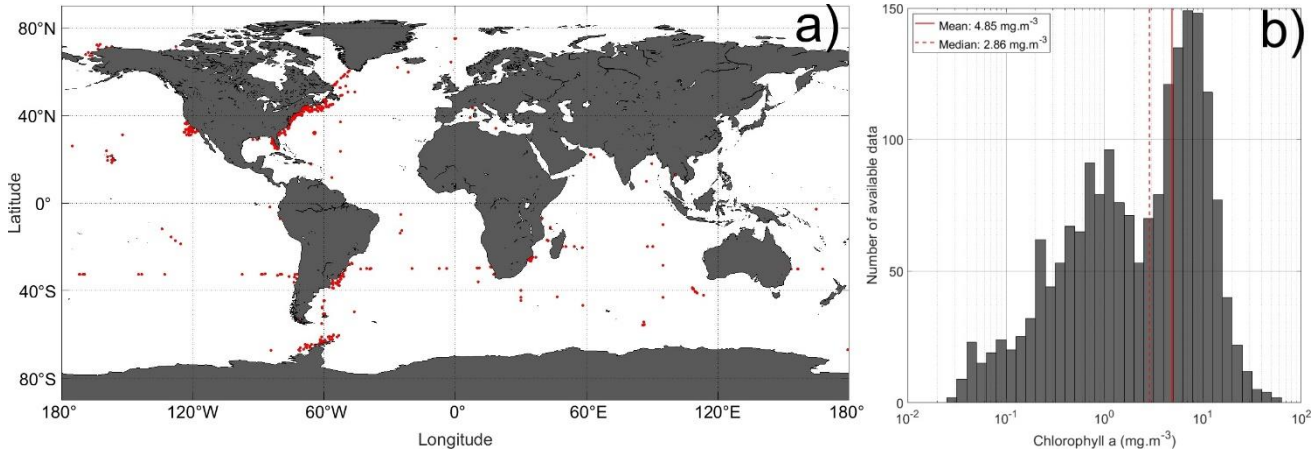

Figure 4.1: (a) Map of *in situ* matchups from the SeaBASS data set and (b) histogram of SeaBASS Chl concentrations. Most samples were collected on the USA East coast.

The NOMAD dataset (Werdell and Bailey, 2005), similarly to SeaBASS, contains high quality matchups between light and *in situ* Chl samples crossing different areas in the world (Figure 4.2a). Rather than having remote sensing light spectra, NOMAD Rrs spectra come directly from in water collection though the use of inherent optical properties and profiling instruments, or from above surface sensors. Its data were collected in mostly case 1 water conditions and 4459 matchups are available. However, because the light samples were collected using different instruments and methods, the dataset does not

contain consistently Chl and all light information for each matchup. Light information is available at 405, 411, 443, 455, 465, 489, 510, 520, 530, 550, 555, 560, 565, 570, 590, 619, 625, 665, 670 nm. Some bands are available more often than others are; for example, the 405 nm band is available 12 times when the 411nm band is available 4293 times. Moreover, Chlorophyll a pigments are reported using two methods inside the dataset (Figure 4.2b), using 1) fluorometry measurements (3392 samples, displayed in red in Figure 4.2b) or 2) HPLC (1381 samples). This diminishes data availability to conduct tests on a bottom of atmosphere dataset for algorithm design. To overcome the data consistency issue, comparisons are made in the results section between neighbouring Rrs bands and the two concentrations to test if they could be merged (Section 4.5.3).

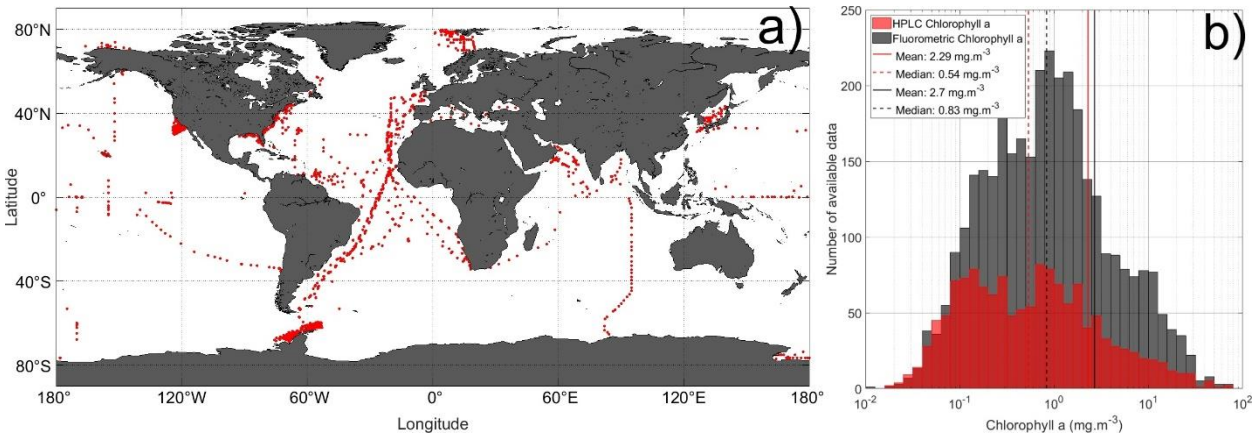

Figure 4.2: (a) Map of *in situ* matchups from the NOMAD data set and (b) histogram of NOMAD chlorophylls concentrations.

4.3. Manual selection of wavebands

The first stage of analysis is to determine optimal NN architectures for each experimental configuration detailed in Table 4.1. For each experiment, the performance of NN architectures ranging in size between 1 to 10 layers and 1 to 20 neurons per layer were systematically explored, following the process outlined in Chapter 3, Section 3.4.1. Each architecture was tested ten times using random resampling of the training data to generate training, test and validation subsets each time. Values of MAD were recorded for training, test and validation analyses for each iteration of the architecture. The final performance metrics for each architecture are the median MADs of the 10 iterations. Figure 4.3 shows results from NNs operating on TOA Rhot inputs, while Figure 4.4 shows equivalent results for NNs operating on BOA Rrs inputs for a selection of experiments. In both cases experiment **A** always provides the highest performance, **D** and **G** are have middling performance and **N** provides noticeably poorer quality results.

For both reflectance signals, there is a general reduction in NN performance as the quantity of wavebands used as input reduces.

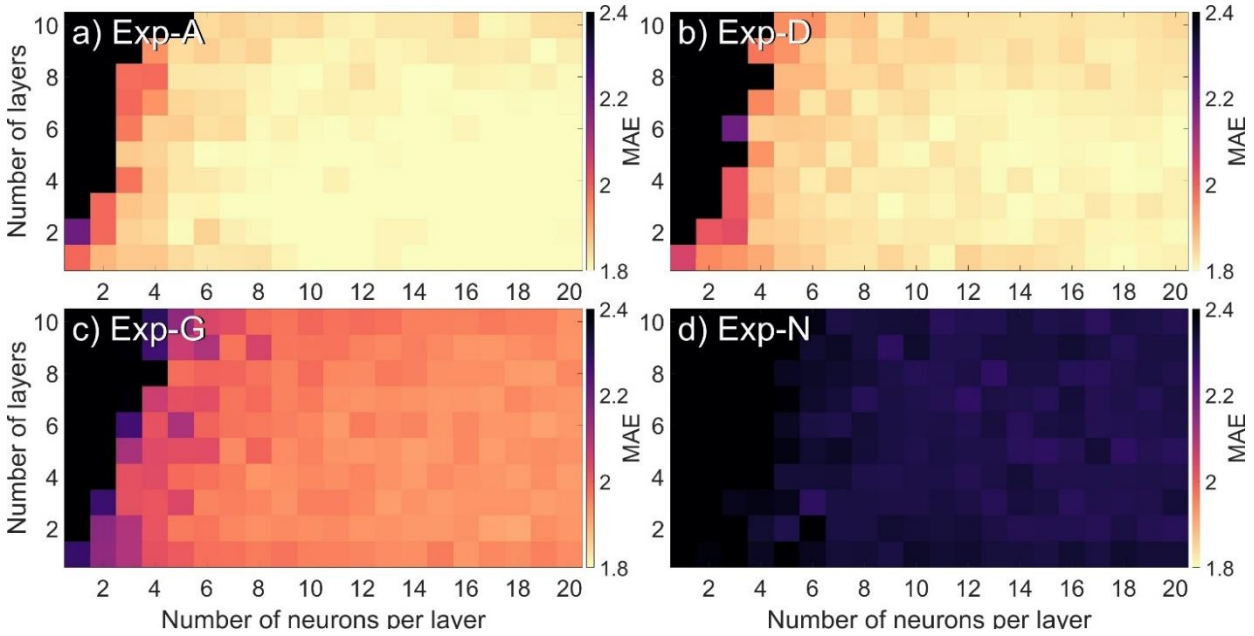

Figure 4.3: Heat maps of the neural network architectures using the test set for **Rhot** experiments. Experiments **A**, **D**, **G** and **N** are broadly representative of 4 clusters identified later.

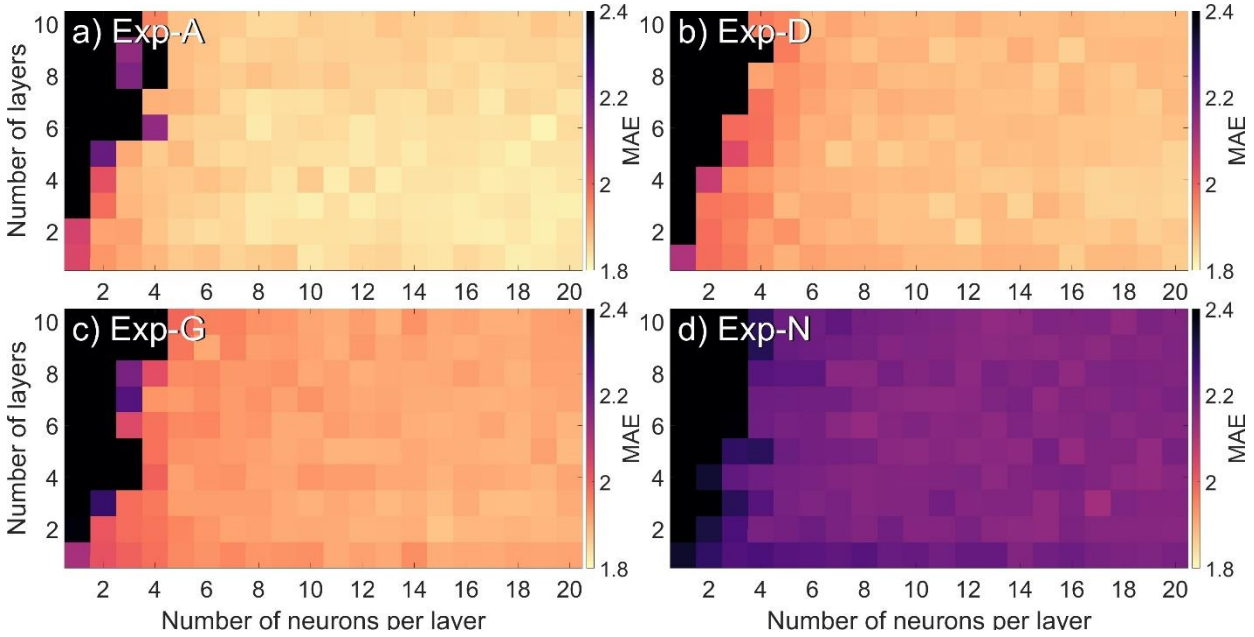

Figure 4.4: Heat maps of the neural network architectures using the test set for **Rrs** experiments. Experiments **A**, **D**, **G** and **N** are broadly representative of 4 clusters identified later.

As soon as approximately six neurons per layer are used, there are only small differences in performance for the test set across the range of architectures. There appears to be no significant benefit to extending structure to greater than 5 layers, with experiment **A** returning slightly worse results for >5 layers. In Chapter 3, we selected an architecture of 3 layers and 15 neurons per layer when using 15 wavebands. This seems to be broadly representative of near optimal performance in each of the cases shown in Figures 4.3 and 4.4, even though these represent a wide range of varying input wavebands. The dark regions in each plot occur when NN have very low numbers of neurons per layer and increasing numbers of layers, producing poor quality results in every case. This is a region of NN architecture space that should always be avoided.

One important observation concerns a limiting feature of using MAD as the performance metric: MAD does not have the capacity to highlight underfitting issues affecting smaller architectures. For example, an architecture using only 1 layer of 3 neurons for the **N** experiment that contains 3 input bands returns similar performances to other architectures according to the MAD score. However, underfitting issues become clearly apparent when networks struggle to provide consistent distribution around the 1:1 line (Figure 4.5d). The reason this particular network underfits comes from the limited number of neurons used. Selecting the same number of neurons per layer as inputs leads to underfitting issues when this

number is small. As soon as 10 neurons were used to create an RGB algorithm for experiment **N**, underfitting issues disappeared, but MAD performance remained very similar.

This feature is further elucidated in Figures 4.5 and 4.6 which represent the test set matchup performances for 3xi NNs using each experiment's waveband combination for Rhot and Rrs, respectively, and i number of neurons per layer. As observed in Chapter 3, Rhot performs better than Rrs for every experiment except **N** where there is a clear under-fitting issue (Figure 4.5d) to lead to a total failure for this specific experiment when using Rrs (Figure 4.6d). In this case, the NNs return values around 1-2 $mg.m^{-3}$, close to the median of the training data set. Under-fitting tends to happen when too few neurons per layer are used with not enough inputs. Experiment **N** only uses 3 wavebands as input (3 x 3) but testing a larger architecture (3 x 15, Figures 4.7 and 4.8) did not completely solve the under-fitting issue. While the obvious failure of returning a single value from the network due to its incapacity to converge using these three wavebands disappeared, the underfitting problem is still visible as Chl is still estimated with values similar to the median of the distribution in both cases (4.7d and 4.8d). The conclusion is that the information carried by these three bands is insufficient to make accurate predictions for this data set. Increasing the number of neurons per layer beyond the number of wavebands available did not improve MAD performance, and as soon as the network converges to a global minimum with three layers of 15 neurons. These results are not included in this chapter because the network already converged. Similar results should be expected for bigger architectures.

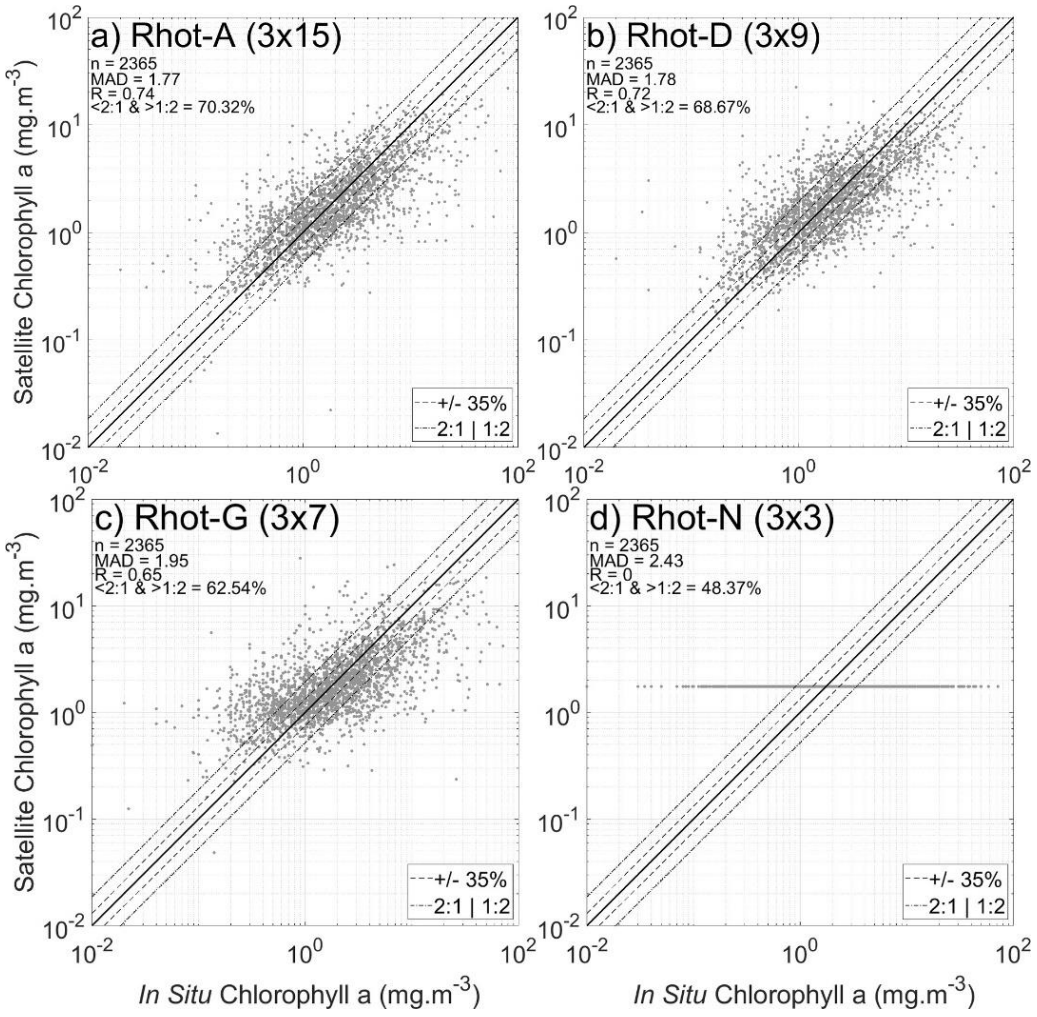

Figure 4.5: Matchup performances of the **Rhot** test sets using the experiment shown in Figure 4.3. The number of neurons is following the number of inputs (3 layers of ith neurons).

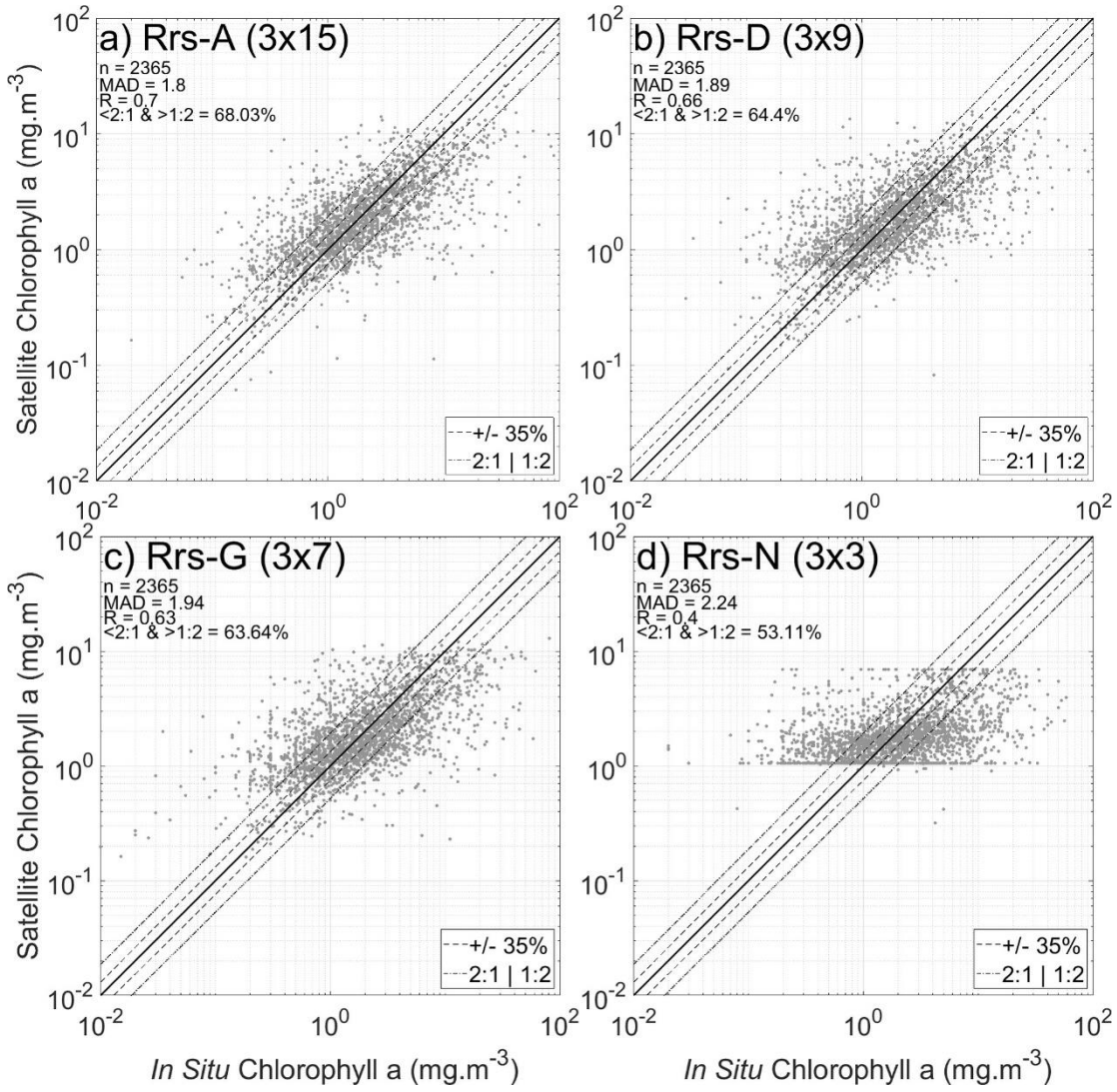

Figure 4.6: Matchup performances of the **Rrs** test sets using the experiment shown in Figure 4.3. The number of neurons is following the number of inputs (3 layers of ith neurons).

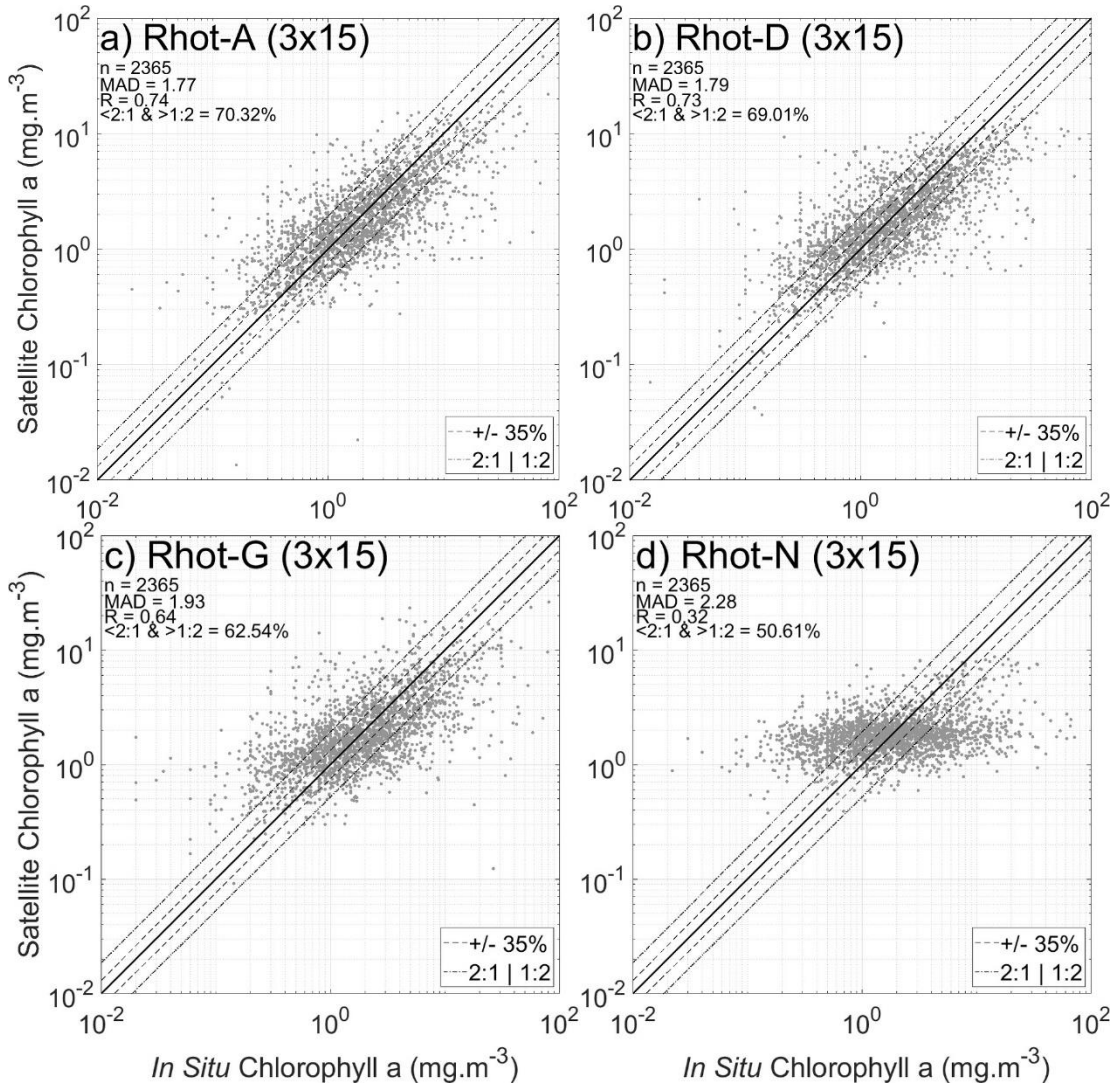

Figure 4.7: Matchup performances of the **Rhot** test sets using the experiment shown in Figure 4.3. The architecture used is shown in parenthesis (3 layers of 15 neurons each time).

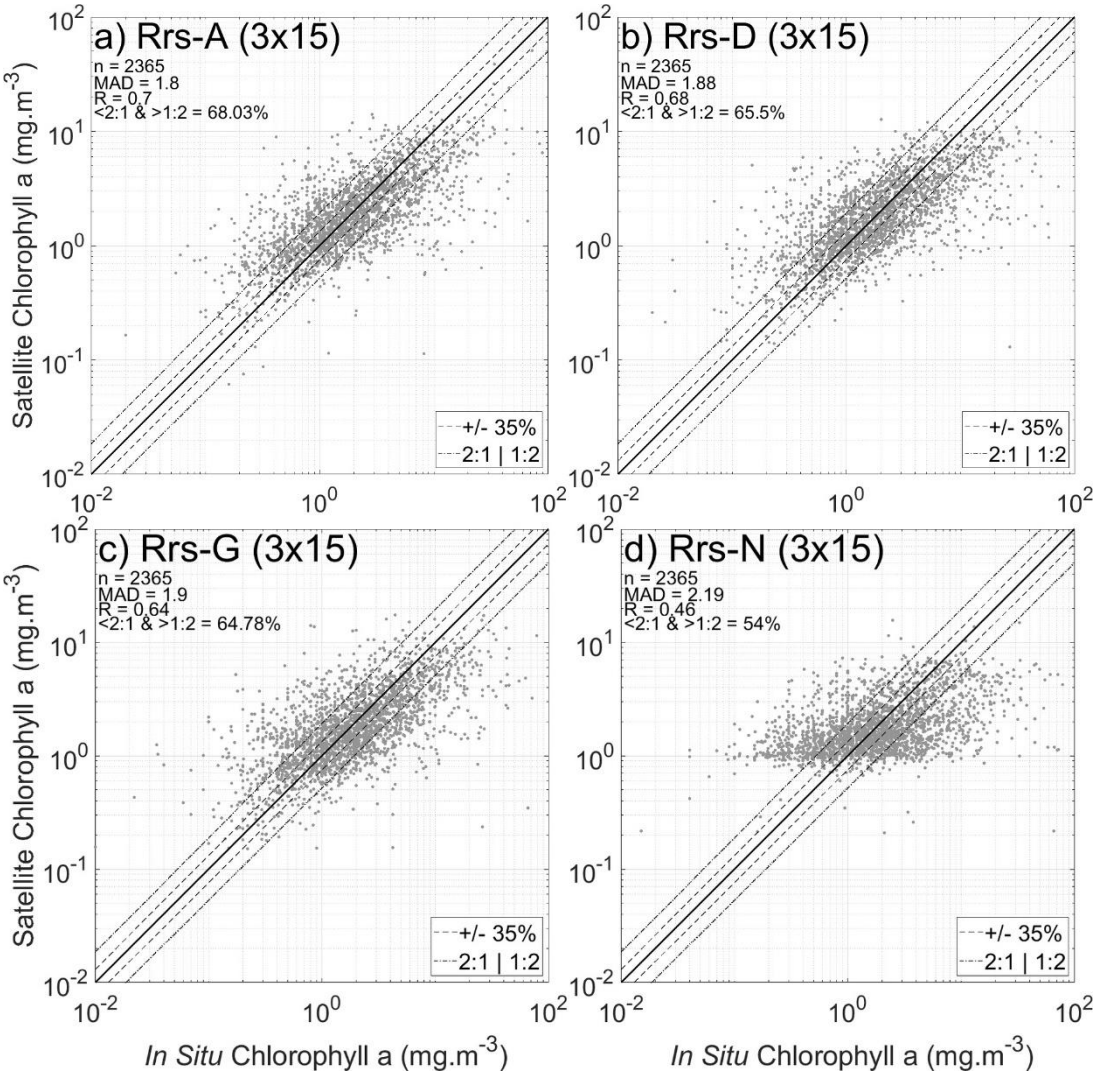

Figure 4.8: Matchup performances of the **Rrs** test sets using the experiment shown in Figure 4.3. The architecture used is shown in parenthesis (3 layers of 15 neurons each time).

A chlorophyll map example for the 14th of September 2014 is shown in Figure 4.9 (using Rhot) and Figure 4.10 (using Rrs) for each of the previous 4 experiments (**A**, **D**, **G** and **N)** using 3 layers and 15 neurons per layer. The first three algorithms (panels a to c) return similar Chl distributions which confirms the proximity in performances illustrated in Figures 4.5 and 4.6. Panels d from Figures 4.7 and 4.8 show a general failure, with reduced variability in returned Chl values across the region. This illustrates the under-fitting problem discussed previously, with these NNs returning values around the median of the distribution: 1 $mg.m^{-3}$.

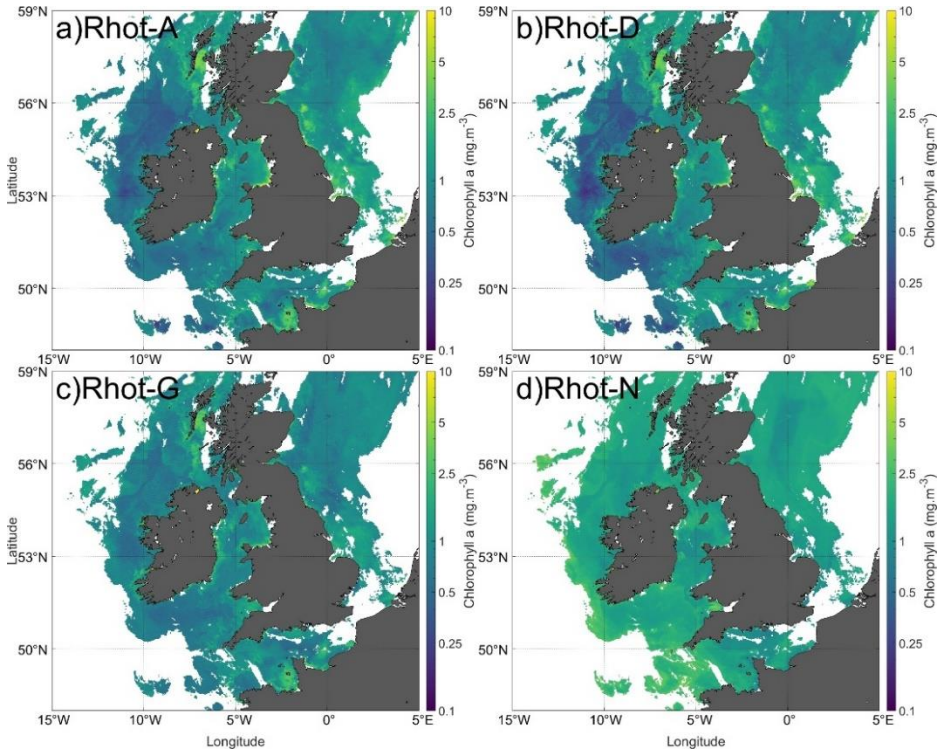

Figure 4.9: Chlorophyll maps created with Rhot NNs using experiments A, D, G and N waveband combinations. MODS Aqua, 14 April 2014.

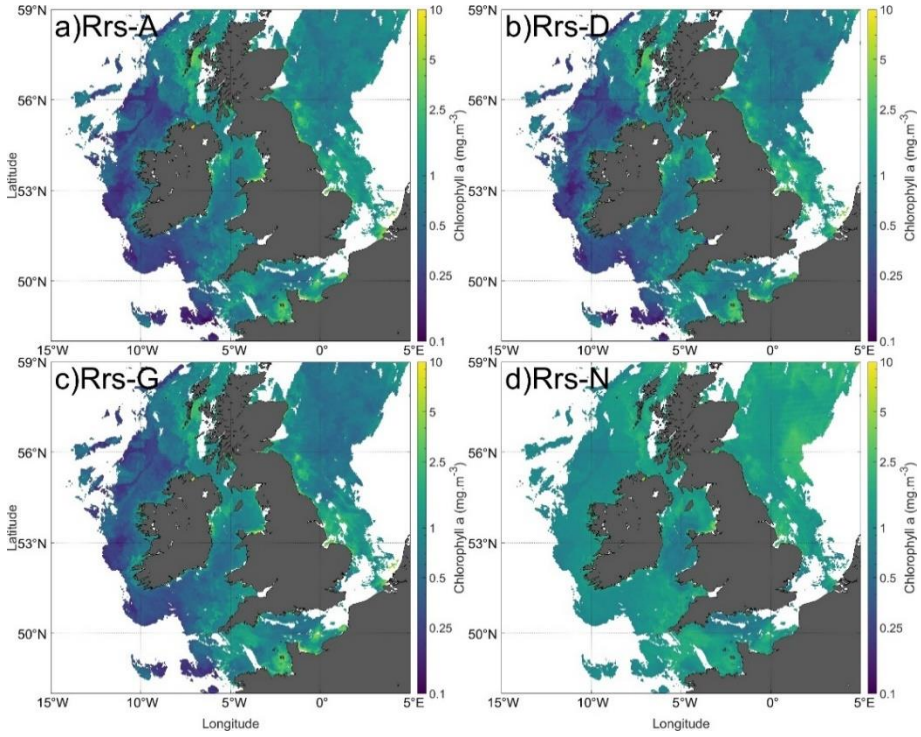

Figure 4.10: Chlorophyll maps created with Rrs NNs using experiments A, D, G and N waveband combinations. MODS Aqua, 14 April 2014.

The median of the MAD values for all NN architectures tested (1 to 10 layers, 1 to 20 neurons per layer) for each experiment (**A** to **S**) are shown in Figure 4.11, for both the test and training sets, and for Rhot and Rrs. The training and test performances are generally close in all cases, with the training sets returning systematically better performances than the test sets. This has already been observed before in Chapter 3 and could highlight some overfitting issues coming from a limited available dataset, not allowing the test set to be entirely representative of the train set with only a few thousand points. In most cases NNs operating on Rhot appear to perform better than Rrs equivalents, with the exception of experiments **N**, **O**, **P**, **Q** which represent 3 and 4 waveband inputs and where performance is generally weakest.

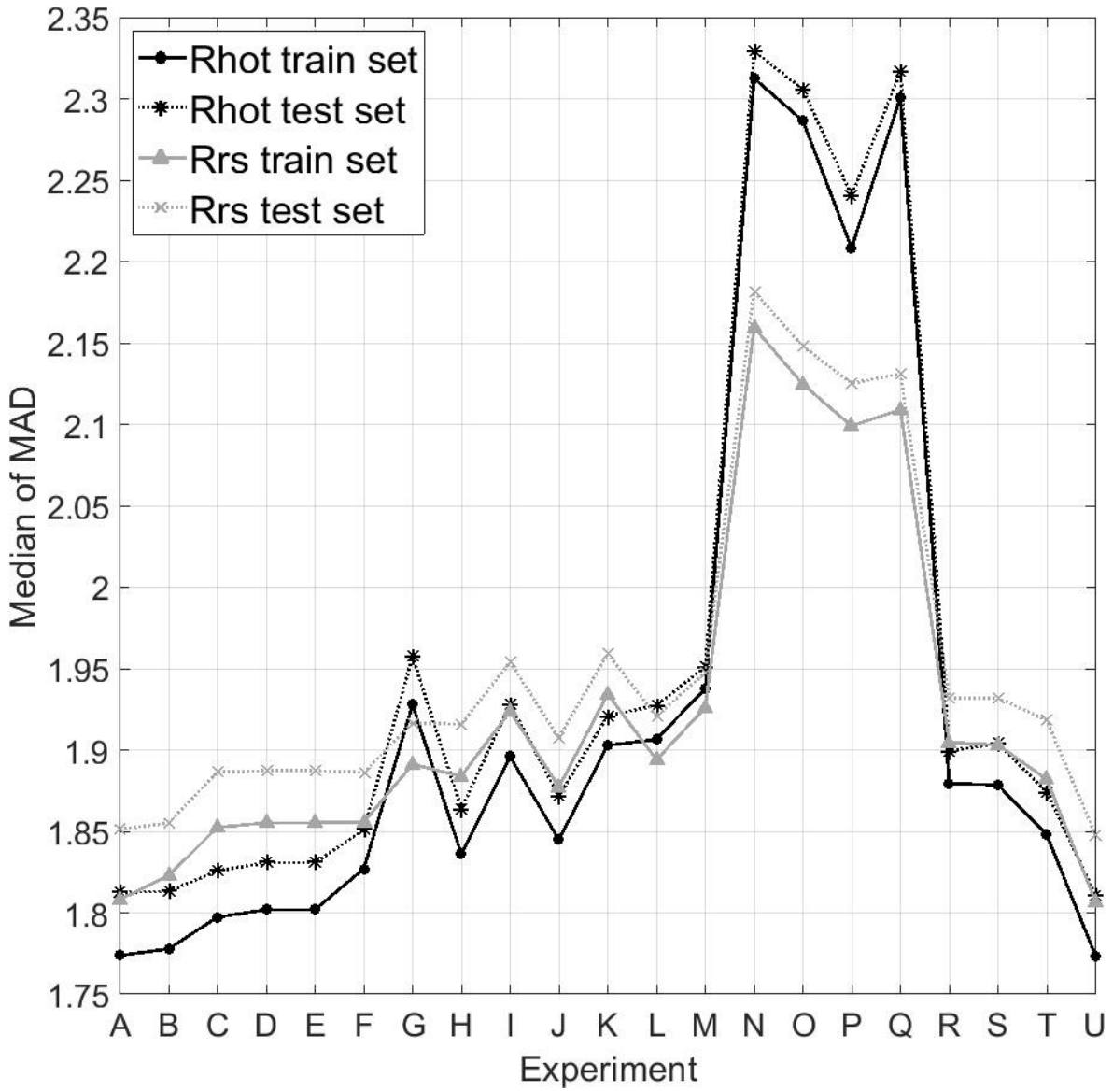

Figure 4.11: Median of the mean absolute differences for all band combinations using Rhot (black) or Rrs (grey) for the train (solid line) and test sets (dashed line).

Results shown in Figure 4.11 can be clustered into 4 groups, with **A**, **B** and **U** being the best performers; **C** to **F** showing marginally degraded performances; **G** to **M** and **R**, **S** and **T** producing further degraded quality performance; and **N** to **Q** producing poorest quality data with limited operational value. Considering Rhot results in the first instance (as they are generally superior to Rrs equivalents), Figure 4.11 shows that using the maximum number of wavebands available (15, Experiment **A**) provides the best performance (MAD close to 1.8). Further improvement in performance can be obtained by selecting a specific, nearer-optimal architecture such as the 3 layers x 15 neurons architecture used in Chapter 3. Experiment **B** (removing the SWIR bands) produces results that are very close to **A**, highlighting that inclusion of the 1240 and 2130nm band produces only marginal benefits across this data set. However, it should be noted that this may be because the data set has relatively few very high turbidity stations where the influence of SWIR bands is expected to be most beneficial. It may be the case that inclusion of SWIR bands provides useful benefits in specific places, but that this is not well reflected in the general metrics for the entire data set. A data set that provides more information on the other optically significant water constituents, e.g. minerals / sediments, is needed in order to establish this.

The next cluster of experiments (**C** to **F**) involves successive removal of various wavebands. Experiment **C** starts the process by removing low SNR bands that have nearby, higher SNR bands available. Interestingly this causes a noticeable increase in the MAD (+0.03) suggesting that adding more light information improves performance, even if similar bands are already present and that the lower SNR bands may still be useful for ocean colour purposes. Experiment **D** removes the 469nm band from **C** causing a further deterioration of (+0.02 MAD) in performance. There is no significant difference between **D** and **E** (removal of the 748nm band) which means that either the signal carried by the 748 nm band is already present somewhere (presumably the 667, 678 or 869 nm bands), or that the 748 nm band is not improving the results for some other unknown reason. Moreover, switching from the use of 748 to 869 nm (experiment **F**) further reduces NN performance, which may support the suggestion that the information provided by 748 nm waveband is largely present in one or more of the red wavebands.

Experiments **G** to **M** are classified into a third cluster of performers, mostly because they use fewer bands than **A** to **F** experiments. **G** removes the “deep” blue bands from the inputs (412 to 469 nm) and performance deteriorates significantly compared to **D** (+0.1 MAD). Experiments **H** to **L** are trying to simulate other sensors using the available MODIS Aqua bands and show results close to MAD = 1.9. It is important to note that in reality these sensors provide more bands than those available here, especially a band near 510-520 nm that is not available for MODIS. Consequently, I anticipate that development of

NNs using the appropriate data set for each of these sensors would perform better than the results presented here suggest. Experiment **T**, where a pseudo 510 nm band was added to the SeaWiFs-like waveband set, supports this by improving the MAD by 0.05.

The 4 RGB experiments (**N** to **Q**) provide distinctly lower quality results, with **P** being better than the 3 other, mostly because it uses a 4th band in the NIR. The last three experiments **Q**, **R** and **S** were conducted after the results from the ADBH method were available and are discussed in next section. Experiments operating on Rrs rather than Rhot broadly follow the same patterns as discussed above with the exception that the Rrs RGB experiments (**N** to **Q**) perform better than equivalent Rhot versions. As discussed above, Figures 4.5 to 4.8 show that the NNs are under-fitting for these waveband combinations so there is no particular interest in further pursuit of these combinations as viable algorithms going forward. The other major observation for Rrs is that experiment **G** (no blue bands) performs better than the Rhot version. This suggests that the Rhot NNs benefit from inclusion of the 412 and 443 nm wavebands, while the Rrs NNs are less affected by their removal, possibly because the BOA signals are less reliable due to limitations in the atmospheric correction.

4.4. Ranking waveband influence using a machine learning approach

In this section, the Adaptive Distance-Based Band Hierarchy method (Sun et al., 2020) was used to cluster and rank the importance of available Rhot wavebands. There is no parameter to tune for the ADBH method, which is a strength from a machine learning perspective and the same classification is returned each time data are processed. The resulting waveband ranking is presented in Table 2. The ranking represents decreasing impact on signal retrieval i.e. the most important waveband is 859 nm ranked first, followed by 547 nm, then 412nm and so on.

Table 2: Results from the Adaptive Distance-Based Band Hierarchy clustering method.

	Wavelength														
Order of selection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rhot	859	547	412	443	667	1240	469	748	531	645	488	2130	555	678	869

The 859 and 869 nm NIR bands are the first and last band selected, suggesting a degree of redundancy though it is surprising that the lower SNR 859 nm band was ranked highest. The green (547 nm) and deep blue (412 nm) bands were selected in second and third position respectively, with the 412 nm band ranked higher than the adjacent 443 nm band, and the red 667 nm band coming next. Interestingly, the 1240 nm waveband is in sixth position while we saw earlier in Section 4.3 that it did not seem to improve the results (no real difference between experiment **A** and **B**). Similarly, the low SNR 469 nm band impacted by atmospheric content is returned before the 748 nm near infrared band, used for atmospheric correction. The low SNR 645 nm waveband is also ranked before the 488 and 678 nm wavebands. The latter is used to identify sun induced chlorophyll fluorescence signals that is expected to be a useful discriminator in coastal waters, more than the 645 band.

Results of the inclusion of each band are shown in Figure 4.12. Following the approach used in Figure 4.11, the MAD values reported here are medians for all network architectures tested (1 to 10 layers, 1 to 20 neurons per layer). The addition of more bands consistently improves NN performance. However, the degree of improvement is variable. Inclusion of bands ranked 4^{th}, 5^{th} and 9^{th} (443, 547, 531 nm respectively) generate massive performance improvements of -0.1 to -0.2 MAD. It is interesting to note that the degree of sequential NN performance improvement is not necessarily aligned with the ABDH ranking. This could be because the ADBH method assigns importance in a manner that is not necessarily consistent with the way that the NNs use the spectral information, or it could be the case that the 9^{th} waveband is only this significant if the previous 8 wavebands are already incorporated into the NN. This is tested in experiment **R** (Figure 4.11) where the 6^{th} ABDH waveband (1240 nm) is replaced by the band in 9^{th} position (531 nm) to produce a median MAD of ~1.9 which is significantly better than the value of 2.03 found with the ABDH top 6 ranked values (Figure 4.12, 6 bands used).

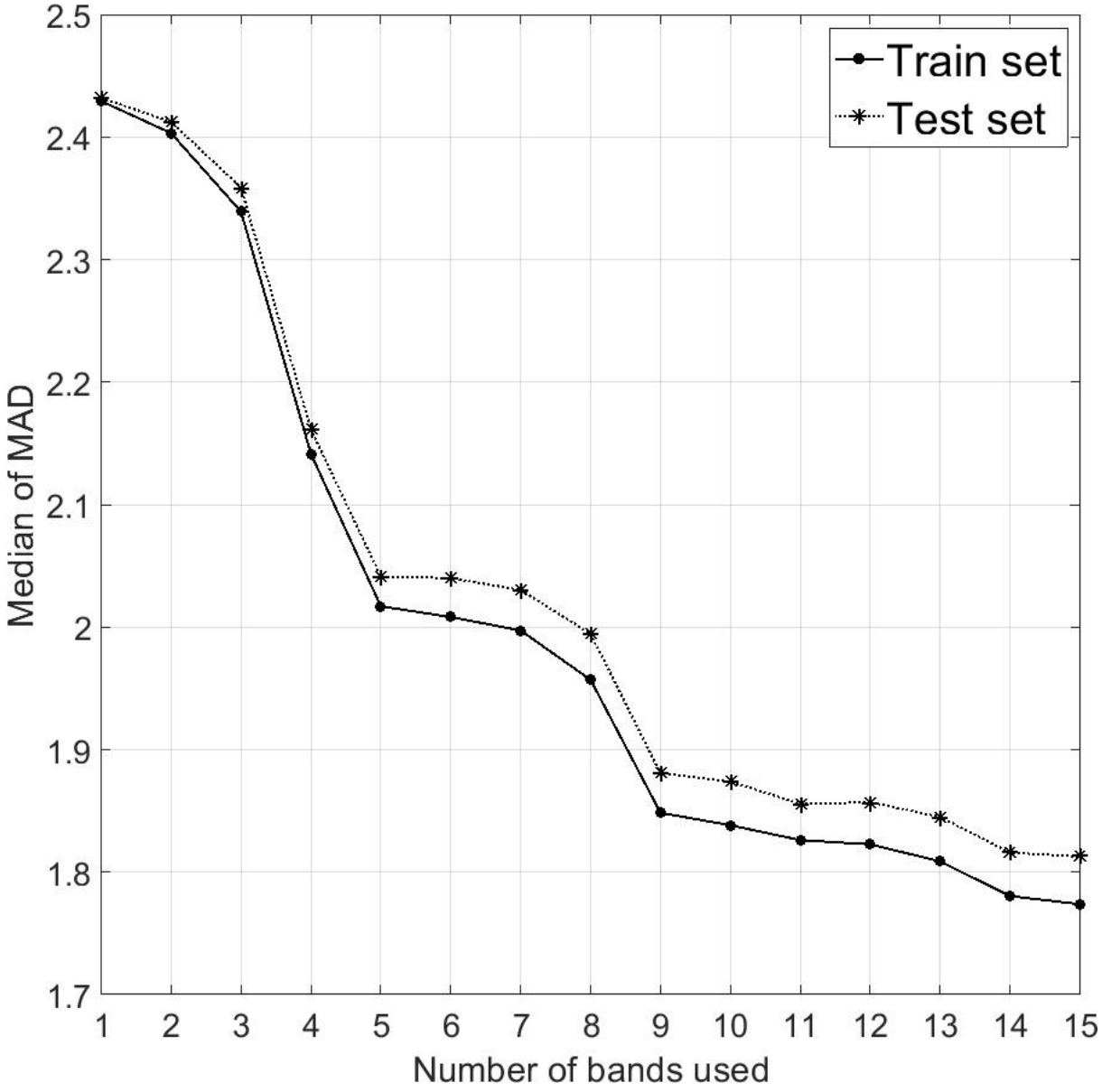

Figure 4.12: Impact on NN performance of sequential addition of each waveband in the order outlined in Table 2.

The number one ranking for a low SNR near infrared waveband (859 nm) and bottom ranking of the adjacent, high SNR 869 nm waveband is potentially a sign of redundancy in the data set. In Figure 4.11, the 859 nm waveband was used in experiment **R** but was replaced by the 869 nm waveband in experiment **S** but NN performance was virtually unchanged, suggesting there is little difference in information content.

The 443, 547 and 531 nm bands, selected in 4th, 5th and 9th position following the ABDH method showed the greatest increases in NN performance in Figure 4.12. However, when only these bands are

used in experiment **Q**, the performance is not particularly higher than other RGB combinations such as experiments **N** or **O** (Figure 4.11). The inclusion of these bands greatly improves the prediction of the network only if sufficient additional bands are already included. The waveband set used in experiment **S** forms an almost mandatory base set of wavelengths required for coastal algorithm development. The **R** and **S** 6-band NNs performed slightly better than other NNs using 6 bands such as OC-CCI (**L**) or SeaWiFS-like (**I**), supporting the idea that careful selection of wavebands is beneficial over and above simply increasing the number of wavebands in total.

4.5 Assessing impact of increased waveband availability on retrieval of chlorophyll for independent datasets

4.5.1. SeaBASS dataset

In this section, Rrs inputs from SeaBASS dataset (an open waters dataset, mostly) are used to create NNs. Networks using 3 layers of 15 neurons per layer are used for each algorithm to ease comparability. Three distinct band combinations are tested:

- OC3-NN, which is a NN using the three wavebands included in the OC3 algorithm, the current algorithm applied in open waters for MODIS (O'Reilly et al., 1998)
- 10-Rrs-NN, which uses all 10 wavebands provided within the SeaBASS data set
- All-NN, which uses all 10 SeaBASS wavebands plus additional non-optical information including the solar zenith angle, sensor zenith angle and windspeed.

Figure 4.13 shows the performance of these 3 variations of NN along with the NASA standard OC3 algorithm. It can be seen that even restricting the waveband set to the three used by OC3, the NN approach produces lower MAD scores than OC3. There is a significant further reduction in MAD when all 10 available wavebands are used in the NN. The inclusion of additional non-optical data did result in a further small improvement in performance but a much larger and more global data set would be required to fully establish the potential significance of adding this type of information to the NN training. Importantly, the results shown in Figure 4.13c demonstrate the NN approach reaching performance levels that improve upon the use of Case 1 algorithms on this data set previously presented by Seegers et al. (2018). Seegers et al. obtained MAD values of ~1.7 when a variety of Case 1 algorithms were applied to this data set. Restricting the data set to oligotrophic samples only, they found a best performance MAD of 1.4 using the GSM algorithm. Here the NN produces a MAD of 1.4 for the entire data set, generally

exceeding Case 1 algorithm performance. The NN performance on the entire data set is equivalent to the performance of the best Case 1 algorithm operating on the clearest water type.

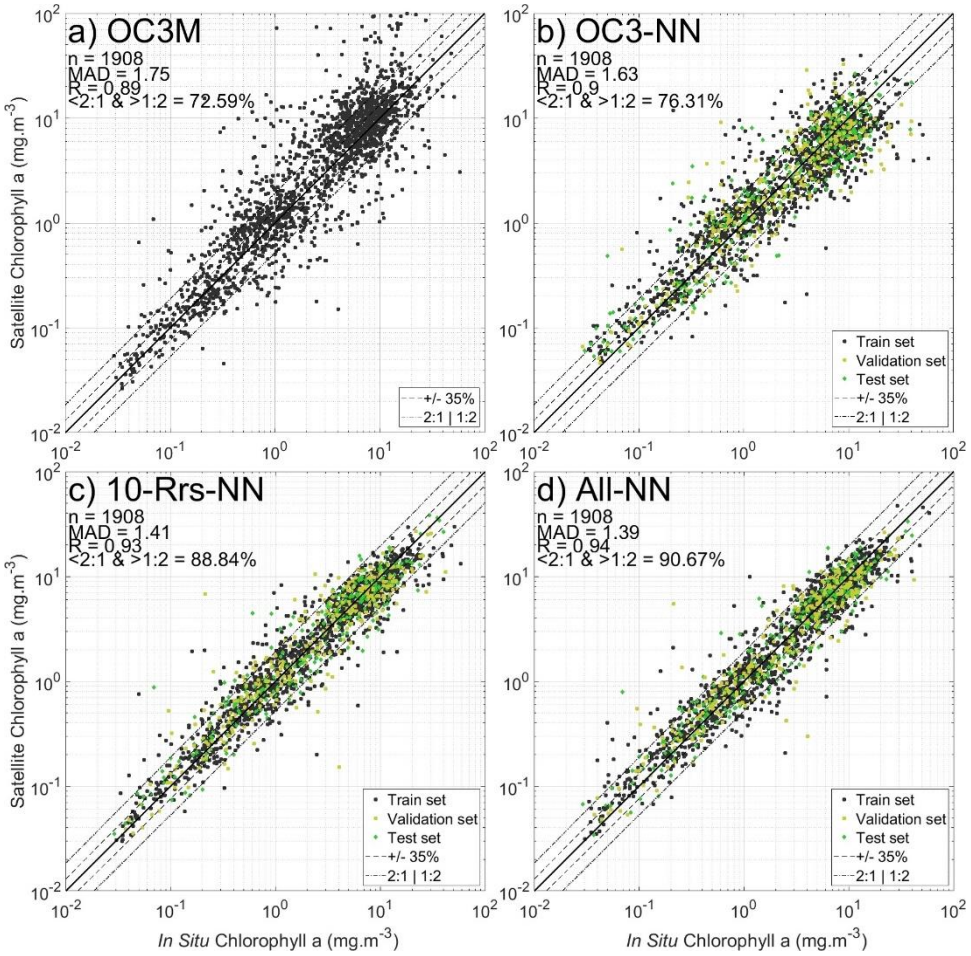

Figure 4.13: Performances over the SeaBASS dataset for a) OC3 algorithm, and a neural network algorithm trained with either b) the same 3 bands used for OC3 algorithm, c) all 10 bands available from the dataset and d) all 10 bands and the wind speed, solar and sensor zenith angles.

4.5.2. NOMAD dataset

NOMAD dataset is an *in situ* matchup dataset, therefore the temporal and spatial difference between light and concentration measurements is minimal and no impact from the atmosphere is present, which eliminates some of the issues that affect remote sensing matchups. This is the dataset in use to tune the algorithms (in open waters conditions), therefore OC3/OC4 were built using NOMAD data. Multiple sensors exist to collect *in situ* measurements, and the consistency of data available in NOMAD is variable. For example, both chlorophyll measurements come from two distinct methods, fluorometric and HPLC (Chapter 2, Section 2.1). Radiometric sensors deployed inside the ocean of just above the surface

also provide a variety of wavebands and performance, resulting in less consistency than would normally be the case for satellite matchups.

The availability of sufficient training data is a major limiting factor for NN development. Here we attempt to maximize the value of the NOMAD data set for NN training purposes by seeking to establish equivalence between the different chlorophyll samples and between adjacent wavebands in order to maximize data availability.

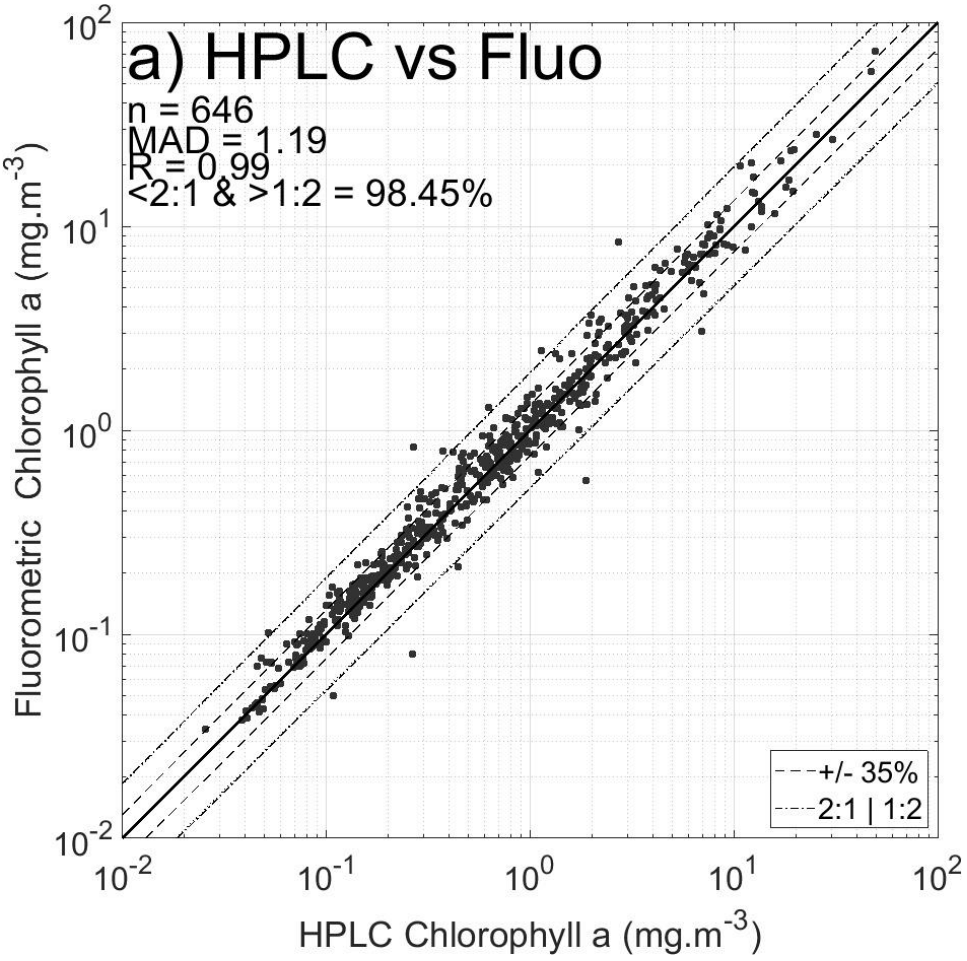

Figure 4.14: Comparison between concurrent fluorometric chlorophyll and HPLC chlorophyll a measurements from the NOMAD dataset.

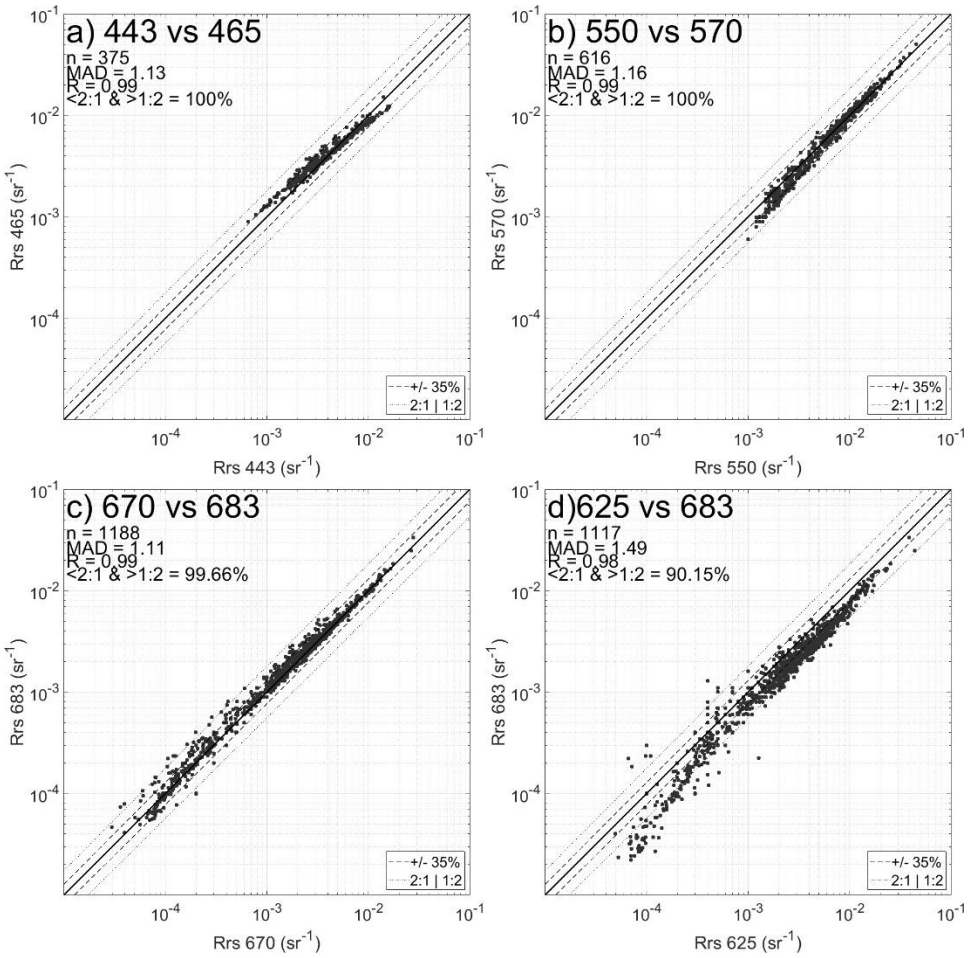

Figure 4.15: Comparison between concurrent *in situ* light measurements at different wavelengths from the NOMAD dataset. Samples may have been measured using an *in situ* or above water sensor.

Both Figures 4.14 and 4.15 use log-log scales. Figure 4.14 demonstrates that fluorometric and HPLC chlorophylls can be broadly approximated with no particular over or underestimation in the relationship in this dataset. For light measurements, different comparisons were conducted between all bands available from the NOMAD dataset and 4 of them are shown here in Figure 4.15. Figure 4.15a shows comparison between Rrs 443 and 465nm highlighting similar distribution meaning that these bands can be averaged with no major impacts when one of them is missing from the dataset. The same answer stands for bands at 550 and 570nm (Figure 4.15b) or 670 and 683nm (Figure 4.15c), with potentially slight tendencies depending on the range observed. Eventually, Figure 4.15d shows a similar distribution for bands at 625 and 683 nm with an obvious offset. Merging in such condition could potentially be harmful (by having up to one order of magnitude difference for low values), and was not included. While there is no particular metric to use to support or avoid merging, visual inspections were conducted for all bands

with their neighbours and a final merging of 8 bands was conducted. Each merging happened when MAD was lower than 1.3 between the different candidates (this is an arbitrary threshold as it provides decent correlation between the different bands). The final 8 merged bands are: 411; 443/455/465; 489; 510/520/530; 550/555/560/570; 590; 619/625 and 665/670/683 nm.

While this simple merging approach looks crude at first glance, we need to keep in mind that neural networks developed here do rely on a normalised version (between 0 and 1) of the log transformed Rrs. From Figure 4.15, relationships between bands falls between a 2:1 or 1:2 log domain for the vast majority of points. Therefore, a band with a Rrs value of 3.10^{-3} sr^{-1} merged with a neighbouring band with a value of 6.10^{-3} sr^{-1} would only make a small difference once log-transformed (-2.52 against -2.22). It will be even smaller when normalised between 0 and 1, depending on the ranges of distribution of the data. This normalisation process is required to allow NNs to be considered evenly the different spectral information and to avoid bias by relying too much onto a specific band. The different NNs are developed using these eight artificial bands and the results are shown in Figure 4.16 below. The eight merged Rrs are used for NN design to access more data (panels b to d). However, OC3 performances shown in panel (a) uses the original bands OC3 has been designed with, and not the eight averaged bands which do not really have a central waveband value.

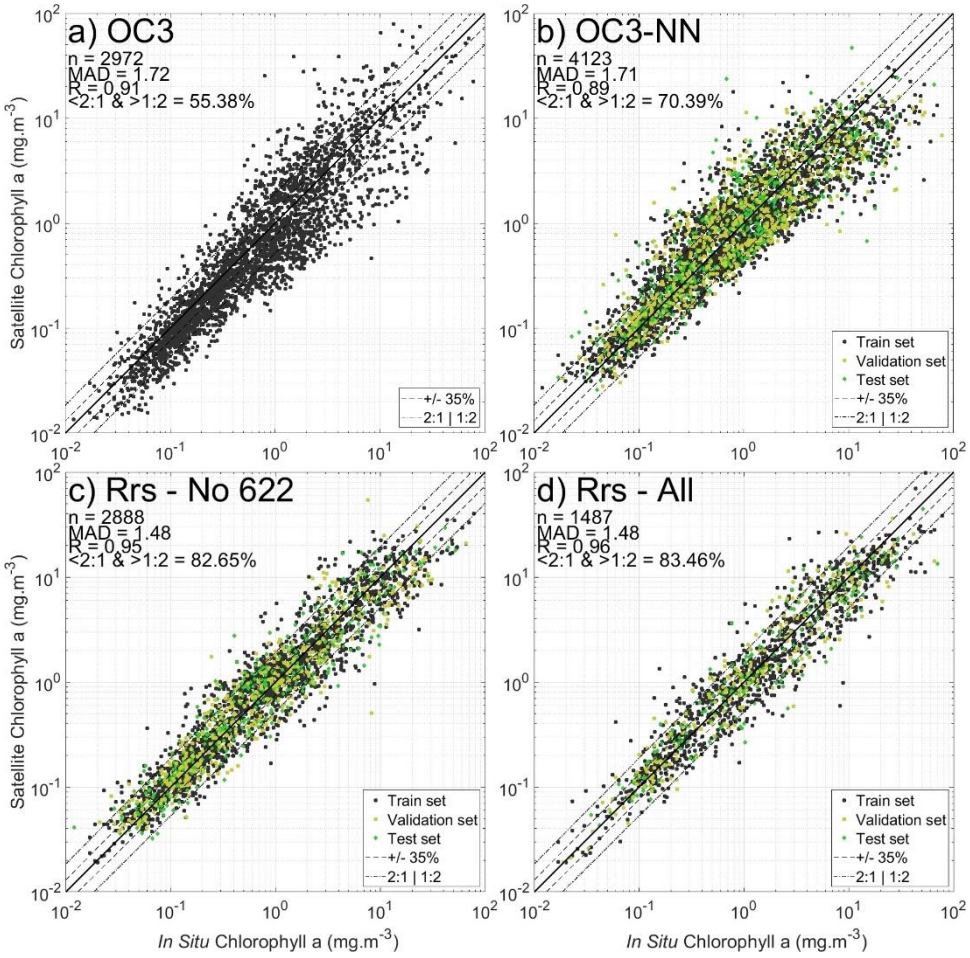

Figure 4.16: XY performances obtained on NOMAD dataset from a) OC3 algorithm, or a neural network using b) 3 merged Rrs bands from OC3 algorithm, c) all 7 merged Rrs bands not including the 622 nm band, and d) all 8 merged Rrs bands. The different parts of the dataset are shown with different colours.

First, we see that the OC3 algorithm performs similarly compared to the one obtained earlier for SeaBASS dataset with MAD of ~1.7 mg m^{-3}. The OC3 and OC3-NN algorithms return similar performances for the NOMAD data set. In SeaBASS data set, better performances were obtained for the NN, probably due to the large number of high concentrations (peak at 10 mg m^{-3}) that are easier to estimate with a fine-tuned algorithm like the NN, while OC3 was initially created without having access to many high concentration data. This explains why estimations at these ranges become increasingly scattered for OC3 (Figures 4.13 and 4.15). It highlights the need for switching to another algorithm in such conditions for open waters (O'Reilly and Werdell, 2019, Figure 3). With information carried by the blue bands being range dependent, the maximum ratio comes from the 510nm band for a SeaWIFS sensor over higher Chl ranges, but MODIS Aqua is not equipped with this band, which explains why performance decreases.

Two other algorithms were created. The first obvious choice is the one shown in panel d, using all merged bands available. However, it comes with a limited amount of data, with only 1487 points available, which explains why an extra algorithm was designed after removal of the merged 622 nm band (619 and 625 nm). Less data is available for these two wavebands, and NNs need enough training examples to reach good performances. That said, they both reach similar performances on a single NN example with MAD < 1.5. However, Figure 4.17 shows boxplots of the performances obtained after repeating the operation 10 times, illustrating that including the 622nm band tends to improve performance and suggesting that this waveband provides useful additional information. These two algorithms encapsulate the main outcome of this Chapter. Having access to more information leads to better performances on the train set in any scenario (Figure 4.17a), with the median of the distributions reaching lower MAD scores. However, because the amount of data is limited for the "RrsAll" scenario, performances obtained on test and validation set are worse than the train set when compared to bigger datasets, and application of such algorithms would lead to overfitting issues. The aim is therefore to obtain numerically large, information-rich training data sets, though caution is needed to avoid over-fitting too.

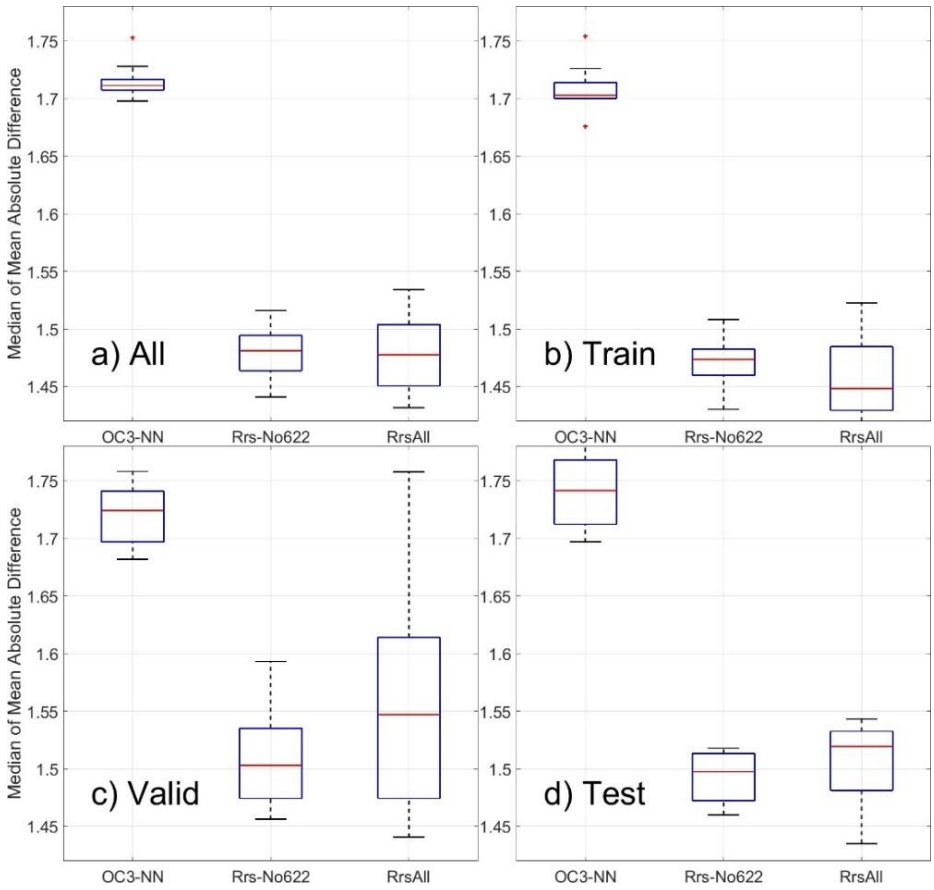

Figure 4.17: Median of mean absolute differences obtained over different subset of the initial NOMAD dataset for a) all data available, b) the train dataset, c) the validation dataset and d) the test dataset. The band combination shown on the X axis are mentioned above (Figure 4.16).

4.6. Discussion

The NN approach to retrieve Chl in coastal waters developed in Chapter 3 was founded on the basis of an *a priori* assumption that inclusion of all available wavebands would improve NN performance. The aim in this Chapter has been to test this assumption and to establish a ranking for wavebands contributing to the NNs. Within the context of the 15 multispectral wavebands considered in this study, it is clear that the original assumption that inclusion of all available wavebands would produce best results is correct. This represents a potentially useful step in the evolution of ocean colour algorithms to retrieve Chl, building on the progression from 3-6 wavebands used in empirical band ratios (OC3 to OC6 algorithms) through the 11 wavebands used by Hieronymi et al. (2017). With 15 bands available, it becomes almost impossible to test all possible combinations (2^n -1 possibilities, with n = 15 represents 32766 possibilities). As a result, two different strategies have been adopted to attempt to establish a

better understanding of the relative importance of each available waveband. The first method was based on community knowledge of the dependence of reflectance spectra on interactions between light and optically significant constituents in seawater and the format of historic ocean colour sensors that have contributed to the global ocean colour time series and involved testing by selective removal of specific wavebands. The second approach relied on a recently developed machine learning method using data densities and capacity to cluster the information to establish ranking of importance. It is clear that these two approaches do not produce entirely consistent results, probably reflecting differences in how the ABDH method treats information content compared to the NNs. However, it is also clear that both approaches provide useful information on the relative importance of different wavebands and that ultimately, it is probably best to use all available wavebands for multispectral NNs.

The NN architectures tested in this study are genuinely small compared to those used in other research fields where NNs using billions of parameters are not unusual. For the number of layers and neurons per layer tested here for Chl prediction in coastal waters, and for the case where all layers use the same number of neurons, the architecture size is not that important as soon as enough neurons per layer are used, approximately 6 for this example. This is interesting for the development of future NN multi-spectral algorithms with variable numbers of wavebands: there is little need to worry over the choice of a specific neural network dimensions so long as there is more than this threshold level of neurons per layer available. I recommend at least 2 or 3 layers to provide sufficient flexibility to account for the range of complexity found in natural waters, single hidden layers tend to underfit. There is a need for visual inspection to avoid underfitting by small NNs. However, caution is advised before extrapolation of these results to hyperspectral signals. It is as yet unclear if the extension of the number of neurons per layer to match the number of available wavebands from a hyperspectral sensor will continue to improve performance significantly or produce marginal improvement at the expense of significantly increased computational cost. Data dimensional reduction methods have been developed for other hyperspectral fields (Zhong et al., 2008; MartÍnez-UsÓ et al., 2007; Khodr and Younes, 2011) and these may become a feature of future hyperspectral NN development efforts.

In many fields researchers tend to use at least the same number of neurons per layer as the number of inputs because the inputs are more or less independent between each other. This is not true for ocean colour reflectance signals, particularly in the case of hyperspectral signals. Ocean colour reflectance signals generally vary slowly with wavelength, reflecting their dependence on equally slowly varying absorption and scattering properties of the optically significant constituents. In this study we have

seen signs of potential redundancy between closely spaced spectral bands, particularly those originally intended for different applications (e.g. 859 and 869 nm). At the same time, this Chapter has shown that the inclusion of all available multi-spectral wavebands produced our best performing NNs. For the small NN architectures discussed here, it is likely that the difference in computational efficiency associated with using all available wavebands compared to a more careful selection is not terribly relevant, and that improved performance is justification for using all available wavebands. On the other hand, given the slowly varying nature of ocean colour spectra, it seems likely that hyperspectral reflectance spectra will provide significant redundancy that will need to be carefully managed to avoid significant computational inefficiency.

It is also important to consider the results presented in this chapter in terms of geographical extent of the NW European data set and what this might mean in terms of interpreting results for specific wavebands. For example, the 469 nm band was removed from experiment **D** with little impact compared to **C**. As previously seen with the 1240 and 2130 nm wavebands, the 469 nm waveband is not mandatory to reach good performance, but its inclusion in the NN does improve prediction accuracy. This band is sensitive to the ozone concentration in the air column between Earth and the satellite sensor (Glantz and Tesche, 2012). Because the data considered in this study is from a single region and only covers approximately 1% of Earth's surface it is possible the NNs presented here do not seem to rely much on this input but that this might be more significant for development of a global NN algorithm.

Redundancy from the 748 nm band compared to either the 6xx nm or 8xx nm bands from MODIS could explain why removing it from the training did not change performances of the network's prediction. It is supported by experiment **F** (MODIS visible only) compared to **E** when the 869 nm band is replaced by the 748 nm and performance decreases. The signal from the infrared carries more significantly different information than the signal at the edge of the visible spectra. The removal of the 412 and 443 nm wavebands (experiment **G**) following previous work by El Habashi et al. (2019) and Gilerson et al. (2021) caused a significant reduction in performance. The suggestion to remove these bands is driven by poor performance of atmospheric correction methods for these wavebands in turbid waters. Use of TOA Rhot as input to the NN is a real benefit in these cases as there is no requirement to apply an atmospheric correction and these wavebands, which contain important information about optical water quality, can therefore be safely used.

Experiments **H** to **K** use different band combinations available for other sensors. They do show good performances, close to but slightly worse than the best performer **A**. Addition of the 678 nm band

related to the Chl fluorescence (experiment **I** against **J**) shows an improvement in performance with MAD reduced by 0.05. These additional satellite sensors have access to bands that were not available for this study as they are not part of the MODIS waveband set, which limits the validity of this type of comparison for this study. However, the absence of the 531 nm, which is specific to MODIS, appears to be significant as its inclusion produces a significant improvement in performance (Figure 4.12, 9th position). Inclusion of a pseudo 510 nm waveband (which is available in many other sensor data sets) was found to be marginally significant (**T** and **U**) and suggests that a properly tuned SeaWiFS NN would likely perform similarly to the results presented here. It is also interesting to consider the potential benefit of including wavebands at 704 or 709 nm used in red edge algorithms to estimate Chl from MERIS and OLCI (Clevers et al., 2001, Gons et al. 2002). It may be interesting to use these previously carefully selected waveband options as a guide to select data reduction strategies for future hyperspectral NN algorithm development.

The OC5 algorithm (Gohin et al., 2002) was specifically developed for operation in the turbid coastal waters of Northwestern European shelf seas and represents the current state of the art for OCx-style algorithms. There are multiple flavours of this algorithm provided by different data suppliers. In Chapter 3 we tested the performance of two versions of OC5 and found MADs of 2.1 and 2.8 for the PML and ACRI-ST variants respectively for this data set. Experiment **M** returned a MAD of 1.98 from a NN using the same waveband set used by OC5. However, OC5 LUTs if retrained on our specific dataset should probably reach similar performances, and our test set is not entirely independent as it is just a subset of the training set. To highlight that this is not a number of band limitation but a question of information content, the CZCS-like NN (**K**) also used 5 wavebands and produced better performance (MAD = 1.93). The 5 waveband CZCS-like NN (**K**) even outperformed a NN using the 6 OC-CCI merged wavebands (**L**). The difference comes from the availability of band 748nm for the CZCS like algorithm, while the OC5 / OC-CCI methods only include 555 or 670 nm as the longest wavelength. While 748 nm band had little impact when 670 and 859 were available earlier (**F** and **E**), it does help when infrared information is missing. Clearly there is merit in use of longer red / near infrared wavebands for NN retrieval of Chl in coastal waters.

Looking to the future, it seems likely that this NN approach could in theory be applied to data from any specific ocean colour sensor. The major limitation would be generation of sufficiently large and varied matchup data sets to give global coverage, and the time and effort required to generate them. Undoubtedly this would provide the best quality performance, but it would require major *in situ* field campaigns that are not currently part of mission planning. One possibility coming directly from the

machine learning field would be to train algorithms for specific sensors and fine tune the relationship learned to apply these algorithms on other sensors. A similar approach is currently used in ocean colour with empirical algorithms, when a relationship is found (for example the blue-green ratio in open waters) the same algorithm can be applied for a new sensor as soon as enough data were made available to update the different coefficients of the algorithm. An alternative route forward is to try to generate merged satellite reflectance products from different sensors along the lines proposed by the OC-CCI project. This provides an opportunity to incorporate all of the historically available *in situ* data into training data sets for NN generation. At the moment the OC-CCI project produces a limited number of merged BOA Rrs products. The work presented here points to the opportunity that extending the number of wavebands and potentially merging at TOA offers to provide enhanced training data sets for development of NN and other ML algorithms.

These results also suggest that smaller satellites (such as CubeSats designed to be compact which carry less sensors and therefore less spectral information) can usefully contribute to the Earth observation assessment even if they do not include all the bands available from the more expensive dedicated ocean colour sensors. It is possible to imagine future scenarios where data from small satellites provides helpful supplementary data to primary ocean colour satellite systems. The machine learning method used here to reduce and cluster the data (ADBH, Sun et al. 2020) shows interesting features. The first six wavebands identified by this process span the full deep blue to near infrared spectral range of this MODIS data set. However, the selection of the 531 nm band in 9th position seems late compared to its impact on NN Chl prediction. This is probably representative of the fact that the ABDH method uses a different approach to estimate information content than the NN approach. Indeed it is certainly the case that different machine learning approaches will identify different features in any given data set. There is unlikely to be any way to firmly establish a truly optimal single algorithm. As access to computational power continues to improve we may very well find that there is scope to use ensemble approaches where estimates of Chl are produced using a wide range of algorithms (empirical, semi-empirical NNs, random forest etc.) and that final estimates are based on intelligent interpretation of combined data sets. The results presented here suggest that there is scope to usefully incorporate all of the spectral information currently available from multi-spectral sensors, but that some wavebands are more informative than others for retrieval of Chl. This may prove to be useful for future development of computationally efficient hyperspectral NN algorithms.

Finally, the main idea of using more spectral information produces better Chl estimation has been applied to independent datasets to test that it was not a biased observation coming from the available dataset created for Chapter 3. The same behaviour was observed for both the SeaBASS and NOMAD dataset, with more information (light or variables that impact the light signal) leading to improved performances. The development of machine learning and associated fields increases the urgency to make global repositories for any type of data collected for Earth observation freely and easily accessible in order to support further algorithm development. Support for generating *in situ* ground truth data and producing quality controlled data for repositories should be a vital part of mission planning.

It is also important to consider what types of data to include in these repositories. Data from the SeaBASS dataset, for example, are MODIS Aqua matchups but cannot be used to create a better algorithm by merging it with the one used in Chapter 3, because only bottom of atmosphere reflectances have been stored. One outcome from the work presented here is the potential to operate successfully on TOA Rhot. The community should urgently consider the potential merit of merging satellite data at TOA rather than BOA, immediately eliminating the impact of variable AC performance on the merging process and opening the door to producing globally consistent, merged TOA reflectances that could form the basis for a suite of new algorithms with potential to be applied across multiple satellite sensors. Production of globally consistent and extensive (global) training data sets is an urgent imperative that requires a reanalysis of future information needs taking into account the types of results that have been presented in this chapter using machine learning techniques.

4.7. Conclusion

This aim of this chapter was to understand if the method developed in Chapter 3 was optimal, i.e. using all information available from a multispectral sensor to design algorithms, leaving the neural network to deal with potentially useless or less useful bands. Two methods for assessing the contribution of different spectral wavebands to Chl retrieval using NNs have been tested over the same dataset. The first method used a process of waveband elimination based on availability of specific wavebands in historic missions and knowledge of bio-optical properties of natural waters. Removal of specific bands highlighted interesting features for development of NNs for Chl prediction in coastal waters.

The main observation is that addition of more bands generally produced better results (Figures 4.11 and 4.12). Secondly, for any experiment, as soon as enough layers (2 or 3) and neurons per layer (6 or more) were used, performance more or less plateaus and there is no significant improvement by

extending to much higher dimensions of architecture based on analysis of MAD metrics (Figure 4.3 and 4.4). However, visual inspection is required in order to avoid architectures that underfit the signal when low numbers of neurons are used. Interestingly, underfitting issues return similar MAD performances for smaller architectures, usually when they fail at producing accurate enough algorithms above a below a specific value. This highlights the need for better metrics to evaluate neural network performances and ability to detect under or overfitting issues not only using visual inspection for cases when the failure is not obvious or for untrained eyes. Removal of specific wavebands or even sets of wavebands (e.g. in the deep blue or SWIR) had identifiable impacts on performance which can be used to improve understanding of their relative significance. Reduction in the number of wavebands to RGB-style observations had a particularly negative impact. Excluding that case, it is interesting to note that NNs operating on TOA Rhot generally outperformed NNs operating on BOA Rrs as soon as enough bands were used, but performances with fewer bands were subject to underfitting issues for most cases.

The second method established waveband significance using machine learning approaches to cluster (E-FDPC algorithm; Jia et al., 2015) and sort (ADBH algorithm; Sun et al., 2020) the wavebands, and provided ranks that identified the importance of full spectral range (deep blue to infrared). There were interesting anomalies such as prioritisation of the low SNR 859 nm waveband over the higher SNR 869 nm waveband which might be attributable to the degree of redundancy in such closely spaced wavebands, or because the 859nm band is also wider than the 869 one, i.e. contains more information. It seems likely that this technique interprets information content in a manner that is different to that used by NNs, resulting in discrepancies between the two approaches, as significantly different enough bands will be selected first by the ADBH method, while networks only try to minimise the error of all estimations. For example, it seems likely that the 531 nm waveband is particularly important for NN performance but was only ranked 9th by the ADBH technique, probably because the 547 nm band was already used and shows a very similar signal.

Finally, the idea of taking all meaningful information available to improve performance have been applied to two independent datasets, SeaBASS and NOMAD. In both cases, NNs using all available spectral information have shown improved performances compared to the standard OC3 blue-green algorithm, but in doing so generally confirm the direction the community has taken towards use of more spectral data during the past 40 years, which started with OC2 algorithms during the CZCS era (O'Reilly et al., 1998) to OC6 nowadays (O'Reilly and Werdell, 2019) or even 7 bands (Zhang et al., 2019). When restricted to the same number of wavebands, NNs did not outperform the empirical OC3 algorithm, meaning that the

physical approach already benefits from most of the information included in the signal. However, the specific band ratio approach used by OCX algorithms clearly does not capture the full range of spectral variability found in optically complex waters, which may explain why NNs outperform OCx algorithms as soon as more bands are used. Designing empirical algorithms based on multiple relations is harder when more variables come into consideration. NNs on the other hand are specifically designed to handle multiple relationships and are easily optimized through the backpropagation method.

Overall, the conclusion that inclusion of all available multi-spectral bands is optimal for NN development can be made and is consistent with the approach originally proposed in Chapter 3. This work further confirms that use of TOA Rhot is viable for prediction of Chl in coastal waters using relatively small NN architectures. Whilst inclusion of all available multi-spectral wavebands is beneficial, there is considerable leeway in choice of NN architecture. Finally, there is probably a degree of data redundancy in this approach that is not problematic for the small NN sizes considered here, but there are clearly benefits to careful selection of wavebands for future hyperspectral NNs where extensive data redundancy could lead to significant computational inefficiency. In the next chapter will try to elucidate if a modelled hyperspectral dataset based on the upcoming PACE sensor can bring higher performance compared to a MODIS Aqua like algorithms.

Take home message: Chapter 4

After Chapter 3, one of the main question concerned the relevance of using the 15 MODIS Aqua bands to make more accurate estimates of chlorophyll a.

- Chapter 4 tests different band combinations used as inputs to build a neural network algorithm using two methods.
- The first method manually selects the bands based on literature
- The second method uses a machine learning classification technique to order bands by importance.
- The main result shows that using the full light spectrum provides the best estimates, and more bands (from various parts of the spectrum) tend to improve the estimates.
- The use of redundant information (bands that are close to each other such as 547 and 555 nm) barely improve performances, but consistently improve them and would be recommended to use them for best estimates.
- Neural networks do not outperform physical approaches when they follow the same limits (restricting the number of bands used). They really benefit from access to more information from other light bands.
- The improvement is tested and applied to independent datasets coming from mostly case 1 waters, and such approach improves current performances by 30% on the MAD (1.4 vs 1.7 for OC3 algorithm)
- There is an opportunity to use more complex input (such as a band ratio rather than both bands) that was not tested in this thesis and could potentially improve performances further. It was briefly tested by creating an interpolated 510 nm bands, and a NN trained with 16 bands returned consistently better performances.

Understanding in Chapter 4 that a full spectrum provides the best estimates, a hyperspectral dataset is built in Chapter 5 (using models) to anticipate the development of future hyperspectral sensors. Having access to the freedom provided by a model, different hypothesis based on the data distribution are tested to evaluate the importance of *in situ* sampling strategies and their potential help to target under sampled data, at the edges of the data distribution.

5. Neural networks to retrieve in water constituents applied to radiative transfer models simulating coastal water conditions

5.1. Introduction

In Chapter 3 and 4, a neural network algorithm has been developed to estimate chlorophyll a (Chl) in coastal waters. Compared to state of the art algorithms, the performance achieved using a neural network trained with top of atmosphere reflectances are improved. However, the data availability constrained the development and is expected to be the main limiting factor during winter (Section 3.4) and for both high and low concentrations (<1 & >10 mg m^{-3}, Section 3.3). Retrieving Chl in coastal waters conditions from remote sensing is a challenging task due to the presence of other constituents, Colour Dissolved Organic Matter (CDOM) and Mineral Suspended Sediments (MSS). These two additional constituents alter the light signal and complex interactions arise that make a Chl estimate harder than in open water conditions (not taking the potential atmospheric correction failure in consideration). Moreover, collected *in situ* data of the two other constituents CDOM and MSS are scarce, with only a few hundreds of points in modern datasets in Europe (Nechad et al., 2010) up to a few thousands mostly in the United States with some data across the globe (Pahlevan et al., 2022). CDOM absorbs light in the visible with a decreasing exponential relationship from ultraviolet to infrared (Bricaud et al. 1981). It impacts the light signal used to retrieve Chl in coastal waters and leads to failure of Chl algorithms (Darecki and Stramski, 2003; D, Pittarch et al., 2016). MSS absorb light in the blue and backscatter in the green-red, are relatively easy to estimate with good confidence from remote sensing algorithms (Nechad et al., 2010; Neil et al., 2011). High sediment concentrations impact the atmospheric correction process that converts the signal measured by a satellite spectroradiometer at the top of atmosphere into a water leaving remote sensing reflectance (Rrs) which most algorithms rely on. With limited to inexistent datasets containing all three constituents at the same time, testing different hypothesis is difficult. One of the potential approach consists of the creation of a controlled dataset with the use of radiative transfer model that can simulate light interactions with various mediums.

Hydrolight is a software that can simulate light spectra leaving the ocean surface, and requires knowledge of inherent optical properties (IOPs, absorption, attenuation and backscattering of the different constituent of a medium). In this case we need to be able to relate IOPs to constituent concentrations using a bio-optical model operating on material-specific IOPs (SIOPs). Relatively few complete sets of SIOPs have been presented in the literature. The dataset presented by Bengil et al. (2016)

for optically complex waters in the Ligurian Sea, comprising both Case 1 and Case 2 water types (Morel and Prieur, 1977), provides the SIOPs needed to support rigorous exploration of the optical variability associated by freely varying Chl, MSS and CDOM concentrations. By being able to simulate surface remote sensing reflectance signals for a wide range of constituent combinations, we can test several hypotheses related to neural network development. Efforts are made to incorporate realistic estimates of measurement noise in both light and optical constituent concentrations in order to better simulate real world conditions. Hydrolight simulations of hyperspectral Rrs were used to produce the 13 MODIS Aqua bands available from 412 to 869 nm and used for most parts of this study, as well as being used to study the potential of hyperspectral data for future ocean colour missions e.g. the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE, Gorman et al., 2019).

The first hypothesis (H1) to be tested is that NNs will be able to provide accurate estimates of all three optical constituents across a wide range of constituent concentration combinations. This hypothesis is the control group of the analysis and informs us on what type of performance can be achieved with the dataset in ideal conditions. This is an apparently simple test, but has to be considered within the context of the limits of real world data sampling. The distribution of data sampled in natural waters typically follows log normal distributions, reflecting a tendency to under-sample extreme scenarios of very high and very low concentrations of any given constituent (true for all three constituents, Pahelvan et al., 2022). NNs require more data than empirical methods to learn robustly, especially if the signal contains complex non-linear interactions and depends on other factors, which are numerous in ocean colour (sun angle, temporal window used, resolution etc.). The reduced amount of data at both low and high ends of the data distribution is expected to negatively impact NN development when applied to such ranges (Chapter 3). The second hypothesis (H2) is that training with an evenly distributed 'flat' data distribution will produce higher quality performance over the range of variability than is possible from a log-normal data distribution. If found to be true, this would point to potential benefits of directing future *in situ* sampling effort to more carefully attempt to cover the full range of optical variability found in coastal waters.

NNs typically require substantial datasets to support training and limited availability of clear sky matchups between *in situ* and remotely sensed data is a limiting factor on the development of NNs. To date most NN algorithms such as the algorithms developed in Chapter 3 and 4 remain regional with limited application to global scale or under represented conditions. Using the radiative transfer model, Hydrolight 5.2, we can simulate remote sensing light fields for a wide variety of optical constituent combinations and

create artificial data to test different hypothesis, thereby overcoming data availability issues and generating an opportunity to establish the real limits of NN development for coastal water remote sensing. The main reason why NNs are not used is their capacity to overfit signals by remembering the training examples rather than learning properly the relationship between inputs and the target. This type of artefact is mostly due to reduced number of data available from ocean colour matchup datasets with only several thousand examples for the biggest datasets in the literature, while a single MODIS Aqua image can contain multiple millions of 1km² pixels. Multiple techniques exist to avoid overfitting issues, including multi-task learning (MTL). MTL occurs when NNs are trained to produce multiple related targets at the same time, with the main objective being to improve their performance, robustness and reduce overfitting problems (see Ruder, 2017 for a recent overview of different techniques available). Optical signals sampled in coastal waters are a good candidate to evaluate MLT as all three constituents contribute to the light signal. The third hypothesis (H3) is that simultaneous retrieval of all three constituents using MTL will perform better than individual retrievals by helping to constrain NN construction.

To date the majority of ocean colour NN development has been done in the context of data from multispectral sensors. A number of hyperspectral radiometers onboard satellites have been launched in the past including EO-1 and PROBA-1 (2001), with others added as an additional sensor to the ISS (International Space Station), including HICO the hyperspectral imager for the coastal ocean in 2009 (Corson et al., 2008) and HISUI the Hyperspectral Imager Suite in 2020 (Iwasaki et al., 2013). This development has continued with the launches of PRISMA (PRecursore IperSpettrale della Missione Applicativa, Loizzo et al., 2018) in 2019 and EnMap (Environmental Mapping and Analysis Program, Guanter et al., 2015) in 2020. There is a clear trend towards future ocean colour missions being equipped with hyperspectral sensors. However, increased spectral resolution is a technical challenge that is usually achieved by compromise with other mission parameters. For example, all of the sensors mentioned above have high spatial resolution (30-100m) which comes with the side effect of a reduced temporal resolution (usually an image of the full Earth every 16 days) and signal to noise ratios are usually lower than for multispectral systems, reducing their effectiveness for deep ocean observations. These factors greatly reduce their impact for global scale algorithm development even though they provide access to much higher spectral information content and explains the absence to date of publicly available hyperspectral remote sensing matchup datasets. A further limiting factor stems from the challenge of accurate atmospheric correction for hyperspectral sensors (Ibrahim et al., 2018). The first sensor fulfilling global scale and time overpass requirements, PACE is planned to be launched in the near future by NASA.

Providing a neural network with additional relevant information should typically lead to improved performance, so it is reasonable to expect that NNs operating on hyperspectral data should perform better than those operating on multispectral data. Radiative transfer simulations can be performed with hyperspectral resolution that can be subsequently re-sampled at multispectral resolution, in this case corresponding to the wavebands used by MODIS. There is, of course, the potential for hyperspectral data to contain an element of information redundancy as there is likely to be some degree of correlation between adjacent or nearby spectral bands. By resampling the hyperspectral reflectance data produced by simulations we can test a fourth hypothesis (H4) that NNs operating on hyperspectral data will perform better than those operating on multi-spectral data. At the moment and until such time as there has been opportunity to collect sufficient volumes of matchup datasets for PACE, the only way to test the hypothesis that NNs will benefit from availability of hyperspectral data is with the use of modelled data.

5.2. Materials and Methods

5.2.1 Hydrolight radiative transfer simulations

All remote sensing reflectance data used in this study were generated using Ecolight 5.2, part of the Hydrolight 5.2 software package(Sequoia Scientific Ltd). EcoLight 5.2 was used for the creation of reflectance spectra rather than Hydrolight mainly due to the processing time involved in creation of such extensive datasets (10,000 constituent combinations each). It took approximately 2 weeks to produce the simulated above surface remote sensing reflectance (Rrs0+) spectra on a dedicated server. Comparison of light spectra with the more accurate model Hydrolight was not conducted here but is expected to be very similar (Lefering et al., 2016) and satisfies requirements for this study.

Simulations were set up with a uniform water column, a solar zenith angle of 0˚, zero cloud cover, wind speed 9 $m.s^{-1}$, a refractive index of 1.34, water temperature of 20°C and salinity of 35 PSU. The light signal was saved every 5nm from 390nm to 895nm. 13 MODIS Aqua wavebands from the visible and infrared spectrum were simulated by averaging the hyperspectral signal using their full wavebands width provided by NASA (https://modis.gsfc.nasa.gov/about/specifications.php, last access 26th of March 2022) at 412, 443, 469, 488, 531, 547, 555, 645, 667, 678, 748, 859, 869 nm. Two datasets of 10 000 hyperspectral light spectra each were created. A bespoke Matlab script was used to generate IOPs using constituent data distributions and a bio-optical model described below, with data being presented to Hydrolight in the form of simulated AC and BB instrument files.

5.2.1.1 Constituent data distributions

Two constituent concentration data distributions were generated in order to test the hypothesis that evenly distributed training data would lead to NNs that outperform those trained with log-normal training datasets (H2). Chl, CDOM and MSS constituents were created following two different approaches. Both approaches use a random distribution of values for all three variables and return two datasets of 10 000 values each. The first dataset uses a log-normal (LN) distribution and crosses several orders of magnitude with limits summarized in Table 5.1 for each variable. These kinds of distributions are commonly found in reports of sampling campaigns from natural waters (e.g. Babin et al., 2003, Pahlevan et al., 2022) and can be observed in Figure 5.1 (a, b and c). The second dataset was created using a log-flat (LF) distribution, applying the same logarithmically spaced intervals as LN, shown in Figure 5.1 (d, e and f)**.** While medians between the normal and flat distributions remain the same, there are significant difference in the mean values for each distribution type.

Table 5.1: In situ constituent concentration ranges.

Variable	Range from *in situ* samples	Range used for model creation	Units
Chlorophyll a	0.29 - 3.31	0.01 - 100	$mg.m^{-3}$
CDOM	0.021 - 0.11	0.01 - 1	m^{-1}
MSS	0.13 - 3.7	0.1 - 100	$g.m^{-3}$

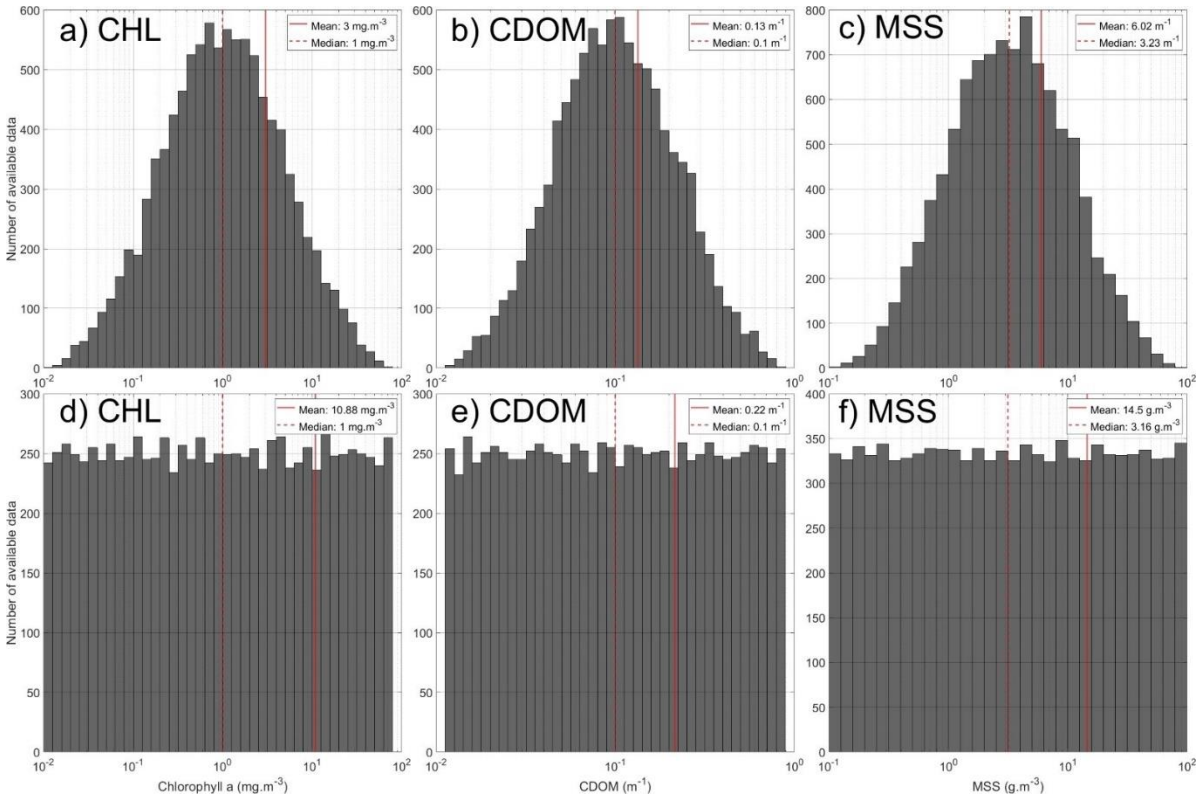

Figure 5.1: Histogram of each constituent concentration used for application of the radiative transfer model. First row shows the log normal distribution of Chl, CDOM and MSS respectively, second row shows the log flat distribution.

5.2.1.2 Bio-optical model used

In order to simulate reflectance spectra for different combinations of optical constituents, the radiative transfer simulation requires selection of a bio-optical model to allow prediction of IOPs from constituent concentrations. Bengil et al. (2016) presented a bio-optical model for the Ligurian Sea that was adopted here. Full details are provided in Bengil et al. (2016) and are briefly summarized here. Chl, CDOM and MSS samples and IOP profiles were collected during a cruise campaign in the Ligurian Sea from 13 to 26 March 2009 off the northwest coast of Italy on board NR/V Alliance. Absorption and attenuation profiles were collected with a 25 cm pathlength AC-9 (WetLabs Inc.) operating at 9 wavebands (10 nm FWHM) centred on 412, 440, 488, 510, 532, 555, 650, 676 and 715 nm. The AC-9 was calibrated using ultrapure water (Milli-Q, Millipore) before and during the cruise, with corrections applied for the temperature and salinity dependence of pure seawater. Absorption data were corrected for scattering errors using the proportional correction method (Zaneveld et al., 1994) Backscattering profiles were collected using a WETLabs BB9 operating at 9 wavebands centred on 412, 440, 488, 510, 532, 595, 650,

676 and 715 nm. Backscattering data were interpolated to AC-9 wavelengths and measurements were corrected according to the BB-9 manual (WETLabs Manual, 2013). See Lefering et al. (2016) for more details. The absorption of all dissolved and suspended components minus water was measured using a Point Source Integrating Cavity Absorption Meter (PSICAM; Rottgers & Doerffer, 2007; Rottgers et al., 2005, 2007). A 1 m liquid waveguide capillary cell (LWCC) with an Ocean Optics USB2000 mini-spectrometer was used to measure absorption by CDOM. Total particulate absorption was also measured using the quantitative filter pad method (Ferrari & Tassan, 1999). Samples were placed directly in front of the optical windows of a Shimadzu UV-2501 PC spectrophotometer. Absorption by phytoplankton was determined by bleaching samples, measuring the absorption of non-algal particles, and subtracting this from total particulate absorption. Path length amplification factors and scattering offset corrections were determined using a linear regression approach (Lefering et al., 2016; McKee et al., 2014) and corresponding PSICAM particulate absorption data. The resulting filter pad corrections were subsequently applied to both bleached and unbleached filter pad absorption spectra.

Chlorophyll concentration was measured using standard HPLC measurements on samples filtered through GF/F filters, stored in liquid nitrogen and transported to laboratories for later analysis. Chl data presented here were collected by colleagues from Management Unit of the North Sea Mathematical Models (MUMM). Triplicate HPLC samples were analyzed by the Marine Chemistry Laboratory of the MUMM using a reversed phase, acetone-based method with a C18 column and a Jasco FP-1520 fluorescence detector. Total suspended solids concentrations (TSS) were obtained by colleagues from MUMM by filtering samples through pre-ashed, rinsed and pre-weighed 47 mm GF/F filters. Samples were rinsed with several aliquots of ultrapure water, taking care to rinse the edge of the filter to minimize salt retention. Filters were stored frozen and returned to the lab where they were dried and reweighed. All samples were measured in triplicate and final values expressed as averages. TSS in northeaster stations was numerically decomposed into organic (OSS) and mineral (MSS) components using the technique outlined in Bengil et al. (2016).

34 stations were available after quality control (Figure 5.2). Stations were partitioned into onshore and offshore sub sets, with deep clear case 1 waters in the southwestern part and shallower clear to turbid case 2 waters in the northeastern part. Figure 5.2 shows that the northeastern, onshore area is partly influenced by the Arno River plume and generally shows higher sediment concentrations near the coast. IOPs measured in the south western part of the area (Figure 5.2) were used to establish chlorophyll specific SIOPs for absorption, scattering and backscattering. These Chl-specific SIOPs were then used to

partition particulate IOPs in the northeastern sector in order to derive mineral specific SIOPs (again, for absorption, scattering and backscattering). Absorption by CDOM was directly measured in both sectors.

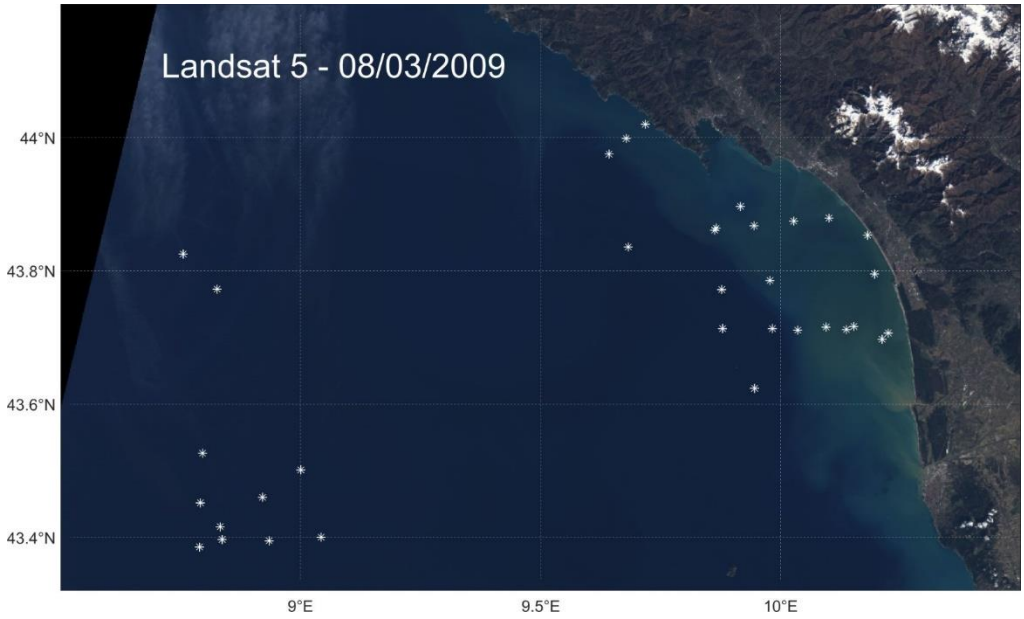

Figure 5.2: Repartition of the 36 *in situ* stations (displayed as white stars) where light and constituent concentrations were collected during the Ligurian cruise campaign in March 2009 displayed onto the true colour Landsat 5 image of the 8th of March, 2009.

SIOP spectra were generated from IOP measurements spanning the visible range (400 – 715 nm). In order to fully represent the range of wavebands provided by MODIS, SIOP spectra were extended out to 895 nm by linear extrapolation. Figure 5.3 shows the final set of SIOP spectra used to form the bio-optical model used for Ecolight simulations. Figure 5.4 shows remote sensing reflectance spectra obtained from Ecolight simulations using both LN and LF constituent distributions. These reflectance spectra together with their associated input constituent concentrations form the basis for training and testing NNs in this chapter. On Figure 5.3, a is the absorption, b the scattering, b_b the backscattering, * means that the relationship is constituent specific, PH stands for phytoplankton, BD for biogenic detritus.

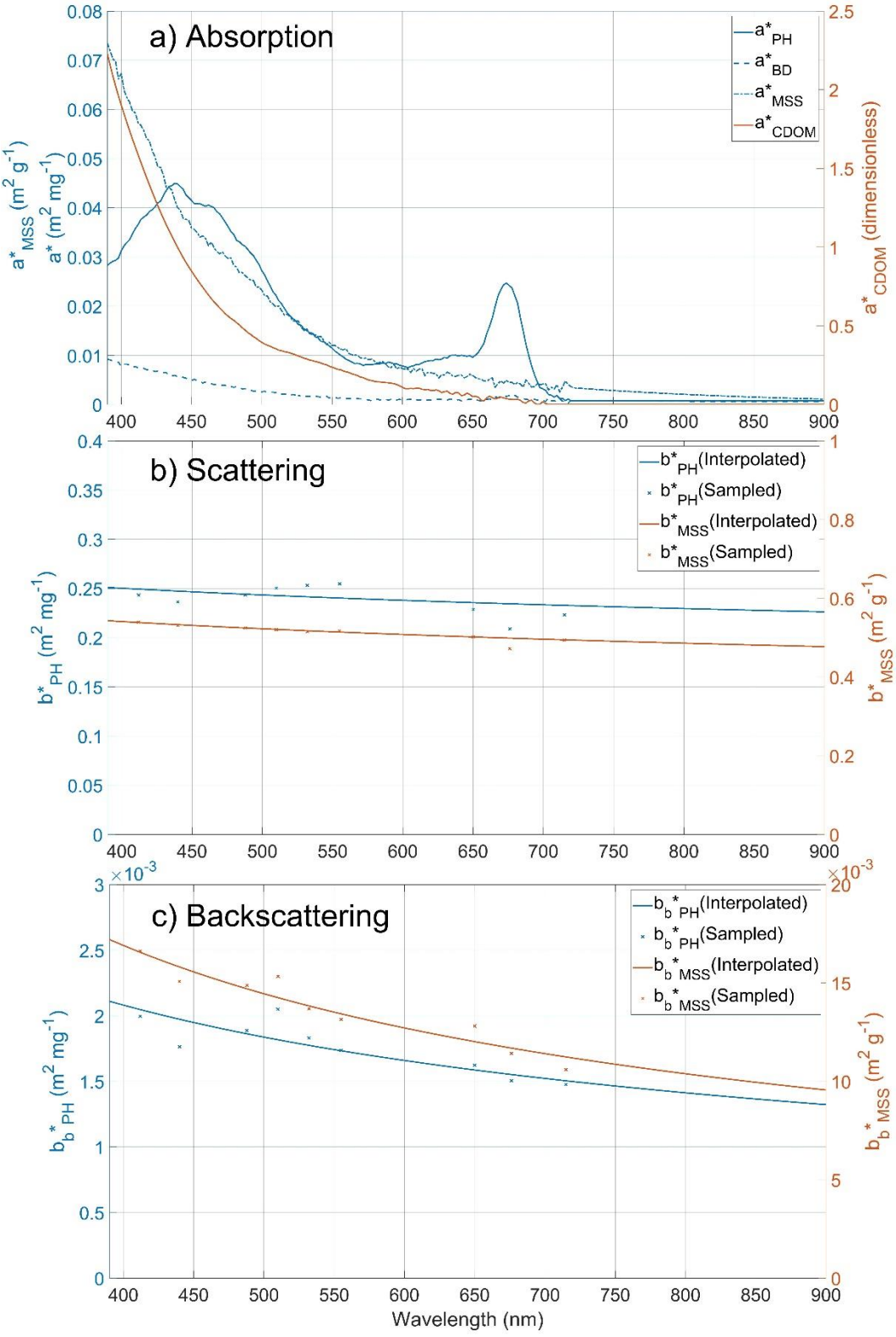

Figure 5.3: SIOP spectra used in radiative transfer simulations.

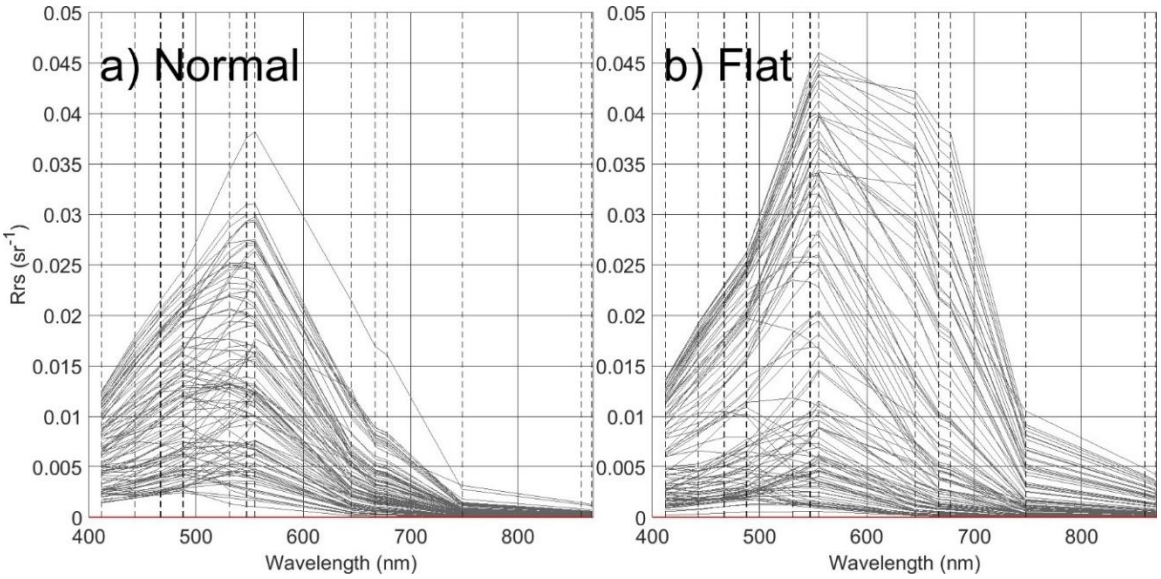

Figure 5.4: Rrs spectra for log Normal and log Flat constituent distributions.

5.2.1.3. Simulation of radiometric noise and constituent measurement uncertainty

Simulated data from model outputs are essentially error-free and not impacted by noise compared to real Earth Observation data. In reality, measurement uncertainties will impact both remote sensing reflectance signals and measurements of constituent concentrations, both of which go into training and testing of NNs. In order to better simulate real world conditions, artificial noise was added to both the Rrs and constituent data prior to NN training

Mélin et al. (2016) evaluated noise impacting the MODIS Aqua sensor data and found a wavelength dependent relationship, with shorter wavelengths returning higher measurement uncertainties. Figure 5.5 shows the error measurements for 5 MODIS Aqua bands following their work. Here we have interpolated the Mélin et al. (2016) results using a power law relationship to provide estimated measurement uncertainties for Rrs on a hyperspectral basis. These values provide the standard deviation of measurement uncertainty for each wavelength, with noise being assigned to each wavelength of simulated Rrs using a random normal distribution operating on these predicted standard deviations.

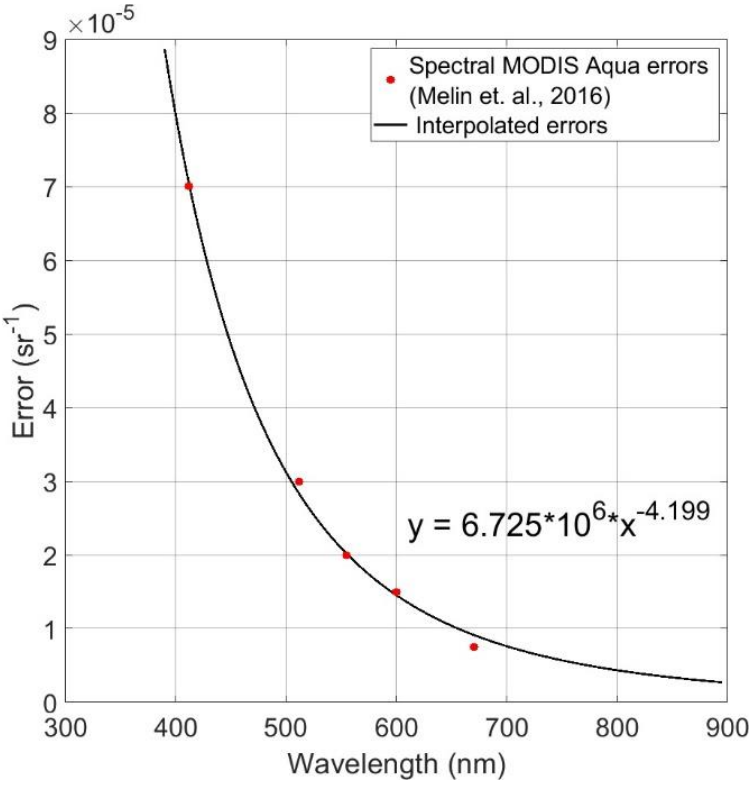

Figure 5.5: MODIS Aqua spectral noise and its hyperspectral interpolation using Mélin et al 2016.

Constituent concentration measurements collected during fieldwork campaigns are sensitive to errors for several reasons, including errors related to the water sample filtration, sensor calibration, method specific or human errors, etc. Estimates of noise related to Chl sampling range from +/-10% (Claustre et al., 2004) to +/-80% standard deviation (Sørensen et al., 2007; Tilstone et al., 2012), depending on the method used for sampling and the degree of quality control applied. There is less information available in the literature for estimates of uncertainty in MSS measurements, so we have assumed that errors will be similar to those found for Chl as both techniques operate on filtered samples. For CDOM, Dall'Olmo et al. (2017) respectively found absorption measurements accuracy and precision of 0.0004 m^{-1} and 0.0025 m^{-1} when compared with independent data at 440 nm. For consistency Gaussian errors were applied to Chl and MSS following a standard deviation of 20% and were assumed to be proportional to the concentration. Uncertainties for CDOM were determined using random normal distributions with a standard deviation of 0.0025 m^{-1}.

5.2.2 Neural network development

For this study, feed forward neural networks with backpropagation of the error until convergence was reached were developed using Matlab's train function. An architecture of 3 hidden layers and N

neurons in each layer was selected for each networks, with N being the number of inputs. For example, N was set to 13 when NNs were created using the 13 MODIS Aqua-like bands available with both datasets. Selecting 3 hidden layers is sufficient to avoid underfitting issues and is computationally efficient. The Rectified Linear Unit activation function was selected and the error was evaluated using the MSE error function. Light and constituents concentrations were log transformed and then normalised between 0 and 1 prior to training. The train set represented 70% of available data, and validation and test sets 15% each, all randomly selected for each training. For the last results section when hyperspectral NN were developed, the number of neurons per layer was selected to be the number of bands available for each experiment. Figure 5.6 shows a schematic diagram of a NN. It contains 4 inputs, 2 hidden layers of 4 neurons each (following the number of inputs as mentioned above), and can make estimations of all three constituents at the same time, Chl, CDOM and MSS as used in multi-task learning. When a single constituent is estimated, the output layer contains only 1 node associated with the desired constituent.

NN performance will be evaluated with the Mean Absolute Error (MAE) using the Seegers et al. (2018) formula (Chapter 2, Section 2.4), which is a MAE applied to log transformed values to the model and observation parameters prior to application as shown in equation 1 below. For example, a MAE of 1.3 represents a relative measurement error of 30%.

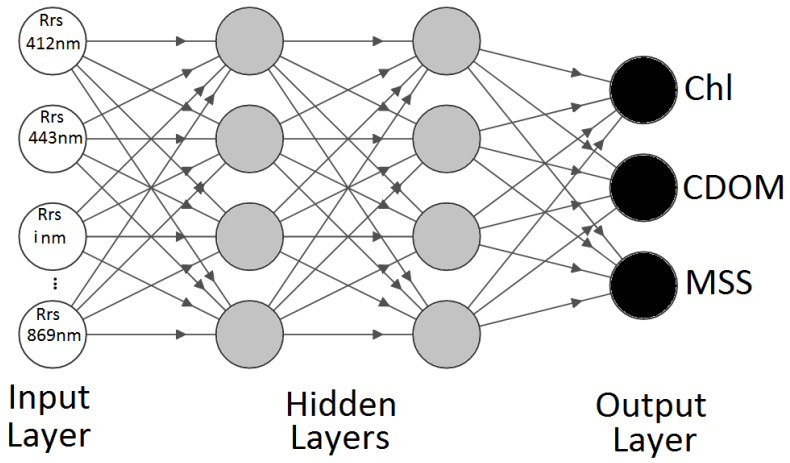

Figure 5.6: Neural network diagram as used for multi-task learning. Hidden layers always contain a number of neurons equal to the number of inputs. Output layer is returning a single constituent when MLT is not used. Bias unit not shown.

The purpose of neural network development is to provide sufficient training data to allow the NN to establish robust statistical relationships that enable accurate prediction of the target parameter from potentially complex input data. The train part of the total dataset is used to train the network, the validation part is used to stop the network training when it stops improving, and the test part is used to evaluate the performance of the resulting NN. Figure 5.7 shows data for train, validate and test dataset for Chl prediction using the LN distributed dataset without inclusion of noise. All three datasets show very similar performances, and the same observation was made during the analysis of results section. To avoid showing similarly repetitive diagrams in the results section, only the independent test set results will be shown going forward.

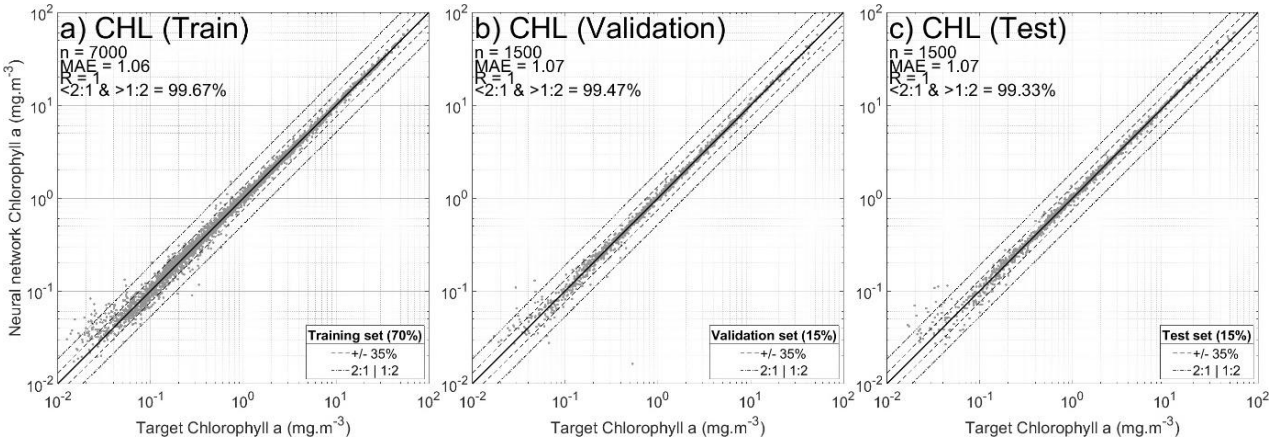

Figure 5.7: Neural network performances to estimate Chl based on the 13 MODIS Aqua bands, using the log normal distribution of data without addition of noise.

5.3. Results

5.3.1. Neural network retrieval of constituents in optically complex waters (H1)

The first set of experiments is designed to test the hypothesis that NNs should be able to accurately retrieve individual constituent concentrations (Chl, CDOM and MSS) across the broad range of optical water conditions found in coastal waters (H1). Therefore for this section, NN were trained to produce a single constituent at a time. Figure 5.8 (a, b and c) shows performance obtained for the test sets for each constituent concentration, for the LN dataset, without addition of noise. All three constituents can be predicted with very high performances under these idealized conditions, with MAE vales close to 1 and more than 99% of data falling within a factor of 2 of the 1:1 line. Adding realistic estimates of noise to both the reflectance and constituent datasets has a significant impact on NN performance. Figure 5.9 (a, b and c) shows that retrieval of Chl, CDOM and MSS is still largely successful,

but there is a noticeable increase in the spread of data for each parameter, with MAEs reaching as high as 1.25 for Chl, though more than 96% of data still falls within a factor of 2 of the 1:1 line.

These results clearly demonstrate that NNs have the capacity to overcome the optical complexity of coastal waters with freely varying constituent concentration combinations. This is perhaps unsurprising in the case of noise-free data, but it is reassuring to see that inclusion of noise in the system does not irreparably impair performance. We can therefore conclude that hypothesis 1 (H1) is demonstrated to be correct.

5.3.2 Impact of data distribution on neural network performance (H2)

The results presented in section 3.1 were produced using the log-normal (LN) dataset where the distribution of data has been organized to broadly replicate datasets found in the literature. In this section we test the hypothesis (H2) that NN performance will improve if the training dataset is more evenly distributed to better capture extreme events at both high and low concentrations. Figure 5.8 (d, e and f) shows NN performance using the log-flat (LF) data distribution. NN performance for the LF dataset is generally slightly worse than for the LN dataset, with MAEs increasing very slightly for CDOM and MSS, but more markedly for Chl (MAE = 1.11). It is noticeable the greatest deterioration in performance appears to be for low Chl values. This is slightly surprising at part of the interest in testing the LF distribution was specifically to address the question of less commonly occurring scenarios at the extremes of the concentration ranges. It may be the case that although the LF training dataset has increased the proportion of low concentration training data, there is an intrinsic problem in trying to estimate very low concentrations of Chl in the presence of potentially high concentrations of other constituents. This could simply be attributable to the Chl making an insignificant contribution to the optical signals under these circumstances.

Figure 5.9 (d, e and f) shows the impact of incorporating noise into the LF NNs. As found previously with the LN dataset, introduction of realistic measurement uncertainties negatively impacts NN performance for all three constituents, with Chl more strongly affected than CDOM and MSS. In the latter cases although MAEs increase to 1.13 and 1.2, approximately 99% of points still fall within a factor of 2 of the 1:1 line. In contrast, performance of the Chl NN deteriorates significantly with a MAE of 1.49 and the fraction of points falling within a factor of 2 of the 1:1 line dropping to 83%. Chl performance is again most notably affected for low concentrations where it would appear that introduction of measurement uncertainties has made it even harder to resolve the small contribution of Chl to the optical signals. This level of Chl retrieval is close to the levels found with real *in situ* observations (Hadjal et al., 2022,

submitted; Pahlevan et al., 2022). Retrieval of CDOM and MSS is fairly robust under all of the circumstances tested here. This is unsurprising in the case of MSS which has previously been robustly determined using even single red wavebands (Nechad et al., 2010; Neil et al., 2011).

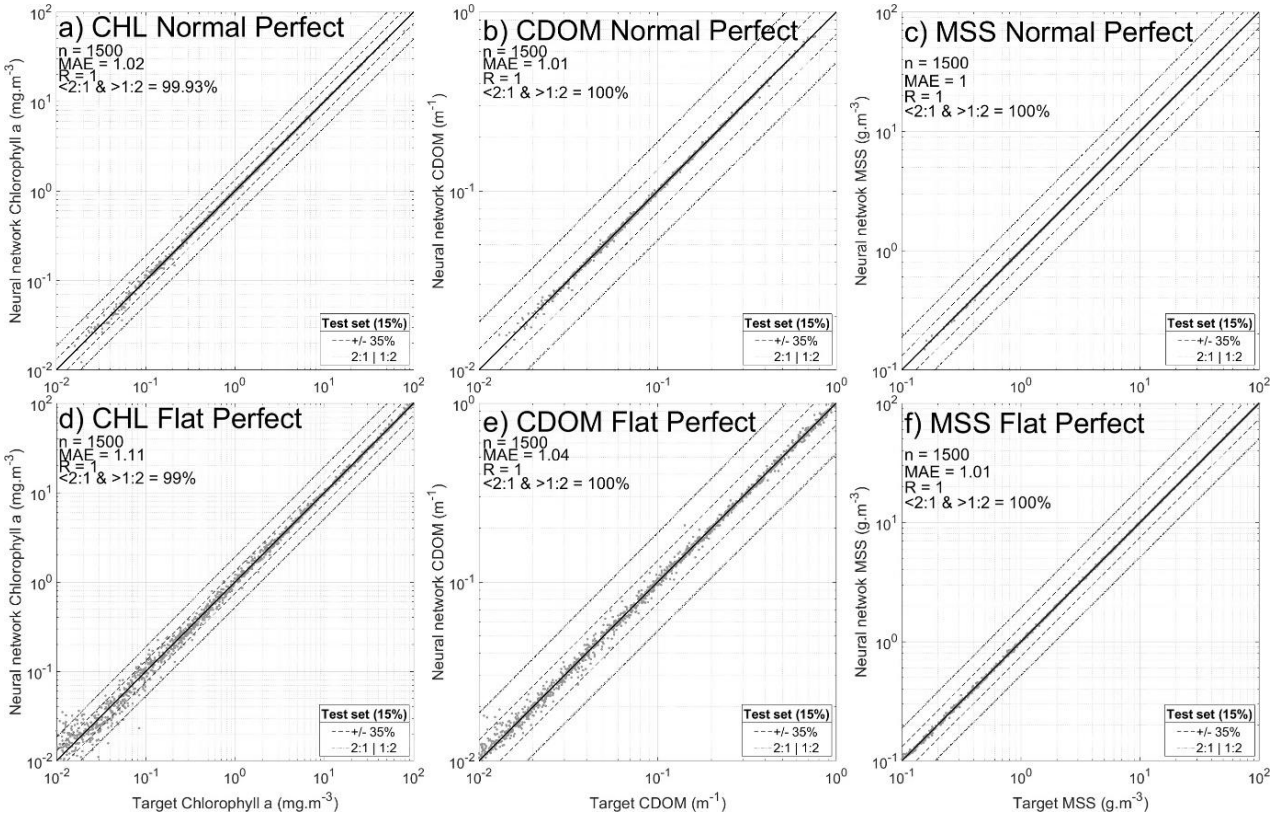

Figure 5.8: Neural network performances obtained for each constituent using a log **normal** (top row) or log **flat** (bottom row) distribution of data without addition of noise (raw model output).

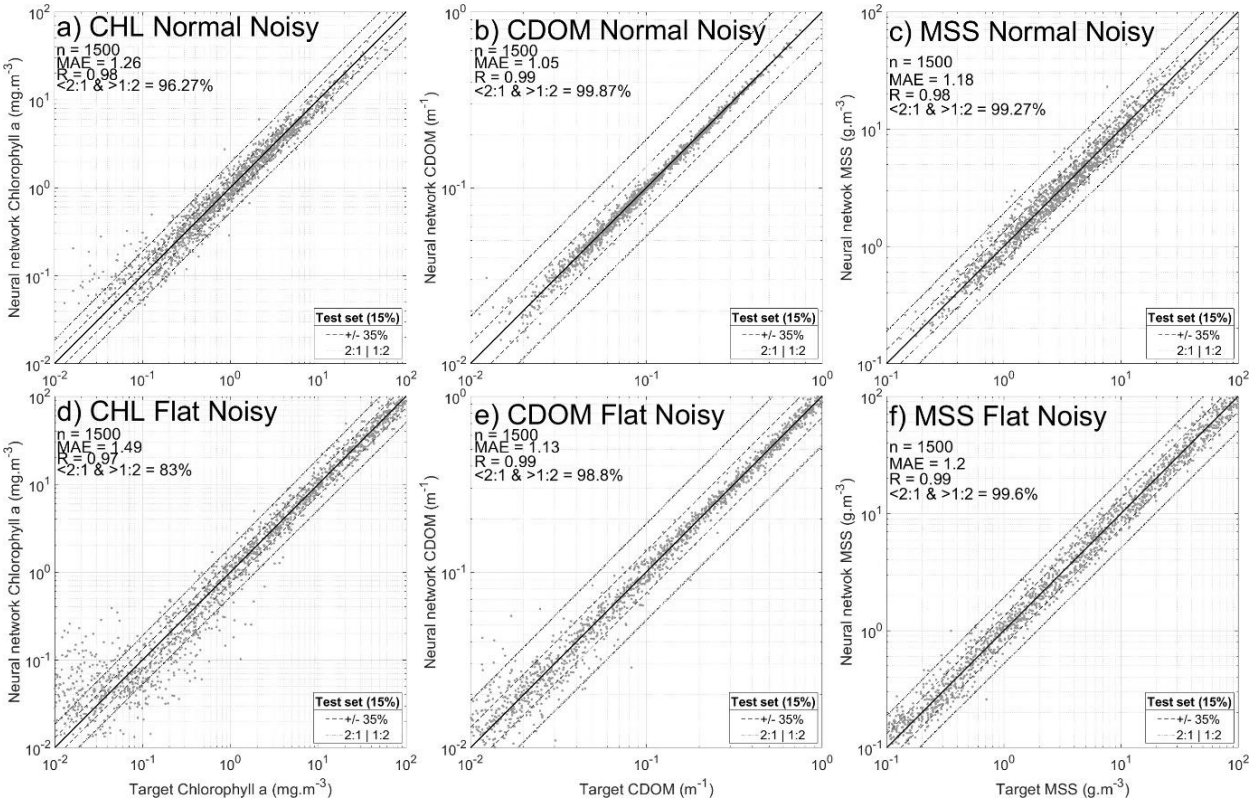

Figure 5.9: Neural network performances obtained for each constituent using a log **normal** (top row) or log **flat** (bottom row) distribution of data with addition of noise.

The results presented in Figures 5.8 and 5.9 refute the hypothesis (H2) that a more evenly distributed dataset will tend to improve NN performance. In fact results shown here tend to suggest the opposite is true, with superior performance being obtained with the LN dataset. This is a surprising result and caution is advised as there is a possibility that results are skewed to some extent by the greater occurrence in the LF dataset of scenarios where the Chl is simply too low to meaningfully impact on the reflectance signal. Nonetheless, across the full range of variability of the three constituents there is no evidence to suggest that the LF dataset is producing superior performance. Thus it seems unlikely that either subsampling existing datasets to artificially produce log-flat distributions or targeting sampling effort to achieve it in future will lead to any improvement in performance.

5.3.3 Multitask learning: simultaneous estimation of Chl, CDOM and MSS (H3)

Multitask learning (MTL) is a type of machine learning method (Caruana, 1997) that tries to improve neural networks performance by compelling networks to learn how to estimate multiple, potentially correlated variables simultaneously. There are multiple reports of successful applications from

different fields in the literature (Collobert and Weston, 2008; Deng et al., 2013; Girshick, 2015; Ramsundar et al., 2015). In order to test the potential benefits of MTL one needs to have access to a set of data containing both the reflectance signals and all three optically significant constituent concentrations. Additionally the dataset needs to be sufficiently large and representative to be suitable for NN training. Unfortunately there are relatively few publicly available *in situ* datasets where all of these parameters are simultaneously recorded. Here, because we use modelled datasets based on user-defined ranges of constituent concentrations and a complete set of SIOPs, we have sufficient flexibility to produce a dataset that can be used to test the hypothesis that MTL will improve determination of constituent concentrations using NNs (H3).

The NNs developed in this section estimate all 3 constituent concentrations (Chl, CDOM and MSS) simultaneously in the output layer as shown in Figure 5.6. Figure 5.10 displays the performance reached for each variable for both the LN and LF distributions, with noise included in both cases. MTL performance levels are broadly comparable with single parameter retrievals (Figure 5.9) in all cases. There is no evidence to suggest that MTL has improved retrieval of any of the constituents and in the case of CDOM there is even some degradation in performance compared to single parameter retrieval. Whilst we cannot rule out the possibility that MTL may have benefits if used with more complex NN architectures or with real world data, at this point we can only draw the conclusion that there is currently no evidence to support the hypothesis (H3) that MTL will improve NN retrieval of Chl, CDOM and MSS.

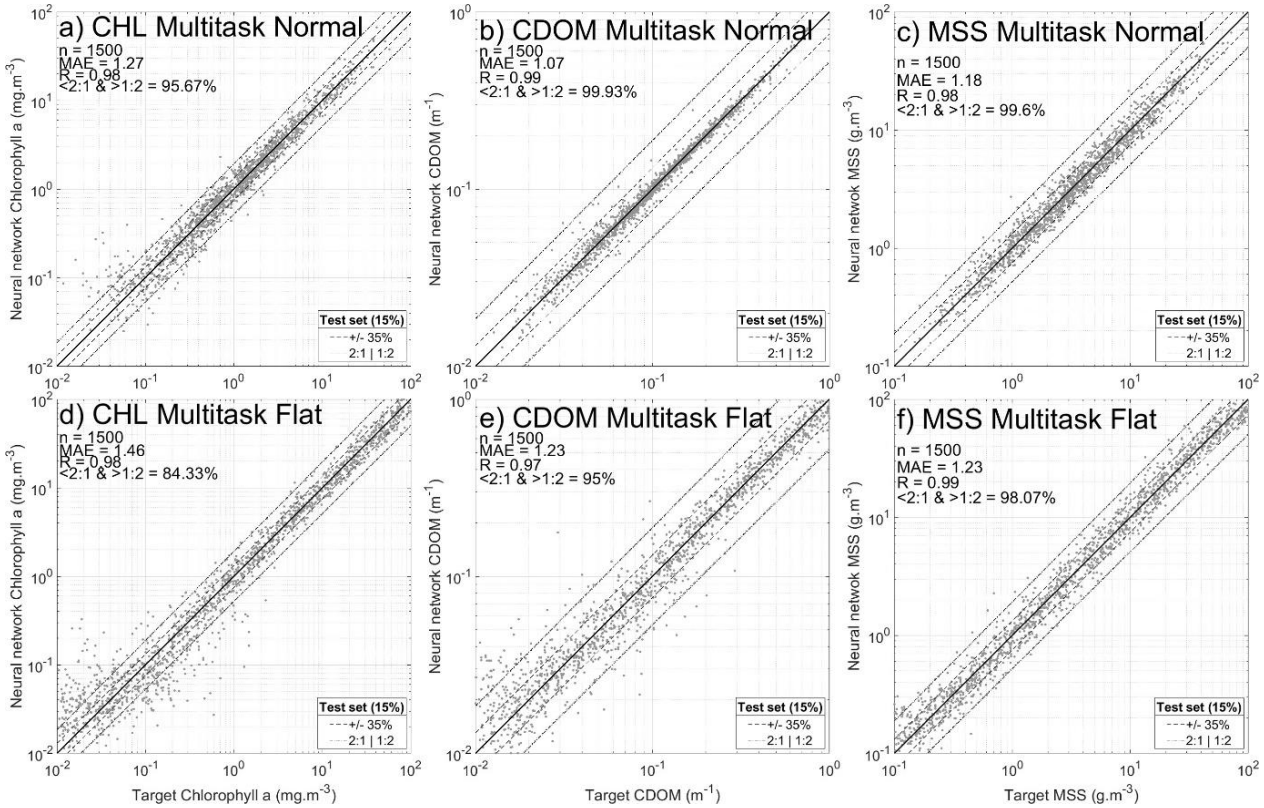

Figure 5.10: Performances obtained at estimating Chl, CDOM and MSS at the same time using a single 3 layers of 13 neurons each network architecture based on the light signal.

5.3.4 Comparison of hyperspectral vs multispectral neural nework performance (H4)

The final experiment presented in this study concerns evaluation of the potential for hyperspectral reflectance data to significantly improve the performance of NNs over existing multispectral capabilities (H4). The work presented in previous sections was conducted using 13 wavebands that were selected to mimic MODIS signals. The Ecolight simulations produced a total of 102 wavebands. Using all available wavebands would be computationally expensive and there is good reason to believe that such an approach would be superfluous due to information redundancy between adjacent bands. Instead we systematically explore the impact of increasing the number of bands available for the network. In order to be methodical, bands were selected using even spacing. For example, when 2 bands were used, bands 33 and 66 (550 and 715 nm respectively) were selected among the 102 available. When 3 bands were used, bands 25, 50 and 75 were selected. This approach does not attempt to optimize performance by selecting the best performing bands for each subset, but rather treats the data in a systematic manner operating on an assumption that each band has similar information value. Here between 1 and 20

wavebands were selected and resulting NNs were tested for both the LN and LF datasets, with noise included in all cases. Each NN is composed of 3 layers with the number of neurons per layer being equal to the number of wavebands used, and separate NNs being developed for each constituent (no MTL).

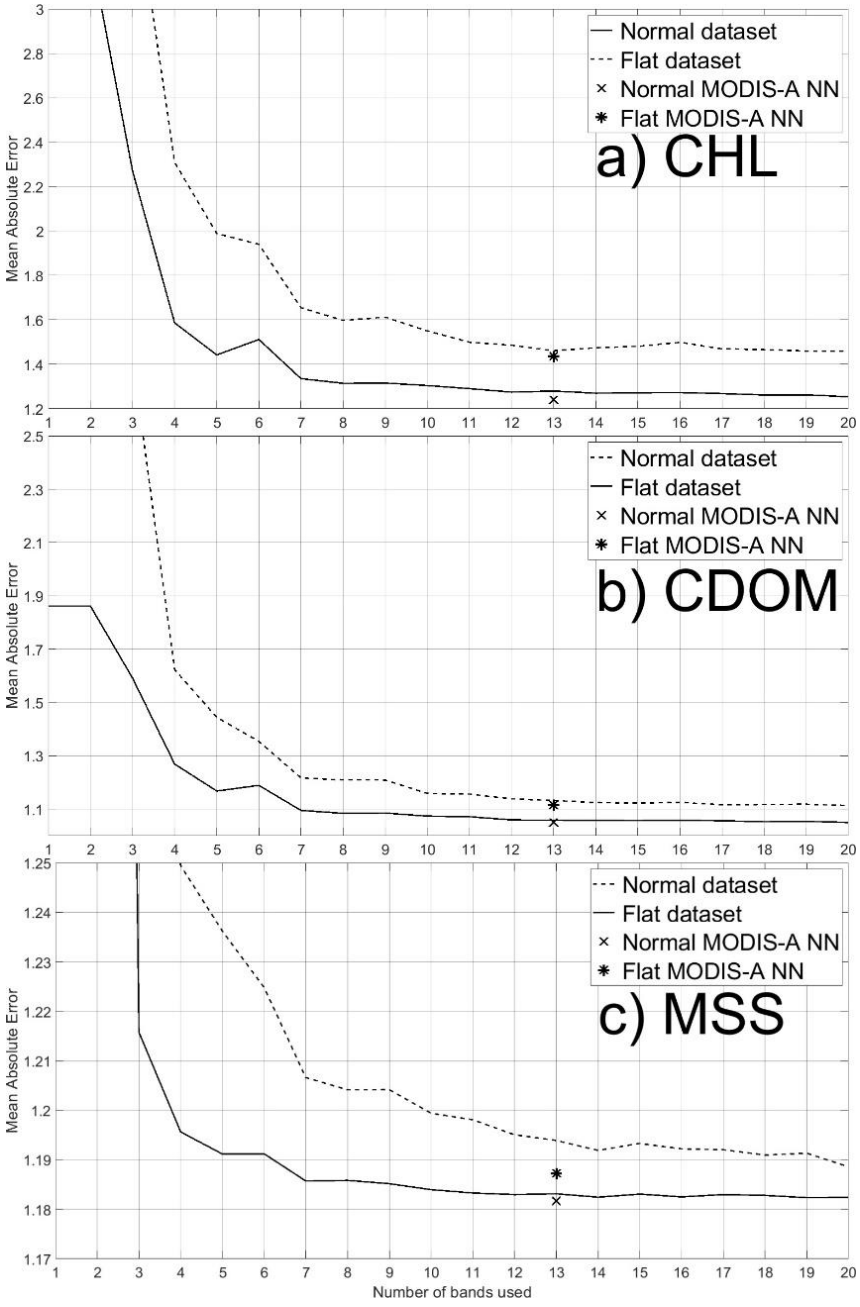

Figure 5.11: Median of mean absolute difference obtained for 10 neural networks averaged designed using the specified number of band used to create an evenly-spaced algorithm. The black star and cross represents performance obtained for the 13 MODIS Aqua NN shown in Figure 5.8.

Figure 5.11 shows the median MAE obtained for neural networks trained with 1 to 20 bands evenly spaced from the full hyperspectral signal. To improve consistency for each band combination, an ensemble approach was used (Hadjal et al., 2022, submitted). 10 neural networks were created for each band combination and the value shown on the y axis represents the median of the 10 different MAE obtained, one for each neural network, as if they were merged.

As expected, there is a clear improvement of NN performance with increasing waveband availability. Improvements are most significant for small numbers of wavebands and then in most cases a region of much slower improvement is reached once approximately 7-10 bands are used. The LN distributed dataset converges faster than the LF dataset. In all cases the MODIS Aqua NNs slightly outperform evenly-spaced algorithms with equivalent numbers of bands suggesting that careful selection of specific wavebands may be slightly beneficial compared to evenly spaced wavebands. Further testing of 25 to 50 evenly spaced wavebands (not shown) provided little further improvement in NN performance. These results generally refute the hypothesis (H4) that ever greater spectral resolution will improve retrieval of Chl, CDOM and MSS in optically complex coastal waters. This may reflect the fact that the optical properties of the water constituents vary slowly with wavelength and associated reflectance spectra offer only limited spectral information content.

5.4. Conclusion

The potential for NNs to provide improved quality ocean colour products for optically complex coastal waters has been discussed for many years (Doerffer and Schiller, 1994; Buckton et al. (1999); Gross et al., 1999). The advent of hyperspectral ocean colour sensors with genuine global spatio-temporal capabilities and the availability of affordable computational resources provides growing impetus to further explore this potential. However limited data availability for training and testing NNs is a serious impediment to development of this approach. Here we have developed realistic radiative transfer simulations in order to generate training datasets that span the range of constituent concentrations needed to test NN performance across the range of variability encountered in coastal waters. This modelling approach has allowed us to test a number of fundamental hypotheses relating to development of NN algorithms for coastal ocean colour applications.

Neural networks have shown capacity to accurately retrieve Chl, CDOM and MSS when all three constituents are free to vary independently from one another over concentration ranges spanning several orders of magnitude (H1). NN performance is affected by inclusion of realistic measurement uncertainties, but the fundamental conclusion remains the same that relatively small NN architectures are capable of

handling the levels of optical complexity encountered in coastal and shelf seas. These results are broadly consistent with recently presented research by Pahlevan et al. (2022) who have demonstrated ability to retrieve all three constituents using Mixture Density Networks. The simulated datasets presented here could usefully be used to test approaches of this nature and other machine learning algorithms. Whilst NN return almost perfect results with noise-free simulations, their performance appears to be strongly linked to the uncertainty in the *in situ* training data. With 20% (StdDev) noise added to both Chl and MSS, MAEs close to 1.2 were reached for the normally distributed dataset. Similarly, when 40% (StdDev) noise was added (not shown) MAEs close to 1.4 were obtained for both products. Because CDOM measurements are less noisy and based on a net value, it did not show the same behaviour. The performance of NN estimates is directly linked with *in situ* constituents quality and probably the main limiting factor here. Except for the very low values of Chl and CDOM, NN have shown the capacity to make excellent estimates of the constituents. Therefore, the quality of the retrievals obtained from *in situ* is one of the main limiting factor in terms of performance achieved.

Various strategies to improve NN performance have been developed over a wide number of research fields. One of the more commonly discussed approaches is multitask learning (MTL) which is immediately of interest in ocean colour remote sensing in coastal waters as the reflectance signals is inherently dependent on more than one optical constituent. The ability to determine constituent concentration ranges used in radiative transfer simulations provides an opportunity to systematically test the potential merit of MTL. In this case we have clear evidence that simultaneous retrieval of all three optical constituents does not improve upon single parameter retrievals and in fact tends to reduce overall performance (H2). For a pure performance approach, MTL should not be considered, at least with simulated data. However, MTL is also being used to help generalisation of neural networks in other fields, but this hypothesis was not testable here because we rely on simulated coastal data.

One of the most common criticisms of NNs (and other machine learning approaches) is supposed limitation to the training dataset provided. Whilst there is indeed an element of truth to this, it should also be recognized that if a training dataset is genuinely representative of prevailing circumstances then there is good scope for a NN to be able to provide general predictive power for that system. Many of the criticisms based on training set limitations are similarly true for empirical and semi-analytical algorithms. In all cases datasets for algorithm development are subject to the vagaries of *in situ* sampling effort and impact of cloud cover on matchup realisation. The NN approach discussed in this Chapter was first developed using an *in situ* dataset to predict Chl (Chapter 3 and 4, Hadjal et al., 2022). One of the concerns

identified in that work was the log-normal nature of the data distribution in the assembled training dataset, with concern that both high and low concentration scenarios were under-represented. The simulation approach developed here has allowed us to compare results from datasets with both log-normal and log-flat constituent distributions. Somewhat surprisingly, there does not appear to be any benefit to having a more evenly spaced training dataset and in fact the performance of Chl retrieval was of lower quality for the flat dataset at low concentrations. It seems likely that there is a fundamental limit on accurate retrieval of any constituent when its contribution to the reflectance signal becomes sufficiently insignificant. There is naturally interest in trying to retrieve Chl concentrations at very low concentrations such as are found in oligotrophic offshore waters (Signorini et al., 2015) . However, in the case of optically complex coastal water it may be much more difficult or even impossible to achieve the same level of Chl retrieval at low concentrations due to the confounding influence of CDOM and MSS which would typically either be absent or found at very low concentrations in case 1 waters. That said, these results are helpful in so much as they illustrate that the normal distributions, which are similar to those generally obtained from large field campaigns, are capable of producing high quality results across the full range of concentrations for each constituent, and there is no obvious merit in trying to further manipulate them to manage over- or under-representation across the dataset.

Development of the hyperspectral PACE mission has brought renewed interest in establishing the potential for hyperspectral remote sensing to improve the quality of ocean colour products for optically complex coastal waters. This is particularly relevant for NNs and other data-hungry machine learning approaches that have potential to exploit additional information content to improve product quality. Here we have tested the hypothesis that NNs trained on simulated hyperspectral reflectance data will produce better quality estimates of Chl, CDOM and MSS than is possible with multispectral data (H4). Results presented in Figure 5.11 suggest that there is in fact a practical limit to NN performance and that there is little further improvement in algorithm performance with higher numbers of wavebands. It should be noted that these results were obtained using evenly spaced hyperspectral wavebands and that there is clearly scope for further optimization by careful selection of specific combinations of wavebands which is an option with hyperspectral data. Indeed, in all cases NNs operating on the MODIS Aqua waveband set outperformed evenly spaced hyperspectral data, illustrating the potential benefit of carefully selected waveband subsets. Nonetheless, these results strongly suggest that simply increasing spectral resolution will not of itself improve determination of Chl, CDOM and MSS in coastal waters. However, there may be many other benefits to use of hyperspectral data such as identification of specific spectral features associated with e.g. cyanobacterial blooms. The main improvement from a remote sensing point of view

could in fact come from the capacity of these neural network algorithms to deal with natural sources of signal contamination (e.g. sun glint, thin clouds, etc.). The NN method developed in Chapter 3 using TOA signals to retrieve constituents directly could benefit from inclusion of additional bands providing information on sources of signal disruption. For good quality Rrs data, expectations for significant improvement in product quality across the board would be misplaced. Additional factors such as signal to noise ratio, atmospheric correction performance and quality of spatio-temporal matching will significantly impact product performance as well.

Take home message: Chapter 5

In Chapter 3, one of the limit was the quality of data estimated at the edges of the distribution. In Chapter 4, we saw that training a neural network to estimate chlorophyll using the full light spectrum returns the best performances. Chapter 5 focused on testing the impact of data distribution, multi task learning and the potential benefits of hyperspectral information for such methods.

- There is not particular benefit in building a non-Gaussian and more evenly distributed dataset as it does not improve robustness over the edges of the data distribution.
- A neural network trained on a specific dataset always outperforms other neural networks trained on similar datasets.
- Therefore, it is not possible to build a more representative algorithm by selecting data that are known to be limited and hard to estimate (edges), as soon as any data is available, it should be included in the training.
- Multi-task learning did not improve performances (for the metrics tested). It may help generalisation but it was not possible to test it with a modelled dataset.
- Multi-task learning is greatly limiting the amount of data available and should not be pursued in future until *in situ* data collection is standardised and regroup all constituents at the same time.
- Hyperspectral signals may barely improve performance achieved (by a few percent) compared to a multi spectral approach.
- The improved results from hyperspectral signals could be more impactful for a real dataset impacted by all sorts of noise (atmosphere, lands, etc.), that was not simulated here.

In Chapter 6, the methods developed in Chapter 3 and 4 are applied to remote sensing estimates of the sea surface temperature which allows us to evaluate if some of these methods can be applied to any remote sensing inverse problem.

6. Retrieval of Sea Surface Temperature using neural networks

6.1 Introduction

This Chapter concerns estimation of sea surface temperature (SST) from remote sensing using the methods developed previously in this thesis from Chapters 3 to 5. The main idea here is to replicate different tests that have been developed previously and evaluate if they return similar observation with a different variable, here the temperature. By observing similar behaviours, generalisation of the method could be developed in future. Moreover, the availability of temperature is much greater and available globally, which allows us to challenge the development of current temperature algorithms.

The temperature of the sea surface is one of the most important variables to understand the Earth's climate system. The heat exchange between the ocean and the atmosphere controls the global temperature on Earth. Because of Earth's sphericity, the temperatures distribution on Earth's surface evolve with latitude. Heat is constantly exchanged following a gradient rule, from highly heated areas (at the equator or tropics during summer) to low heated ones, at the poles. Following the same rule, the ocean and atmosphere also exchange heat when one is colder than the other. Therefore, the temperatures of the atmosphere and the ocean are linked. Anthropogenic greenhouse gases emissions influence the atmosphere's heat absorption (which increases the temperature of the atmosphere) and directly change the temperature of the system by allowing the atmosphere to absorb more energy. Variation in temperature of the ocean for long periods has a direct impact on species sensitive to it and associated with other factors can lead to extinction, for example corals (Carpenter et al., 2008). Growing of plants is dependent on the temperature and some species can adapt their photosynthesis processes to maximize their absorption using temperature acclimation (Yamori et al., 2014). Within the ocean, temperature differences lead to creation of currents responsible of heat exchanges and build the actual climate state (Boyle and Keygwin, 1987; Clark et al, 2002). An evolution of the surface's temperature repartition would disturb the mechanisms and could alter the different oceanic currents responsible for the climate. A direct example is the Atlantic Meridional Overturning Circulation (AMOC) that has slowed between 1900 and 2014 with oceanic temperature increase (Latif et al., 2022). Moreover, the El Niño phenomenon impacts the world's climate and is directly linked to the temperature of the Pacific Ocean (Philander et al., 1984; Ropelewski and Halpert, 1987; Trenberth, 1997). Therefore, having access to high quality daily global SST maps of the entire Earth is a key tool to improve our understanding of the whole system.

With an emissivity of approximately 0.95, the surface of the ocean can be considered as an imperfect blackbody (when the emissivity equals 1) and the ocean emits long wave infrared light based on its temperature (Wu and Smith, 1997). A blackbody's brightness temperature can be retrieved using Planck's function (Equation 6.1). To include the fact that the ocean is an imperfect blackbody, an emissivity Ɛ was defined and needs to be applied (wavelength and viewing angle dependent). The relationship depends on the wavelength, and therefore the temperature (on land or sea) can be estimated by measuring the energy of light received at a specific wavelength. The amplitude of the signal is maximal at around 8-10 micrometres and decreases with increasing wavelength (see Figure 6.1).

$$B_{\lambda}(T) = \frac{2hc^{2}\lambda^{-5}}{e^{\frac{hc}{k\lambda T}}-1} \quad (6.1)$$

with $B_{\lambda}(T)$ the brightness temperature expressed in Kelvin, Planck's constant h = $6.626068.10^{-34}$ $m^{2}.kg.s^{-1}$, λ the wavelength in m, Boltzmann's constant k=$1.38066.10^{-23}$ $m^{2}.kg.s^{-2}.K^{-1}$, c=$2.99792.10^{8}$ $m.s^{-1}$ the speed of light in vacuum and T the object temperature in Kelvin. Therefore, the temperature of an object can be obtained by measuring its brightness temperature. Figure 6.1 shows different atmospheric transmission values for different wavelengths between 1 and 14 µm with variations depending on the thickness of the atmosphere, which is latitude dependent. Polar seas have a thinner and more emissive atmosphere, while the atmosphere at the Equator is thicker. Bottom panel shows the relationship between the brightness temperature of the ocean and the wavelength (smoothed lines) for 0, 10, 20 and 30°C (0° = 273.15K). MODIS bands 20 (3750 nm), 22 (3959 nm), 23 (4050 nm), 31 (11 000 nm) and 32 (12 000 nm) are also displayed.

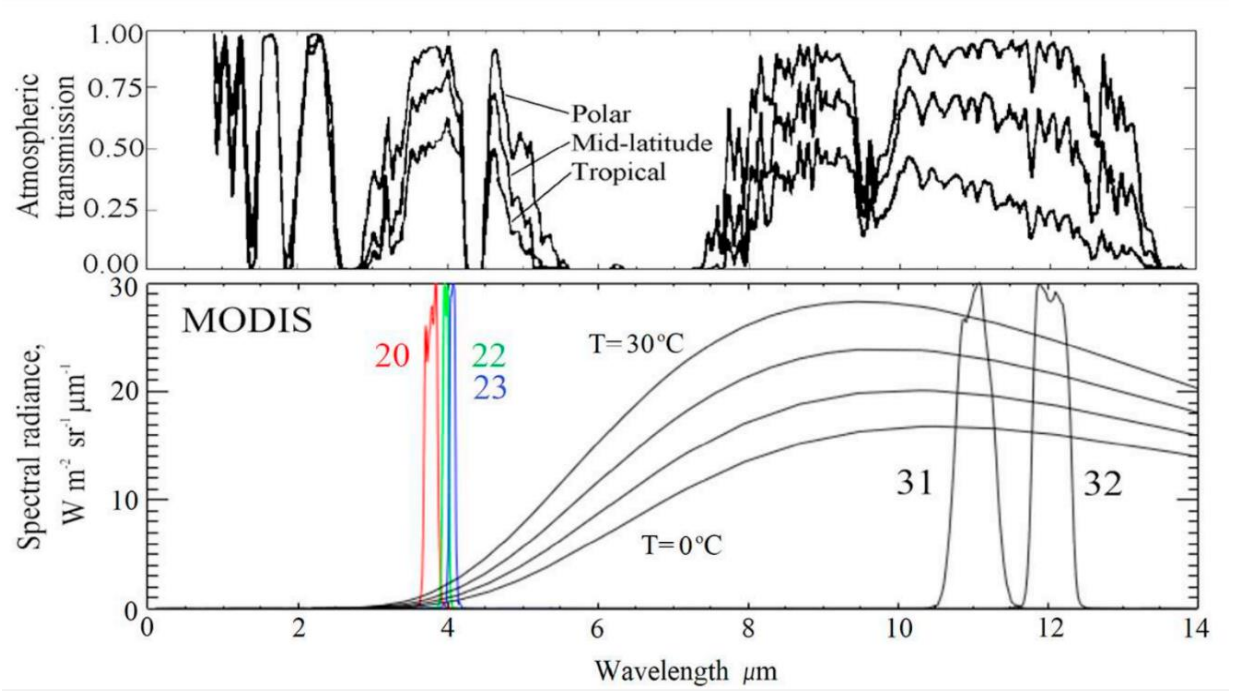

Figure 6.1. Top panel: Variation of the atmospheric transmission in the Infrared for Polar, Mid-latitudes and Tropical atmospheres. Bottom panel: Relative spectral response for the MODIS infrared bands 20, 22, 23, 31 and 32 (signal they can measure). The smoothed lines of bottom panel are the Plank's function between 0 and 30 °C (Minnett & Barton, 2010). The top panel is adapted from Llewellyn-Jones et al. (1984) and the bottom panel is taken from Kilpatrick et al. (2015). Both come from Minett et al., 2019 (Figure 4).

Remote sensing techniques allow retrieval of SST using different parts of the light spectrum, either the infrared (4 or 11 µm) or microwave wavelengths (30 cm to 3cm). While the light spectrum is continuous, these specific infrared bands were selected because they are less impacted by the atmosphere due to a higher emissivity in these regions. For example, for a wavelength λ = 10 µm, the atmospheric transmissivity ε = 0.970 with a viewing angle of 40°, and ε = 0.985 with a viewing angle of 55° (Hanafin and Minnett, 2005). Observation of SST from space started in 1964 with the first American sensor Nimbus-1 equipped with radiometers (Nordberg, 1965). The development of further radiometers continued, and various infrared and microwave sensors are available nowadays (Minett et al., 2019). For this study, the Moderate Resolution Imaging Spectroradiometer (MODIS) on Aqua spacecraft launched in 2002 and already used previously in Chapter 3 and 4 will be used. Infrared sensors are usually equipped on sun synchronous satellites and come with a resolution at nadir of approximately 1 km (MODIS and AVHRR) to 100m (Landsat 8 & 9 with a reduced swath) but are highly sensitive to clouds and can be impacted by sun glint during daylight. Conversely, clouds are almost invisible to microwave bands, hence

microwave sensors have access to an increased visibility, especially at high latitudes where cloud cover is dense (5 to 10% of pixel availability in north Atlantic). However, microwave sensors come with a reduced pixel resolution of approximately 25km for the best sensors due to the lower signal in the microwave part of the spectrum and thereby cannot make accurate predictions in coastal waters. They are also more sensitive to waves produced by wind (Nalli et al., 2008a; Nalli et al., 2008b). Both techniques come with their own limits and infrared sensors are usually used for coastal observation while microwave sensors can generate global daily images of the Earth's surface temperature not impacted by clouds.

Algorithms that are used to retrieve SST in the infrared rely on a combination of two close pairs of bands at 4 or 11 µm (Kilpatrick et al., 2015). Algorithms relying on the 4 µm bands are used during night time due to their higher sensitivity to sun glint (smaller signal magnitude compared to the 11 µm bands but varying faster than it). In contrast, 11 µm bands are used for daylight estimations of SST due to their higher signal magnitude, and greater resistance to glint. One important point of the SST definition is that it cannot be defined as a single variable. It is supposedly associated to the temperature within the mixed layer of the ocean (usually 0 - 10 m). But the temperature signal observed by sensors comes from the very top surface of the ocean, i.e. the first 10 micrometers for infrared sensors, the first mm for microwave sensors (Minnett and Kaiser-Weiss, 2012). Therefore, different definitions have been attributed depending on where the signal was coming from based on the sensor. Infrared sensors measure the skin sea surface temperature (SSTskin) while the microwave sensor measure the sub-skin temperature (SSTsubskin). However, it is almost impossible to measure the temperature for these precise layers of the ocean, and most *in situ* data coming from drifting buoys usually sample the temperature at approximately 20 cm below sea level, under the influence of mixing. Hence, SST estimations depend on the wind speed (Harris et al., 1994). Recently, visible bands of the 10m resolution satellite Sentinel-2 have shown capacity to estimate SST (Medina-Lopez and Urena-Fuentes, 2019) with relatively lower performances compared to dedicated infrared sensors, but with a much higher resolution, using neural networks.

Neural networks have been used in the past for different Earth Observation tasks, see review by Zhu et al. (2017) for more details. Current state of the art algorithms rely on Physics (Planck's law) and several corrections are made to make accurate estimation when non-optimal conditions happen due to external factors (sun glint, haze, high wind speed). The main objective of this Chapter is to develop neural network algorithms using information available from SST matchup datasets for infrared sensors, evaluate if they can overcome the issues affecting current algorithms and return improved performance. Two datasets have been collected for the MODIS Aqua sensor. The first uses the 15 MODIS Aqua bands

processed for Chapter 3 and 4 (visible and infrared bands, 412 to 2130 nm), and include four brightness temperature bands in the infrared (3959, 4050, 11000 and 12000 nm) over European northwest shelf seas; it will be called the European dataset. For convenience, these four bands will be denominated as long wave infrared bands (LWIR), rather than separating into medium wave infrared bands (MWIR) for the 4000 nm and LWIR for the 11000 nm bands. Similar to Chapters 3 and 4, this dataset contains matchups between temperature, salinity and MODIS Aqua light information made over the 2002 to 2020 period, coming from the Copernicus Marine Environment Monitoring Service (CMEMS), International Council for the Exploration of the Sea (ICES) and Marine Scotland datasets. The idea of developing SST or SSS algorithms emerged from a potential link established by Medina-Lopez and Urena Fuentes (2019) between temperature, salinity and visible bands and Medina-Lopez (2020). Similarly to Chapter 4, this dataset allows evaluation of potential benefits from inclusion of visible and infrared bands for improved SST estimations. There is also the capacity to evaluate a potential visible-infrared salinity algorithm, while current algorithms use microwave bands.

The second dataset is the SeaWiFS Bio-optical Archive and Storage System (SeaBASS) global matchup datasets for MODIS Aqua processed by NASA, for the same period. It matches level 2 MODIS Aqua with temperature measured by drifting buoys worldwide. The main difference with the initial dataset is that the SeaBASS data set was created for a SST purpose, i.e. including other variables such as the wind speed, different sun and sensor angles. However, it only contains two visible bands because the visible part of the spectrum is currently not used for SST algorithms. For the same reasons, extra data such as the wind speed or sensor viewing angles were not processed for the European dataset because to date, nobody has used these variables in a Chl algorithm. It is now clear from Chapter 4 that the decision of processing or not a particular type of data should be made after having tested its capacity to be relevant to retrieve a specific variable. Matchup datasets should be generic to allow design of more complex algorithms such as neural networks, as these have the capacity to benefit from more information in general, rather than limiting it to information contained within the most up to date Physics formulation. Unfortunately, the time required to process data does not allow us to reprocess the SeaBASS or European datasets to include the missed metadata within the timeframe of the PhD project, and therefore performances will only be tested using available information.

By having access to different information, these two matchup datasets are complementary and allow exploration of different aspects for SST algorithm design using neural networks. Neural networks are the main method used in this study. The first dataset gives the capacity to evaluate the impact of

including visible bands for SST algorithm design using the European dataset. It grants the possibility for creation of a visible-only algorithm to be compared with literature. Because the clouds are white and backscatter the light across the full spectrum, visible bands inform about their presence. Retrieval of the salinity using visible and infrared bands can also be tested using the same information. Moreover, having access to both temperature and salinity at the same time gives the opportunity to estimate both variables at the same time (multitask learning); or separately, including the other as an input. Secondly, using the SeaBASS dataset, a similar approach can be taken by evaluating the inclusion of different information such as infrared bands, sensor and solar angles, or the wind speed during daylight availability. Night SST retrievals can also be assessed and a similar approach can be applied with a restricted amount of information available (no visible bands). Eventually, an "atemporal" algorithm (applied to both day and night data) can also be assessed. For a consistent and impartial evaluation of the different algorithms developed in this study, only independent data (not used for the training of the algorithm will be considered for performance assessment. Neural network generalisation capacities will therefore be only briefly shown in the last section.

6.2. Materials and methods

6.2.1 *In situ* samples

For this study, *in situ* samples come from different institutions. The European dataset contains *in situ* sea surface temperature data collected through the Copernicus Marine Environment Monitoring Service (CMEMS); International Council for the Exploration of the Sea (ICES); and Marine Scotland Science. All the available *in situ* data collected between the surface and a 2 meters depth, from July 2002 to January 2020 were used. The SeaBASS dataset contains *in situ* sea surface temperature collected by automatic drifting buoys. The data come from a 20cm depth on average, from July 2002 to January 2021. The temperature of the water has been measured using two different techniques in the past, bathy-thermometers or conductivity-thermometers. At the surface, most temperature samples come from conductivity measurements operated by drifting buoys (Kilpatrick, 2015). Accuracy of the *in situ* sample estimations highly depend on the method used, and a difference of 0.01K to 0.1K can be expected for these methods. For convenience, temperature will be expressed in °C (0°C = 273.15K).

Salinity is a measure of the amount of dissolved salts in a water body and is expressed in Practical Salinity Units (PSU), with 35 PSU equal to 35 grams of salt per 1 kilogram of seawater. Conductivity measurements can also return a salinity concentration of the water sample, and both salinity and

temperature were commonly measured during the past 20 years in the European region. Therefore, because the data was also available from the same dataset in Europe, it has been decided to include salinity in the process to briefly test the feasibility of developing a visible and infrared algorithm to estimate it. *In situ* salinity is available from the European dataset only, the SeaBASS dataset is purely a temperature dataset and the drifting buoys deployed for this task do not measure the salinity. It is the most common value expected in global ocean (Hosoda et al., 2009). NaCl (sodium chloride) is the most common salt present in seawater by several orders of magnitude (Huber et al., 2000). While different methods exists to measure salinity, the actual operational method uses the water conductivity as a proxy (Qian et al., 2018). The conductivity is linked to the amount of chloride ions inside the water body and can therefore be transformed into salinity. The conductivity of a water body is also impacted by other factors, such as the temperature or refractive index of the water and needs to be corrected for these effects prior to making an appropriate salinity measurement. *In situ* salinity and temperature measurements for the European dataset were collected using mostly CTDs (conductivity, temperature, depth) for the same water volume.

6.2.2. Satellite data

The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Aqua spacecraft has produced images since early July 2002. MODIS data were downloaded as L1A products from the National Aeronautics and Space Administration (NASA) ocean colour servers using the R2018 calibration and processed to level 2 for the SeaBASS dataset, or level 3 for the European dataset. The difference between level 2 and level 3 concerns the interpolation applied to level 3 data; while level 2 keep the projection captured by the satellite and more metadata are available (such as the flags that inform from the presence of other light contributors). The SeaBASS dataset was processed separately by NASA and is available as a matchup dataset directly. For the European dataset, processing details can be found in the materials and methods section of the thesis (Section 2.3).

6.2.3 The European matchup dataset

The European matchup dataset contains *in situ* temperature and salinity, matched with visible and infrared MODIS Aqua bands (412 to 2140 nm, with inclusion of the 4 LWIR bands at 3959, 4050, 11,000 and 12,000 nm). All MODIS-Aqua images available between 25°W – 13°E and 48°N – 65°N during daylight only from July 2002 to 2020 that were processed for Chapter 3 were used to build the European matchup dataset. The visible bands were processed at top-of-atmosphere (TOA) and therefore not atmospherically

corrected as it has shown improved performance in Chapter 3. NASA algorithms SST4 and SST11 were also saved during the process (details about the algorithms given in Section 6.2.4). *In situ* data and satellite pixels were co-located using a temporal window of 3 hours between the satellite images and *in situ* samples using the exact pixel, similarly as Chapter 3. 19,513 matchups with both a temperature and salinity sample were collocated with MODIS Aqua and are available following this process, presented in Figure 6.2. The solar zenith angle was processed later relying on geospatial and temporal information of the pixel's matchup information using a solar position algorithm (Reda and Andreas, 2004). The histogram of the 19,513 *in situ* temperature and salinity samples are shown on panel b and c, respectively. The map highlights that most data were sampled by cruise campaigns rather than drifting buoys. The temperature in the region varies from 0°C in winter to 25 °C in late August. Salinity of the seawater is close to the global value of 35 PSU, with some coastal areas influenced by freshwater inputs sometimes reaching 0 PSU.

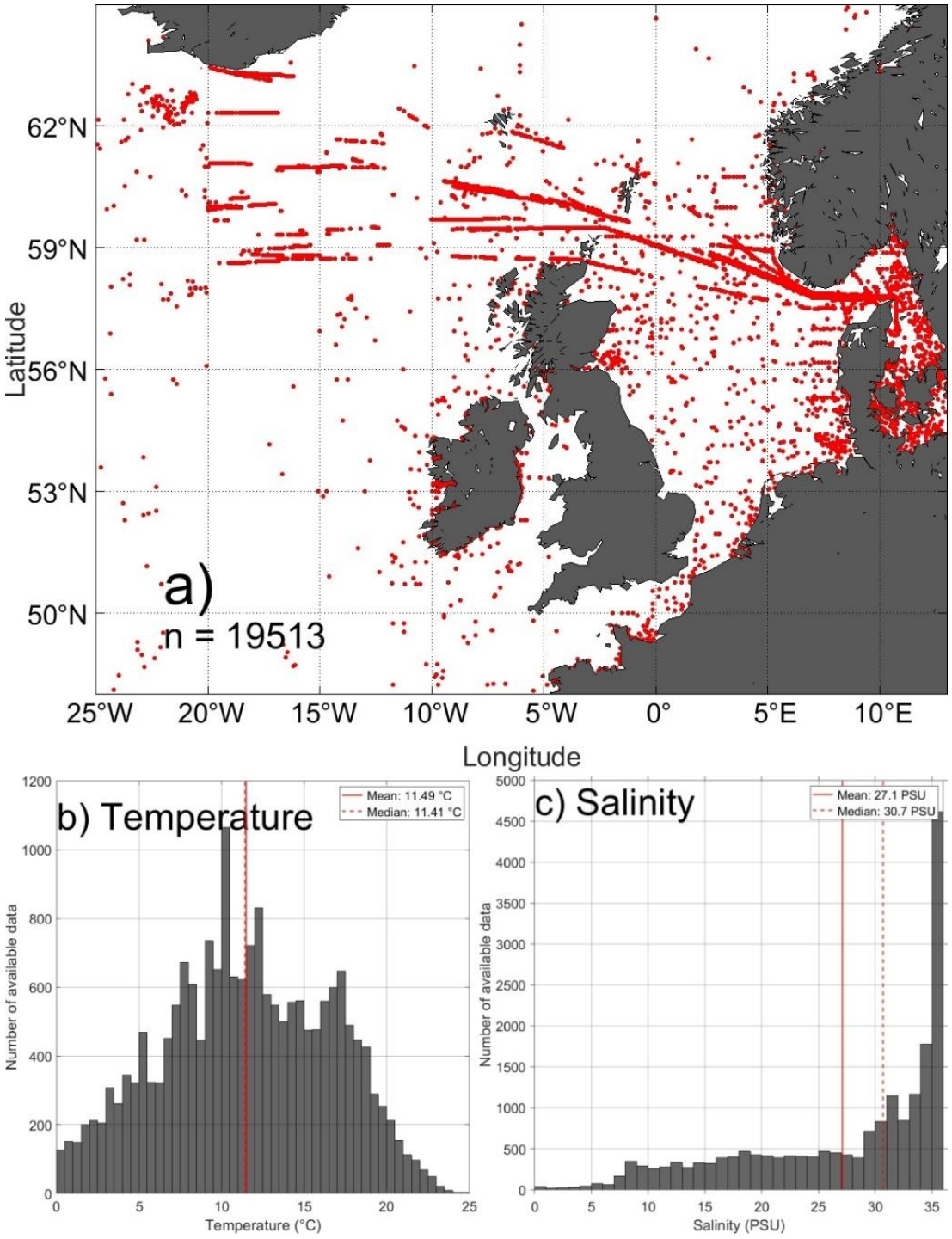

Figure 6.2: a) MODIS Aqua matchup map for the SST within the first 2 meters of the ocean, between July 2002 and 2020, using a +/- 3 hours temporal window and an exact pixel condition for the European region. b) *In situ* temperature and c) salinity histograms associated to the matchups.

6.2.4. The SeaBASS dataset

The SeaBASS SST matchup data set is publicly available for MODIS Aqua and other sensors from the Ocean Biology distributed archive system (OB.DAAC) SeaWiFS Bio-optical Archive and Storage System (SeaBASS). This data set contains 24 million matchups all over the world for the MODIS Aqua sensor at level 2, with coincident *in situ* sea surface temperatures and satellite available light signal using a ±30 min temporal window within a 10 km area, for day and night (Figure 6.2a). Nine MODIS Aqua bands are available in the SeaBASS data set: the four LWIR bands at 3950, 4059, 11000 and 12000 nm mentioned earlier which operational SST algorithms rely on; three bands in the LWIR at 6715, 7325 and 8265 nm and two bands in the visible at 678 and 748 nm at top-of-atmosphere. More spectral bands are not processed by the group at the moment. *In situ* samples from this data set come from drifting buoys measuring temperature at the surface of the ocean at a depth of 20 cm below the sea surface (the real depth is not reported). Other data are also available, such as the windspeed, both solar and sensor zenith or azimuth angles and information concerning the flags for the satellite pixel. Lastly, a day and night flag is available to identify whether a pixel belongs to the day or night. Polar areas are also less sampled due to ice coverage during winter. The temperature ranges from –1.8°C (for 35 PSU, a moving water body can reach approximately –1.91°C) to 33°C. Above 32°C, the water evaporates faster than it is heated, and permanent warmer waters bodies in open waters are not found. More buoys are released in tropical waters, which explains why more temperate temperature (25–30°C) are available compared to the polar ones, covered by ice during winter.

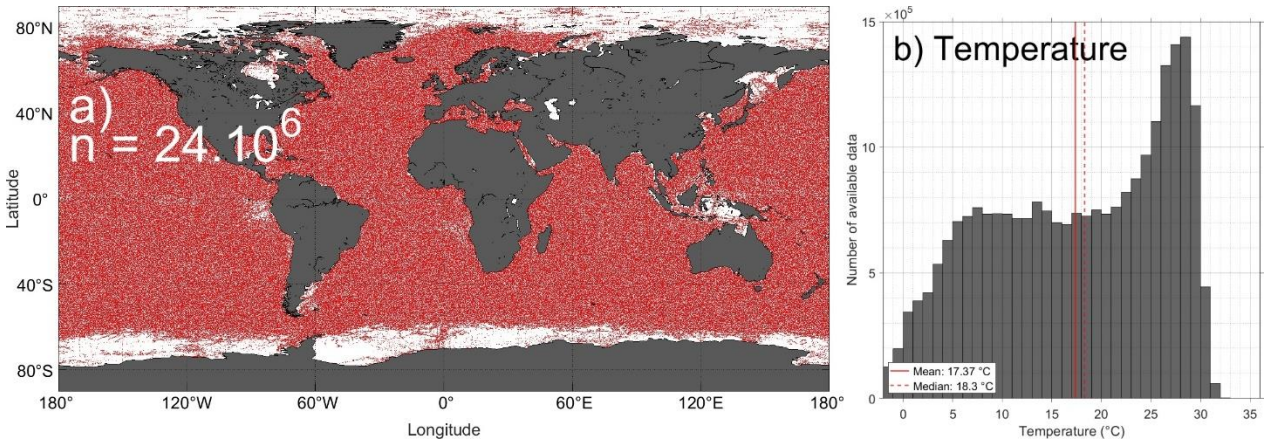

Figure 6.3: a) MODIS Aqua SeaBASS matchup map, between July 2002 and 2020, using a +/- 30 min temporal window and a +/- 10km pixel condition. b) *In situ* temperature histogram associated to the matchups (1°C bin).

Both of these matchup datasets are impacted by atmospheric correction at different levels. The radiative energy of LWIR bands is measured at top of atmosphere, corrected by effects from the atmosphere and later transformed into a brightness temperature measurements using Planck's law (Luo et al., 2019). Imperfect atmospheric correction has been suspected to be an issue for ocean colour data in certain conditions, and it turns out it is also the case for temperature measurements. For the approximately 24 million matchups available, 12 million are available during daylight and 12 million during nighttime. For the 12 million samples available during daylight, less than 7 million are available after cleaning the data for negative values for any of the LWIR bands used in SST algorithms (LWIR-4 and 11), i.e. almost 40% of the matchups were overcorrected by the atmospheric correction process and the atmosphere absorption has been overestimated. Similarly to Chapter 3, having a non-negative brightness temperature for all LWIR bands does not mean that the signal has not been degraded, and some of the values can be physically realistic but not entirely linked to the water leaving signal. The atmospheric correction process is also directly responsible for data flagging when thick clouds, straylight or lands are detected. Therefore, the results shown in this chapter directly depend on performance obtained by the correction method applied and cannot be evaluated as a TOA raw product. Data would have to be reprocessed globally to assess this question, which would require processing capacities above those available for this thesis.

6.2.5. Sea surface temperature algorithms

Two SST algorithms developed by NASA are available for both datasets. Both use non-linear relationships between LWIR bands and the sensor zenith angle, and are based on an updated version of the Walton algorithm (Walton et al., 1998), see Eq. 6.2 and 6.3, taken from Kilpatrick et al. (2015). The first algorithm relies on bands at 3959 and 4050 nm and is called SST4. The spectral information near 4000 nm is smaller but evolves faster between the two 4000 nm bands when the temperature increases compared to the algorithm using two 11,000 nm bands. This lower signal induces a sensitivity to sun glint (Kilpatrick et al, 2015), therefore this algorithm is generally used during nighttime. The second one uses bands at 11,000 and 12,000 nm and is called SST by NASA but will be named SST11 in this study. It is the algorithm used during daylight due to its greater robustness against sun glint, which is due to a higher signal magnitude compared to SST4. Because the satellite dataset gathered for this study contains visible light and tries to elucidate the potential benefits from their inclusion for SST retrieval, it can only be performed during daylight (for the European dataset) and the SST11 algorithm will be the one to be

compared to. However, it is also interesting to observe performance reached for SST4 under these conditions.

$$\mathrm{SST4} = c_1 + c_2 * T_{3.9} + c_3(T_{3.9} - T_4) + c_4 * (\sec(\theta) - 1) \quad (6.2)$$

$$\mathrm{SST11} = c_1 + c_2 * T_{11} + c_3(T_{11} - T_{12}) * T_{\mathrm{sfc}} + c_4 * (\sec(\theta) - 1) * (T_{11} - T_{12}) \quad (6.3)$$

where SST4 or SST11 are the retrieved surface temperature, $T_{3.9}$, T_4, T_{11} and T_{12} are the brightness temperatures for the 3959, 4050, 11000 and 12000 nm bands, T_{sfc} is an estimate of the surface temperature, θ is the satellite zenith angle and c_1 to c_4 "constants" that are fine-tuned with empirical data related to drifting buoys and satellite matchups on a month-of-year basis (Luo et al., 2019), inherited from the non linear SST algorithm developed for AVHHR (Walton et al., 1998).

The T_{sfc} is a blended remote sensing SST coming from a 0.25° spatial resolution product (Reynolds et al., 2002; Reynolds et al., 2007) taken on a sliding window weekly basis. Compared to Chl algorithms, SST algorithms evolve on a month-of-year basis and are not constant, at least for the period 2005 to 2014 (Kilpatrick et al., 2015). These algorithms are limited by four main issues, a) the absorption of water vapour which is latitude dependent (hidden in the c_x coefficients); b) the thickness of the atmosphere and therefore the viewing angles; c) the requirement to have access to an estimate of the surface temperature prior making an estimate (T_{sfc}) and d) the aerosol contents during volcanic events (Kilpatrick et al., 2015). Since the MODIS Aqua reprocessing of 2019, the algorithm mentioned in Eq. 6.2 and 6.3 changed and now contain more information (Luo et al, 2019). It now incorporates slightly more information of which MODIS mirror is used, the sensor zenith angle is made negative for pixels in the first half of the scan line, coefficients c_1 to c_4 are now month-of-year and latitude dependent and the T_{sfc} is the SST derived by the Canadian Meteorological Center Global Foundation Sea Surface temperature product (CMCSST). Incorporation of latitude was added to retrieve change of the water vapour content of the atmosphere that is maximum at the equator where more solar energy is received and minimum at the poles. This reprocessed version of the SeaBASS dataset was not available at the time of writing, data were downloaded using the R2018 version collection 5. The version 6 algorithm can be seen in the technical notes (Kilpatrick et al., 2014) and Luo et al., (2019).

6.2.6. Neural networks

The neural networks used in this study have proven efficiency in Chapter 3 and are used here, they are multi-layered perceptrons that estimate the temperature based on different input bands. Neural networks have the capacity to learn how to fit a complex signal. They are made of layers of artificial

neurons. A neuron transforms different inputs into a single output using an activation function, with the Rectified Linear Unit (ReLU, y=x if x>0, else y=0) function used for this study. Once the last output layer is reached, an error is measured using the Levenberg-Marquart function. The error is then backpropagated to the first input layer, the weights connecting each layers of neurons are updated, and if possible, an optimal combination of the weights is finally found to provide the minimal possible error of the prediction after several iterations. Different variables such as the light signal at a given wavelength or the solar zenith angle can be used as inputs. Similarly to what has been done previously in Chapter 4, different band combinations are tested to understand the benefits from each part of the spectra, i.e, visible, infrared, long wave infrared, angles, salinity and windspeed, when available. Based on observations made previously (Chapter 3 to 5), an architecture size of 3 layers and between 10-20 neurons per layer returns nearly optimal performance while reducing the capacity of the network to overfit the signal (remember the training dataset and perform poorly onto independent test data) and underfit (when the algorithm is not complex enough). Hence, **all neural networks used in this study consist of 3 layers of 20 neurons each**. This allows us to accommodate for the varying number of inputs that can go from 2 (when only 3959 and 4050 nm bands are used for example) to 21 for the European dataset (19 wavebands available, salinity and solar zenith angle) without having to tweak the size of the architecture each time. In the end, between 921 weights when 2 inputs are used and 1301 when 21 inputs are used need to be tuned by the network. Using a low amount of inputs and a high amount of neurons can cause overfitting that can be hard to detect, but this is avoided by showing only the independent test datasets for the different performance assessments.

6.2.7. Metrics used to evaluate performance

Various statistical metrics are used to evaluate performance of different SST algorithms and details are available in Section 2.4 of this thesis. Pearson correlation coefficient is used (Eq. 2.40). The Mean Absolute Difference (MAD, Eq. 6.4) is slightly different from equations used for ocean colour in the method section of this thesis as it does not log transform the model and obsevation (Eq. 2.41). The reason comes from the several orders of magnitude chlorophyll can span to, while temperature ranges linearly from -2 to 32 on average. The root mean square deviation RMSD is commonly used in the SST field (Eq. 2.42). A more robust absolute difference insensitive to negative values temperature can reach is used, the weighted absolute percentage difference (WAPD, Eq. 2.43) along with the BIAS (Eq. 2.44).

$$\mathrm{MAD} = \frac{\sum_{i=1}^{N} |Mi - Oi|}{N} \tag{6.4}$$

6.3. Results

6.3.1. SST NASA algorithms performances

Before assessing neural network performance, current state of the art algorithms need to be evaluated for both the European and SeaBASS datasets, which provides a basis for the later comparison. NASA SST algorithms are available as an operational product as daily images on NASA website (https://oceancolor.gsfc.nasa.gov/, accessed on 11 July 2022). These empirical algorithms were developed using the SeaBASS dataset. The SeaBASS dataset contains matchups between drifting buoys and MODIS Aqua with a 30 min temporal window, a 5x5 pixel box, and comes with quality controlled flags. The objective of this chapter is to test the hypothesis developed in Chapter 4 at a global scale: inclusion of more spectral information that provides a meaningful signal for either the retrieval of the surface variable or the different contributors to the light signal allows creation of more accurate neural network algorithms. Hence, the SeaBASS dataset will be separated into smaller datasets with increasing data quality. Two set of flags that can be set to mask a pixel when the quality is questionable are available. The first one, l2flags, is used in the ocean colour field and applied to chlorophyll products (using information in the visible and infrared, previously mentioned in Chapter 3 Section 3.5.4). The second one, qual_sst is specific to the SST field, and uses infrared information to make decisions about the potential contributor to a pixel. More information about this flag is available on the NASA website (https://oceancolor.gsfc.nasa.gov/atbd/sst/flag/, accessed on 11 July 2022). The official products rely on this qual_sst flag, while the l2flags product is provided as an additional available information that can be used during daylight only due to the absence of visible information during the night. The first section of the results (6.3.1.1) shows the quality achieved by NASA SST4 and SST11 algorithms at an operational level (with the qual_SST flags sets to 0, no flags present). The second section (6.3.1.2) shows the performances achieved with different subsets of the daylight SeaBASS dataset and for the European dataset when different flags are applied to improve data quality. The first section helps to understand the real performances of the products, while the second section allows tests of improvement under poorer conditions when other contributors degrade the water leaving light signal (which we will call noise). Because different datasets or subset versions are used in this Chapter, a table will summarise the different metrics obtained at the end of the results section to ease readability.

6.3.1.1 Performances in operational conditions

Performance of the SST4 algorithm is displayed for nighttime observations (Figure 6.4a) and the SST11 algorithm is displayed for daytime observations (Figure 6.4b). Figure 6.4 and every following matchup figure shown in this Chapter represent the density of matchups per 0.25°C bin. When more data fall inside a bin, the bin gets darker. This was used to avoid the huge numbers of matchups available in this Chapter (3 million during the day or the night each). Both algorithms shown in Figure 6.4 return good performances in operational conditions, with more than 98.87% of the matchups provided within a ±2°C difference, and the majority being very close to the 1:1 line (darker area). MAD of 0.45 and 0.48 are obtained and they both return a bias, -0.4°C during night, -0.24°C during the day, respectively. These are the standard performances for the SST field and will be compared later with a neural network algorithm.

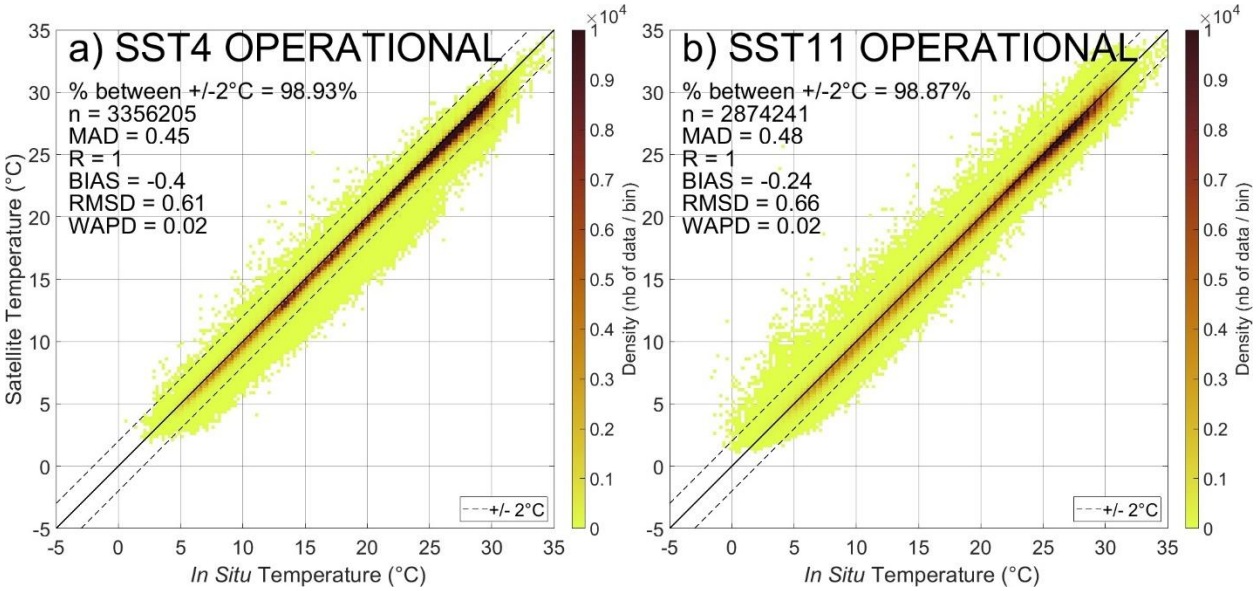

Figure 6.4: Performances of the nighttime algorithm SST4 (a) and the daylight algorithm SST11 (b) applied in operational conditions (qual_sst flag set to 0).

6.3.1.2. Performances in non-operational conditions during daylight

Now that the operational performances are known, the performances under poorer conditions can be assessed for both the European dataset and different SeaBASS subsets, using daylight data only. Figure 6.5 represents performances achieved for the SST4 (panels of the left column) and SST11 (panels of the right column) algorithms for both datasets and for different subsets of the SeaBASS dataset. Both

algorithms return different performances. As expected, the SST11 algorithm provides the most accurate estimations for this data set. These are daylight datasets, and therefore contaminated by glint which is known to more significantly impact the bands used by the SST4 algorithm compared to the bands used for the SST11 algorithm, and is the main reason of the existence of the two algorithms. The European data performances are relatively good because pixels where atmospheric correction has been found to failing have been masked out which is unusual for a SST product. There is a clear need to clean the SeaBASS raw dataset (any data available, Figure 6.5, panels c and d) as some matchups are obviously impacted by specific flags. The two main flags identified that cause the problem are the CLDICE and STRAYLIGHT flags. The CLDICE flag (CLouD and ICE) is set when a pixel is suspected to be contaminated by clouds (thick, haze or thin cirrus) or sea ice due to a higher backscattered reflectance. The STRAYLIGHT flag is set when a lot of straylight is detected, meaning that information received is impacted by a direct reflection of the sun and less correlated to the ocean water leaving signal. After elimination of these two flags from the SeaBASS raw dataset (flag information is not available for the European dataset as it is a level 3 product), a better evaluation of SST4 and SST11 can be performed using this less permissive dataset (Figure 6.5e and f). This subset SeaBASS dataset will be called the “noisy” dataset, as it still impacted by other flags related to other contributors and, therefore, noise., SST11 shows the capacity to deal with most of the remaining noise by providing 97.88% of the data between the +/-2°C limits. The size of the raw dataset was halved after removal of the two flags to create the noisy dataset (6.8 million to 3.4 million).

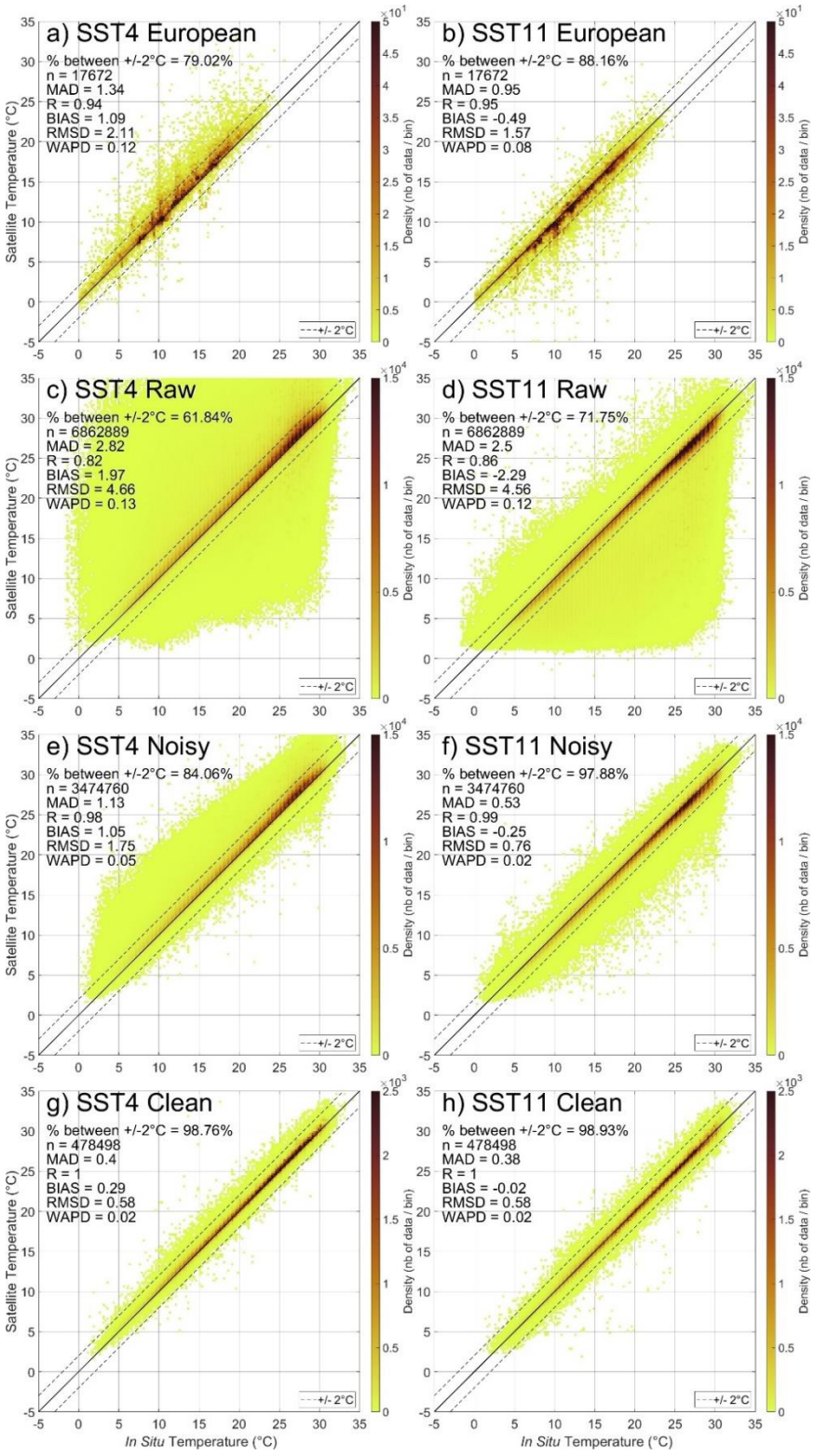

Figure 6.5: Comparison between *in situ* (x axis) and satellite (y axis) temperature measurements for the SST4 (left column) and SST11 (right column) algorithms, for the European waters (first row), SeaBASS raw (second row), SeaBASS noisy (third row), and SeaBASS clean (fourth row) datasets.

This chapter is about trying to improve SST algorithms with the use of neural networks, i.e. testing inclusion of more information to deal with noise when it occurs, when other algorithms fail. Therefore, the SeaBASS noisy dataset is the most interesting dataset to develop a NN with. A second subset of the SeaBASS raw dataset is created when no flags are allowed using the l2flags product. It is called the SeaBASS "clean" dataset (Figure 6.5, g and h). This dataset is more restrictive and smaller than the noisy dataset, with approximately 500,000 matchups available. Performances of both SST4 and SST11 are good and similar for this subset, with SST4 still showing a positive bias due to some minor sun glint contamination as it is a daylight dataset (+0.29°C), elevating the brightness temperature at these bands, while SST11 is resistant to it, which supports the use of the SST11 algorithm during daylight. Performances of the clean dataset are better than the operational one (Figure 6.4) due to the reduced number of available data of higher quality. The European satellite dataset (Figure 6.5a and b) has been designed initially for Chl estimation and shows less apparent noise than the SeaBASS noisy subset, with only some matchups far from the +/- 2°C dashed lines, but returns poorer metrics compared to the noisy dataset. This observation linked with the greater number of available matchups led to the use of density to represent these matchups rather than all points available compared to previous chapters. While visually the European data estimation (Figure 6.5b) look better than the noisy one (Figure 6.5f), it is only because this dataset is 200 times bigger than the European dataset. It allows for more noisy data to visually "take space" while the majority of the data are very close to the 1:1 line. The clean dataset is only a slight improvement of the noisy dataset, with MAD going from 0.53 to 0.38°C while RMSD improves from 0.76 to 0.58°C, for 7 times less data. No particular trends depending on the temperature can be observed, and most data (98%) follow the 1:1 line within a +/-2°C window.

6.3.2. Development of neural networks

It has been shown previously in Chapter 3 that neural networks can deal with noise (Jim et al., 1993). As mentioned in the method section (Section 6.2.4), all neural networks used in this Chapter use 3 layers of 20 neurons each, and the ReLU activation function. All the results shown are test data to avoid overfitting uncertainties. Except the European dataset, all networks were trained using 50,000 training points, with the remaining data being presented as test data. Note: when compared with Figure 6.4 and 6.5 that show the official algorithm performance, 50,000 points will be missing from the NN approach, i.e., not the exact same data are compared for independent conditions. For Figure 6.6, a neural network algorithm was trained using the SeaBASS raw dataset during daylight (to be compared with Figure 6.5, c and d). The neural network was developed using only both pairs of LWIR-4 and LWIR-11 bands with

inclusion of the sensor zenith angle (5 inputs in total). Most data belong between a ± 2°C limit (84%), the poorer estimates seem evenly distributed from both sides of the 1:1 line. Performance obtained are not of high quality for this specific dataset, and NNs return better metrics than the product using the daylight SST11 algorithm (MAD of 1.1 against 2.5, Figure 6.5d). While the densities of data found near the 1:1 line is decent, a lot of noise is visible within a ± 20°C window and the whole algorithm cannot really be considered as working, 16% of the samples are poorly estimated. The underestimations comes from the presence of clouds which are mostly colder. The overestimates could potentially come from stray light contamination, but are not present in Figure 6.5d. This difference with Figure 6.6 provides an interesting observation related to neural networks in general: rather than estimating the clear data efficiently and the noisy data with huge underestimations like the SST11 algorithm is doing in Figure 6.5d (and it is physically right, clouds are colder), neural networks are designed to minimise the global error and only show over and underestimations when the amount of noisy data is substantial. Being able to deal with noise for chlorophyll when designing an algorithm gives access to more data. However, data availability is not an issue for SST due to the deployment of numerous buoys and stricter datasets can be created. This observation promotes the need to clean data prior to implementation so that neural networks can reach better performance. This is why the two additional subsets of the raw dataset were created in Figure 6.5, to allow neural networks to deal with genuinely noisy dataset (SeaBASS noisy) and a cleaner version where no detected remote sensing noise is allowed (SeaBASS clean).

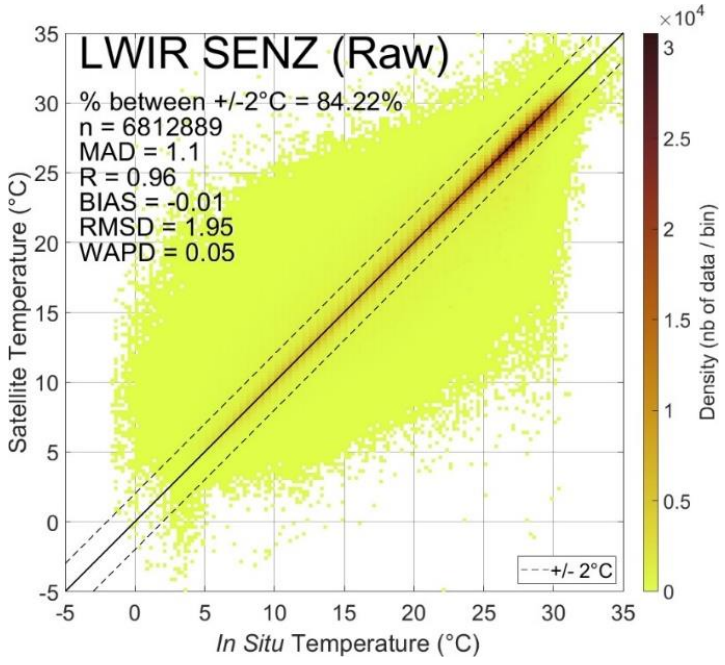

Figure 6.6: Neural network performance using the Raw SeaBASS dataset.

Building from observations made for chlorophyll in Chapter 4, we can test several combinations of inputs that alter the light signal and evaluate the performance reached by each. For this task, rather than showing the train, validation or test sets results, we will train the algorithm with 10,000 points of the European dataset that contains 19,513 matchups in total; or 50,000 points of each of the SeaBASS datasets that contains 3.4 million matchups for the noisy dataset and 0.5 million for the clean dataset, and ≈3 million for the operational dataset. The results are displayed using the remaining data, entirely independent and not seen by the network. This approach can be taken because the number of available data is numerous thanks to the automated efforts made by the community in the past. While the amount of data for a NN to find an accurate solution is unknown, the easier the relationship is between an input and a target is, the less a NN will require data to produce consistent estimations. Due to the nature of the SeaBASS matchup datasets with numerous available data, the training datasets represent between 10% (for the clean dataset) and 1.5% (for the noisy dataset) of the total amount of available independent data. Under such conditions, there is no particular need to evaluate generalisation capacities of the neural networks. All data shown are independent and the results can be considered as the general performances achieved by the NN algorithm.

Neural networks were developed for the European dataset (Figure 6.7), the SeaBASS noisy dataset (Figure 6.8) and the SeaBASS clean dataset (Figure 6.9) and the SeaBASS operational dataset (Figure 6.10). Results obtained for each NN are summarised in Table 1 at the end. For the European dataset (Figure 6.7), algorithms have been developed using a mixture of either LWIR-4 bands (3950 and 4050 nm); LWIR-11 bands (11,000 and 12,000 nm); addition of the solar zenith angle (SOLZ); visible and infrared bands (412 to 2130 nm, called 'Visible' on the figure 6.7f for convenience) or any of these data available including the salinity (All, Figure 6.7h). For the SeaBASS datasets (Figures 6.8 to 6.10), algorithms have been developed using a mixture of either LWIR-4 bands; LWIR-11 bands; the sensor zenith angle (SENZ); other angle information available (solar and sensor azimuth or zenith angles, SOLA, SOLZ, SENA, SENZ); two Rhot bands (678 and 748 nm); the intermediate LWIR bands (6500, 7500 and 8500 nm, called "All Light") and the windspeed (All, including any previous input). Each combination is summarised in table 1 below. It has to be noted that SST4 or SST11 information from NASA algorithms are not available for 100% of the European available light matchups, with approximately 10% missing due to masks applied by NASA quality control processes while the visible light signal is available. Similarly, the European MODIS Aqua product was initially built for chlorophyll estimation, and some pixels that have been masked (application of the ATM FAIL flag, Chapter 3) that might otherwise be available for an SST product. This means that a proper

SST satellite product for the European region could return more matchups if the European dataset had been initially created for a temperature analysis.

Table 1: Different input combinations used to train the neural networks for both datasets.

Information used	3959 and 4050 nm	11,000 and 12,000 nm	Solar Zenith angle (SOLZ)	Solar Azimuth Angle (SOLA)	Sensor Zenith Angle (SENZ)	Sensor Azimuth Angle (SENA)	VISIBLE bands (412-2130 nm)	Intermediate Infrared bands (6500, 7500, 8500 nm)	Wind speed	Salinity
European dataset										
LWIR-4	x									
LWIR-11		x								
LWIR4-SOLZ	x		x							
LWIR11-SOLZ		x	x							
LWIR SOLZ	x	x	x							
Visible							x			
All light SOLZ	x	x	x				x			
All	x	x	x				x			x
SeaBASS dataset										
LWIR-4	x									
LWIR-11		x								
LWIR-4 SENZ	x				x					
LWIR-11 SENZ		x			x					
LWIR SENZ	x	x			x					
LWIR ANGLES	x	x	x	x	x	x				
ALL LIGHT Angles*	x	x	x	x	x	x		x		
ALL	x	x	x	x	x	x		x	x	

6.3.2.1. The European dataset

Performances obtained for the European dataset in Figure 6.7 can be compared against SST4 and SST11 algorithms in Figure 6.5a and 6.5b. Except the Visible algorithm (Figure 6.7f, not using LWIR bands), all algorithms developed here outperform the state-of-the-art SST11 algorithm and more than 92% of retrievals fall between the ± 2°C dashed line, compared to 88% earlier for SST11 (Figure 6.5b). Even algorithms using only LWIR-4 or LWIR-11 bands (Figure 6.7a and b) outperformed their pairs relying on light and the sensor zenith angle, with the latter missing from the European dataset. This observation highlights that the sensor zenith angle is important for physical retrieval of temperature but does not seem mandatory for a neural network approach. Relying on the light emission signal only (not using metadata linked to the different viewing angles) already produces good retrieval, at least for a small region like northwest European shelf seas. Inclusion of the solar zenith angle does slightly improve retrieval for all metrics (+~5%, Figure 6.7c and 6.7d). Using both LWIR-4 and LWIR-11 bands along with the solar zenith angle increased performances again (Figure 6.7e). While it does seem to show some kind of correlation, the non-LWIR algorithm (Figure 6.7f) did not return optimal performance compared to LWIR algorithms and performances are closer to what is observed for Chl with tendency to overestimate low values and underestimate high values. Using LWIR, solar zenith angle and visible bands (6.6g) improves performance again. Addition of salinity as an input (6.7h) marginally improved the performances. It is hard to evaluate how each input separately helped the network, except a clear fundamental dependence on the LWIR bands. However, it is clear that addition of any of this additional information helped the network to better deal with the effect of other light contributors. The main reason why the European dataset was created was to test the feasibility for MODIS Aqua to produce accurate SST retrievals using visible and near infrared information (i.e. not including any LWIR data). The results for this (Figure 6.7f) are much worse than any algorithm including a LWIR band. A higher pixel resolution could help getting better results, but greater accuracy similar to the other algorithms tested should not be expected.

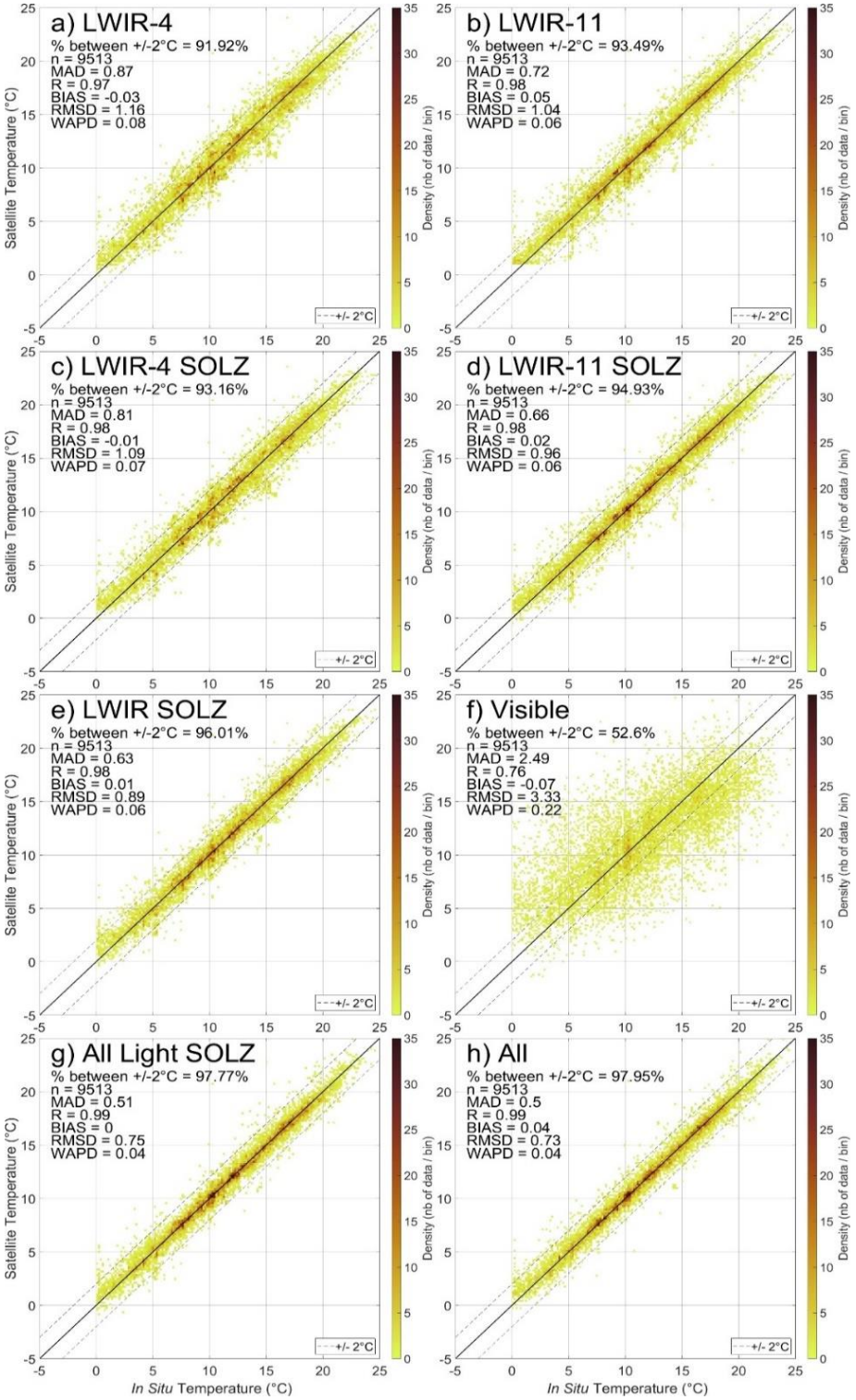

Figure 6.7: Independent comparison between the European *in situ* and satellite temperatures during daylight estimated by a neural network algorithm using, as inputs, a) LWIR-4 bands; b) LWIR-11 bands; c) LWIR-4 and the solar zenith angle; d) LWIR-11 and the solar zenith angle; e) LWIR-4, LWIR-11 and the solar

zenith angle; f) the visible and infrared bands from 412 to 2130 nm; g) all the light available for the dataset (412 to 12000 nm) and h) all light and salinity.

6.3.2.2. Neural networks developed for the SeaBASS noisy dataset (daylight)

Neural networks using different inputs (Table 1) are conducted for the global SeaBASS noisy data set (Figure 6.8). Performances obtained need to be compared with Figure 6.5f. The previous observation concerning the need for the sensor zenith angle made for the European dataset in the previous paragraph is only partly true when going from a regional to a global dataset. Performances obtained using only LWIR–4 bands (Figure 6.8a) are greatly improved when the sensor zenith angle is included (Figure 6.8c, MAD almost halved), while its inclusion only slightly improves the LWIR–11 algorithm (Figure 6.8d). Interestingly, the SST11 algorithm (Figure 6.5f) and its twin neural network version (Figure 6.8d) return almost the same performances. The physical approach for daylight SST reached optimal performance relying on three main sources of information: LWIR at 11,000 nm, 12,000 nm and the sensor zenith angle, and the NN does not outperform it. However, this is exactly where the machine learning approach can take the lead and highlight its superiority for complex tasks by being able to use more information easily. Using both LWIR-4 and LWIR–11 bands with the sensor zenith angle improves metrics again compared to the SST11 algorithm (Figure 6.8e; MAD of 0.39 against 0.53°C, RMSD of 0.58 against 0.76°C, 99% of the data between the ± 2°C limit versus 98%). Addition of other angles information (solar and sensor, zenith azimuthal angles, Figure 6.8f) only slightly improves the MAD (0.34). Likewise, addition of more light bands (the two Rhot at 678, 748nm and the 3 LWIR bands from 6000 to 8000nm, Figure 6.8g) returns similar MAD, but RMSD moved from 0.39 to 0.71 or 1.1°C. Eventually, inclusion of windspeed (Figure 6.8h) seems to be somehow harmful to the performance consistency. Despite reaching the best MAD with 0.33, the RMSD returned is 1.57 with some outliers estimated above 35°C, not visible from the figure. After careful visual evaluation, the same behaviour happens to some of the NN using Figure 6.8f and 6.8g conditions, which explains their higher RMSD. It seems that inclusion of intermediate LWIR bands, azimuth angles and the windspeed causes issues to the neural network (a few values are estimated up to 50°C which disturbs the RMSD each time, sensitive to outliers) and is therefore harmful. Using the windspeed along with the LWIR bands and the sensor zenith angle (not shown) only leads to consistent MAE of 0.36 and RMSD of 0.53, slightly better than the more robust algorithm shown in panel (e). However, non-light contributors have already been tested in Chapter 3 and lead to better metrics but poorer representability. The situation is mixed for the wind speed. While the wind is independent from the light measurement, it affects the signal retrieved by the network by changing how flat the ocean is, and contributes to mixing

of the surface ocean which directly influences temperature measurements. Because the improvement are not massive when wind speed is included, including it as an input should be avoided in any algorithm due to the complexity of the interactions, despite showing independent test data only.

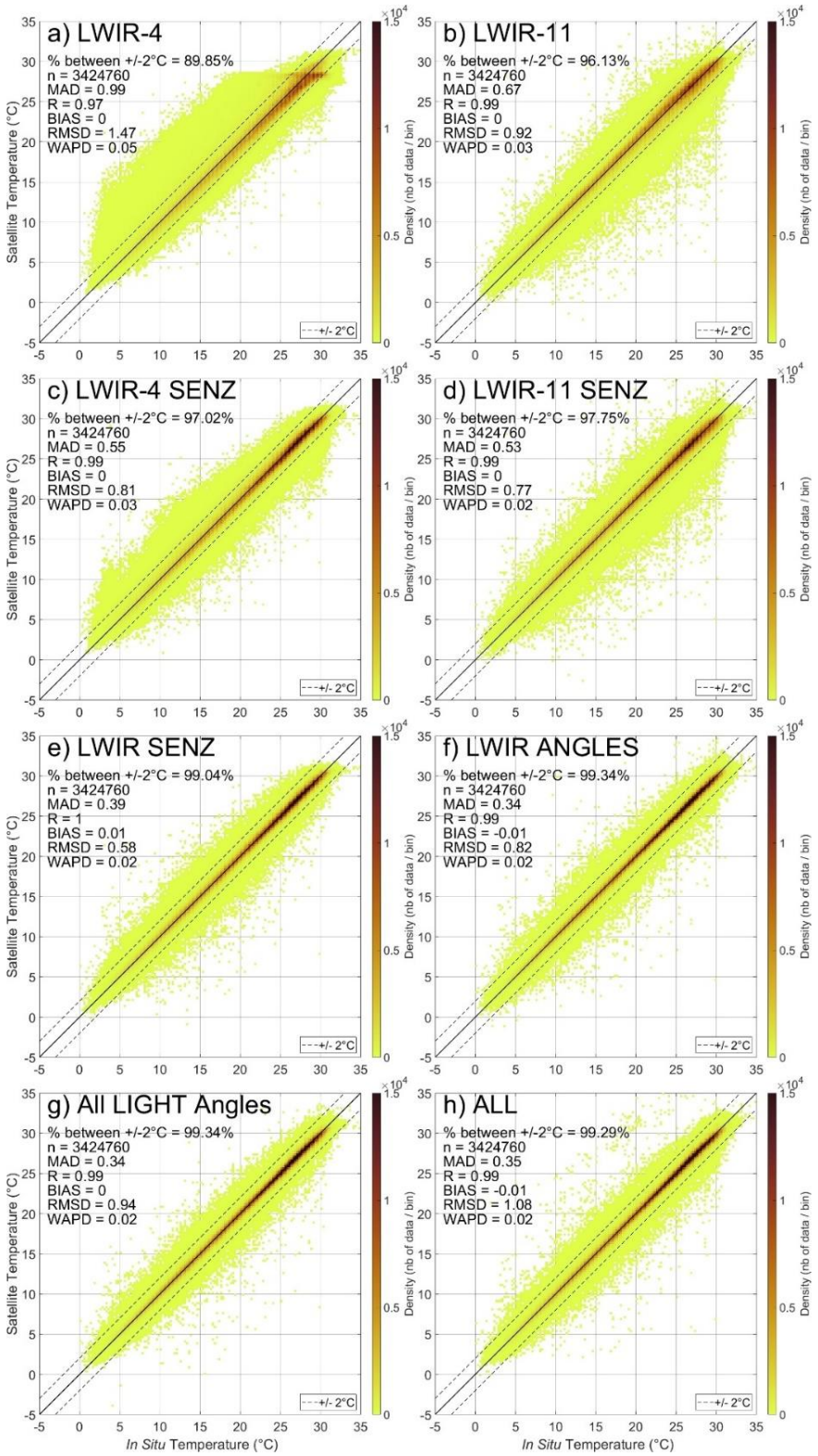

Figure 6.8: Independent comparison between the SeaBASS noisy *in situ* and satellite temperatures during daylight estimated by a neural network algorithm using, as inputs, a) LWIR-4 bands, b) LWIR-11 bands, c) LWIR-4 and the sensor zenith angle, d) LWIR-11 and the sensor zenith angle, e) LWIR-4 and LWIR-11 and

the sensor zenith angle, f) LWIR-4 and LWIR-11 bands all angles (solar and sensor, azimuth and zenith), g) all light and angle information available for the dataset (678 to 12000 nm) and h) all information available (light, angle, windspeed).

6.3.2.3. Neural networks developed for the SeaBASS clean dataset (daylight)

The same algorithms have been developed using the improved quality SeaBASS clean dataset (with no flags allowed) shown below in Figure 6.9. This figure can be compared with Figure 6.5h. For physics based algorithms, MAD of 0.38–0.4°C and RMSD of 0.58°C were observed for both SST4 and SST11 algorithms, with the former showing a positive bias (+0.3°C). With no glint heavily contaminating the signal for the clean dataset, networks designed with LWIR–4 bands (Figure 6.9a and c) show better performances compared to their LWIR–11 equivalents (Figure 6.9b and d). This is due to the higher sensitivity of the LWIR–4 bands to the surface temperature compared to the LWIR–11 bands. As observed previously in Figure 6.8, the LWIR-11 and SENZ algorithm (Figure 6.8d) returns the same performances as the physical SST11 algorithm (Figure 6.5h), meaning that all information has probably been fully used by both the empirical and neural network algorithms. However, the LWIR–4 NN version outperformed the SST11 algorithm which should be due to the absence (or highly reduced presence) of glint that is more harmful to LWIR-4 rather than LWIR-11 bands. Inclusion of sensor zenith angle only slightly improved performances for this dataset (less than 5%) and is therefore less impactful when noise is absent from the pixels. The algorithm designed using both LWIR band pairs and the sensor zenith angle (Figure 6.9e) is again showing better performance with MAD down to 0.28°C and RMSD of 0.46°C. Inclusion of more information (Figure 6.9f to 6.9g) similarly to what has been observed in Figure 6.8, returns slightly lower MAD (0.27 or 0.26) but can produce outliers for a few estimations and return higher RMSD and is therefore not recommended. MAD is even increasing back to 0.27 when all available information are used, highlighting harmful inclusion of the windspeed as an additional inputs (6.9h).

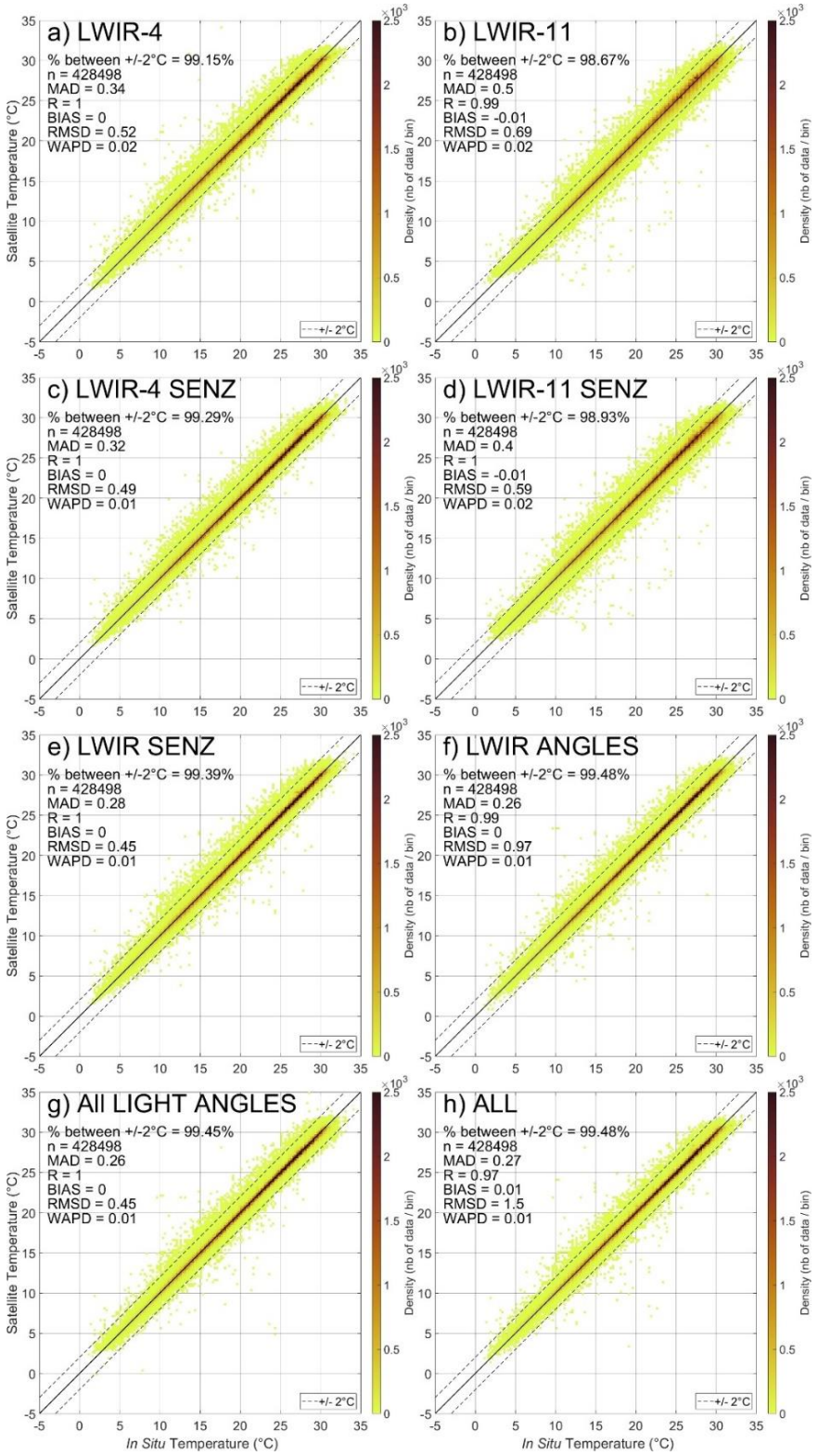

Figure 6.9: Independent comparison between the SeaBASS clean *in situ* and satellite temperatures during daylight estimated by a neural network algorithm using, as inputs, a) LWIR-4 bands; b) LWIR-11 bands; c) LWIR-4 and the sensor zenith angle; d) LWIR-11 and the sensor zenith angle; e) LWIR-4, LWIR-11 and the

sensor zenith angle, f) LWIR-4, LWIR-11 and all angles (solar and sensor, azimuth and zenith); g) all light and angle information available for the dataset (678 to 12000 nm) and h) all information available (light, angle, windspeed).

6.3.2.4. Neural networks developed for the SeaBASS operational dataset (day and night)

The last dataset available is the operational dataset officially used by NASA. Neural networks developed with this dataset are shown in Figure 6.10 for daylight and can be compared to Figure 6.5b with the SST11 algorithm. The networks using the intermediate LWIR bands (panels f and g from Figures 6.8 and 6.9) were not processed as they produced poorer performances and outliers that would require further quality control. Six input combinations are tested in Figure 6.10 below. Again, similarly to what has already been discussed in previous subsection for the SeaBASS dataset, neural networks return similar performance to the operational algorithm in Figure 6.5b when they use the same amount of inputs. The main difference is the negative bias present in the empirical method (-0.24°C), while all neural networks tend to have almost no bias (or 0.01°C). The algorithms using all four LWIR bands and the sensor zenith angle provide better performances. Inclusion of wind speed as an input again hardly provide improved statistical metrics, and are not mandatory. A pure light product as shown in Figure 6.10e is the best option.

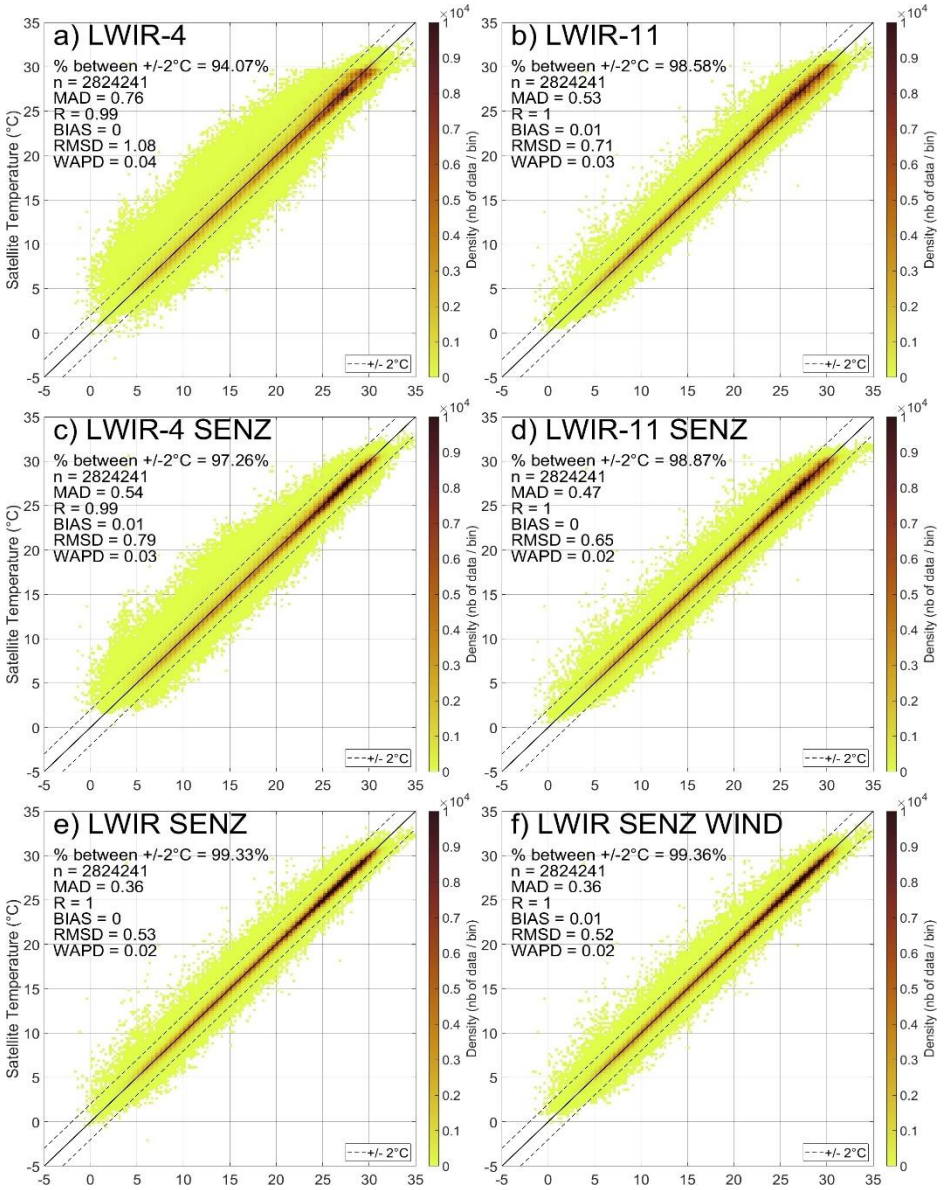

Figure 6.10: Independent comparison between the SeaBASS operational *in situ* and satellite temperatures during daylight estimated by a neural network algorithm using, as inputs, a) LWIR-4 bands; b) LWIR-11 bands; c) LWIR-4 and the sensor zenith angle; d) LWIR-11 and the sensor zenith angle; e) LWIR-4, LWIR-11 and the sensor zenith angle and f) LWIR-4, LWIR-11, the sensor zenith angle and the wind speed.

The process has been repeated for the night-time data in Figure 6.11 below and can be compared to Figure 6.5a with the SST4 algorithm. LWIR–4 bands combination provide the best performances as expected at night with their higher sensitivity, slightly better than the LWIR–11 bands. Except the LWIR–11 algorithm only, any of the developed neural network algorithm outperform the official SST4 algorithm at night which returns a bias (-0.4°C) compared to almost none from the neural networks. Wind inclusion

does not provide better performance at night, and I am unsure if there is a difference on the quality of wind retrieval compared to during daylight. The LWIR and sensor zenith angle (Figure 6.11e) would again be the recommended algorithm.

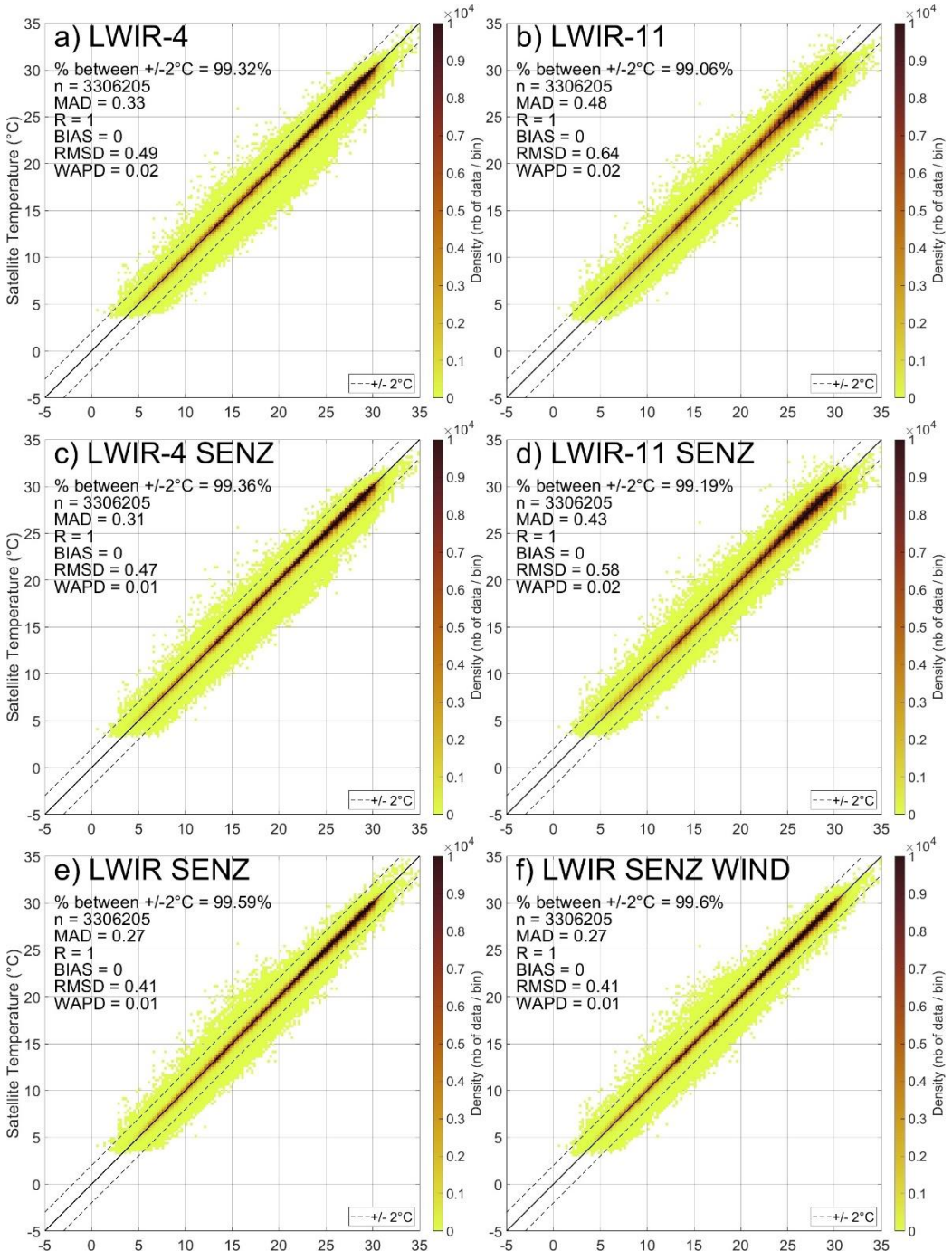

Figure 6.11: Independent comparison between the SeaBASS operational *in situ* and satellite temperatures during the night estimated by a neural network algorithm using, as inputs, a) LWIR-4 bands; b) LWIR-11

bands; c) LWIR-4 and the sensor zenith angle; d) LWIR-11 and the sensor zenith angle; e) LWIR-4, LWIR-11 and the sensor zenith angle and f) LWIR-4, LWIR-11, the sensor zenith angle and the wind speed.

6.3.3. A unique algorithm developed for both day and night

One again, the same input combinations have been tested for an operational dataset that was merged for the exercise. Both day and night operational datasets were assembled. The whole dataset contains 3.3 million samples (53.4%) from the night and 2.8 million samples (47.6%) from the day, with some samples in between with a very low sun angle when the sun is rising or setting. The neural network training steps were applied to the full dataset (normalisation, test/train subsets, etc.). The input selection shown in panels (e) from figures 6.8 to 6.11 (LWIR–4, LWIR–11 and the sensor zenith angle) provided the best trade-off between a robust algorithm (no outlier creation) and good performances. Therefore, this combination was tested to create a unique algorithm that could be applied during both day and night. Figure 6.12 shows the performance achieved by this algorithm over the remaining 6 million test data. A MAD of 0.37°C, no BIAS and a RMSD of 0.53°C are returned for the more than 6 million points, with 99.3% of the data between the ±2°C limit. These performances are moderately poorer than the individual day or night versions (MAD of 0.31°C at night, 0.27°C during the day). However, this algorithm has the advantage of being atemporal and can be applied anywhere on Earth at any time of the day, and is therefore a global algorithm. The other input combinations (not shown) provided poorer results. Inclusion of the windspeed did not produce a different figure than Figure 6.12 below, therefore wind speed should not be included as an input as it may contain a potential day/night bias.

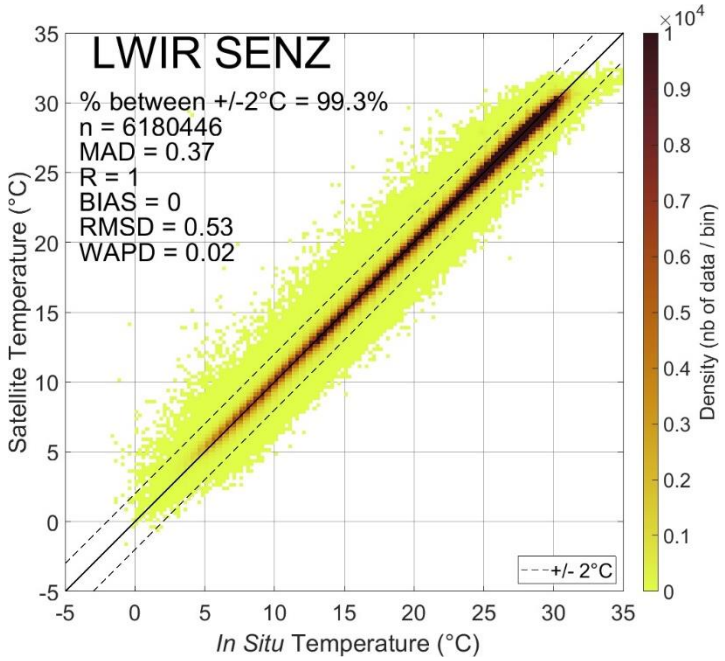

Figure 6.12: Independent comparison between the SeaBASS daytime and night-time noisy merged dataset *in situ* temperature and satellite temperature measurement for an algorithm using LWIR-4, LWIR-11 and the sensor zenith angle.

6.3.4. Temporal drift of the signal

In this section, a single test is conducted to evaluate impact of data availability for long periods over the entire Earth. It has been mentioned earlier that current operational SST algorithms are available on a month-of-year basis. There is no particular need to develop such algorithms from a NN perspective as they already greatly outperform standard algorithms. In fact, it would be harmful by dividing the amount of available data by 12, especially during the initial development of a dataset when data availability is scarce. However, one more question remains: *when having access to a worldwide matchup datasets for almost 20 years, is there a negative impact of applying the same algorithm for too long? Do algorithms need to be year-specific, or month-specific as is currently the case?* The 20 years old SeaBASS matchup dataset was not available 20 years ago, and exists only due to the increased efforts to deploy drifting buoys in the ocean during the 20th century. However, after a year and half of MODIS Aqua lifetime in December 2003, the SeaBASS matchup dataset in operational conditions (when the quality flag is set to 0) contained approximately 200,000 matchups, 100,000 during the day and night each. For the next evaluation shown in Figure 6.13, a network was trained with 50,000 samples from the day and night operational dataset of the 2002-2003 period and applied to the available dataset coming from 2004 onwards. It used the LWIR–4, LWIR–11 and sensor zenith angle as inputs (similar to Figure 6.12). The

network shown in Figure 6.13 below. Except a small bias (-0.03°C), it produces the exact same performances as the atemporal algorithm developed in Figure 6.12. Therefore, there is no particular temporal bias induced by the network (in terms of years and sensor drift). In this case the neural network algorithm developed using only a specific period of time, is robust over time and can be used for decades, presumably because the quality of the satellite signals has been maintained by the vicarious calibration team. This algorithm could have been developed in January 2004, and would provide the same level of information since 2004.

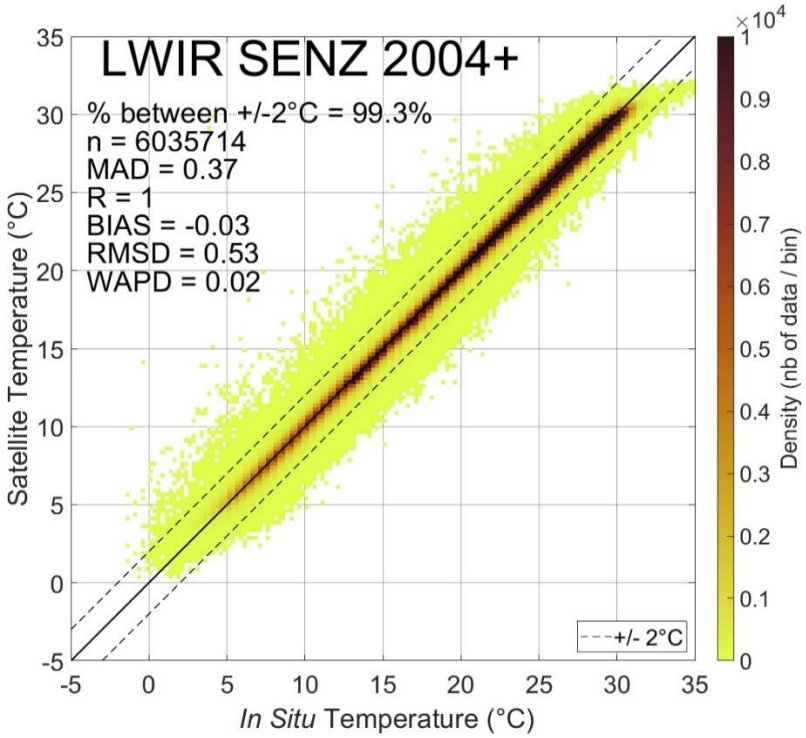

Figure 6.13: Independent comparison between the SeaBASS daytime and night-time operational merged dataset *in situ* and satellite temperature estimated by a neural network algorithm using LWIR–4, LWIR–11 and the sensor zenith angle as inputs. The training data were limited to the 2002-2003 period, and the data shown represent the 2004+ period.

6.3.5. Multitask learning and salinity estimates for the European dataset

In this section, two tests are conducted using the daylight European dataset. The first test tries to make an estimate of salinity using visible and infrared bands following the same methods outlined previously in this Chapter. The second test is another attempt at performing a multi-task learning evaluation using real data this time, to estimate temperature. Multi-task learning was previously tested on a modelled data in Chapter 5 (Section 5.3.3). Salinity and temperature are independent from each other, both affect the density of the water and can be sampled *in situ* using a CTD measurement. Both can be estimated by remote sensing techniques with relatively good performances using microwave bands. While temperature uses either infrared or microwave bands, salinity relies only on L-bands (1.4 GHz, approximately 20cm). MODIS Aqua is not equipped with radiometers sensitive to microwaves. The L band resolution ranges from 25 to 50 km, hence thin mesoscale surface features linked to salinity cannot be seen except from high temporal *in situ* samples. Therefore, having access to a salinity and temperature matchup dataset is an opportunity to evaluate if 1) salinity can be retrieved from visible and infrared bands and 2) if temperature retrieval can be improved when salinity is available. Medina-Lopez (2020) have shown the possibility to retrieve salinity from visible bands for the Sentinel-2 sensor, with potential links with blue bands in coastal waters related to colour dissolved organic matter (CDOM). High concentrations of CDOM lead to higher absorption of blue light (Bricaud et al., 1981) mostly linked with freshwater inputs from rivers, therefore associated to a reduced salinity. Figure 6.14a shows the salinity estimates using all 19 bands available from the European dataset and the temperature as an input. Figure 6.14b shows a neural network trained to estimate both temperature and salinity at the same time. Only the temperature estimation using this technique is shown in Figure 6.14b, because the salinity performances are worse than Figure 6.14a. Salinity was estimated using the 19 bands (Visible and LWIR, Table 1), the solar zenith angle and the temperature as an input. The temperature from the multi-task learning approach shown in 6.14b was trained using the 19 bands and the solar zenith angle.

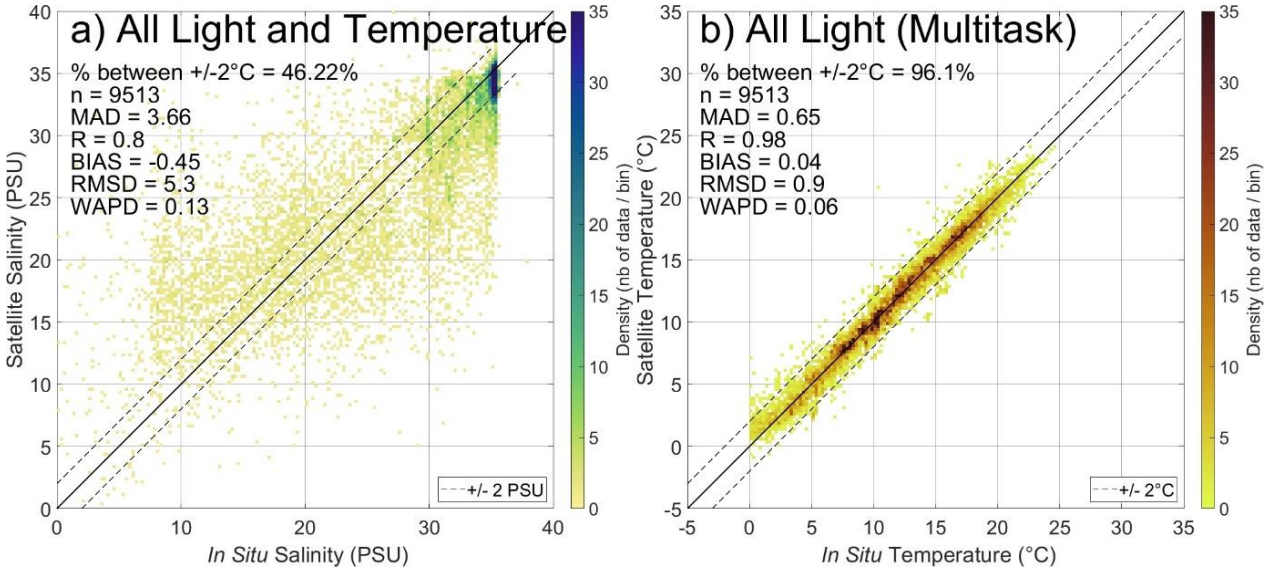

Figure 6.14: Estimation of a) salinity and b) temperature using multitask learning

Figures 6.14a shows that salinity is poorly estimated from visible and infrared bands. The density of the dataset is clearly biased towards open waters with salinity of 35 PSU in the European region and the network can only find a weak relationship between light and salinity. Smaller values cannot be expected to be estimated well, the networks tend to overestimate salinity at lower values, probably due to the reduced amount of data. The test was just a brief exercise to try to retrieve salinity using the basic concepts applied to either temperature or chlorophyll. The requirement for microwave bands seems necessary for accurate salinity estimations compared to visible and infrared bands only. Estimating salinity using multitask learning method (not shown) returns the same results, hence cannot be considered as "working".

For temperature (Figure 6.14b), poorer performances are obtained when a multi-task approach is used compared to expected performances using the exact same inputs (Figure 6.7g). Multitask learning did not work with modelled data, and here does not work with sampled data. We know from the test conducted earlier in Figure 6.7h that inclusion of salinity along with the 19 wavebands and the solar zenith angle improved the neural network performance by a small amount (2%). Here, estimating these two variables at the same time decreased performance expected (MAD of 0.51°C for Figure 6.7g compared to 0.65°C here). However, this failure can be expected because the NNs do not show capacities to make accurate estimations of salinity using visible and infrared. Including salinity as an additional input did slightly help the network previously (MAD of 0.5). Multitask learning is probably dependent on performances obtained for all the tasks required to potentially improve performances, but here

performances for temperature estimations decreased compared to a single task algorithm, therefore applying this method in this way is not recommended.

6.3.6. Map example

Finally, to understand where the overall improvements were made and to evaluate the neural network generalisation capabilities to provide a coherent temperature signal, a sea surface temperature map is shown below in Figure 6.15. The image was captured by MODIS Aqua during the 17th of February, 2003, at 12:20, in the south west North Sea. Figure 6.15 shows the temperature retrieved by three algorithms: a) the NASA SST11 daylight algorithm; b) a LWIR algorithm using the 3959, 4050, 11,000 and 12,000 bands; and c) the Visible-LWIR algorithm (similar to Figure 6.5g, using all 19 bands available from the European dataset, 412 to 12,000 nm without the solar zenith angle). This day was selected because the SST11 algorithm returns underestimates in the European dataset when compared to *in situ* samples (due to the presence of cloud/haze). All three algorithms provide similar general estimates. Lower values are visible near the coast in the East, which highlights the presence of either colder freshwater inputs or an upwelling. Warmer temperatures are found in the central and southern North Sea parts of the image, up to 8°C. Differences can be observed near the clouds (assumed to be the white pixels and main reason of masking): lower temperature values are visible from the SST11 algorithm, while the two neural network algorithms do not seem affected by the cloud presence. This is more visible from the bottom left cloud near 53°N, with nearby pixels consistently estimated at low values. The neural networks appear to be more robust to other contributors due to inclusion of all four LWIR bands and with addition to visible bands for the algorithm shown in Figure 6.15c.

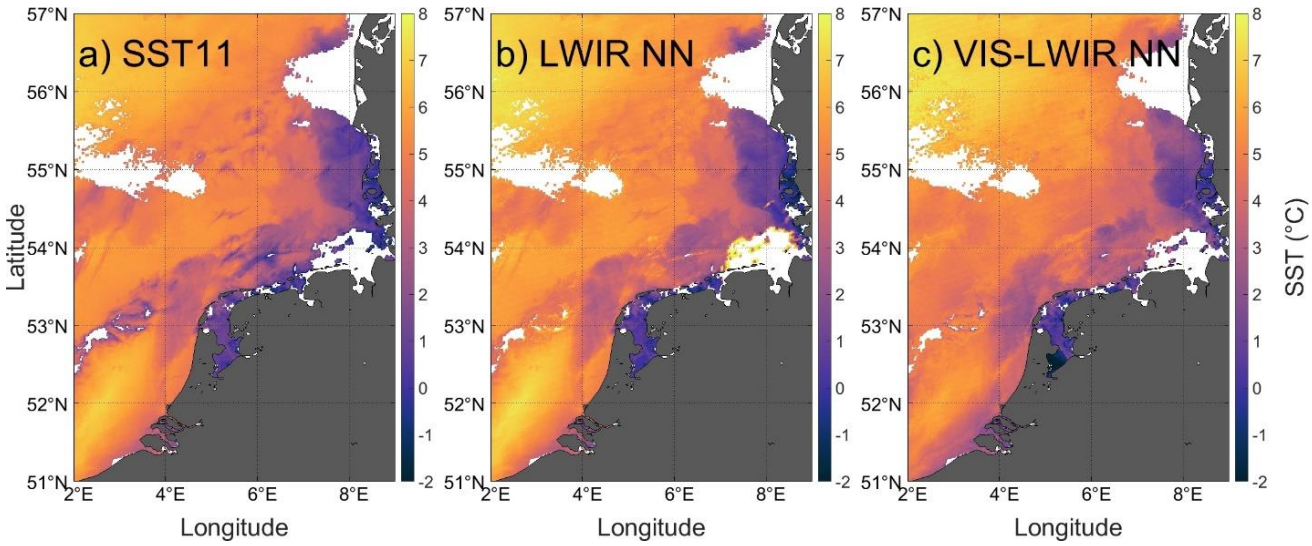

Figure 6.15: Sea surface temperature map example for the 17th of February, 2003.

Table 2 below presents the statistical metrics obtained in different conditions shown in Figure previously (Figure 6.4 to 6.13). For details of exact input combination used, please refer to Table 1.

Table 2: Metrics of all algorithms for different datasets presented. NASA algorithms are used in appropriate conditions.

Dataset and algorithm used	n	Metrics					
		MAD (°C)	R	BIAS (°C)	RMSD (°C)	WAPD (%)	± 2°C limit (%)
SeaBASS Operational (Daylight)							
SST11	2,874,241	0.48	**1**	-0.24	0.66	**0.02**	98.87
LWIR-4	2,824,241	0.76	0.99	**0**	1.08	0.04	94.07
LWIR-11	2,824,241	0.53	**1**	0.01	0.71	0.03	98.58
LWIR–4 SENZ	2,824,241	0.54	0.99	0.01	0.79	0.03	97.26
LWIR–11 SENZ	2,824,241	0.47	**1**	**0**	0.65	**0.02**	98.87
LWIR SENZ	2,824,241	**0.36**	**1**	**0**	0.53	**0.02**	99.33
LWIR SENZ WIND	2,824,241	**0.36**	**1**	0.01	**0.52**	**0.02**	**99.36**
SeaBASS Operational (Night)							
SST4	3,356,205	0.45	**1**	-0.4	0.61	0.02	98.93
LWIR-4	3,306,205	0.33	**1**	**0**	0.49	0.02	99.32
LWIR-11	3,306,205	0.48	**1**	**0**	0.64	0.02	99.06
LWIR–4 SENZ	3,306,205	0.31	**1**	**0**	0.47	**0.01**	99.36
LWIR–11 SENZ	3,306,205	0.43	**1**	**0**	0.58	0.02	99.19
LWIR SENZ	3,306,205	**0.27**	**1**	**0**	**0.41**	**0.01**	99.59
LWIR SENZ WIND	3,306,205	**0.27**	**1**	**0**	**0.41**	**0.01**	**99.6**
SeaBASS Noisy (Daylight)							
SST11	3,474,760	0.53	0.99	-0.25	0.76	**0.02**	97.88
LWIR-4	3,424,760	0.99	0.97	**0**	1.47	0.05	89.85
LWIR-11	3,424,760	0.67	0.99	**0**	0.92	0.03	96.13
LWIR–4 SENZ	3,424,760	0.55	0.99	**0**	0.81	0.03	97.02
LWIR–11 SENZ	3,424,760	0.53	0.99	**0**	0.77	**0.02**	97.75
LWIR SENZ	3,424,760	0.39	**1**	0.01	**0.58**	**0.02**	99.04

LWIR ANGLES	3,424,760	**0.34**	0.99	-0.01	0.82	**0.02**	**99.34**
ALL LIGHT ANGLE	3,424,760	**0.34**	0.99	**0**	0.94	**0.02**	**99.34**
ALL	3,424,760	0.35	0.99	-0.01	1.08	**0.02**	99.29
SeaBASS Clean (Daylight)							
SST11	478,498	0.38	**1**	-0.02	0.58	0.02	98.93
LWIR-4	428,498	0.34	**1**	**0**	0.52	0.02	99.15
LWIR-11	428,498	0.5	0.99	-0.01	0.69	0.02	98.67
LWIR–4 SENZ	428,498	0.32	**1**	**0**	0.48	**0.01**	99.29
LWIR–11 SENZ	428,498	0.4	**1**	-0.01	0.59	0.02	98.93
LWIR SENZ	428,498	0.28	**1**	**0**	**0.45**	**0.01**	99.39
LWIR ANGLES	428,498	**0.26**	0.99	**0**	0.97	**0.01**	**99.48**
ALL LIGHT ANGLE	428,498	**0.26**	**1**	**0**	0.45	**0.01**	99.45
ALL	428,498	0.27	0.97	0.01	1.5	**0.01**	**99.48**
European (Daylight)							
SST11	17,672	0.95	0.95	-0.49	1.57	0.08	88.16
LWIR-4	9,513	0.87	0.97	-0.03	1.12	0.08	91.92
LWIR-11	9,513	0.72	0.98	0.05	1.04	0.06	93.49
LWIR–4 SOLZ	9,513	0.81	0.98	-0.01	1.09	0.07	93.16
LWIR–11 SOLZ	9,513	0.66	0.98	0.02	0.96	0.06	94.93
LWIR SOLZ	9,513	0.63	0.98	0.01	0.89	0.06	96.01
Visible	9,513	2.49	0.76	-0.07	3.33	0.22	52.6
ALL LIGHT SOLZ	9,513	0.51	**0.99**	**0**	0.75	**0.04**	97.77
ALL (Light, SOLZ, Salinity)	9,513	**0.5**	**0.99**	0.04	**0.73**	**0.04**	**97.95**
Operational (Day and Night)							
LWIR SENZ	6,180,446	**0.37**	**1**	**0**	**0.53**	**0.02**	**99.3**
LWIR SENZ (2002–2003)	6,035,714	**0.37**	**1**	-0.03	**0.53**	**0.02**	**99.3**

6.4. Conclusion

This chapter is an attempt to apply the methods developed in previous chapters to a different remote sensing inversion problem. The assessment of the potential to use neural networks and inclusion of additional remote sensing information to improve upon existing algorithm performance is conducted. By using the same technique developed for coastal chlorophyll to produce SST estimates from remote sensing, better performing algorithms than the ones currently in use and distributed by NASA are obtained. Neural networks did not outperform NASA SST11 official algorithm when they only had access to the same information for SeaBASS dataset. The carefully selected empirical method already made the best estimate possible, and a neural network would be less useful for the process if no more information would be available. This is due to the potential overfitting issues it can produce when low amounts of data are available, and a harder readability (the "black box" issue). However, following observations made in Chapter 4, neural networks really benefit by exploiting more of the available information, including data that is not immediately linked through physical models or known to fail under specific conditions. Inclusion of more information that is linked with the target signal improves accuracy of the retrieval. The LWIR bands at 4 and 11 µm combined with the sensor zenith angle returned consistent and high quality estimates of the sea surface temperature. This algorithm includes information used by the two SST algorithms used by the MODIS Aqua SST community, SST4 and SST11. LWIR-4 bands clearly show higher links with the temperature variation compared to LWIR-11 due to their position on the curve (Figure 6.1), where small temperature changes cause greater signal changes. However, while SST4 is sensitive to glint during daylight, inclusion of LWIR-11 bands and LWIR-4 bands together eliminates the effect and the combined algorithm was able to return good performance for any dataset tested here.

Compared to chlorophyll, temperature is generally much easier to estimate, with no particular oceanic area of interest that seems problematic unlike pigment estimation in coastal waters. Temperature measurements are highly sensitive to the atmosphere content, such as clouds and probably more sensitive to haze than chlorophyll due to the higher impact of absorption for the specific LWIR bands used to make the estimation. However, once the two most common flags were applied (Cloud, ice and straylight contamination), estimates reached very good performance (MAD below 0.4°C, more than 99% of the data within a ±2°C limit) with only small fractions of remaining outliers. Including more inputs when designing the neural network helps to deal with noise produced by other contributors to the light signal. While the estimations of these specific noisy points are improved by including more information such as the visible spectrum or other variables (salinity, windspeed), some algorithms designed with non-light signals did show outliers that are not realistic (SST above 32°C) and should not be used in future. Of course, there are still issues that affect

comparison of remote sensing and *in situ* data. For example, the satellite pixel resolution is still 1km² compared to water sampled using a few millilitres. Moreover, the SeaBASS dataset uses a light signal averaged from a 25km² area (5 by 5 pixels around the matchup). The temporal window also alters performance and ranges from ±30 min from SeaBASS dataset to ±3h for the European dataset. Moreover, the data is sampled at a 20 cm depth from the buoys in good conditions. This value depends on how rough the sea can be, and the real provenance is probably within the first meter under high wind conditions, which in turn produces uncertainties.

Estimating surface temperature using only visible and short wave infrared bands shows poor performances compared to LWIR bands (Figure 6.7f, MAD of 2.72°C and 47% of the data within the 2°C limit). However, it still seems possible, and rather than an expected ±2°C limit, an expected window of ±6°C would be more realistic for such a method. Performances obtained in Medina Lopez and Urena Fuentes (2019) with Sentinel-2 were very similar to the one obtained here. While Sentinel-2 has a 10 m resolution compared to 1 km here, it does not contain as many useful bands as MODIS Aqua, which is an ocean colour dedicated radiometer. It is therefore hard to evaluate if a high resolution MODIS Aqua-like sensor could reach better performances at evaluating SST without access to LWIR bands, and the trade-off of a high resolution sensor equipped with LWIR bands should be more appropriate (like Landsat 8 and 9 sensors with LWIR bands with a resolution of 100m). It was not possible to make accurate estimations of salinity using MODIS Aqua and visible infrared bands. Testing multitask learning did not show particular benefit and temperature is retrieved with lower performances compared to single training potentially due to failure of estimating salinity from using visible and infrared bands.

Considering the results presented above, it is clear SST can be accurately retrieved despite noise contaminating the measured signals, and there is an opportunity to switch to neural network algorithms to make more accurate estimations in the near future. Physics based SST algorithms return good quality performance but require monthly tuning. Because of this monthly requirement, it is usually required to wait until the end of the month (or the month after sometimes) to be able to process the previous month's SST. The approach detailed in this Chapter only requires LWIR and sensor zenith angle information and can be applied as soon as an image is available. Inclusion of visible bands improves performances of a permissive remote sensing product (no flags applied), and reproducing the procedure at a global scale could improve general performance of retrievals and should be tested. A robust algorithm could have been developed during the first year of MODIS Aqua lifetime and would still provide the same quality of temperature retrievals nowadays (Figure 6.13.). It is also clear that the method developed in previous chapters, that consists of using more information can be generalised to more inversion problems as long as the supplementary inputs are linked to the

signal itself, and not the target. This SST study benefitted greatly from having access to a larger global datasets, something that is clearly missing for the chlorophyll community due to the absence of good quality automated sensors. Using this global dataset illustrated that neural networks did not show particular regional bias and the full data distribution is close to the 1:1 line. Using most of the available matchups (90 to 98.5%) as the independent test set helps build confidence that as soon as enough data are used for the training, overfitting is not particularly an issue and the exact methods (activation function, normalisation method, architecture, cost function, etc.) only need little attention compared to producing consistent and good quality datasets. Of course, there is still space for fine tuning algorithms and trying to get more accurate estimations. The current operational algorithms were outperformed in terms of pure statistical metrics in noisy and operational conditions thanks to the use of additional spectral information. Inclusion of the wind speed would not be recommended as it did show only slight improvement while it could lead to overfitting issues.

Neural networks did not take the lead for operational products in remote sensing yet. This is due to the difficulty to visualise and understand them, and a 3 layers of 20 neurons network used in this study contains more than 1300 weights, impossible to be meaningful for a human brain. However, despite being computationally huge, the method is actually much simpler to implement than the current operational algorithms, which require spatial information (latitude), temporal information (month of year) and another remote sensing temperature, to be able to make an estimate (the T_{sfc} from Eq. 6.2 and 6.3). The LWIR and sensor zenith angle neural network developed here is pure observational algorithm and does not require all of these steps to return a more accurate temperature signal, and does not require continual updating. Finally, based on observations from Chapter 3 to 6, the decision to create any remote sensing dataset for community algorithm testing and development should try to be as general as possible, include all spectral information available and provide a top of atmosphere version.

Take home message: Chapter 6

The idea of Chapter 6 was to apply a similar approach developed for chlorophyll retrieval earlier to estimates of sea surface temperature. Is the method valid at global scales and be expanded to temperature? Can machine learning challenge the current approach that relies on physics, and improve the quality or quantity of temperature estimates?

- Neural networks are well suited for temperature estimates because of the huge amount of data available, which gives us confidence on performances achieved on large independent datasets.
- Temperature estimates based on physics relationships are already good, but improved performances are achieved by neural networks.
- The ability to use any potentially relevant information as an input gives the opportunity to build efficient algorithms that can perform in currently limited conditions, mostly near clouds or under hazy areas
- The use of the 4 LWIR bands equipped on MODIS Aqua along with the sensor zenith angle returned good and the most consistent results and should be used in future.
- The inclusion of latitude as an input, similarly to what is conducted by physics equations did not improve performances.
- Limiting the development on a monthly basis is not required and similar performances are achieved when using only the first year of data and applying it to the rest of the dataset. No drifts in performances were observed.
- The neural networks allows an instantaneous access to the estimate of the temperature as soon as the LWIR information is available (compared to the current physics approach that require the user to wait the end of the month).
- A single algorithm using only LWIR bands can be developed to make estimates for both night and day at the same time, while current version propose one algorithm for each.
- A single algorithm performs slightly poorer than a double algorithm approach.

The approach developed in the thesis is valid for temperature, and probably for other variables as long as a correlation with a light signal is present. Chapter 7 concerns the eutrophication status of coastal waters, and algorithms developed in Chapter 3 to 5 will be used and compared with and *in situ* approach currently in use.

7. Eutrophication assessment of OSPAR regions: How can remote sensing contribute?

7.1. Introduction

This is the last results Chapter of this thesis. The initial overall goal of this whole project was to develop a chlorophyll a (Chl) algorithm that could be used by the different organisations in charge of assessing water quality to evaluate if some areas are subject to eutrophication. Eutrophication has different definitions (Nixon, 1995). The most common is an increase in algal biomass over a specific period and area. Nutrient inputs in the water increased with the development of commercial agriculture since the middle of the 20th century (van Bennekom and Wetsteijn 1990; Boesch 2002). Eutrophication usually occurs in coastal areas that are influenced by human activities, but it can also be a natural phenomenon due to its high complexity and dependence on other parameters (currents, species, stratification, etc., Beusekom, 2018). In coastal waters such as the Wadden Sea in the southern North Sea region, eutrophication has been linked in the 1970s to the release of anthropogenic nutrients that induced coastal phytoplankton blooms (De Jong, 2006). Actions were taken to reduce the anthropogenic addition to riverine nutrient concentrations, and they have decreased by 2-3% each year since the 1980s (van Beusekom et al., 2009). The average North Sea surface temperature has risen by 1.6°C between 1974 and 2014, and temperature increases coupled to riverine nutrient inputs are the two major drivers of the eutrophication in the region (Desmit et al., 2019). Intense phytoplankton blooms, being unusually highly concentrated and dense, die and fall in the water column. Bacteria consume the dead phytoplankton cells and by doing so turn the organic matter back into inorganic form. During the process, they consume the O_2 in the water column and make nutrients available again for the surviving plants. The O_2 consumption by bacteria is the main threat for organisms using it to breathe, which include most organisms except plants. Physical interactions (stratification in summer) also contribute to how fast a water column can recover oxygen after a depletion (Grennwood et al., 2010). The impacts of eutrophication may be wide-ranging, from loss of seagrass (Reise et al., 2008) to death of upper trophic level organisms trapped in oxygen limited waters (Joyce et al., 2000).

The OSLO-PARIS Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR) is a legislative treaty regulating international cooperation between 35 European countries (https://www.ospar.org/, accessed on 11 July, 2022), or "contracting parties", on environmental protection in the North-East Atlantic and is responsible for the assessment of the state of eutrophication in the area. Because of scientific and technological developments, the actual

eutrophication assessment process has evolved over the different assessment and reporting cycles and is currently being finalised for the next assessment and reporting period, which will culminate in the Quality Status Report (QSR) 2023. While the original definition of eutrophication was purely Chl based, the current Common Procedure takes into consideration other mandatory indicators such as O_2 deficiency and the different nutrient concentrations that may highlight presence of eutrophication indirectly. Other indicators can also be reported on a "voluntary" basis because they can contribute to an enhanced understanding of the problem, such as Secchi depths to evaluate the transparency of the water, presence of Harmful Algal Blooms (HABs) or presence of some specific macroalgae and benthic species. Because of the complexity of water management impacted by multiple causes, the Common Procedure eutrophication definition is not as trivial as observing a change in Chl concentration only. An assessment tool was developed to process the eutrophication status, called COMPEAT for Common Procedure Eutrophication Assessment Tool. It uses R studio for the processing and is freely available at https://github.com/ices-tools-prod/COMPEAT (accessed on 11 July 2022). It requires a formatted dataset that includes the above-mentioned indicators and produces an eutrophication map for each region. Understanding about how the different indicators interconnect is still evolving and new research continues to contribute to improving our understanding of the phenomenon. More or different indicators may be used in future. Direct measurements of most indicators is the main methodology in use at the moment. For nutrients, models can also be used to map them to help our understanding. The indicators are compared to a reference level that represents the state of the environment prior to being impacted by humans, similar to the temperature at a pre-industrial level for climate change.

Due to its nature, different aspects of the assessment are still open to discussion. A good example is the role of Chl in the problem, which highlights the presence of eutrophication. Some countries proposed to report eutrophication using the mean of all observations made during the growing season over each area. Others proposed to use the 90^{th} percentile to report the same observation, because what matters are the high values that can be harmful (if a bloom is too high or lasts for too long). This was to tackle the issue of very high values observed only once when using the maximum Chl observed value to represent a region. The selection of one method over another is not trivial, especially in an ocean influenced by warming and a CO_2 enriched atmosphere. All three values can be reported during the assessment and which one is used should be specified, with the P90 potentially being more appropriate as it is less sensitive to unrepresentative outliers. The reason for such discussion comes from the high variability of periodic and episodic Chl concentrations changes in the water. Sampling the ocean during a massive bloom may return a few occurrences of potentially

very high concentrations over a short period but there is potential for these to be anomalies against a background of generally low Chl concentrations.

For Chl, different observation methods exist. The original method consists of a direct sample of a water body, which requires huge efforts to cover Northwestern European waters that represent 1% of the Earth's surface. The different methods used to retrieve Chl are described in Chapter 2 section 2.1. The main limiting factor of such an approach is the sparse data availability, especially for offshore regions that are sampled only a few times per year which can barely represent the variations. Small numbers of samples will usually be insufficient to reveal trends over a short period of several years. Recently, dedicated ocean colour satellite sensors have been developed and are available to the public. This sampling method is an indirect observation of Chl based on a direct observation of the light signal. Satellites allow access to Chl on a daily and synoptic basis or more for geostationary satellites (Ryu et al., 2012). OSPAR selected the JMP-EUNOSAT algorithm to be applied over the region to provide Chl satellite concentrations. The JMP-EUNOSAT approach merges the OC5-PML algorithm (evaluated in Chapter 3) with the Gons algorithm (Gons et al., 2002) available for European satellites with an extra waveband available in the red (near 700 nm). The reason of this inclusion comes from the poorer estimates provided by the OC5 algorithm in general in coastal conditions, where the Gons algorithm is substantially better (Van Der Zander et al., 2019). The main idea of the JMP-EUNOSAT method consists of classification based on thresholds of different bands for each pixel to decide when to apply the different algorithms and when to mask data when no good estimates can be proposed. The main flaw of this approach is the requirement to only use European satellites. If the pixel was not observed by such a satellite, the approach cannot be applied. Two sensors can apply the Gons algorithm, MERIS and OLCI. European satellite Envisat-2 carried MERIS from spring 2002 to 2012, and Sentinel-3 satellites now carry OLCI sensors since 2016. This approach cannot be applied before 2002, or between 2012 and 2016 and should be available until 2036, when the Sentinel-3 mission will end.

It has been shown in previous chapters that remote sensing techniques can reach relatively good agreements to the measured Chl concentration in very coastal waters using the NN-Rhot algorithms, on a matchup basis. Except in winter when the algorithm overestimates the Chl concentration and cannot reach very low values (<0.2 $mg.m^{-3}$), the remote estimation is generally accurate and comparable to what is measured *in situ*. Moreover, the NN-Rhot algorithm removed the systematic bias of overestimating Chl in coastal waters, especially in the Baltic part of the area, where CDOM presence negatively impacts performances of the OC5-PML algorithm (Pittarch et al., 2016). The actual eutrophication assessment relies on *in situ* sampling efforts maintained for different sites or collected during cruise campaigns. The future Quality Status Report (QSR 2023) made by OSPAR will use both *in situ* and remote sensing Chl samples for the different regions. For a specific region, an

average (or P90 or maximum value) of any *in situ* data sampled for a specific period is created. The same procedure is applied to the remote sensing Chl. Both Chl values are then averaged, each method contributing to 50% of the final Chl signal. The full satellite dataset from Chapter 3 is not available at the moment of writing (still being processed), therefore a full eutrophication assessment cannot be conducted following the QSR 2023 conditions. Moreover, I do not have an immediate access to the other indicators used by the Common Procedure, but there is a possibility to partially conduct the assessment using Chl only. Rather than trying to reproduce what a dedicated team is already doing, the idea of this Chapter is to show the impact of enhanced data availability from the remote sensing products developed in Chapter 3 and their advantage over an approach that is limited to using only *in situ* data. The aim here is to show the limits of the *in situ* only method and benefits from having access to more data thanks to remote sensing. An attempt at showing an eutrophication trend is conducted in Section 7.3.4 and the main observations discussed in the conclusion.

7.2. Materials and methods

Three datasets are available for this chapter, two related to *in situ* sampling and one to remote sensing. The first *in situ* related dataset contains the matchups obtained from Chapter 3. It consists of collocated data between MODIS Aqua and *in situ* Chl a from July 2002 to January 2020, 08:00 and 16:00, 0 and 10 meters, and is described in Section 3.2. It also contains matchups for the OC5-PML and OC5-ACRI algorithms. The second dataset consists of any *in situ* Chl samples gathered from different agencies using the same temporal and geospatial conditions. These *in situ* data are the data that were used to create the matchup dataset in Chapter 3, with the difference that spatial averages are different, because the matchups were created based on a specific projection using 1 km² from MODIS. Here, we want to compare any *in situ* data from a region, which comes with different geospatial limits. The third dataset is a direct application of the remote sensing algorithm developed in Chapter 3 to MODIS Aqua data, and is applied to a single month, May 2003, to evaluate the difference with a remote sensing approach for eutrophication assessment.

The original *in situ* data were provided by different institutions, such as ICES, CMEMS, Marine Scotland Science or a number of institutes that collected Chl during the past 20 years and made them available. Because of the nature of the data being usually free to access, there may be duplicate data among these datasets. This is especially true for the ICES and CMEMS datasets. These two institutions do not collect data themselves, but act as repositories with different quality control requirements, with ICES requiring stricter conditions to publish the data while CMEMS approach is more permissive (CMEMS clearly states that no particular quality control is applied to their data). This means that multiple instances of the same data can occur (e.g. the original provider, CMEMS and ICES). The

detection of replicate data has a number of challenges resulting from the way the data and metadata are stored by different organisations, e.g. different precision or attributes stored. The problem emerged during the matchup creation but was not important because all data are matched within a 1x1 km² pixel. A ± 10 meters precision error for the same sample duplicated due to a rounding/truncating alteration of the coordinates by one institution is unlikely to result in a matchup with a different satellite pixel to create a falsely duplicated matchup.

In this chapter, I try to evaluate how remote sensing can benefit eutrophication assessments, and the main comparison comes from the number of available observations for a region. It is not possible to perfectly clean *in situ* data prior to the application of such a method due to their nature, and a small overestimation of the amount of *in situ* sampling can arise. After visualising the data, it appears that the CMEMS dataset is more inclined to such behaviour, due to less stringent data control criteria. To avoid such issues, the CMEMS dataset was removed from the available *in situ* dataset used for comparison. Removing duplicates between ICES and the original data providers returns sensibly the right number of unique samples, yet unique observations from CMEMS are not available.

The matchup dataset used in Chapter 3 includes the CMEMS data, therefore it contains more data. Moreover, there is a potential 1 km difference between the position of an *in situ* sample and a pixel by definition. It means that an *in situ* sample can be associated to a specific OSPAR region due to its precise coordinates, while the matchup created using this *in situ* sample can be associated to a neighbouring region, because satellite coordinates are reported for matchup creation. Both methods cannot be strictly compared due to this difference.

For the remote sensing product, the algorithm details are specified in Chapter 3. Because processing 20 years of satellite data is computationally expensive, the NN-Rhot Chl daily product covering the full period is not available at the moment of writing and would require at least two additional months of processing (almost fully processed by the submission date). To produce consistent quality data to compare with other remote sensing providers, the ensemble approach (Chapter 3) based on 10 neural networks was used to create a median value of each image available, each day of May 2003. The different images were then merged as daily products and a single median value is available for each pixel at the end. To avoid outliers, estimations below 0.1 $mg.m^{-3}$ were set to 0.1 $mg.m^{-3}$and estimations above 150 $mg.m^{-3}$ were removed. The relative standard deviation (rSTD) provided by the 10 networks was used to filter out estimations when rSTD was above 20%. While some areas (especially coccolithophore-impacted) are cleaned, this approach also helps to deal with noise induced near clouds, mostly in the Atlantic where data availability is the most limited due to their presence.

7.3. Results

7.3.1. *In situ* sampling efforts since 2002

The different eutrophication assessment regions used by the OSPAR commission were provided by Marine Scotland Science as a shape file and are shown in Figure 7.1. *In situ* data and OSPAR regions are shown together on Figure 7.1a, while Figure 7.1b shows each region with a different colour for ease of visibility. These regions do not include very coastal waters (the first couple of kilometres from the coast, depending on the position), which are instead covered by the Water Framework Directive (WFD). Most *in situ* data available are sampled in very coastal waters, and, in fact, among the 65,000 unique stations of available *in situ* data, only 13,002 (~20%) were sampled inside an OSPAR eutrophication QSR reporting regions. This is a limiting factor, especially since the performances obtained for matchups using the method developed in previous chapters clearly targets improved performance for these very coastal areas. In fact, due to the OSPAR area not including nearshore coastal waters and the absence of the Baltic area, the OC5-PML algorithm would not suffer too much from its extensive flagging procedure. OC5-PML algorithm tends to remove nearshore coastal data, when shallow waters or high loads of sediments/CDOM are present. This is due to the OC5 algorithm itself that automatically hides the estimates when it cannot achieve good estimates, or as a sanity check with extra flagging used by PML to provide high quality data. The NN-Rhot algorithm developed in Chapter 3 did not show any overestimates close to the coast, and in fact, it has been shown that it tends to underestimate such high values. Therefore, having access to a better nearshore coastal algorithm is less impactful when using such defined areas, and an extension to the coastline could be applied.

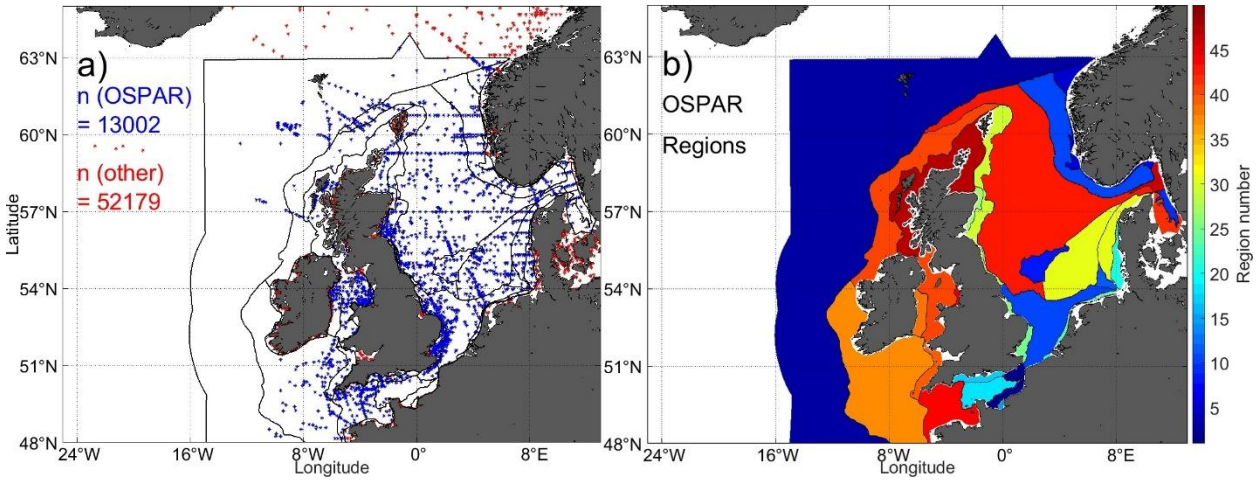

Figure 7.1: a) Map of the different OSPAR regions displayed as black lines, with *in situ* samples displayed as blue stars if they fall inside an area, or red if they do not. b) Map of the different OSPAR Regions shown with different colours. The Baltic, northern Norway and most nearshore coastal data are not included in the OSPAR QSR eutrophication reporting regions.

The Baltic Sea is not included in the OSPAR area because it is part of the HELCOM assessment (Helsinki Commission), similar to OSPAR, an intergovernmental organisation to assess the quality status of the Baltic. As mentioned earlier, these areas are designed based on hydrological and ecological similarities (stratification, salinity and depth). In total, sixty-seven areas are available for European waters down to the Portuguese coast (not shown). Among the 67 areas available initially, 36 fall in the 48°N-65°N limit shown on the map. The total number of *in situ* data available for each region is shown in Figure 7.2. The figure highlights huge discrepancies between the different regions, with 2000 samples available for the southern North Sea region compared to less than 200 for the Atlantic Seasonally Stratified (ASA, in the Celtic Sea, below Ireland), twice its size. Data availability is very unevenly distributed. The different countries may target different areas, and while Ireland provided data to ICES in their coastal waters, almost no data were available from the Atlantic offshore regions surrounding it. While data for these regions may exist, their accessibility is a limiting factor. Many areas in black were sampled less than 200 times over the 20 years (averaging <10 samples per year), and even the most heavily sampled area, the Southern North Sea region averages <100 samples per year. The areal size of each region also needs to be taken into consideration. One might expect fewer samples in smaller regions. However, it is also true that some of the most under-sampled regions are also among the largest. This leaves an important question of to what extent are such a small number of *in situ* samples genuinely representative of the natural variability of such large areas, and to what extent is seasonal variability captured?

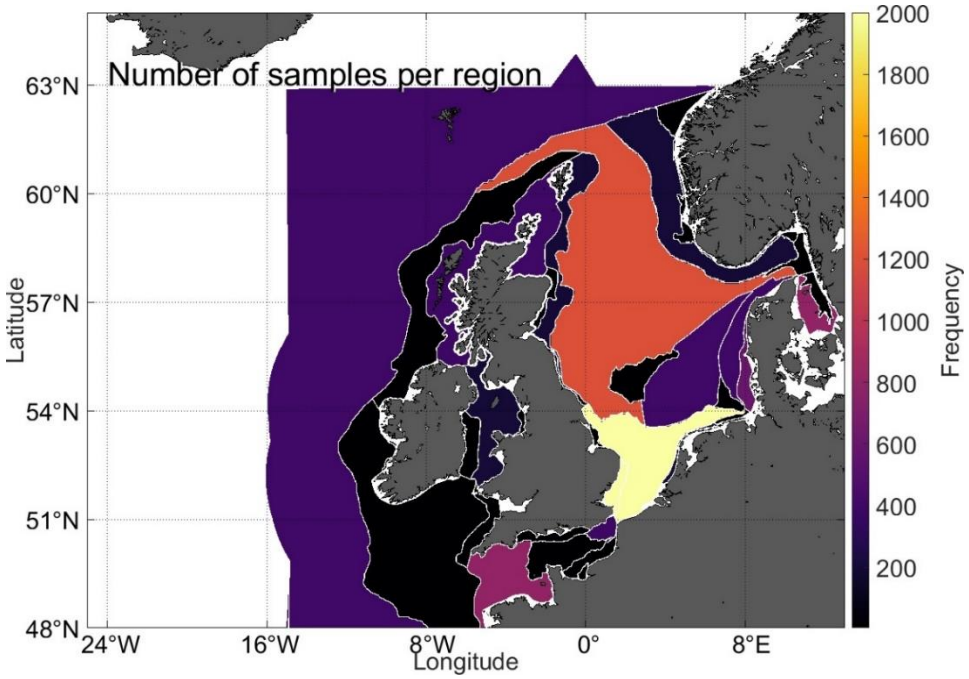

Figure 7.2: Number of times *in situ* samples were collected at the same location and date between 2002 and 2020, 0 and 10 meters.

While more data exist historically, for this exercise, *in situ* sampling is limited to the remote sensing era of MODIS Aqua for comparison (since July 2002), and by the 36 different regions available from the OSPAR commission for the area of interest used in this thesis. In practice, eutrophication is reported using data sampled between March and September (what is called "the growing season"). Hence, only the Chl data available for any of the 36 areas between March and September are considered. For Figure 7.3, all 13 002 surface samples (0-10m averaged) that were collected during the time period of the MODIS time series (July 2002 to Sept 2020); for the months of March to September; are binned in to occurrences per area and per month and the resulting histogram is shown. For example, all samples collected in July 2002 for the Atlantic region (ATL, the biggest OSPAR region visible on Figure 7.1a including the Atlantic, shown in dark blue) represent a unique amount of data collected (none for this example), which becomes a value in the histogram, therefore, zero. This is a good exercise to understand how often each area is sampled per month. A mean of 1.47 *in situ* samples per month, year and area is obtained. Approximately 30% of all areas are sampled at least once for any given month and year (during the growing season), which means that 70% of the time, no data for a specific area, month and year are available. When similar assessments are conducted using remote sensing, a monthly approach is usually applied compared to the growing season for *in situ* samples. The observation of a significant shortage of monthly *in situ* data supports accepting the definition of the growing season from March to September as more data are collected during this period. Figure 7.4 below represents number of *in situ* data sampled for each region, but during the full growing season rather than monthly as shown in Figure 7.3. The number of available data increases above 50%, which validates the use of a growing season definition rather than a monthly approach. Both figures show the same tendency, with a good proportion of regions that are never sampled during a month or the full growing season.

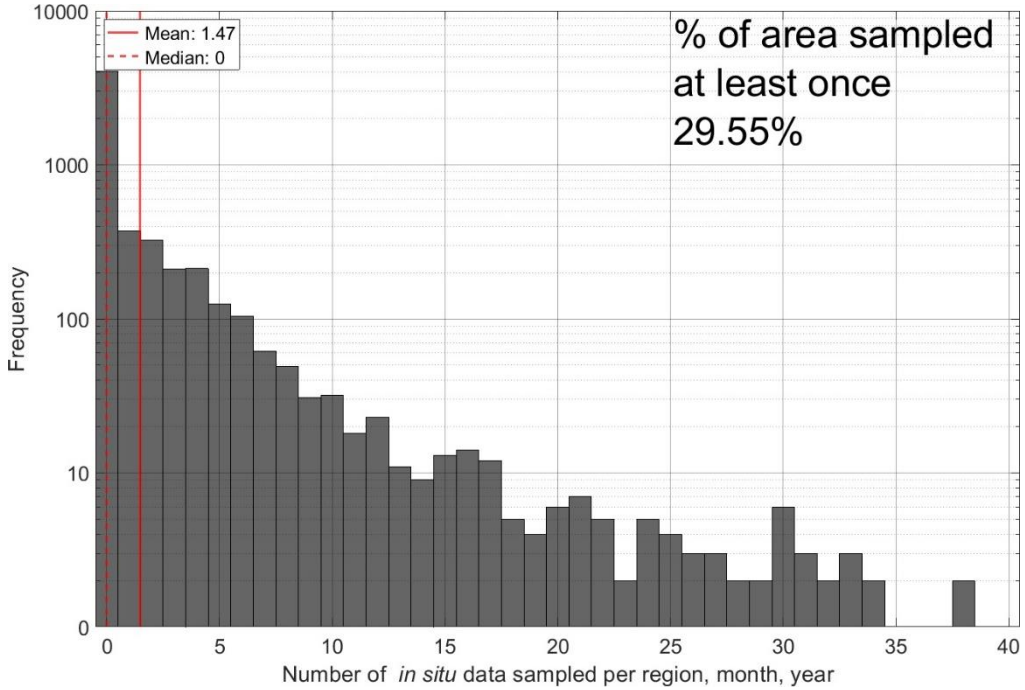

Figure 7.3: Histogram of all *in situ* Chl a samples for each month of each year and each OSPAR region. Note the log y scale.

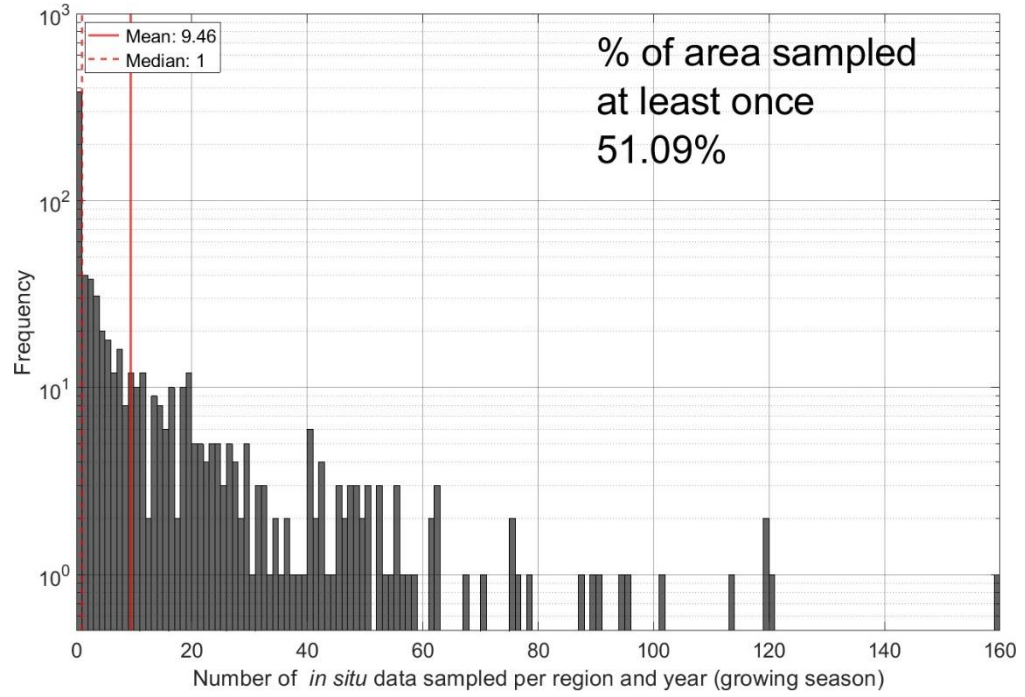

Figure 7.4: Histogram of all *in situ* Chl a samples for the entire growing season (March to September included) of each year and each OSPAR region. Note the log y scale.

7.3.2. Remote sensing versus *in situ* sampling

One of the key strengths of remote sensing is the ability to generate large numbers of observations with high spatial and temporal resolution. Of course, the major issue is then the question of data quality. To what extent does uncertain data quality undermine the value of large data volume? This can only ever be addressed through ground truthing effort using cruise effort and increasingly deployment of longer term *in situ* assets such as moorings and autonomous vehicles equipped with suitable sensors. Whilst there is a significant cost associated with *in situ* sampling, and the availability of abundant free to use satellite data is highly attractive, it is essential to highlight that the need for *in situ* sampling effort will remain ever present and, if anything, the demand for this type of data is likely to increase with the emergence of data hungry ML algorithm developments.

Data availability for *in situ* and remote sensing are compared for the OSPAR regions for a single month. Figure 7.5a shows the monthly Chl concentration for May 2003 from the NN-Rhot approach (monthly median value). Panel b shows how many times each pixel was available during this month (based on daily observations, therefore 31 is the maximum for May). Having access to a median product is more meaningful in the North Sea region when data are available 30% to 70% of the time. In the Atlantic part of the image, pixels are available less than 10 times a month down to zero due to intense cloud cover. How representative each pixel in the Atlantic is can be questionable. However, due to the randomness involved that make the different pixels available at different positions and times, the averaged value over the region should be meaningful despite the lower frequency. Nonetheless, the availability of remote sensing data per OSPAR region for this month is orders of magnitude greater than for *in situ* observations. In open waters the algorithm is suspected to return values near 1.0 $mg.m^{-3}$ when they are closer to 0.5 $mg.m^{-3}$ generally, or lower. The quality obtained in open waters is limited by the quantity of *in situ* training data that was available during the training process and was one of the limiting factors. Despite these minor problems, the whole image should be representative of surface Chl for May 2003.

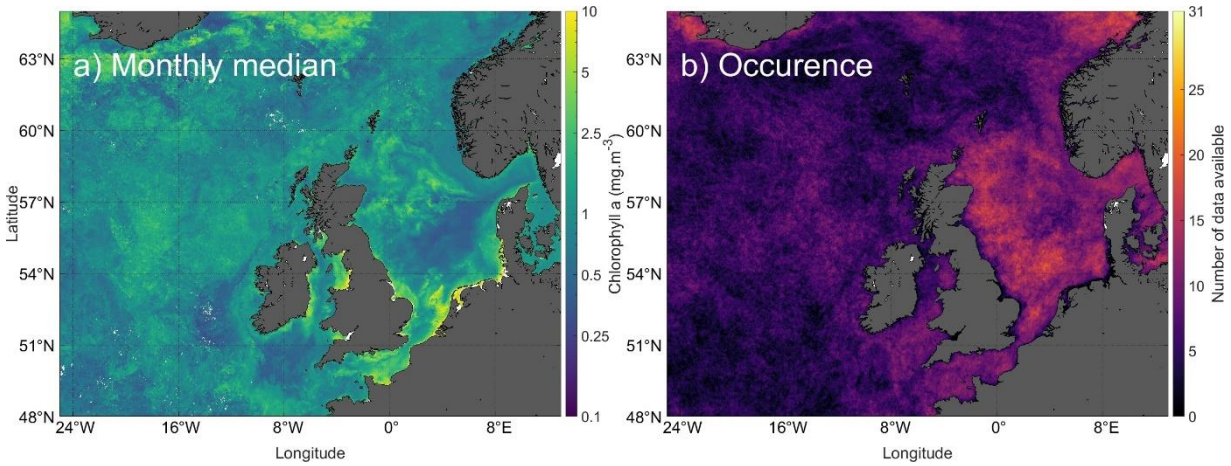

Figure 7.5: a) Monthly median of surface Chl a concentrations for May 2003 observed with the NN-Rhot algorithm applied to MODIS Aqua data. b) Number of data times a pixel was observed for this month.

To evaluate the data availability difference between a pure *in situ* and a pure remote sensing approach, Figure 7.6 and 7.7 below were created. Figure 7.6 shows the histogram of data availability from the remote sensing approach (Figure 7.5b) for each pixel, including the full image. Less than 10,000 pixels are never available, mostly due to the cloud cover in the Atlantic. The quality requirements used for the creation of the image also removed some of the pixels: with 10 networks available, 10 versions of each pixel are also available, and a relative standard deviation (rSTD) is calculated. If the rSTD of a pixel is above 20%, the pixel is flagged out. The remaining pixels are available up to 22 times over the 31 days possible. Figure 7.7 shows the number of times each region was sampled using either the remote sensing approach (7.7a) or the *in situ* sampling approach (7.7b). For remote sensing in 7.7a, the x axis is in a log form. All regions were observed at least 5,000 times, up to 5,000,000 for the Atlantic region due to its size. There is a net difference in data availability, and almost all pixels inside OSPAR regions were seen at least once using the remote sensing approach (99.68% for the full map). On the other hand, *in situ* samples were collected for only 31% of the OSPAR regions, which leaves most areas with no data collected at all for this specific month. In contrast, remote sensing provided large numbers of observations for every OSPAR region during this period. Whilst data quality still needs to be considered, clearly there is an argument that, by providing information where there is zero *in situ* data, remote sensing has potential to plug a major gap in our ability to monitor eutrophication. Whilst there are still some pixels in the image area where there are no remote sensing observations due to persistent cloud cover, considering the problem from the

perspective of OSPAR regions it is clear that there is potential to always have access to data from remote sensing. Moreover, only one month is considered here.

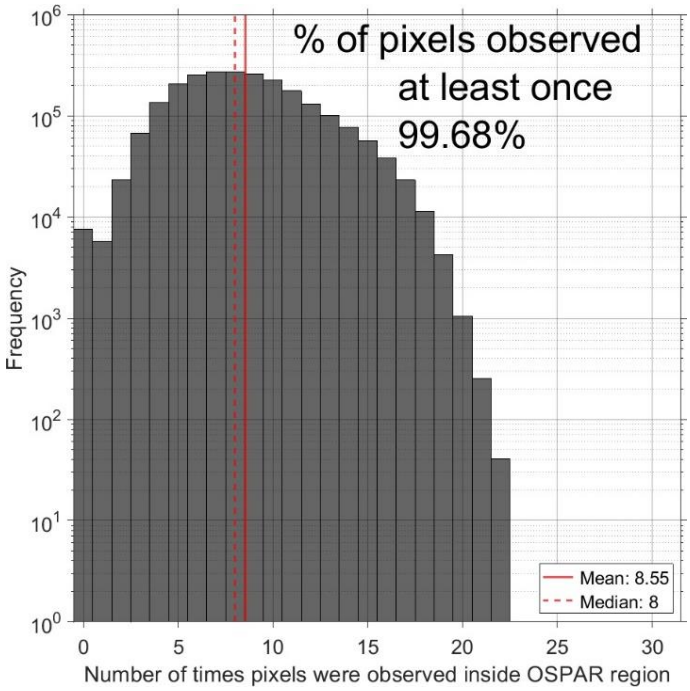

Figure 7.6: Histogram of the number of times all satellite pixels were observed daily inside any OSPAR region in May 2003.

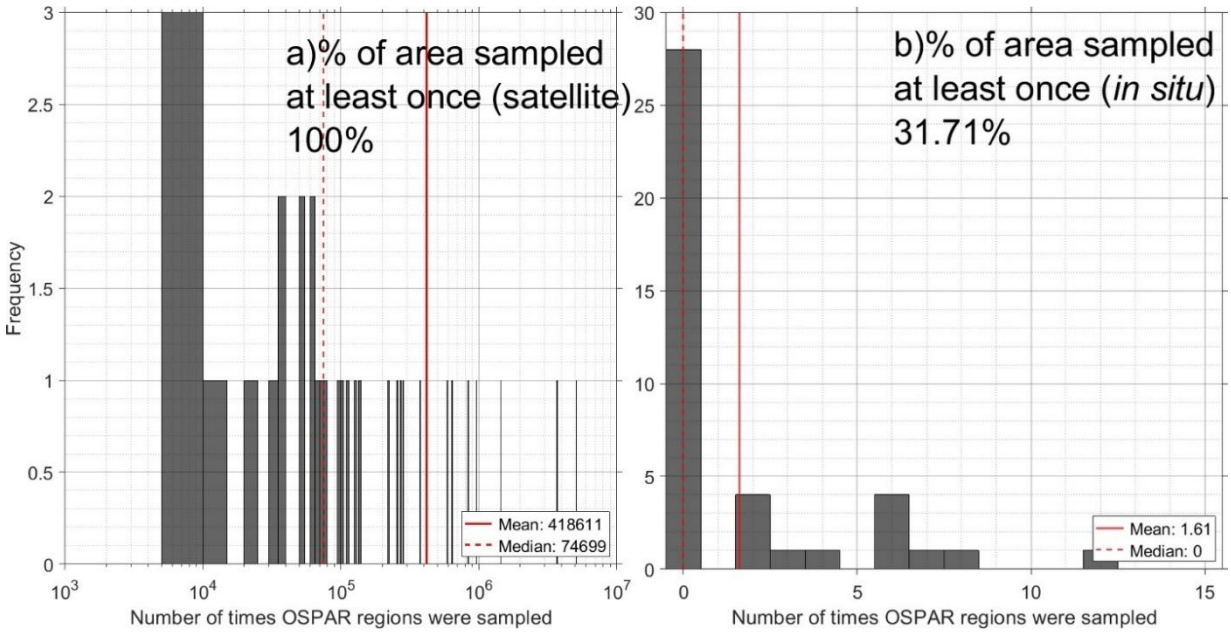

Figure 7.7: Histogram of the number of times pixels were observed each day inside any OSPAR region in May 2003 (using Figure 7.3b). b) Histogram of the number of *in situ* samples collected inside any OSPAR region in May 2003. Note the orders of magnitude difference.

7.3.3. Quality of data from remote sensing for the different OSPAR regions

In Chapter 3, it has been shown that remote sensing algorithms provide variable quality estimates of Chl in coastal waters. The neural network algorithm developed in Chapter 3 advanced performance even further compared to the current state-of-the-art algorithms provided by the Plymouth Marine Laboratory (PML) or ACRI-ST. The actual method to quantify the eutrophication status uses either the median or the 90th percentile (P90) of Chl as the main variable to be reported. Following the current OSPAR areas definition, which excludes nearshore coastal waters, there is limited benefit in having access to the improved algorithm developed in Chapter 3, as other algorithms may perform similarly far from the shore where turbid conditions are seldom present. The WFD group in charge of the assessment of the nearshore waters will benefit more from the improved algorithm developed in this thesis. Ideally, I think that these areas should be included separately as extra regions in the OSPAR Common Procedure assessment, as many *in situ* samples are collected near the coast and likelihood and impact of eutrophication is most likely to be significant in these areas. Figure 7.8 and 7.9 respectively show the median and P90 comparisons between the three satellite algorithms and any *in situ* data available (matchups between both approaches). Each point represents an average for each region, for each month and each year separately. The (a) panels of both figures represent a comparison between *in situ (*any data available*)* and *in situ (*that matched a pixel), which gives an insight at how having access to only certain data due to cloud cover or noise may impact how representative available data are compared to any data collected. The x axis represent any *in situ* data available (true for all panels). The y axis represents the *in situ* data that were used for the NN-Rhot development. These data are the x axis used on the NN-Rhot, on panel (b). It is a way to evaluate how *in situ* data that matched a MODIS Aqua pixel are representative of any *in situ* data available. For example, say only one *in situ* sample matched a pixel for a region during a specific month due to intense cloud cover. If multiple samples were collected and the matchup available is not representative (far from the median or P90), panel (a) will show discrepancies between *in situ* matching a pixel (y axis) vs *in situ* available (x axis). This is used as an indicator of residual variability in assessing the Chl concentration for a region using *in situ* observations. With MAD between 1.31 and 1.36 and R = 0.88–0.86, there is still considerable spread in the data even though we are comparing one subset of *in situ* data with another. This effectively provides an upper limit to the level of performance that we can reasonably expect remote sensing data to achieve.

Figures 7.8 and 7.9 illustrate the impact of spatio-temporal averaging on algorithm performance. Compared to pixel-specific performance (Figure 3.9, Chapter 3, also displaying the median value) we can see that there is a general improvement in statistical performance. For Rhot NN MAD = 1.75 for all 15,765 matchups (Figure 3.9d), while here the regionally averaged version has MAD

= 1.63 (Figure 7.8b). This represents a significant convergence towards the 1:1 line and needs to be compared with the comparison of *in situ* data sets (Figure 7.8a) where the MAD = 1.31. Spatio-temporal averaging also improves statistics for OC5-PML and OC5-ACRI products with MAD values reducing from 2.1 to 1.79 for OC5-PML and 2.85 to 1.95 for OC5-ACRI. Once again, Rhot NN is the best performer compared to OC5-PML and OC5-ACRI algorithms, and spatio-temporal averaging for the OSPAR regions further reduces uncertainty in the final product. At the same time, all algorithms still reflect the general performance seen in Chapter 3 with overestimations over the full range for PML and ACRI products, and overestimations for low values for the NN-Rhot product. The same observation is true for the P90 version (Figure 7.9), with slightly worse performances compared to Figure 7.8. Again, comparing a subset of *in situ* values (used for matchups) against the remaining *in situ* data set reveals significant residual variability that illustrates real sensitivity to under-sampling issues. There is genuine scope for rapid and significant variability in the Chl concentration for any given location, and tremendous scope for single local samples to be poorly representative of overall variability across a large region. In this context, it is clear that the level of data availability provided by remote sensing provides much needed insight into intra-regional variability. Moreover, the results presented in Figure 7.8 and 7.9 show that the level of data quality provided by the Rhot NN algorithm is broadly comparable with that obtained from *in situ* sampling (though still subject to limitations discussed previously e.g. relatively poor performance for coccolithophore blooms). The combination of improved data quality and massively increased data availability is a very strong argument for inclusion of remote sensing data in future eutrophication assessments.

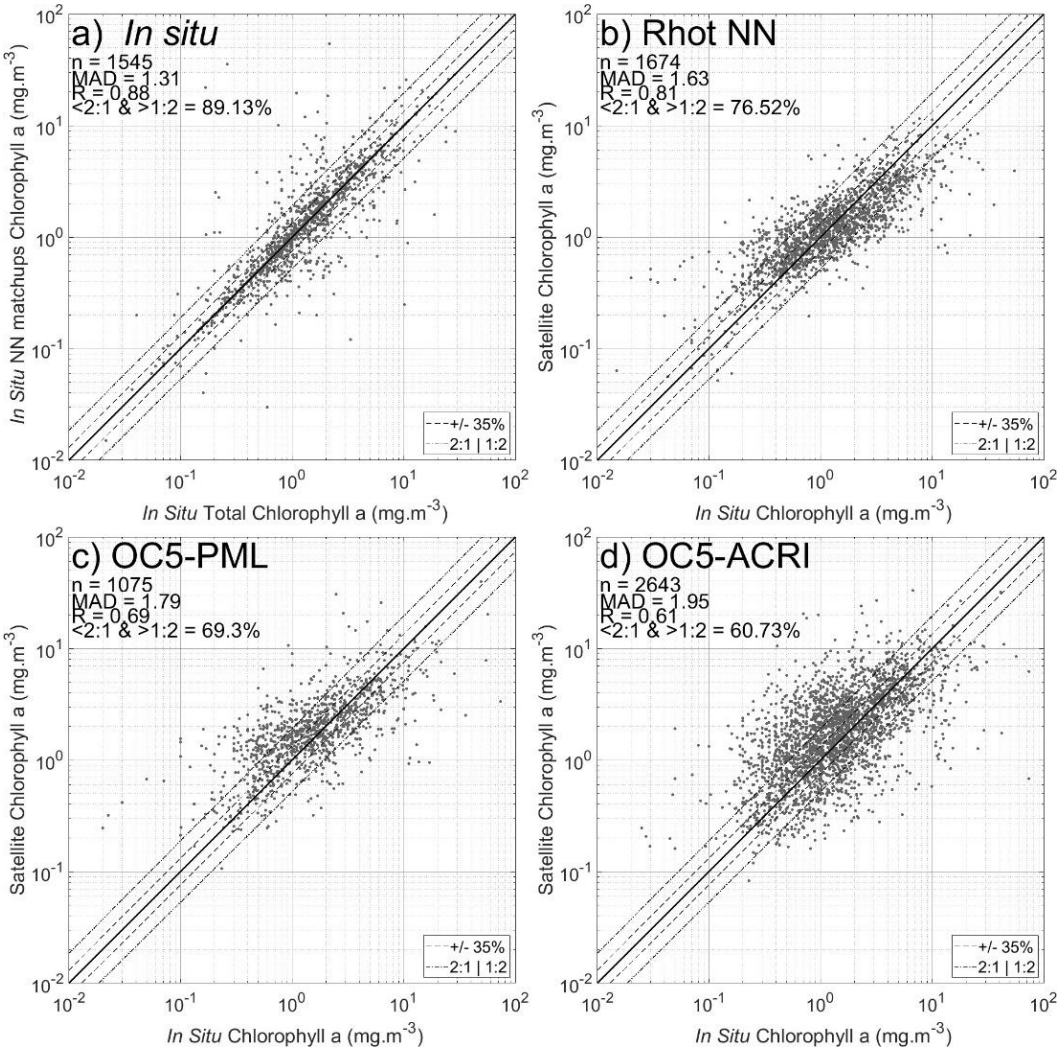

Figure 7.8: Performance obtained for each month of each year and over each region for different products using the **median**. It compares the *in situ* available values (x axis) against what is returned from the product. For panel a), the y axis uses *in situ* Chl matchups used to develop the neural network shown in panel b) while the x axis contains any available *in situ* data for the same conditions (region and time). Performance obtained for b) the Rhot NN algorithm, c) the OC5-PML algorithm, d) the OC5-ACRI algorithm.

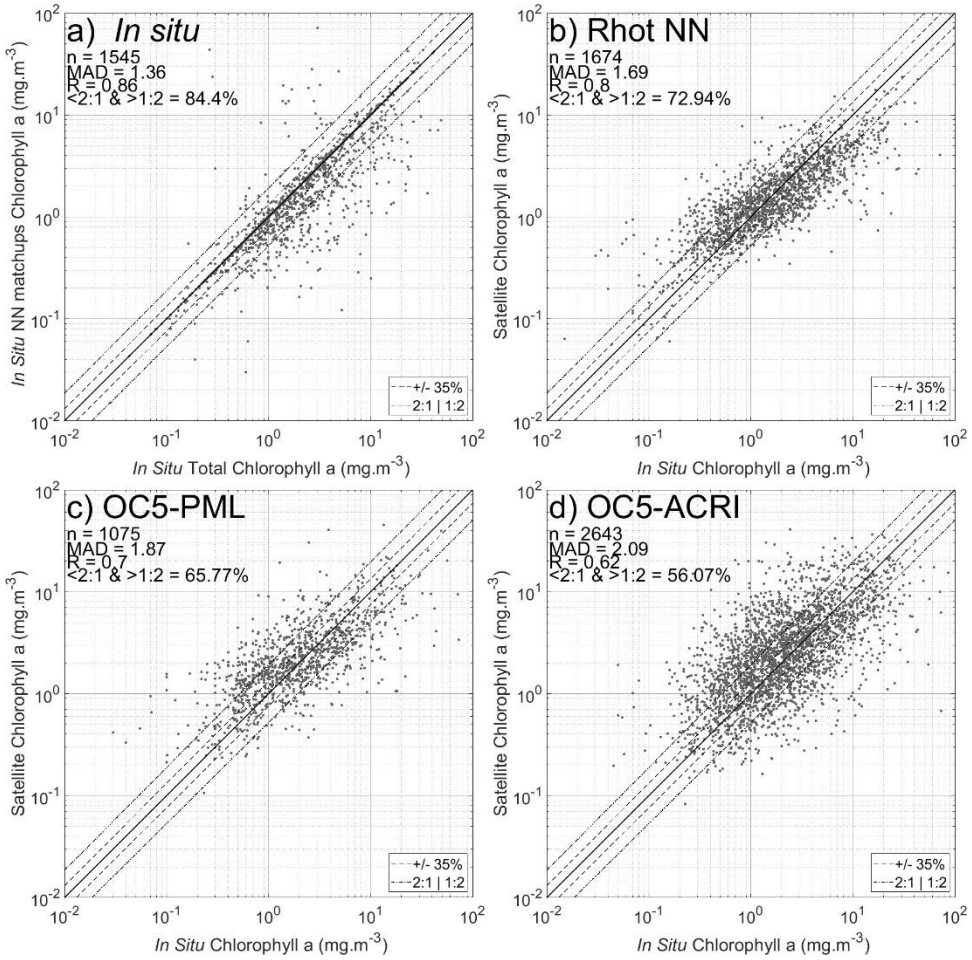

Figure 7.9: Performances obtained for each month of each year over each region for different products using the **P90** of Chl. It compares the *in situ* available values (x axis) against what is returned from the product. a) *In situ* Chl matchups used to develop the neural network compared to any available *in situ* data for the same conditions (region and time). Performances obtained for b) the Rhot NN algorithm, c) the OC5-PML algorithm, d) the OC5-ACRI algorithm.

Using the same May 2003 data, Figures 7.10 and 7.11 respectively display the Chl concentration averaged over the different available OSPAR regions using the median (Figure 7.10) or the 90th percentile (Figure 7.11) for the NN-Rhot remote sensing (a) and *in situ* methods (b). I have opted to present the median in Figure 7.10 as the mean value has potential to be impacted by outliers, which is potentially particularly problematic for the *in situ* data that is generally very low in abundance. Figures 7.10 and 7.11 highlight the impact of data availability on ability to report for each region. Whilst there is full coverage from the remote sensing product, there are multiple OSPAR regions for which there was no *in situ* data available for this month. Where data are available from both approaches, there is some level of consistency, though it is clear that it is not exact. Taking into consideration the performance levels demonstrated in Figure 7.8, and furthermore taking into account the very small amount of *in situ* samples collected during this month for many of the regions

(Figure 7.7), the question that arises is to what extent should the *in situ* data be regarded as 'truth'? With a maximum of 12 *in situ* samples in one OSPAR region in this month, it is difficult to attribute strong statistical confidence to the *in situ* data, particularly P90 values that are even more likely to be impacted by outliers and effects of limited sample numbers. Local spatial heterogeneity means one should have little confidence in such small numbers of samples providing an adequate representation of general properties across such large areas. In contrast, the high data density provided by the remote sensing data, combined with the established levels of performance, suggest that there are good grounds for at least including satellite data in the eutrophication assessment. Whilst it is understandable that there will be caution about adopting remote sensing data due to the indirect nature of the observation, it is equally important to be conscious of both the residual uncertainties associated with *in situ* Chl observations (Figures 7.8a and 7.9a) and the potentially severe limitations on statistical significance caused by lack of data density for *in situ* observations.

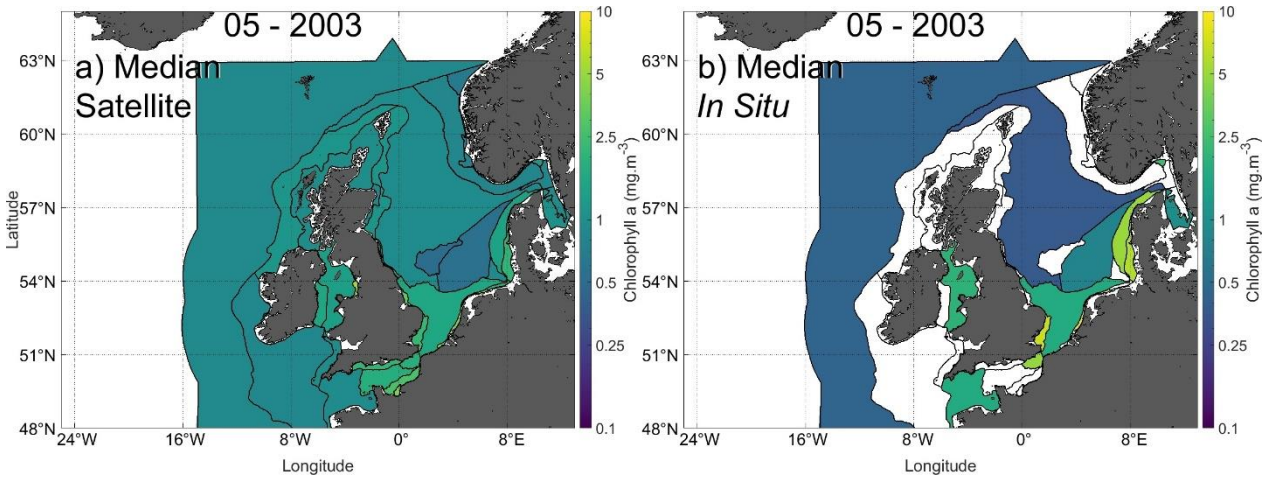

Figure 7.10: a) Monthly median map for all pixels of each OSPAR region using the remote sensing approach. b) Monthly median map for all *in situ* samples collected inside each OSPAR region. May 2003.

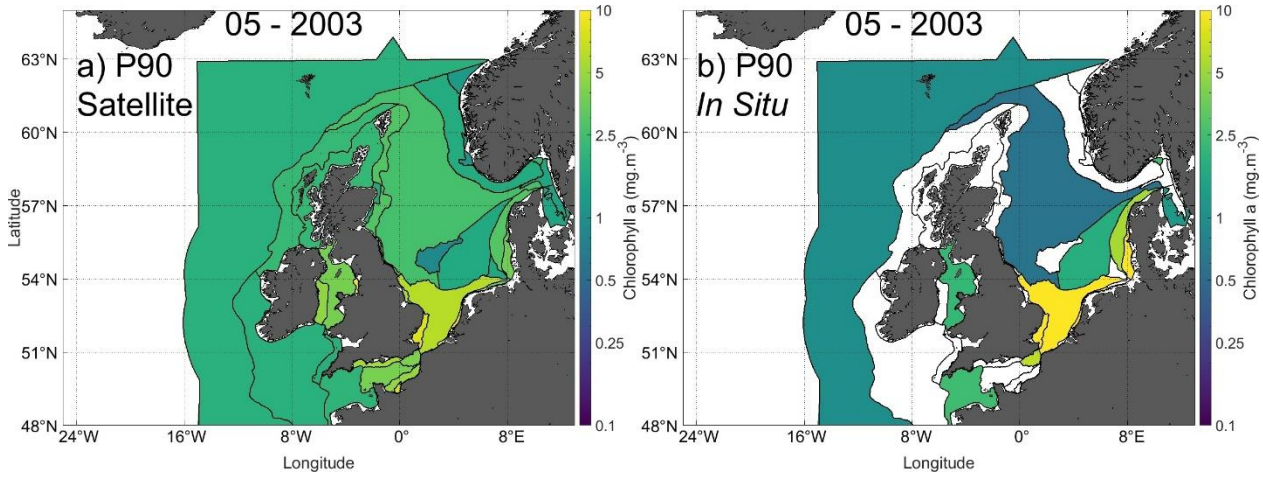

Figure 7.11: a) Monthly P90 map for all pixels of each OSPAR region using the remote sensing approach. b) Monthly P90 map for all *in situ* samples collected inside each OSPAR region. May 2003

7.3.4. Evolution of Chl for the different OSPAR regions

This final section of the thesis tries to show how the Chl concentration of different regions evolved from 2002 to 2020. The satellite product developed in Chapter 3 is not fully available at the time of writing (daily Chl maps). It will be provided to Marine Scotland Science at the end of the studentship. The main objective of the thesis was to develop a Chl product that would benefit the assessment community for eutrophication assessment by making more accurate estimates in coastal waters, where currently available algorithms do not perform particularly well. Because the NN remote sensing algorithm is still being processed, only *in situ* data will be used for this section. The main question that needs to be answered is simple: did the water Chl concentration over the different OSPAR regions increase since 2002, which would highlight eutrophication? While the question is simple, presenting it on a figure to answer it is not, especially because multiple definitions and methods stand for eutrophication (use of the mean, max, P90, inclusion of the satellite data or not, and finally inclusion of additional variables such as dissolved oxygen and nutrients). If we consider only Chl concentration changes, it is still not trivial. Many regions were not consistently sampled during the growing season from 2002 to 2020, and some may have only several samples available a year. We need to make the assumption that the low data availability does not affect the evaluation and that any value obtained during a year is representative of the region. The period for this assessment is restricted to the 17 years between 2003 and 2019, because MODIS produced data since July 2002 only, and no matchups were available after January 2020. More data are available, especially if the *in situ* method is used only, and samples can go back to the 1960's at least.

To create Figures 7.12 and 7.13 below, any value available for each month and each region is converted into a monthly value, median (7.12) or P90 (7.13). Only the months of the growing season have been used (March to September included). The monthly values are converted into a single

median yearly value. The reason for the monthly conversion is to avoid using too much information from a single month that has been sampled more than the others, possibly because a specific cruise campaign had occurred at that time. Different figures were tested, such as representing the concentration itself as a continuous curve, or trying to find a potential linear trend as has been shown previously in the literature (Desmit et al., 2019, Figure 4). None of the approaches tested seemed to provide useful visual information to inform a reader about a potential trend. The high variability in Chl concentrations coupled to the limited amount of data makes it difficult to evaluate. Following what is usually done for temperature, the trends are commonly shown as a difference between a period and the measured value (for example 1900-1981, see NOAA website for examples https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature, accessed on 11 July 2022). This is called an anomaly. I reproduced that approach to try visualising the Chl trends. Using *in situ* samples, a dozen of the available regions have access to a least one Chl value during the growing season continuously each year, from 2003 to 2019. Eight of them are shown below in figure 7.12 and 7.13 (randomly selected). A median of the entire period is calculated for each of the eight regions, and is removed from all yearly Chl concentrations available. Therefore, Figures 7.12 and 7.13 show the yearly Chl concentration for each of the 17 year minus the averaged value of the whole sampling period.

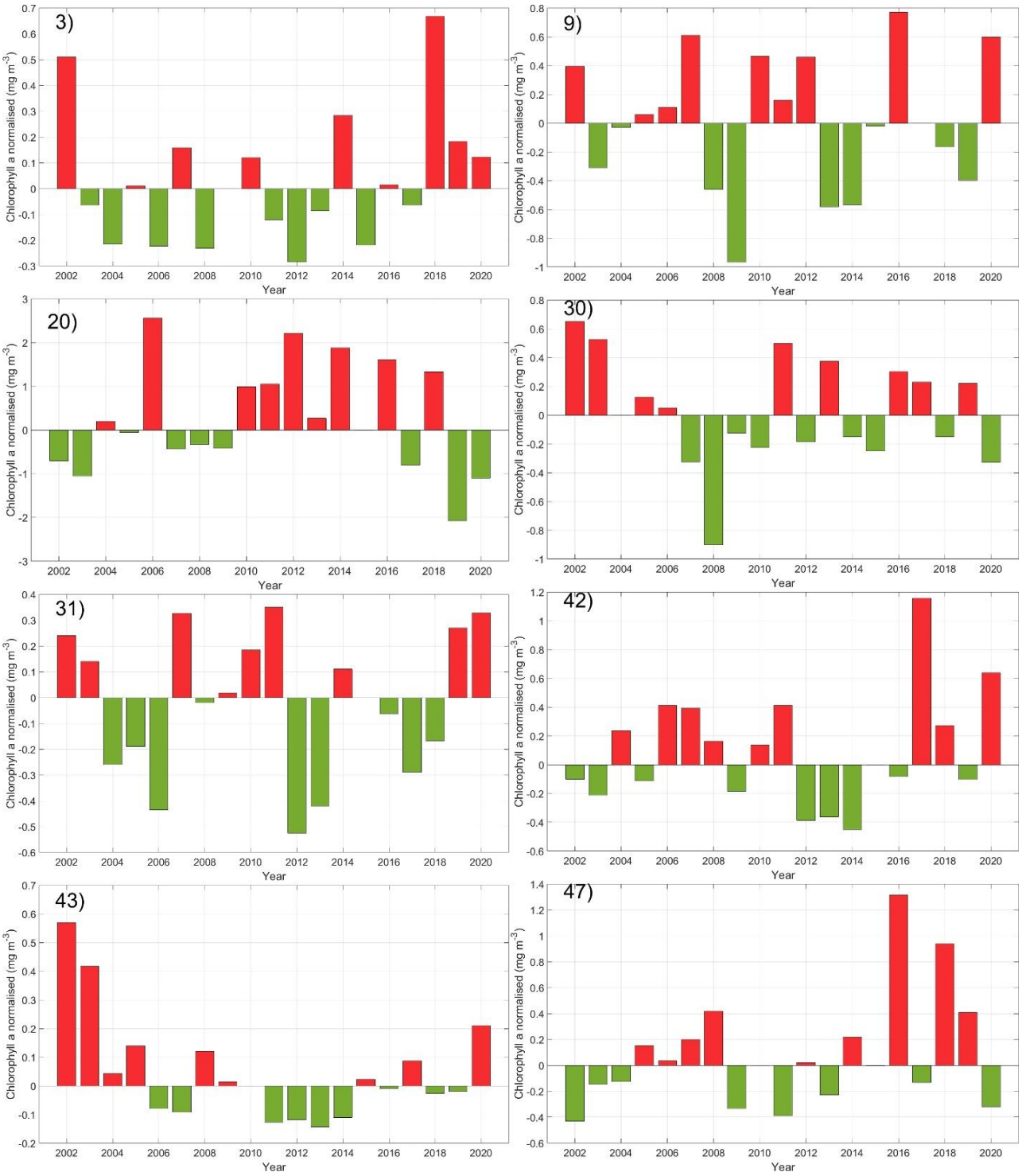

Figure 7.12: Yearly median Chl a concentration (normalised by the 2002-2020 median value) for eight different regions. The number shown is the OSPAR's region identifier: 3 = Atlantic, 9 = Kattegat Deep, 20 = Elbe Plume; 30 = Outer Coastal Germany/Denmark ; 31 = Eastern North Sea; 42 = Kattegat Coastal; 43 = Northern North Sea; 47 = Scottish Sea. Note the different y axis.

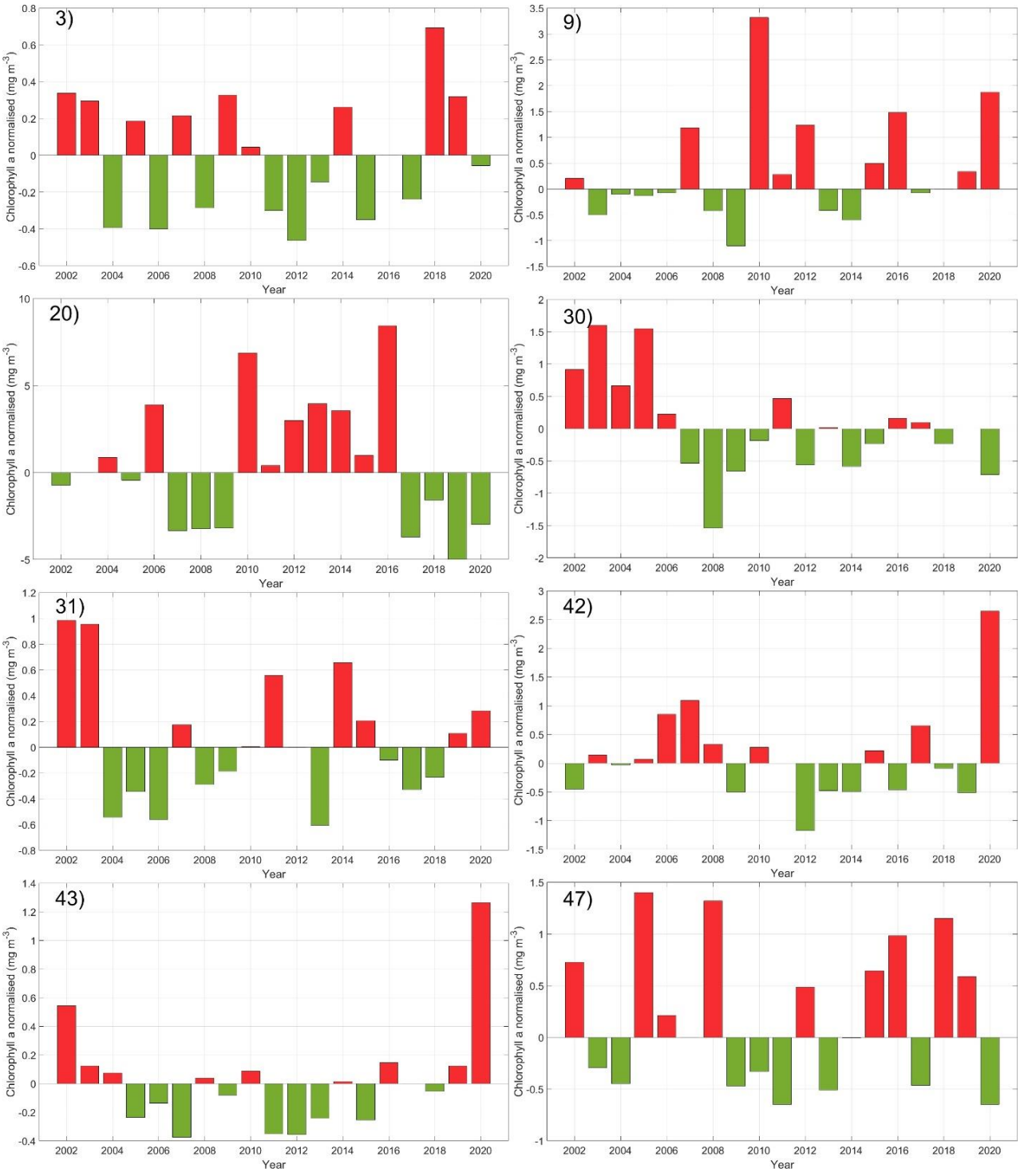

Figure 7.13: Yearly P90 Chl a concentration (normalised by the 2002-2020 median value) for eight different regions. The number shown is the OSPAR's region identifier: 3 = Atlantic, 9 = Kattegat Deep, 20 = Elbe Plume; 30 = Outer Coastal Germany/Denmark ; 31 = Eastern North Sea; 42 = Kattegat Coastal; 43 = Northern North Sea; 47 = Scottish Sea. Note the different y axis.

There is no clear statement that can be made of the trend occurring for the different regions. Chl is driven by huge seasonal variations. Except region 30 "Outer Costal Germany/Denmark" that presents lower values compared to early 2000s (more visible for the P90 approach), it is not clear that any trend is present. The two figures look like an oscillation, similar to an El Nino Southern Oscillation (ENSO, McPhaden et al., 2006) or Atlantic Meridional Oscillation (Srokosz and Bryden, 2015) events. The different regions do not seem linked, as some can increase or decrease at the same time; therefore I will not comment on it further. The same procedure has been repeated with a log transformed version of the Chl to avoid difference in orders of magnitude to hide the potential trend. The same features are visible and only slightly rescaled (not shown as it is very similar to Figures 7.12 and 7.13). If no strong trend is visible, does it mean that the different regions are stable at the moment? Would we get the same answer if we would include the very coastal waters? Is the data available representative enough? It is hard to answer, even if I highly suspect that the later question is potentially the main limiting factor.

7.4. Conclusion

This chapter has illustrated limitations of *in situ* sampling for eutrophication assessment and presented further evidence of the potential to successfully use remote sensing data to support the process. There are a variety of outstanding issues that need to be addressed in order to establish an agreed eutrophication process irrespective of whether remote sensing data are included in the assessment process or not. At the moment, the mean, max or 90th percentile of *in situ* Chl sampled over different regions in European coastal shelf seas are used to quantify how the Chl evolves (OSPAR report, 2017). The original definition of eutrophication indicates that a water body faces an increased concentration of Chl compared to previous observations due to an anthropogenic enrichment of nutrients. OSPAR's exact definition is: "the enrichment of water by nutrients causing an accelerated growth of algae and higher forms of plant life to produce an undesirable disturbance to the balance of organisms present in the water and to the quality of the water concerned, and therefore refers to the undesirable effects resulting from anthropogenic enrichment by nutrients as described in the Common Procedure". However, to evaluate if a water body faces eutrophication, not only Chl but also other Common Indicators (O_2 and nutrients, which may involve indirectly other variables used to build the nutrient models such as the temperature) values are used by OSPAR. The different indicators are compared for successive years, usually six years in a row (Van der Zande et al., 2019) to a reference level supposedly not impacted by any human inputs. If a specific area is affected by eutrophication, it can highlight a potential anthropogenic impact such as an excessive release of nutrients (De Jong et al., 2006). Agriculture use soil nutrient enrichment to improve production. Rain and wind can then transport the nutrients (nitrogen and phosphorus) to the ocean, which can cause eutrophication if the

area is limited by nutrients (Daniel et al., 1998; Conley et al., 2009). Previous policy decisions helped to manage eutrophication problems in the southern North Sea region by forcing a reduction in nutrient release (Schindler, 2006; Smith and Schindler, 2009; Beusekom, 2018). The process is called de-eutrophication because it directly tries to invert the eutrophication process. Desmit et al. (2019) observed that two major variables influence the eutrophication status in the North Sea, the warming of the sea (Hoyer and Karagali, 2016) and the control of the riverine nutrient inputs (Meyer et al., 2018). Changing one of them directly affects the eutrophication status. Our understanding of the phenomenon in continuously improving due to its high complexity and multiple interactions.

If we consider only Chl, the previous assessments made by OSPAR rely on *in situ* sampled data only, which are scarce due to their cost. The OSPAR regions have been selected to be ecologically and hydrologically similar and therefore comparable. Having an area that shows the same behaviour reduces the pressure to sample the different indicators everywhere. This type of *in situ* data based assessment depends on the data that can be sampled and efforts made by the community. Being able to sample every area consistently was not achieved with *in situ* efforts between 2002 and 2020. 30% of the areas were observed at least once a month during the growing season each year (Figure 7.3). This increases to 50% of areas sampled at least once over the whole growing season March-September (Figure 7.4). Data availability is a limiting factor which can be tackled by including remote sensing surface Chl, if the quality is accurate enough for the environmental agencies. The current OSPAR assessment planned for 2023 includes the JMP-EUNOSAT remote sensing Chl algorithms, recently developed by Van der Zande et al. (2019). This new product merges the OC5–PML algorithm (shown in Chapter 3) and the Gons red algorithm (Gons et al., 2002). The JMP-EUNOSAT approach relies on classification as the main method to retrieve more accurate Chl concentration from space by successfully identifying when the OC5-PML fails to produce good quality estimates. With this new approach used by OSPAR, *in situ* and satellite data are equally weighted for the different areas. It directly tackles the data availability issue for the remaining 70% of the areas not sampled monthly and improves the data availability for months that already contained data.

It has been demonstrated that satellites can provide much more data than is currently sampled from cruise campaigns, by up to five orders of magnitude if we consider the whole region and a single month (Figure 7.7). The quality of the retrieval achieved from remote sensing products is getting better with time thanks to new discoveries. When applied to the current OSPAR regions, any of the available remote sensing products mentioned in this thesis (JMP-EUNOSAT, the algorithm developed in Chapter 3, OC5–ACRI or OC5–PML) return similar performances with only small differences (~10–20% difference for the MAD, Figures 7.8 and 7.9). This similarity in performance is due to the non-inclusion of the nearshore coastal waters, handled by other organisations or assessed

through different procedures. The Baltic region is assessed by the HELCOM, which reduces impact of improved algorithms dedicated to solving estimation in conditions met in these waters (high concentrations of CDOM in the western Baltic). The current regions definition does not allow a real benefit from the Chl algorithm developed during this thesis. From a remote sensing approach and to characterise the evolution of Chl only, a consistent grid could be used and include coastal regions if required. The future method (QSR 2023) will transform satellite and *in situ* data to a single value for each region (mean or P90), over a six month interval. With a single sensor like MODIS Aqua, a single pixel in the North Sea can be visible up to 140 times a year from March to September, the actual definition of the eutrophication period. Pixels are available less often for the Atlantic part area due to the increased presence of clouds, with availability reduced to 30-40 times on average over the same period. By nature, the satellites provide information all over the region at different times during the growing season, and clouds are the main source of reduction. It has not been discussed in this thesis due to time of development required, but there is an opportunity to merge satellite images temporally, similarly to the OC5–ACRI product, with various methods (e.g. DINEOF) to increase the data availability, if more data are required (Alvera-Azcárate et al., 2009; Nechad et al., 2011).

The remote sensing approach has its own flaws. Except the algorithm developed in this thesis, other current remote sensing algorithms do not propose a Chl estimate for high latitudes during winter. These data are usually masked due to concerns about impacts of low sun angles on atmospheric correction and production of accurate BOA remote sensing reflectance data. Switching to a top-of-atmosphere algorithm seems to drastically improve estimates (Chapter 3) and may provide access to winter data in future. The quality achieved during winter with the algorithm presented in Chapter 3 cannot be expected to be as good as during the rest of the year due to very low amounts of training samples available, but should not show the huge overestimations proposed by blue–green algorithms. The growing season may have to be redefined in future, especially if climate change impacts the shift of the blooming season (Wasmund et al., 2019; Ardyna and Arrigo, 2020). The OSPAR assessment spanning over 30 degrees of latitude may lead to a redefinition of “growing season”. It may be very difficult to retrieve low chlorophyll concentrations when concentrations of other constituents are very high as the contribution to the reflectance signal may be insignificant, when too much sediments are present for example. *In situ* monitoring should clearly be the main observation method in such conditions, which should be easier than sampling everywhere due to the proximity of the coast. This observation also stands for coccolithophores, which can cover a great surface during the growing season in the North Sea and Eastern Atlantic (Rivero-Calle et al., 2015). Current remote sensing Chl algorithms produce underestimates when coccolithophores are present and usually mask them. The underestimation is more pronounced when the coccolithophores die and released their

shells, which increases the backscattering across the full spectrum (Cazzaniga et al., 2021). Being able to estimate coccolithophores' Chl concentration may require a different approach. Eventually, the remaining limiting factor of a remote sensing approach is the access to a two dimensions signal, which will never be able to replace the depth information gathered by the *in situ* approach.

The OSPAR Common Procedure eutrophication assessment developed since 1997 is closer to a water quality and ecosystem evaluation, rather than a pure Chl evolution. The reason for such a definition comes from the complexity of the problem. For the North Sea, a warming trend has been observed between 1983 and 1987 (Beaugrand and Reid 2003; Edwards et al. 2006; van Aken 2010) and still occurs nowadays (Høyer and Karagali, 2016). Different regions and periods facing stratification will lead to different strategies and dominance of species (Hunter-Cevera et al. 2016, Barton et al. 2018). The nutrient inputs from rivers have been limited since the 1980s, and de-eutrophication of some North Sea regions has been happening since (van Beusekom et al. 2009; Burson et al. 2016; Meyer et al. 2018). There is a potential change of seasonality due to increased temperatures (Atkinson et al. 2015), spring may start earlier, and autumn last longer. Moreover, the nitrogen inputs to the ocean also come from the atmosphere for 10 to 30% of the total anthropogenic input (Dulière et al. 2019), which reduces the control we have on riverine nutrients. An unusual increase in Chl disturbs the upper trophic levels that depend on it, and is just the start of new problems, such as oxygen depletion (Greenwood et al., 2010). In fact, the area is already not in the same environmental conditions as the pre-industrial level for temperature, one of the main impacting factors (Desmit et al., 2019), and the system is constantly evolving. This unstable state is challenging, especially when a specific definition is used, such as "evolution of biomass during a specific period". The OSPAR Common Procedure already changed from the initial definition, and will continue to evolve following science discoveries. The recent inclusion of remote sensing is one of the available solution to tackle the data availability problem and support the *in situ* observations.

Take home message: Chapter 7

In Chapter 7, the main idea was to evaluate how the current eutrophication assessment is conducted using only chlorophyll a and the limitations involved, while showing the potential of remote sensing products to generate more data.

- An approach relying on *in situ* data only to evaluate the eutrophication status is greatly limited by data availability, which support the need to use remote sensing.
- The method that is currently being deployed use both the *in situ* and remote sensing estimates of chlorophyll with the same weight in the final output (50% each), but the later one is available up to five orders of magnitude more often than the former. Such approach greatly limits the remote sensing capabilities and puts pressure on the definition of an "eutrophication region".
- Comparing any available matchups with any *in situ* data collected already creates discrepancies between ground and remote sensing observations.
- Among the areas with samples consistently collected during 2003-2019, none of them showed a sign of eutrophication.

There are open questions concerning the use of remote sensing products and its limitations or advantages over the *in situ* approach. The method developed in Chapter 3 that provides more data than ever may help refining the future definition of eutrophication, and a pure remote sensing approach may be appropriate.

Chapter 8: General Conclusion

8.1. Main results of the thesis

The main objective of this thesis was to improve the retrieval of environmental variables such as chlorophyll a in coastal waters or sea surface temperature using remote sensing. Remote sensing has potential to be a key tool as it provides observation of almost the entire Earth surface every day, depending on the sensors used. However, this data availability is only part of the story; data quality is essential and requires proper data assurance qualification before remote sensing data can be adopted by monitoring agencies for reporting against environmental legislation. For chlorophyll a and temperature, the methods currently used by the remote sensing community require two steps, first correction of the signal measured by the sensor from different effects linked to the atmosphere or surface features, then development of algorithms. Once the signal leaving the water has been isolated, different algorithms can be applied to it. For chlorophyll a, in open waters where the presence of phytoplankton directly controls the light signal, simple relationships between the signal in the blue and green have been found in the past and a review of these algorithms was recently made (O'Reilly and Werdell, 2019). In coastal waters, when non-algal constituents alter the light signal, different algorithms to retrieve chlorophyll a have been created using mostly the red part of the signal (Gons et al., 2002; Gohin et al., 2002; Gitelson et al., 2011; Moses et al., 2012; Moses et al., 2019). These algorithms cannot be applied to all sensors, as they require access to specific wavebands in the red or near infrared. Moreover, there is no currently agreed method on how to perform the atmospheric correction in coastal waters and the light signal is generally of poor quality for turbid water conditions.

Disruptive signals from the proximity of the land (Bulgarelli and Zibordi, 2018a, 2018b, 2020), the depth of the seabed (Canizzaro and Carder, 2006) or aerosols emitted by human cities (Retalis et al., 1998) are factors that further deteriorate atmospheric correction methods. Operational methods currently proposed by the agencies in European polar waters rely on 1) application of extra quality controls prior to using the light signal (IOCCG reports 16, 2015), and 2) classification of the water into similar spectrum type where specific algorithms for open, oligotrophic and coastal waters can be applied (Moore et al., 2014). Due to the issues met in coastal waters (correction of the signal and performance of the algorithm), there is a concern that the quality of ocean colour variables do not meet the performance requirements of environmental agencies for assessing the environment. Having access to better quality remote sensing key variables will greatly help policymakers for future decision making. The main strategy adopted in this thesis consisted of not following the current path of classification and water-type specific algorithms, but rather using machine learning to build a more

robust general algorithm. The main steps were to 1) process satellite and *in situ* data to build matchup datasets, to be able to 2) develop new algorithms that could answer the current issues.

8.2. Chapter 3: Development of a coastal chlorophyll algorithm

In Chapter 3, different tests have been conducted to evaluate if the atmospheric correction procedure, known to fail in coastal waters, could be bypassed. Three similar matchup datasets have been created, using either the top of atmosphere reflectance (Rhot), top of atmosphere partially corrected reflectance (Rhos) or the bottom of atmosphere reflectance (Rrs). With the help of neural network algorithms (multi-layer perceptrons), it has been shown that an algorithm relying on the top of atmosphere signal not only matched but actually outperformed the bottom and partially corrected versions by a small factor (Section 3.4.2). Neural networks belong to the deep learning field, a most active research field in recent years due to their capacity to deal with complex tasks and outperform humans on many tasks. One of the key challenges in developing a neural network is in deciding upon optimal architecture features and considerable effort was put into trying to find statistical justification for each step in that process. First, it has been shown that the size of the architecture used impacted performance (Section 3.4.1). Avoiding high numbers of layers and low numbers of neurons is the correct decision to reach higher performance and to avoid overfitting. However, it was also found that there was typically no obviously better single candidate, so there remains considerable flexibility in this aspect of network design. A recommendation would be to use a similar or slightly higher number of neurons compared to the number of inputs, with 3 hidden layers. Visualisation of the neural network estimations is required to avoid underfitting issues that only slightly impact the statistical metrics. The second step concerned the ensemble approach (Section 3.4.3), which consists of application of multiple iterations of the same architecture to avoid artefacts associated with randomness in the training data set. This ensemble approach provides an opportunity to determine variability in network predictions and enables assessment of how confident the neural networks are between each other. When the same neural network architectures produce non consistent data for a specific pixel, it highlights that hose particular conditions were not well represented in the training data set. Repeating the application of a similar network gives access to standard deviation that can be used to provide a measure of how robust each estimation is.

Neural networks applied to top of atmosphere signals also outperformed existing empirical algorithms to retrieve chlorophyll a in coastal waters (Section 3.5.1). By having been trained using a very permissive matchup dataset including data from winter and limited flags applied (when the standard atmospheric correction fails, mostly when it meets clouds, lands or glint), neural networks return more coverage than standard methods and better estimations. They also provide access to

winter data, though with a poorer quality at the moment due to the limited quantity of training data (Section 3.5.2). It is worth noting however, that access to winter data is currently not of concern to most of the environmental agencies as they base their eutrophication assessments on observation of the growing season (start of Spring to late Summer). Low sunlight in winter is the primary limiting factor to phytoplankton development, and waters usually contain very low concentrations in winter. If the paradigm changes and winter data become of importance because they may trigger what will happen during the growing season (Gonzalez et al., 2022), having access to a remote sensing product in winter will be a decisive tool. Other techniques simply mask their data once a threshold solar zenith angle is reached, because it induces an even poorer light signal which further degrades the atmospheric correction process (IOCCG Report 16, 2015; Li et al., 2017). This issue is problematic for neural networks operating on bottom of atmosphere data; here it was found that using a neural network with bottom of atmosphere light signals did not produce coherent estimations in winter (Section 3.4.2), which is attributed to poorer atmospheric correction with low sun angles. Application of extra flags to further improve the quality of the light signal did improve performances reached (Section 3.5.4) but drastically reduced the numbers of available data and led to overfitting. Neural networks differ from empirical methods mostly by the amount of data required in order to achieve good quality estimations. It is currently more appropriate to rely on a noisier algorithm that can be applied in most conditions rather than having access to limited datasets which would tend to overfit due to the lack of data. Generalisation is one of the first concerns when developing a neural network algorithm. With scarce access to data over a limited area, there is little hope a regional algorithm can generalize, and the method should be extended worldwide before being applied anywhere outside the training area. Inclusion of non-light signals such as the geographical coordinates or temporal information was also tested (Section 3.6). This was a healthy test because performance of such networks was found to be significantly improved ($R > 0.95$, $MAD < 1.2$). However, this apparent improvement in statistical performance was found to be associated with overfitting problems and when applied to real images, the mesoscale features of the surface of the ocean were blurred and smoothed out. Therefore, using a signal that does not affect the measure of the light (such as the coordinates or temporal information) should not be used for directly measured Earth observation products. Note, however, that data filling methods such as DINEOF *can* use these additional data types to good effect.

8.3. Chapter 4: Band combination analysis

Having access to more training data is not the only factor in achieving better performance; having access to more spectral information also contributed. In Chapter 4, different band combinations were tested (Sections 4.2.1.1 and 4.2.1.2). The information carried by each waveband

can be similar for neighbouring bands when they are only a few nm apart from each other. Decisions are required when designing a spectroradiometer that observes the Earth, and for example being specific for the land or ocean observation is of importance to reach good quality data. The main difference is the signal-to-noise-ratio required for the ocean bands because a lower signal is measured due to lower backscattering from the water particles compared to the land ones. From basic RGB combinations, revisiting existing empirical algorithms or testing inclusion of any data, it has been observed that access to more spectral information generally led to better estimations (Section 4.3). Inclusion of redundant bands also improved performance achieved, though with a reduced impact. A machine learning clustering algorithm, called the Adaptive Distance-Based Band Hierarchy method (ADBH, Sun et al., 2020), was applied to select a reduced number of band (Section 4.4). The ADBH algorithm often selected bands designed for land observation instead of the bands designed for the ocean observation. Switching from neighbouring land or ocean bands did not significantly alter performance. Land bands usually come with a wider bandwidth which can help by providing useful signal that is not observed using oceanic bands. Some bands may have poorly contributed to the improvement of performances (469, 1240 and 2130 nm for example), but could help deal with specific conditions such as very turbid conditions for the infrared bands and should be kept. There is a need to analyse the generalisation capacities under high aerosol content, but it will require a proper dataset to knowingly make the decision if a global algorithm is designed in future.

This idea of including more bands in the neural network was applied to independent matchup datasets for mostly case 1 open waters (Section 4.5). In such conditions, using OC3 or a neural network using OC3 bands to retrieve chlorophyll a returned very similar performance (MAD of ≈1.7). It is only when more bands are included that the neural networks can take the lead and outperform empirical methods, because they can use the full light spectrum with no difficulty (example with SeaBASS dataset, Figure 4.13). This is one of the main strengths of neural networks, the capacity to seamlessly incorporate additional information as an input. There is no trivial method to find a physical relationship when using an empirical method when many bands are available, and empirical algorithms rarely use more than two or three bands in practice. The difficulty to build complex empirical algorithms explains why the community relies on classification and water specific algorithms. By being able to identify significantly different areas, empirical algorithms using a few bands can be developed and applied at the right place with decent performance. It is much simpler to develop empirical models with two rather than 15 bands that would be general and applied anywhere. This is a further reason why neural networks are one of the potential solutions to solve issues when the signal is complex in coastal conditions. As soon as the training data available is sufficiently representative of the range of observed conditions, classification is no longer required and general

relationships can be obtained. Both independent datasets tested in this section do not propose the maximum number of band available by either the satellite (Seabass dataset, Section 4.5.1) or *in situ* sensors (Nomad dataset, Section 4.5.2). With the capacity of machine learning method to incorporate more information, there is a clear need for matchup datasets to be impartial and provide any information available from a sensor. The decision on what exact signal to report in a public dataset should not be decided in advance. Moreover, public matchup datasets need to contain the top of atmosphere light signal to allow tests of different atmospheric correction methods, or direct inversion of this signal as it has been shown in this thesis.

8.4. Chapter 5: Modelled datasets

Data availability was suspected to be the main limiting factor when developing the algorithm in Chapter 3. Data was approximately normally distributed on a log scale, with relatively small numbers of low and high concentrations compared to central values. At the same time, network performance was poorest at both extremes of concentration. Chapter 5 tried to answer that question by creating a modelled dataset containing the main water constituents altering the light signal, chlorophyll, coloured dissolved organic matter (CDOM) and mineral suspended sediments (MSS). This modelled dataset is a case 1 and 2 matchup dataset created using a bio-optical model developed with *in situ* optical data. Having access to all three main constituents (Chl, CDOM and MSS) and flexibility when designing the dataset is a strength to conduct further tests that are not currently possible with real data. Two datasets were created, one with a log normal distribution, one with a log flat distribution. Neural networks were created for both datasets (Section 5.3.2) and returned almost perfect estimations for all three constituents (Figure 5.8). In order to better simulate real world conditions, noise coming from either radiometric measurement or *in situ* sampling methods was simulated (Figure 5.9). Chlorophyll results were poorer for the flat distribution which had access to more training data at low and high values. The reason for poor chlorophyll retrieval at low concentrations in this case can be attributed to vanishingly small contribution by chlorophyll to the reflectance signal when chlorophyll concentration is low and concentration of either sediment or CDOM is very high. When high values of sediments or dissolved matter are present, making chlorophyll estimations for such samples is not realistic. Moreover, these two datasets highlighted the fact that comparison between datasets is not trivial. Using the exact same method, results were significantly different for chlorophyll estimations, because extreme examples were present more often for one of the datasets.

While overestimations at low values were also observed for real data, underestimations at high values were specific to the real data and not observed for the modelled one. Natural noise that

would be present in a remote sensing dataset was not simulated, such as foam induced by waves, different aerosols or presence of thin clouds, each of which could be a candidate for failure at high values. Land adjacency or shallow waters also distort the light signal quality. High Chl values (>10 mg m^{-3}) are usually found near the coast when high concentrations of nutrients are present, in eutrophic conditions. Errors added to the *in situ* measurements are still subject to be underestimated, and adding a 20% standard deviation (std) noise may not be enough to be fully representative of a real *in situ* dataset. Increasing the noise did not create underestimations at low values, which therefore questions the quality of the *in situ* samples used in Chapters 3 and 4. Different institutions using different methods collected these samples and 20% std errors is probably underestimating the real noise. Interestingly, the performances obtained for the normal dataset were highly linked to the noise added to *in situ* samples, and an MAE of 1.2 or 1.4 was calculated when std noise of 20 or 40% were added to the modelled *in situ* data for chlorophyll and the sediments. The noise of the *in situ* samples is probably the first limiting factor in terms of performance achieved, supported by the observation that non noisy data were almost perfectly estimated using neural networks. Multi-task learning methods harmed the performance achieved and are not recommended for an ocean colour problem, at least with a modelled dataset (Section 5.3.3). The number of data available from constituent specific matchup datasets with only Chl or MSS is higher than triple constituent matchup datasets, as it will always require multiple methods to be available during a cruise campaign. Hence, developing constituent specific algorithms in the future will be much easier by having access to more data compared to triple constituent matchup datasets, when data availability is a limiting factor. Other machine learning fields are using multi-task learning (Collobert and Weston, 2008; Deng et al., 2013; Girshick, 2015; Ramsundar et al., 2015). However, multi-task learning should not be abandoned yet, as it has only been applied to a modelled dataset. Further tests applied to real datasets such as the one proposed by Pahlevan et al. (2022) are required prior to making a final decision on how useful it could be, or if specific configurations of multi-task learning can lead to retrieval improvements.

The future of light measurements from either satellites or *in situ* is slowly switching from multi-spectral to hyperspectral. Multi-spectral sensors require decisions to select specific bands when they are designed, which relies on previously observed relationships between light and water constituents. Only a few sensors such as MERIS or OLCI respectively carried by the European Envisat or Sentinel-3 satellites come with increased resolution in the red which gives access to development of coastal dedicated algorithms. Past and current hyperspectral sensors provide a high spatial resolution (60-100 m), and were available over a short period. These conditions do not allow development of global algorithms like the OCx ones due to a reduced access to potential matchup datasets that would require more data than conventional multi-spectral algorithms, or a specific

feature selection method (Zhao and Du, 2016). Having access to a hyperspectral signal was tested with the modelled dataset (Section 5.3.4). The general observation is that including more information by using more bands improves performance, though improvement appears to plateau above ~7–10 wavebands. A MODIS-Aqua like algorithm which contained 13 bands in the visible performed similarly to hyperspectral models using 20 bands, suggesting that careful selection of input wavebands remains beneficial. While the results may not support the need for a hyperspectral dataset at first glance because it reaches a plateau, we should also consider the potential to deal with sources of noise that may be present in real world data sets. Alteration of the light signal by external sources is the main reason why the atmospheric correction method exist in remote sensing, and each contributor can influence a different part of the spectrum differently. When hyperspectral sensors and matchup datasets become available in future, there is a good expectation that more spectral information will help to deal with noise, similarly to what has been observed in Chapter 4 (Figure 4.12) and 5 (Figure 5.11).

8.5. Chapter 6: Sea surface temperature (SST)

Whilst chlorophyll estimations from space are known to be difficult in coastal waters, there is no particular oceanic area that causes troubles for temperature (see Luo et al., 2019, Figure 1). Strong relationship exists between long wave infrared bands (LWIR) and temperature of the surface on Earth. The different issues for temperature retrieval comes mainly from interactions with the atmosphere that absorbs the signal based on its water and aerosol content (Deschamps and Phulpin, 1980) or the presence of sun glint (Kay et al., 2009). The current algorithms in use by NASA rely on classification, are day or night specific, and each algorithm uses a different sets of bands. Bands near 11,000 nm are used during the day due to their robustness against glint, one of the main issue for SST retrievals. Bands near 4000 nm are used during the night thanks to their higher sensitivity when the temperature changes. For MODIS Aqua, these state-of-the-art algorithms also rely on more information such as the latitude, linked to the atmosphere thickness; the sensor zenith angle; the mirror of the sensor used or the month of the year (Kilpatrick et al., 2014). Both return good performances in operational conditions (Section 6.3.1.1, Figure 6.4). In Chapter 6 and similarly to Chapter 4 for chlorophyll, different input combinations have been tested to find an optimal SST algorithm using neural networks (Table 6.1). Two datasets were available, a global one (Section 6.3.2.2+), openly accessible provided by NASA and a regional one (Section 6.3.2.1) created with the data processed during the thesis for ocean colour. They both validated the previous observation that using more spectral light information improves performance, subject to inclusion of the bands normally used for SST observations (Table 6.2). An algorithm using only visible wavebands returned poorer performance compared to algorithms

with access to long wave infrared bands (Figure 6.7f). There is little prospect to retrieve accurate temperature when the long wave infrared bands are not available for a sensor.

When neural networks were trained using the same information currently used by operational SST algorithms, they did not outperform the empirical algorithms (Sections 6.3.1 and 6.3.2, Figures 6.7 to 6.11, to be compared with Figure 6.4). This behaviour has been observed twice, for chlorophyll in open waters in Chapter 4 (Sections 4.5.1 and 4.5.2, Figures 4.13 and 4.16), and for SST in Chapter 6. This observation supports the general idea to test inclusion of more information linked to the desired variable or potential sources of noise when designing neural network algorithm to make improved retrievals of a desired variable from space. Compared to what is currently applied to the operational algorithm, inclusion of latitude did not contribute to improved retrievals. An algorithm developed with data collected during the 2002-2003 period did produce the same performances as an algorithm developed with any data available. Inclusion of more information, especially if not spectral, is not trivial when different metrics behave oppositely. Inclusion of overcorrected intermediate bands (LWIR bands between 6000 to 8000 nm) or solar angles did help reach lower mean absolute differences (MAD) but at the same time produced inconsistent estimations for a few samples (temperature above 40°C), and RMSD was much higher than networks not including these additional data (Figures 6.1. MAD is known to be robust against outliers and RMSD sensitive to it. It makes the decision of to whether or not include these bands harder, as MAD is clearly improved when including such information. The fact that RMSD increases may highlight some kind of overfitting, but the failure concern only hundreds of samples over the several million matchups available. Achieving better statistical metrics should not be the only target, and being able to visualise realistic oceanic surface features should be considered, though we do not have any mathematical way to validate such observation at the moment. Visualising images that would create suspicious features and comparing it with images produced by other algorithms is still the main tool. When the dataset is big enough, for example for SST algorithm when it reaches millions of samples, generalisation is not an issue as most data are independent from the training dataset (90 to 98.5% independent data used in this study). In the end, a quadruple band algorithm including the two 4000, 11,000 bands and the sensor zenith angle was found to be the best compromise between performance and overfitting issues. The amount of photons emitted at infrared bands is directly linked to the sensor zenith angle, which explains how this information can contribute to improve performances. This algorithm outperformed the current algorithms in use during day or night separately and provided consistent estimations over the entire Earth. An additional version of the algorithm using the same inputs was developed using both day and night data at the same time rather than separately, to become the first general SST algorithm that would make appropriate estimate of temperature for any water surface. Performances achieved were

also better than current state of the art algorithms, but poorer than individual day or night NN algorithms (on average over the entire dataset, Section 6.3.3). Keeping both version could be more appropriate in terms of pure performance, but using a single algorithm may be more useful, especially for end users.

A further tests included the latitude as an additional input as it is already done by operational algorithms, and did not show any improvement (not shown). Finally, there is some general criticism towards neural networks that they manage to outperform other methods over the data they have access too for the training, but tend to overfit and perform poorer on new data. Whilst this has not been seen in this Chapter, temperature allowed us to challenge this hypothesis. An algorithm was trained using 2002-2003 data only, and applied to the 2004+ remaining data (6 million samples). This algorithms produced the exact same performances (Section 6.3.4), which means that the NNs created following this method are global algorithms. This possibility arise from the incredible data availability due to the *in situ* sampling method (automatic buoys everywhere on Earth). Creations of maps (Section 6.3.6) shown where the actual improvement are expected by the different algorithms developed in this study. The general answer is haze or near clouds, which tend to produce lower SST due to the increased absorption by water from the atmosphere which leads to underestimates by operational algorithms. Training an algorithm with noisy data helped making more accurate estimates onto pixels that did show unusual water front, linked to the clouds. Algorithms using visible bands (Figure 6.15c) produced more realistic estimates where clouds edges may either absorb more infrared light or backscatter more towards the sensor.

Inclusion of visible bands is also a possibility to make improved temperature retrievals and reduce impact of other contributors to the light signal, haze and glint. However, this information is not currently available for most datasets, especially at top-of-atmosphere, and only a few bands linked to aerosol contents are considered presently. For the regional algorithm, inclusion of all MODIS Aqua visible bands helped the network to make better estimations over noisy matchups. An algorithm using only visible bands was tested (Figure 6.7f), because small satellite sensors typically do not contain long wave infrared bands used to measure temperature but do contain visible bands. Having a working algorithm with cheap sensors would increase the coverage of temperature measurements. It is possible to make low quality estimations of SST using visible bands only, which is relevant only if the smaller sensor comes with a higher resolution than 1 km. An extra test was conducted in Chapter 6 with the estimation of salinity (Section 6.3.3). Compared to literature, it was not possible to achieve a functional salinity algorithm for MODIS Aqua. For salinity, the dataset is not evenly distributed. 20,000 samples are available and more than 90% of the data fall between 34 and 35 PSU, the average value for sea water. Other studies usually rely multiple regression models (Wong et al., 2007; Marghany et

al., 2010). Following the general method used for chlorophyll or temperature, there is only a little expectation that salinity can be estimated from space using visible or infrared sensors, at least using the current multi-spectral sensors available. Better salinity retrievals may require higher resolutions and bands near 400 nm to produce coherent estimates. Microwave bands seem mandatory to make good quality salinity estimations. Surprisingly, including salinity as an input for temperature retrievals improved performance obtained using the visible and infrared algorithm at a regional scale. Therefore, there is still extra motivation to find a method to develop salinity retrieval from space at higher spatial resolution than what is currently achieved with microwave sensors.

8.6. Chapter 7: Eutrophication assessment

Both ocean surface chlorophyll and temperature can be better estimated with MODIS Aqua sensor by applying the method developed in this thesis, and the approach can be extended to any dataset and sensor available. Chlorophyll and temperature are the two main variables impacting the eutrophication status of the North Sea (Desmit et al., 2019). Having access to more accurate remote sensing algorithms is the first step to a future enhanced assessment of eutrophication by the environmental agencies. Current eutrophication reports are based on *in situ* observations, and it has been shown in Chapter 7 that switching from *in situ* to remote sensing can improve data availability by five orders of magnitude during a single month, May 2003 (Figure 7.7, Section 7.3.2). Most of the OSPAR areas are not sampled continuously every year (Figure 7.3 and 7.4), and remote sensing seems to be one of the only method to provide observational information over time until more efforts are deployed to sample the ocean *in situ*. Remote sensing is just about to be included in the process using the JMP-EUNOSAT algorithm (Van Der Zande et al., 2019) for the future OSPAR report in 2023. Inclusion of remote sensing products directly tackle the data availability problem. For the next Common Procedure analysis made by OSPAR, *in situ* and remote sensing chlorophylls will be evenly weighted in the 2023 report (a single chlorophyll is created for each region, which is the mean between all *in situ* and all remote sensing chlorophyll concentration). Whilst this is a good start, a problem may arise when a region is sampled only several times during the entire growing season and compared with millions of satellite pixels. How meaningful is it to give 50% of a signal to a 5 *in situ* values, while the remote sensing product have seen the whole region?

It is not clear how chlorophyll concentrations evolved during the past 20 years using the available *in situ* data for the different OSPAR regions. The satellite data are still being processed to create daily chlorophyll maps with the algorithms developed in Chapter 3 when writing the thesis. It has been shown in the past that remote sensing products can lead to opposite conclusions (Dierssen et al., 2010), and the ocean phytoplankton stocks are either increasing (Antoine et al., 2005) or

decreasing (Gregg et al., 2005). The current assessment method is still evolving after each report and new scientific discoveries, and discussions are held by the OSPAR committee to make appropriate decisions. The future OSPAR assessment promotes the use of the JMP- EUNOSAT algorithm based on assumptions such as "MODIS Aqua cannot make accurate estimations in coastal waters due to the lack of red wavebands to perform advanced". While this is true and no real purely red algorithm exists for MODIS, the neural networks developed in this thesis use the entire spectrum to overcome the issue, and MODIS can now produce better estimates in coastal waters. The choice of an appropriate algorithm was the limiting factor, not the sensor used in this case. Other factors are also under discussion, for example should the P90 or the median of chlorophyll be used, which one is the most relevant? With improving performance achieved by algorithms in coastal waters, the current assessment method is expected to evolve in future to benefit fully from information available at a higher spatial and temporal scale and help policymakers in their decision process.

The work conducted in this thesis is exploratory and presents a different path to observe the ocean. Current methods for retrieval of chlorophyll a and temperature use classification as their initial step prior to algorithm development. When I applied for this thesis project, the original title was "Automate classification of ocean colour remote sensing ...". It makes perfect sense to propose a classification thesis project for ocean colour in coastal waters, as it is the main method to deal with the complexity of the signal if we use our understanding of the different physical relationships based on observed band ratios. It is much easier to build an algorithm when the conditions are very similar, for example during the day or night for temperature, or in open water or coastal waters only for chlorophyll. The reason behind that comes from the major difficulty for physical approaches to build bigger algorithms that can retrieve complex nonlinear signal interactions. Understanding how a specific part of a light signal evolves when a specific noise or feature is present is simple in most cases, even if the relationship is not linear. Observing how the full spectrum behaves is more difficult, especially if multiple effects occur at the exact same time. These observations probably explain why neural networks are successful among different fields, due to their capacity to deal with complex signal interactions that we cannot mentally visualise or distinguish in a mixture.

Neural networks come with their own flaws, and the first limitation to develop such techniques are the data availability and quality. Openly available chlorophyll (or other constituents) datasets are the main limiting factor to develop neural networks algorithms, as only restricted numbers of available *in situ* samples are available. Moreover, I had to personally ask some of the data providers because the data were not shared publicly on a public platform or provided by an organisation such as ICES or CMEMS. There is a growing interest to make any data available on public platforms similarly to what ICES or CMEMS currently do for European waters. When the datasets exist

(SeaBASS for example), they limit development of new algorithms by 1) not providing the raw signal at top of atmosphere, which would give freedom to the atmospheric correction; and 2) not providing all information available, such as all available wavebands or ancillary data a sensor has access to. The decisions when datasets are built are based on previously observed relationships and scientific confidence of what is useful and what is not. This overconfident approach makes the test new hypothesis difficult and is counterproductive. As an example, inclusion of SWIR bands used for coastal atmospheric corrections did not show massive increase in chlorophyll a retrieval performance in Chapter 4. However, this is not a reason why they should not be part of future datasets, as these bands could help estimations for more turbid waters, if they were under represented in our datasets. Being limited by the data availability during this thesis directly lead to the use of a very permissive approach in Chapter 3, not applying flags to other contributors to the light signal. Data processing capacities was another limiting factors during this thesis, and data from a single sensor were processed over the European northwest shelf seas, with only 15 of the 36 bands it provides. Nonetheless, this was enough to show the capacities of neural networks to overcome current issues.

Main take home message for each Chapters

- Chapter 3: It is possible to invert directly the top of atmosphere signal to make estimates of water related variables, which bypass the need for an atmospheric correction scheme.
- Chapter 4: The use of the entire light spectrum returns the best performing chlorophyll algorithms.
- Chapter 5: A flatter data distribution is not helping estimates at the edges, and the benefits of having access to a hyperspectral signal over a multispectral one may be limited and produce only slight improvements in future for the three main constituents, Chl, CDOM and MSS.
- Chapter 6: Estimating temperature directly from infrared bands for both day and night gives accurate estimates without inclusion of other information (location, time, angles), and a machine learning approach outperforms the physics one.
- Chapter 7: A pure remote sensing approach to observe water variables greatly improves the observational capabilities for the different OSPAR regions by several orders of magnitude with equivalent performances.

Perspective

The next step following the work initiated in this thesis would be to reproduce the development of neural networks algorithms at a global scale, initially using any data available for a single sensor like MODIS Aqua. It would require a global *in situ* dataset which has not yet been assembled. *In situ* data are not easy to access worldwide, as only few groups make it freely available in Europe (ICES, CMEMS) or in developed countries (USA, Australia, etc.). More data are available but require direct contact with the producer, which is a limiting factor, and a single platform where any collected data would be centralised is the most efficient method. It would require more efforts to access data from countries without an operational ocean observation network, with only scarce data from different collaborations and sporadic cruise campaigns. Data availability would highly benefit from applying the same technique as currently used for salinity and temperature, relying on drifting buoys making automatic measurements which are more cost effective. Such a network already partially exists, though with significantly fewer buoys. The BGC-ARGO floats measure chlorophyll fluorescence (and other biogeochemical variables) at different depths (live map accessible at: https://maps.biogeochemical-argo.com/bgcargo/, accessed on 11 July 2022). This is a good start for data quantity improvement especially in case 1 waters, but *in situ* fluorescence measurements are not of the same quality as extracted chlorophyll from spectrophotometry or chromatography. Moreover, estimating Chlorophyll in coastal waters requires non drifting buoys. It requires constant sampling efforts such as is currently done by Marine Scotland Science at Loch Ewe and Stonehaven coastal stations, with a sample collected each week.

The future of Earth observation is hyperspectral radiometers. It has been shown in three distinct chapters that more light information provides enhanced performance thanks to the use of neural networks. Therefore, switching from multi-spectral to hyperspectral should further improve the overall accuracy, tough our results suggest there is an upper end to the benefits of additional wavebands. The initial objective of hyperspectral sensors for ocean colour is meant to improve the development of phytoplankton functional group specific algorithms. Having access to a higher resolved spectral signal will provide information from pigments that are currently not “visible”. Chlorophyll a is only the main pigment directly linked to primary production. Other pigments exist, and the community is slowly switching from chlorophyll a to pigment specific algorithms that hint at the presence of phytoplankton species that can cause damage, such as red tides in Florida (Kirkpatrick et al., 2004) or diatom blooms on the Californian coast (Lefebvre et al., 1999). Algorithms to identify specific groups already exists, when the group signal signature is different enough and impacts available bands (Nair et al., 2008). Some species such as the diatom *Pseudo-nitzschia* produce toxins

that can be deadly for higher trophic levels and humans (Fire et al., 2010). However, it should be noted that the spectral signature of toxin-producing species is generally indistinguishable from other species with similar pigment compositions. A coupled approach between ocean observation from space and *in situ* automatic monitoring is probably the future for ocean monitoring. Chlorophyll a is the main pigment researchers are currently interested in due to its link to primary production and because it is the main indicator to quantify the eutrophication status of an area. The method used to assess the eutrophication is not trivial, and multiple factors can be included in the process. While the main reason for eutrophication is anthropogenic and comes from increased quantities of nutrient released, the first direct observable effect is an increased algal biomass. Using satellite products greatly increases data availability where current sampling effort may not access daily. The quality of retrieval from space is getting closer to the quality of *in situ* sampling, and this thesis is a good example. There is a need to further characterize the uncertainty of *in situ* data, as it is currently considered as the ground "truth", and yet it is subject to considerable and variable uncertainty in practise (Jaccard et al., 2018). The debate on how remotely sensed chlorophyll compares to *in situ* observations needs to take proper account of the uncertainty in the *in situ* data, a feature that is currently poorly understood and generally underestimated. The remote sensing products may be closer to the reality than is currently believed. The general method for the assessment of eutrophication conducted by OSPAR relies on specific regions and will continue to evolve after each report. With improved data availability from space, the temporal and spatial area definition may need to evolve to fully benefit from inclusion of satellite imagery.

Assessing water eutrophication is one of the problems a more accurate remote sensing algorithm can help to contribute to. The other main outcome concerns climate change in general. Chlorophyll a is used as a proxy of biomass or primary production (Ryther and Yentsch, 1957), which is afterwards used to build climate models. Being able to evaluate a more accurate proportion of these components is decisive under the current global assessment. Whilst coastal waters represent approximately 7% of the ocean surface (Gattuso et al., 1998), in some regions they contribute to 25-30% of the total primary productivity (Ducklow et al., 2001; Muller-Karger et al., 2005; Salgado-Hernanz et al., 2022). In one of the most cited works concerning primary production on Earth, oceanic primary production is estimated using both chlorophyll and temperature measured from satellites (Field et al., 1997). It has been shown in Chapter 3 and the community is aware that old chlorophyll algorithms such as OC3 clearly overestimate the signal in coastal turbid waters. Yet, there is generally agreed global chlorophyll product. It would be interesting to repeat the primary production assessment while applying a more appropriate algorithm that works everywhere, such as a global neural network. The ocean absorbs 30% of anthropogenic emitted CO_2 (Gruber et al., 2007), which

has a direct impact on its chemical composition and decreases the pH of the water. Predicting how species will adapt in an acidifying ocean is crucial (Sunday et al., 2014). The same question stands for the temperature (Thomas et al., 2012). The temperature is probably the most used variable to quantify the anthropogenic impact on the climate, as it is a direct consequence of greenhouse gases enrichment of the atmosphere. Having access to accurate ocean temperature estimations where *in situ* buoys are not present is crucial to understand where and how the ocean is changing.

The approach developed in this thesis is currently limited to the MODIS Aqua sensor, which is probably the best sensor available at the moment for developing machine learning algorithms due to its longevity and correspondingly large numbers of match up data available for algorithm training. It has been shown that there is no need for a distinct atmospheric correction scheme for this coastal dataset at least (Chapter 3). Extending the test to a global dataset would be one of the next experiments to conduct. It would be interesting to test inclusion of atmospheric constituents that are known to alter the light signal (information about meteorological conditions, ozone concentration and the surface of the water, De Vis et al., 2022), to evaluate if they improve performance, or if a top of atmosphere raw signal performs similarly. The MODIS calibration team managed to maintain high quality signals from most bands for more than 20 years, which is also part of the success. The Aqua spacecraft still has enough fuel to correct its trajectory in 2022, but will run out of fuel in the relatively near future and will be deactivated. Therefore, the algorithm developed in this thesis will no longer be applicable to new data as it cannot be easily transferred to a different sensor. It is a good start to have access to more than 20 years of satellite imagery that provide better chlorophyll estimations in coastal waters. However, there is a crucial need to apply the same technique to other more recent sensors with access to higher spatial resolution, such as the OLCI sensors carried by Sentinel-3 satellites that are planned to be available until 2036. However, as the OLCI sensors were launched in 2016 and 2017, there has been insufficient time to accumulate enough match-ups to support development of an equivalent neural network algorithm for those instruments. An *in situ* sample usually requires a few years after it has been collected prior being available on public datasets. Being able to transfer the knowledge earned from MODIS experience to other sensors would be the most important tool to access. OCx algorithms have the advantage of knowing the physical relationships between different bands, and only require a few datasets to be fine-tuned when applied to a new sensor that contains the same wavebands. The MODIS Aqua algorithm used here contains 15 bands because it has been shown that best performance are achieved when all information is included. The time required to assemble a sensor-specific training data set is a general issue for the development of machine learning algorithms for ocean colour applications and should be a stimulus for developing waveband transfer approaches that permit emulation of standard waveband sets from one sensor to

another. This would allow a ML developed using MODIS data to be applied to OLCI data that had been transformed in MODIS-equivalent wavebands.

Three methods come to my mind to transfer the knowledge learned by MODIS. First, the most common technique used in the deep learning field would consist of fine-tuning weights of neural networks and apply them to another sensor, this approach is used for other complex tasks such as image classification and is called transfer learning (Pan et al., 2008; Weiss et al., 2016). The second possibility would require the creation of a merged satellite product including contributions from any ocean colour sensors available, similarly to the OC-CCI product. Only 6 bands are available in the current OC–CCI dataset, despite more bands being available but not used for algorithm design. Such products would greatly benefit from having different versions at top and bottom of atmosphere to allow exploration of different atmospheric correction schemes. There is a risk of producing unphysical light signals when merging images taken at different times due to the movement of the sea and clouds, which would probably require more quality controls. The third method is the easiest one, it requires training current available sensors on MODIS Aqua chlorophyll a images directly as a way to tackle the data availability. It was briefly tested during this PhD (not shown) and training an algorithm using OC3 data from the open ocean helped the algorithm to reach lower values initially missing from the training set. It became unnecessary as soon as Norwegian *in situ* samples were available. This last method allows creation of an algorithm right after several images of a new sensor are available, along with a chlorophyll a product from another sensor, even before the atmospheric correction process of the new sensor becomes operational. It can be used for both Case 1 and Case 2 waters. This is a very short time solution that cannot produce estimates better than the original algorithm used for, and would not benefit addition of the potential improved spectral information / resolution of the new sensor. The quality achieved needs real *in situ* data to be assessed and improved, which means that continuous efforts are required to sample the ocean consistently. Eventually when the sensor would reach a mature stage, data from the old sensors could be slowly replaced by real data, constantly increasing. Each of these methods are independent and could also be coupled.

For the temperature, the future is probably different. Data quantity, with millions of available points, and quality, with 0.01K difference, are not a limiting factor for temperature algorithm design and high-quality global datasets are available. The improvement of temperature retrieval conducted in this work came from inclusion of useful information that help deal with noise mostly. Inclusion of visible bands was shown to produce more accurate retrievals for the European dataset. Their inclusion in the SeaBASS dataset at the top of atmosphere would probably be the next step. Algorithms using more information related to the light signal perform better, and slight improvements can probably still be made in temperature retrieval. Moreover, the brightness temperatures used are sensitive to

atmospheric correction, and it has been observed that it causes poorer estimations for ocean colour. The temperature field would benefit from such a study, and a joint bottom and top of atmosphere signal including all bands available by a sensor should be prioritized when building a dataset. Currently, most limits come from the nature of the datasets. Comparing a 5x5 1km² pixel box with a few mL of water where the temperature was measured by a buoy is clearly a limiting factor, and could be improved by increasing the spatial resolution from satellites. The depth provenance of the data at 20 cm below the sea level in average is known to be an issue too, as the physical signals comes from the very thin layer of the ocean, a few millimetres. The temporal difference in the dataset is also a candidate. A lot can happen in 30 minutes for a water body, such as advection or vertical mixing, and appropriate synchronisation between the buoy and the different satellite would be helpful. Eventually, the noise coming from different interactions of the photons prior to reaching the sensor is not always solvable, and while neural networks seem to be able to manage most parts, sometimes the signal is just not correlated to the observed variable. These observations are common to Earth observation in general. Remote sensing is a fantastic tool to observe what happens at a global scale, in two dimensions. Yet, the ocean is more complex than a surface. There is a need to understand what happens below the surface. This supports the need for continuous and extended network observation of the ocean, at multiple depths, positions and continuously, to be able to fully understand and predict how human activity is altering the environment.

References

Alvera-Azcárate, A., Barth, A., Sirjacobs, D. and Beckers, J.M., 2009. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Science*, *5*(4), pp.475-485.

Antoine, D., Morel, A., Gordon, H.R., Banzon, V.F. and Evans, R.H., 2005. Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *Journal of Geophysical Research: Oceans*, *110*(C6).

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

Ahmad, Z., Franz, B.A., McClain, C.R., Kwiatkowska, E.J., Werdell, J., Shettle, E.P. and Holben, B.N., 2010. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans. Applied optics, 49(29), pp.5545-5560.

Ardyna, M. and Arrigo, K.R., 2020. Phytoplankton dynamics in a changing Arctic Ocean. *Nature Climate Change*, *10*(10), pp.892-903.

Atkinson, A., Harmer, R.A., Widdicombe, C.E., McEvoy, A.J., Smyth, T.J., Cummings, D.G., Somerfield, P.J., Maud, J.L. and McConville, K., 2015. Questioning the role of phenology shifts and trophic mismatching in a planktonic food web. Progress in Oceanography, 137, pp.498-512.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A. and Thingstad, F., 1983. The ecological role of water-column microbes in the sea. Marine ecology progress series, pp.257-263.

Babin, M., Stramski, D., Ferrari, G.M., Claustre, H., Bricaud, A., Obolensky, G. and Hoepffner, N., 2003. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. Journal of Geophysical Research: Oceans, 108(C7).

Bailey, S.W. and 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), pp.12-23.

Bailey, S.W., Franz, B.A. and Werdell, P.J., 2010. Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. Optics express, 18(7), pp.7521-7527.

Barber, J., Malkin, S. and Telfer, A., 1989. The origin of chlorophyll fluorescence in vivo and its quenching by the photosystem II reaction centre. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 323(1216), pp.227-239.

Barcelos e Ramos, J., Schulz, K.G., Voss, M., Narciso, Á., Müller, M.N., Reis, F.V., Cachão, M. and Azevedo, E.B., 2017. Nutrient-specific responses of a phytoplankton community: a case study of the North Atlantic Gyre, Azores. Journal of Plankton Research, 39(4), pp.744-761.

Bardeen, J. and Brattain, W.H., 1948. The transistor, a semi-conductor triode. Physical Review, 74(2), p.230.

Barnes, W.L., Pagano, T.S. and Salomonson, V.V., 1998. Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. IEEE Transactions on Geoscience and Remote Sensing, 36(4), pp.1088-1100.

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

Beaugrand, G. and Reid, P.C., 2003. Long-term changes in phytoplankton, zooplankton and salmon related to climate. *Global Change Biology*, *9*(6), pp.801-817.

Benardos, P.G. and Vosniakos, G.C., 2007. Optimizing feedforward artificial neural network architecture. Engineering applications of artificial intelligence, 20(3), pp.365-382.

Benediktsson, J.A., Swain, P.H. and Ersoy, O.K., 1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. International Journal of Remote Sensing, 14(15), pp.2883-2903.

Bengil, F., McKee, D., Beşiktepe, S.T., Calzado, V.S. and Trees, C., 2016. A bio-optical model for integration into ecosystem models for the Ligurian Sea. *Progress in Oceanography*, *149*, pp.1-15.

Berge, T., Daugbjerg, N., Andersen, B.B. and Hansen, P.J., 2010. Effect of lowered pH on marine phytoplankton growth rates. Marine Ecology Progress Series, 416, pp.79-91.

van Beusekom, J.E., 2018. Eutrophication. In *Handbook on Marine Environment Protection* (pp. 429-445). Springer, Cham.

van Beusekom, J.E., Loebl, M. and Martens, P., 2009. Distant riverine nutrient supply and local temperature drive the long-term phytoplankton development in a temperate coastal basin. *Journal of Sea Research*, *61*(1-2), pp.26-33.

Berry, J., 2018. 3.10 solar induced chlorophyll fluorescence: Origins, relation to photosynthesis and retrieval. Comprehensive remote sensing, 3, pp.143-162.

Björn, L.O., 2009. The evolution of photosynthesis and chloroplasts. Current Science, pp.1466-1474.

Boesch, D.F., 2002. Challenges and opportunities for science in reducing nutrient over-enrichment of coastal ecosystems. *Estuaries*, *25*(4), pp.886-900.

Bouali, M. and Ladjal, S., 2011. Toward optimal destriping of MODIS data using a unidirectional variational model. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(8), pp.2924-2935.

Boyce, D.G., Lewis, M. and Worm, B., 2012. Integrating global chlorophyll data from 1890 to 2010. Limnology and Oceanography: Methods, 10(11), pp.840-852.

Boyle, E.A. and Keigwin, L., 1987. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature*, *330*(6143), pp.35-40.

Bresnan, E., Cook, K.B., Hughes, S.L., Hay, S.J., Smith, K., Walsham, P. and Webster, L., 2015. Seasonality of the plankton community at an east and west coast monitoring site in Scottish waters. Journal of Sea Research, 105, pp.16-29.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S., 2020. Language models are few-shot learners. Advances in neural information processing systems, 33, pp.1877-1901.

Bricaud, A., Morel, A. and Prieur, L., 1981. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. Limnol. Oceanogr, 26(1), pp.43-53.

Bricaud, A., Thiria, S., Chazottes, A., Crépon, M. and Ras, J., 2007. Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method Astico View project ESA SEOM project SynSenPFT View project Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S. and Ruescas, A., 2016, August. 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. In Living Planet Symposium (Vol. 740, p. 54).

Brody, S.S. and Rabinowitch, E., 1957. Excitation lifetime of photosynthetic pigments in vitro and in vivo. Science, 125(3247), pp.555-555.

Buckton, D., O'mongain, E.O.N. and Danaher, S., 1999. The use of neural networks for the estimation of oceanic constituents based on the MERIS instrument. International journal of remote sensing, 20(9), pp.1841-1851.

Bulgarelli, B. and Zibordi, G., 2018a. On the detectability of adjacency effects in ocean color remote sensing of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI. *Remote sensing of environment*, *209*, pp.423-438.

Bulgarelli, B. and Zibordi, G., 2018b. Seasonal impact of adjacency effects on ocean color radiometry at the AAOT validation site. *IEEE Geoscience and Remote Sensing Letters*, *15*(4), pp.488-492.

Bulgarelli, B. and Zibordi, G., 2020. Adjacency radiance around a small island: implications for system vicarious calibrations. *Applied Optics*, *59*(10), pp.C63-C69.

Burson, A., Stomp, M., Akil, L., Brussaard, C.P. and Huisman, J., 2016. Unbalanced reduction of nutrient loads has created an offshore gradient from phosphorus to nitrogen limitation in the N orth S ea. *Limnology and Oceanography*, *61*(3), pp.869-888.

Cannizzaro, J.P. and Carder, K.L., 2006. Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters. Remote Sensing of Environment, 101(1), pp.13-24.

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M. and Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sensing of Environment, 248, p.111974.

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

Carpenter, K.E., Abrar, M., Aeby, G., Aronson, R.B., Banks, S., Bruckner, A., Chiriboga, A., Cortés, J., Delbeek, J.C., DeVantier, L. and Edgar, G.J., 2008. One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science*, *321*(5888), pp.560-563.

Caruana, R., 1997. Multitask learning. *Machine learning*, *28*(1), pp.41-75.

Cazzaniga, I., Zibordi, G. and Mélin, F., 2021. Spectral variations of the remote sensing reflectance during coccolithophore blooms in the Western Black Sea. *Remote Sensing of Environment*, *264*, p.112607.

Charlier, R.H., Morand, P., Finkl, C.W. and Thys, A.C., 2007. Dealing with green tides on Brittany and Florida coasts.

Charlier, R.H., Morand, P. and Finkl, C.W., 2008. How Brittany and Florida coasts cope with green tides. *International Journal of Environmental Studies*, *65*(2), pp.191-208.

Chen, B. and Liu, H., 2010. Relationships between phytoplankton growth and cell size in surface oceans: Interactive effects of temperature, nutrients, and grazing. Limnology and Oceanography, 55(3), pp.965-972.

Chen, J., Quan, W., Cui, T., Song, Q. and Lin, C., 2014. Remote sensing of absorption and scattering coefficient using neural network model: development, validation, and application. Remote Sensing of Environment, 149, pp.213-226.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S. and Schuh, P., 2022. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311.

Clark, P.U., Pisias, N.G., Stocker, T.F. and Weaver, A.J., 2002. The role of the thermohaline circulation in abrupt climate change. *Nature*, *415*(6874), pp.863-869.

Clarke, G.L., Ewing, G.C. and Lorenzen, C.J., 1970. Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration. *Science*, *167*(3921), pp.1119-1121.

Claustre, H., Hooker, S.B., Van Heukelem, L., Berthon, J.F., Barlow, R., Ras, J., Sessions, H., Targa, C., Thomas, C.S., van der Linde, D. and Marty, J.C., 2004. An intercomparison of HPLC phytoplankton pigment methods using *in situ* samples: Application to remote sensing and database activities. Marine Chemistry, 85(1-2), pp.41-61.

Clevers, J.G., De Jong, S.M., Epema, G.F., Van Der Meer, F., Bakker, W.H., Skidmore, A.K. and Addink, E.A., 2001. MERIS and the red-edge position. *International Journal of Applied Earth Observation and Geoinformation*, *3*(4), pp.313-320.

Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., Lucas, R., Armston, J. and Moghaddam, M., 2014. A python-based open source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables. Remote Sensing, 6(7), pp.6111-6135.

Collobert, R. and Weston, J., 2008, July. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160-167).

Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Havens, K.E., Lancelot, C. and Likens, G.E., 2009. Controlling eutrophication: nitrogen and phosphorus. *Science*, *323*(5917), pp.1014-1015.

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., d'Ortenzio, F., Gentili, B. and Schmechtig, C., 2021. Deep chlorophyll maxima in the global ocean: Occurrences, drivers and characteristics. Global Biogeochemical Cycles, 35(4), p.e2020GB006759.

Corson, M.R., Korwan, D.R., Lucke, R.L., Snyder, W.A. and Davis, C.O., 2008, July. The hyperspectral imager for the coastal ocean (HICO) on the international space station. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 4, pp. IV-101). IEEE.

Cunningham, M. and Regan, M.A., 2015, October. Autonomous vehicles: human factors issues and future research. In Proceedings of the 2015 Australasian Road safety conference (Vol. 14).

D'Alimonte, D. and Zibordi, G., 2003. Phytoplankton determination in an optically complex coastal region using a multilayer perceptron neural network. IEEE Transactions on Geoscience and Remote Sensing, 41(12), pp.2861-2868.

Dall'Olmo, G., Brewin, R.J., Nencioli, F., Organelli, E., Lefering, I., McKee, D., Röttgers, R., Mitchell, C., Boss, E., Bricaud, A. and Tilstone, G., 2017. Determination of the absorption coefficient of chromophoric dissolved organic matter from underway spectrophotometry. *Optics express*, *25*(24), pp.A1079-A1095.

Daniel, T.C., Sharpley, A.N. and Lemunyon, J.L., 1998. Agricultural phosphorus and eutrophication: A symposium overview. *Journal of environmental quality*, *27*(2), pp.251-257.

Darecki, M. and Stramski, D., 2004. An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote sensing of Environment*, *89*(3), pp.326-350.

De Jong, F., 2007. Marine eutrophication in perspective: on the relevance of ecology for environmental policy. Springer Science & Business Media.

De Vis, P., Mélin, F., Hunt, S.E., Morrone, R., Sinclair, M. and Bell, B., 2022. Ancillary Data Uncertainties within the SeaDAS Uncertainty Budget for Ocean Colour Retrievals. *Remote Sensing*, *14*(3), p.497.

Deng, L., Hinton, G. and Kingsbury, B., 2013, May. New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8599-8603). IEEE.

Deschamps, P.Y. and Phulpin, T., 1980. Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 μm. *Boundary-Layer Meteorology*, *18*(2), pp.131-143.

Desmit, X., Nohe, A., Borges, A.V., Prins, T., De Cauwer, K., Lagring, R., Van der Zande, D. and Sabbe, K., 2020. Changes in chlorophyll concentration and phenology in the North Sea in relation to de-eutrophication and sea surface warming. Limnology and Oceanography, 65(4), pp.828-847.

Desortová, B., 1981. Relationship between chlorophyll-α concentration and phytoplankton biomass in several reservoirs in Czechoslovakia. Internationale Revue der gesamten Hydrobiologie und Hydrographie, 66(2), pp.153-169.

Devol, A.H., 1978. Bacterial oxygen uptake kinetics as related to biological processes in oxygen deficient zones of the oceans. Deep Sea Research, 25(2), pp.137-146.

Dierssen, H.M., 2010. Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. *Proceedings of the National Academy of Sciences*, *107*(40), pp.17073-17078.

Doerffer, R. and Schiller, H., 1994, October. Inverse modeling for retrieval of ocean color parameters in Case II coastal waters: an analysis of the minimum error. In Ocean Optics XII (Vol. 2258, pp. 887-893). SPIE.

Doerffer, R. and Schiller, H., 2007. The MERIS Case 2 water algorithm. International Journal of Remote Sensing, 28(3-4), pp.517-535.

Dogliotti, A.I., Ruddick, K.G., Nechad, B., Doxaran, D. and Knaeps, E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. Remote sensing of environment, 156, pp.157-168.

Ducklow, H.W., Smith, D.C., Campbell, L., Landry, M.R., Quinby, H.L., Steward, G.F. and Azam, F., 2001. Heterotrophic bacterioplankton in the Arabian Sea:: Basinwide response to year-round high primary productivity. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(6-7), pp.1303-1323.

Dulière, V., Gypens, N., Lancelot, C., Luyten, P. and Lacroix, G., 2019. Origin of nitrogen in the English Channel and Southern Bight of the North Sea ecosystems. *Hydrobiologia*, *845*(1), pp.13-33.

Duysens, L.N.M. and Sweers, H.E., 1963. Studies on microalgae and photosynthetic bacteria. Jpn Soc Plant Physiol Univ Tokyo Tokyo, 353.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A. and Litchman, E., 2016. Phytoplankton growth and the interaction of light and temperature: A synthesis at the species and community level. Limnology and Oceanography, 61(4), pp.1232-1244.

Einstein, A., 1905. The photoelectric effect. Ann. Phys, 17(132), p.4.

El-Habashi, A., Ahmed, S., Ondrusek, M.E. and Lovko, V., 2019. Analyses of satellite ocean color retrievals show advantage of neural network approaches and algorithms that avoid deep blue bands. Journal of Applied Remote Sensing, 13(2), p.024509.

Falkowski, P. and Kiefer, D.A., 1985. Chlorophyll a fluorescence in phytoplankton: relationship to photosynthesis and biomass. Journal of Plankton Research, 7(5), pp.715-731.

Falkowski, P.G. and Wilson, C., 1992. Phytoplankton productivity in the North Pacific ocean since 1900 and implications for absorption of anthropogenic CO2. Nature, 358(6389), pp.741-743.

Falkowski, P.G. and Raven, J.A., 2013. Aquatic photosynthesis. Princeton University Press.

Fan, Y., Li, W., Gatebe, C.K., Jamet, C., Zibordi, G., Schroeder, T. and Stamnes, K., 2017. Atmospheric correction over coastal waters using multilayer neural networks. Remote Sensing of Environment, 199, pp.218-240.

Ferrari, G.M. and Tassan, S., 1999. A method using chemical oxidation to remove light absorption by phytoplankton pigments. *Journal of Phycology*, *35*(5), pp.1090-1098.

Field, C.B., Behrenfeld, M.J., Randerson, J.T. and Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *science*, *281*(5374), pp.237-240.

Fire, S.E., Wang, Z., Berman, M., Langlois, G.W., Morton, S.L., Sekula-Wood, E. and Benitez-Nelson, C.R., 2010. Trophic Transfer of the Harmful Algal Toxin Domoic Acid as a Cause of Death in a Minke Whale (Balaenoptera acutorostrata) Stranding in Southern California. *Aquatic Mammals*, *36*(4).

Garnesson, P., Mangin, A., Demaria, J., Bretagnon, M. and Hembise Fanton d'Andon, O., 2021, April. First release of the CMEMS Global coastal OLCI 300 meters Chlorophyll-a Product. In EGU General Assembly Conference Abstracts (pp. EGU21-6239).

Gattuso, J.P., Frankignoulle, M. and Wollast, R., 1998. Carbon and carbonate metabolism in coastal aquatic ecosystems. *Annual Review of Ecology and Systematics*, pp.405-434.

Gavin, H.P., 2019. The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems. *Department of Civil and Environmental Engineering, Duke University*, *19*.

Gentemann, C.L., Wentz, F.J., Brewer, M., Hilburn, K. and Smith, D., 2010. Passive microwave remote sensing of the ocean: An overview. Oceanography from Space, pp.13-33.

Gholizadeh, M.H., Melesse, A.M. and Reddi, L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors, 16(8), p.1298.

Gieskes, W.W.C. and Kraay, G.W., 1975. The phytoplankton spring bloom in Dutch coastal waters of the North Sea. Netherlands Journal of Sea Research, 9(2), pp.166-196.

Gilerson, A., Malinowski, M., Herrera, E., Tomlinson, M., Stumpf, R. and Ondrusek, M., 2021, April. Estimation of chlorophyll-a concentration in complex coastal waters from satellite imagery. In Ocean Sensing and Monitoring XIII (Vol. 11752, pp. 64-71). SPIE.

Girshick, R., 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

Gitelson, A.A., Gao, B.C., Li, R.R., Berdnikov, S. and Saprygin, V., 2011. Estimation of chlorophyll-a concentration in productive turbid waters using a Hyperspectral Imager for the Coastal Ocean—the Azov Sea case study. *Environmental Research Letters*, *6*(2), p.024023.

Glantz, P. and Tesche, M., 2012. Assessment of two aerosol optical thickness retrieval algorithms applied to MODIS Aqua and Terra measurements in Europe. *Atmospheric Measurement Techniques*, *5*(7), pp.1727-1740.

Gohin, F., Druon, J.N. and Lampert, L., 2002. A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. International journal of remote sensing, 23(8), pp.1639-1661.

Gohin, F., 2011. Annual cycles of chlorophyll-a, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters. Ocean Science, 7(5), pp.705-732.

Gons, H.J., Rijkeboer, M. and Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. Journal of Plankton Research, 24(9), pp.947-951.

González-Gil, R., Banas, N.S., Bresnan, E. and Heath, M.R., 2022. The onset of the spring phytoplankton bloom in the coastal North Sea supports the Disturbance Recovery Hypothesis. *Biogeosciences*, *19*(9), pp.2417-2426.

Gordon, H.R. and McCluney, W.R., 1975. Estimation of the depth of sunlight penetration in the sea for remote sensing. Applied optics, 14(2), pp.413-416.

Gordon, H.R., 1978. Removal of atmospheric effects from satellite imagery of the oceans. Applied Optics, 17(10), pp.1631-1636.

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

Gordon, H.R., 1981. Surface water phytoplankton concentration derived from satellite measurements of ocean color (A). Journal of the Optical Society of America (1917-1983), 71, p.1643.

Gordon, H.R. and Wang, M., 1994a. Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors. Applied optics, 33(33), pp.7754-7763.

Gordon, H.R. and Wang, M., 1994b. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Applied optics, 33(3), pp.443-452.

Gordon, H.R., 2019. Physical Principles of Ocean Color Remote Sensing.

Gorman, E.T., Kubalak, D.A., Patel, D., Mott, D.B., Meister, G. and Werdell, P.J., 2019, October. The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission: an emerging era of global, hyperspectral Earth system remote sensing. In *Sensors, Systems, and Next-Generation Satellites XXIII* (Vol. 11151, pp. 78-84). SPIE.

Goyens, C., Jamet, C. and Schroeder, T., 2013. Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters. Remote Sensing of Environment, 131, pp.63-75.

Graham, M.D. and Vinebrooke, R.D., 2009. Extreme weather events alter planktonic communities in boreal lakes. Limnology and Oceanography, 54(6part2), pp.2481-2492.

Greenwood, N., Mills, D.K., Howarth, M.J., Proctor, R., Pearce, D.J., Sivyer, D.B., Cutchey, S.J. and Andres, O., 2006. High frequency monitoring in Liverpool Bay; variability of suspended matter, nutrients and phytoplankton. European Operational Oceanography: Present and Future, p.413.

Greenwood, N., Devlin, M.J., Best, M., Fronkova, L., Graves, C.A., Milligan, A., Barry, J. and Van Leeuwen, S.M., 2019. Utilizing eutrophication assessment directives from transitional to marine systems in the Thames estuary and Liverpool Bay, UK. *Frontiers in Marine Science*, *6*, p.116.

Gregg, W.W., Casey, N.W. and McClain, C.R., 2005. Recent trends in global ocean chlorophyll. *Geophysical research letters*, *32*(3).

Gross, L., Thiria, S. and Frouin, R., 1999. Applying artificial neural network methodology to ocean color remote sensing. Ecological Modelling, 120(2-3), pp.237-246.

Gruber, N., Clement, D., Carter, B.R., Feely, R.A., Van Heuven, S., Hoppema, M., Ishii, M., Key, R.M., Kozyr, A., Lauvset, S.K. and Lo Monaco, C., 2019. The oceanic sink for anthropogenic CO2 from 1994 to 2007. *Science*, *363*(6432), pp.1193-1199.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C. and Straif, C., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, *7*(7), pp.8830-8857.

Haber, F. and Rossignol, R.L., 1913. The Production of Synthetic Ammonia. *Industrial & Engineering Chemistry*, *5*(4), pp.328-331.

Hansen, P.J., Bjørnsen, P.K. and Hansen, B.W., 1997. Zooplankton grazing and growth: Scaling within the 2-2,-µm body size range. Limnology and oceanography, 42(4), pp.687-704.

Hanafin, J.A. and Minnett, P.J., 2005. Measurements of the infrared emissivity of a wind-roughened sea surface. Applied Optics, 44(3), pp.398-411.

Harris, A.R., Brown, S.J. and Mason, I.M., 1994. The effect of windspeed on sea surface temperature retrieval from space. Geophysical Research Letters, 21(16), pp.1715-1718.

Hebb, D.O., 1949. The first stage of perception: growth of the assembly. The Organization of Behavior, 4, pp.60-78.

Helm, K.P., Bindoff, N.L. and Church, J.A., 2011. Observed decreases in oxygen content of the global ocean. Geophysical Research Letters, 38(23).

Hervo, M., Quennehen, B., Kristiansen, N.I., Boulon, J., Stohl, A., Fréville, P., Pichon, J.M., Picard, D., Labazuy, P., Gouhier, M. and Roger, J.C., 2012. Physical and optical properties of 2010 Eyjafjallajökull volcanic eruption aerosol: ground-based, Lidar and airborne measurements in France. Atmospheric Chemistry and Physics, 12(4), pp.1721-1736.

Hieronymi, M., Müller, D. and 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, p.140.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), pp.82-97.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

Hooker, S.B., 1992. SeaWiFS technical report series: An overview of SeaWiFS and ocean color.

Hosoda, S., Suga, T., Shikama, N. and Mizuno, K., 2009. Global surface layer salinity change detected by Argo and its implication for hydrological cycle intensification. *Journal of Oceanography*, *65*(4), pp.579-586.

Hovis, W.A. and Leung, K.C., 1977. Remote sensing of ocean color. Optical Engineering, 16(2), pp.158-166.

Høyer, J.L. and Karagali, I., 2016. Sea surface temperature climate data record for the North Sea and Baltic Sea. *Journal of Climate*, *29*(7), pp.2529-2541.

Hu, C., Lee, Z. and Franz, B., 2012. Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. Journal of Geophysical Research: Oceans, 117(C1).

Hu, C., Carder, K.L. and Muller-Karger, F.E., 2000. Atmospheric correction of SeaWiFS imagery over turbid coastal waters: a practical method. Remote sensing of environment, 74(2), pp.195-206.

Hu, C., Feng, L. and Guan, Q., 2020. A machine learning approach to estimate surface chlorophyll a concentrations in global oceans from satellite measurements. IEEE Transactions on Geoscience and Remote Sensing, 59(6), pp.4590-4607.

Huang, B., Liu, C., Freeman, E., Graham, G., Smith, T. and Zhang, H.M., 2021. Assessment and intercomparison of NOAA daily optimum interpolation sea surface temperature (DOISST) version 2.1. Journal of Climate, 34(18), pp.7421-7441.

Huber, C., Klimant, I., Krause, C., Werner, T., Mayr, T. and Wolfbeis, O.S., 2000. Optical sensor for seawater salinity. *Fresenius' journal of analytical chemistry*, *368*(2), pp.196-202.

Hülse, D., Arndt, S., Wilson, J.D., Munhoven, G. and Ridgwell, A., 2017. Understanding the ocean's biological carbon pump in the past: Do we have the right tools?

Hunter-Cevera, K.R., Neubert, M.G., Olson, R.J., Solow, A.R., Shalapyonok, A. and Sosik, H.M., 2016. Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. *Science*, *354*(6310), pp.326-329.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S. and Claustre, H., 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies?. Biogeosciences discussions, 4(2), pp.707-745.

Ibrahim, A., Franz, B., Ahmad, Z., Healy, R., Knobelspiesse, K., Gao, B.C., Proctor, C. and Zhai, P.W., 2018. Atmospheric correction for hyperspectral ocean color retrieval with application to the Hyperspectral Imager for the Coastal Ocean (HICO). *Remote Sensing of Environment*, *204*, pp.60-75.

Ilori, C.O., Pahlevan, N. and Knudby, A., 2019. Analyzing performances of different atmospheric correction techniques for Landsat 8: Application for coastal remote sensing. Remote Sensing, 11(4), p.469.

Ioannou, I., Gilerson, A., Gross, B., Moshary, F. and Ahmed, S., 2011. Neural network approach to retrieve the inherent optical properties of the ocean from observations of MODIS. Applied Optics, 50(19), pp.3168-3186.

Iwasaki, A. and Yamamoto, H., 2013, July. Data product of hyperspectral imager suite (HISUI). In *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS* (pp. 4407-4410). IEEE.

Jaccard, P., Hjermann, D.Ø., Ruohola, J., Marty, S., Kristiansen, T., Sørensen, K., Kaitala, S., Mangin, A. and Pouliquen, S., 2018. Product user manual. For Global Ocean Reprocessed in-situ Observations of Biogeochemical Products.

Jackson, T., Sathyendranath, S. and Mélin, F., 2017. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. Remote Sensing of Environment, 203, pp.152-161.

Jamet, C., Loisel, H. and Dessailly, D., 2012. Retrieval of the spectral diffuse attenuation coefficient Kd (λ) in open and coastal ocean waters using a neural network inversion. Journal of Geophysical Research: Oceans, 117(C10).

Jeffrey, S.T. and Humphrey, G.F., 1975. New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. Biochemie und physiologie der pflanzen, 167(2), pp.191-194.

Jerlov, N.G., 1976. Marine optics. Elsevier.

Jerlov, N.G., 1977. Classification of sea water in terms of quanta irradiance. ICES Journal of Marine Science, 37(3), pp.281-287.

Jia, S., Tang, G., Zhu, J. and Li, Q., 2015. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), pp.88-102.

Jim, K.C., Giles, C.L. and Horne, B.G., 1996. An analysis of noise in recurrent neural networks: convergence and generalization. *IEEE Transactions on neural networks*, *7*(6), pp.1424-1438.

Joyce, S., 2000. The dead zones: oxygen-starved coastal waters. *Environmental health perspectives*, *108*(3), pp.A120-A125.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., 2021. Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), pp.583-589.

Kalaji, H.M., Schansker, G., Ladle, R.J., Goltsev, V., Bosa, K., Allakhverdiev, S.I., Brestic, M., Bussotti, F., Calatayud, A., Dąbrowski, P. and Elsheery, N.I., 2014. Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. Photosynthesis research, 122(2), pp.121-158.

Kalaji, H.M., Schansker, G., Brestic, M., Bussotti, F., Calatayud, A., Ferroni, L., Goltsev, V., Guidi, L., Jajoo, A., Li, P. and Losciale, P., 2017. Frequently asked questions about chlorophyll fluorescence, the sequel. Photosynthesis Research, 132(1), pp.13-66.

Karlik, B. and Olgac, A.V., 2011. Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems, 1(4), pp.111-122.

Kay, S., Hedley, J.D. and Lavender, S., 2009. Sun glint correction of high and low spatial resolution images of aquatic scenes: A review of methods for visible and near-infrared wavelengths. *Remote sensing*, *1*(4), pp.697-730.

Keeling, R.F., Körtzinger, A. and Gruber, N., 2010. Ocean deoxygenation in a warming world. Annu. Rev. Mar. Sci, 2(1), pp.199-229.

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

Khodr, J. and Younes, R., 2011, October. Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas. In *2011 4th international congress on image and signal processing* (Vol. 4, pp. 1875-1883). IEEE.

Kilpatrick, K., Podesta, G., Walsh, S., Evans, R. and Minnett, P., 2014. Implementation of Version 6 AQUA and TERRA SST processing. White Paper.

Kilpatrick, K.A., Podestá, G., Walsh, S., Williams, E., Halliwell, V., Szczodrak, M., Brown, O.B., Minnett, P.J. and Evans, R., 2015. A decade of sea surface temperature from MODIS. Remote Sensing of Environment, 165, pp.27-41.

Kirk, J.T., 1994. Light and photosynthesis in aquatic ecosystems. Cambridge university press.

Krause, G.H. and Weis, E., 1984. Chlorophyll fluorescence as a tool in plant physiology. Photosynthesis research, 5(2), pp.139-157.

Kutser, T., Pierson, D.C., Tranvik, L., Reinart, A., Sobek, S. and Kallio, K., 2005. Using satellite remote sensing to estimate the colored dissolved organic matter absorption coefficient in lakes. Ecosystems, 8(6), pp.709-720.

Lacroix, G., Ruddick, K., Gypens, N. and Lancelot, C., 2007. Modelling the relative impact of rivers (Scheldt/Rhine/Seine) and Western Channel waters on the nutrient and diatoms/Phaeocystis distributions in Belgian waters (Southern North Sea). Continental Shelf Research, 27(10-11), pp.1422-1446.

Latif, M., Sun, J., Visbeck, M. and Hadi Bordbar, M., 2022. Natural variability has dominated Atlantic Meridional Overturning Circulation since 1900. *Nature Climate Change*, *12*(5), pp.455-460.

Latimer, P., Bannister, T.T. and Rabinowitch, E., 1956. Quantum yields of fluorescence of plant pigments. Science, 124(3222), pp.585-586.

Lavigne, H., Van der Zande, D., Ruddick, K., Dos Santos, J.C., Gohin, F., Brotas, V. and Kratzer, S., 2021. Quality-control tests for OC4, OC5 and NIR-red satellite chlorophyll-a algorithms applied to coastal waters. Remote Sensing of Environment, 255, p.112237.

Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L. and Davis, C., 2007. Euphotic zone depth: Its derivation and implication to ocean-color remote sensing. Journal of Geophysical Research: Oceans, 112(C3).

Lefering, I., Bengil, F., Trees, C., Röttgers, R., Bowers, D., Nimmo-Smith, A., Schwarz, J. and McKee, D., 2016. Optical closure in marine waters from *in situ* inherent optical property measurements. *Optics express*, *24*(13), pp.14036-14052.

Lefebvre, K.A., Powell, C.L., Busman, M., Doucette, G.J., Moeller, P.D., Silver, J.B., Miller, P.E., Hughes, M.P., Singaram, S., Silver, M.W. and Tjeerdema, R.S., 1999. Detection of domoic acid in northern anchovies and California sea lions associated with an unusual mortality event. *Natural toxins*, *7*(3), pp.85-92.

Lewis, M.R., Kuring, N. and Yentsch, C., 1988. Global patterns of ocean transparency: Implications for the new production of the open ocean. Journal of Geophysical Research: Oceans, 93(C6), pp.6847-6856.

Li, H., He, X., Bai, Y., Shanmugam, P., Park, Y.J., Liu, J., Zhu, Q., Gong, F., Wang, D. and Huang, H., 2020. Atmospheric correction of geostationary satellite ocean color data under high solar zenith angles in open oceans. *Remote Sensing of Environment*, *249*, p.112022.

Lin, C., Wu, C.C., Tsogt, K., Ouyang, Y.C. and Chang, C.I., 2015. Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery. Information Processing in Agriculture, 2(1), pp.25-36.

Litchman, E., Klausmeier, C.A., Schofield, O.M. and Falkowski, P.G., 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. Ecology letters, 10(12), pp.1170-1181.

Liu, H., Zhou, Q., Li, Q., Hu, S., Shi, T. and Wu, G., 2019. Determining switching threshold for NIR-SWIR combined atmospheric correction algorithm of ocean color remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 153, pp.59-73.

Llewellyn-Jones, D.T., Minnett, P.J., Saunders, R.W. and Zavody, A.M., 1984. Satellite multichannel infrared measurements of sea surface temperature of the NE Atlantic Ocean using AVHRR/2. *Quarterly Journal of the Royal Meteorological Society*, *110*(465), pp.613-631.

Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C. and Varacalli, G., 2018, July. PRISMA: The Italian hyperspectral mission. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 175-178). IEEE.

Luo, B., Minnett, P.J., Gentemann, C. and Szczodrak, G., 2019. Improving satellite retrieved night-time infrared sea surface temperatures in aerosol contaminated regions. Remote sensing of environment, 223, pp.8-20.

Luther, R. and Nikolopulos, A., 1913. Über die Beziehungen zwischen den Absorptionsspektren und der Konstitution der komplexen Kobaltamminsalze. Zeitschrift für Physikalische Chemie, 82(1), pp.361-384.

Maas, A.L., Hannun, A.Y. and Ng, A.Y., 2013, June. Rectifier nonlinearities improve neural network acoustic models. In Proc. icml (Vol. 30, No. 1, p. 3).

Maciel, F.P. and Pedocchi, F., 2022. Evaluation of ACOLITE atmospheric correction methods for Landsat-8 and Sentinel-2 in the Río de la Plata turbid coastal waters. International Journal of Remote Sensing, 43(1), pp.215-240.

MacIntyre, H.L., Kana, T.M., Anning, T. and Geider, R.J., 2002. Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria 1. Journal of phycology, 38(1), pp.17-38.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F., 2017, October. Sen2Cor for sentinel-2. In Image and Signal Processing for Remote Sensing XXIII (Vol. 10427, pp. 37-48). SPIE.

Majozi, N.P., Salama, M.S., Bernard, S., Harper, D.M. and Habte, M.G., 2014. Remote sensing of euphotic depth in shallow tropical inland waters of Lake Naivasha using MERIS data. Remote sensing of environment, 148, pp.178-189.

Mao, K., Yuan, Z., Zuo, Z., Xu, T., Shen, X. and Gao, C., 2019. Changes in global cloud cover based on remote sensing data from 2003 to 2012. Chinese Geographical Science, 29(2), pp.306-315.

Marghany, M., Hashim, M. and Cracknell, A.P., 2010, March. Modelling sea surface salinity from MODIS Satellite data. In *International Conference on Computational Science and Its Applications* (pp. 545-556). Springer, Berlin, Heidelberg.

MartÍnez-UsÓMartinez-Uso, A., Pla, F., Sotoca, J.M. and García-Sevilla, P., 2007. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(12), pp.4158-4171.

Marujo, R.F., Fronza, J.G., Soares, A.R., Queiroz, G.R. and Ferreira, K.R., 2021. Evaluating the impact of lasrc and SEN2COR atmospheric correction algorithms on LANDSAT-8/OLI and SENTINEL-2/MSI data over aeronet stations in Brazilian territory. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3, pp.271-277.

McClelland, J.L., Rumelhart, D.E. and Hinton, G.E., 1986. The appeal of parallel distributed processing. MIT Press, Cambridge MA, 3, p.44.

McCulloch, W.S. and Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), pp.115-133.

McDowell, J.C., 2018. The edge of space: Revisiting the Karman Line. Acta Astronautica, 151, pp.668-677.

McKee, D., Cunningham, A. and Dudek, A., 2007. Optical water type discrimination and tuning remote sensing band-ratio algorithms: Application to retrieval of chlorophyll and Kd (490) in the Irish and Celtic Seas. *Estuarine, Coastal and Shelf Science*, *73*(3-4), pp.827-834.

McKee, D., Cunningham, A., Wright, D. and Hay, L., 2007. Potential impacts of nonalgal materials on water-leaving Sun induced chlorophyll fluorescence signals in coastal waters. *Applied optics*, *46*(31), pp.7720-7729.

McKee, D., Röttgers, R., Neukermans, G., Calzado, V.S., Trees, C., Ampolo-Rella, M., Neil, C. and Cunningham, A., 2014. Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. *Journal of Geophysical Research: Oceans*, *119*(12), pp.9013-9025.

McPhaden, M.J., Zebiak, S.E. and Glantz, M.H., 2006. ENSO as an integrating concept in earth science. *science*, *314*(5806), pp.1740-1745.

Medina-Lopez, E. and Ureña-Fuentes, L., 2019. High-resolution sea surface temperature and salinity in coastal areas worldwide from raw satellite data. Remote Sensing, 11(19), p.2191.

Medina-Lopez, E., 2020. Machine learning and the end of atmospheric corrections: A comparison between high-resolution sea surface salinity in coastal areas from top and bottom of atmosphere sentinel-2 imagery. Remote Sensing, 12(18), p.2924.

Mélin, F., Sclep, G., Jackson, T. and Sathyendranath, S., 2016. Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets. Remote Sensing of Environment, 177, pp.107-124.

Mélin, F. and Vantrepotte, V., 2015. How optically diverse is the coastal ocean?. Remote Sensing of Environment, 160, pp.235-251.

Meyer, J., Nehmer, P., Moll, A. and Kröncke, I., 2018. Shifting south-eastern North Sea macrofauna community structure since 1986: a response to de-eutrophication and regionally decreasing food supply?. *Estuarine, Coastal and Shelf Science*, *213*, pp.115-127.

Minnett, P.J. and Barton, I.J., 2010. Remote sensing of the earth's surface temperature. *Radiometric temperature measurements. Ii applications*, *43*, pp.333-391.

Minnett, P. and Kaiser-Weiss, A., 2012. Near-surface oceanic temperature gradients. *GHRSST Discussion Doc*, *7*.

Minnett, P.J., Alvera-Azcárate, A., Chin, T.M., Corlett, G.K., Gentemann, C.L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E. and Santoleri, R., 2019. Half a century of satellite remote sensing of sea-surface temperature. Remote Sensing of Environment, 233, p.111366.

Minnett, P.J., Kilpatrick, K.A., Podestá, G.P., Evans, R.H., Szczodrak, M.D., Izaguirre, M.A., Williams, E.J., Walsh, S., Reynolds, R.M., Bailey, S.W. and Armstrong, E.M., 2020. Skin sea-surface temperature from VIIRS on Suomi-NPP—NASA continuity retrievals. Remote Sensing, 12(20), p.3369.

Mobley, C.D., 1994. Light and water: radiative transfer in natural waters. Academic press.

Mobley, C.D., Werdell, J., Franz, B., Ahmad, Z. and Bailey, S., 2016. Atmospheric correction for satellite ocean color radiometry (No. GSFC-E-DAA-TN35509).

Mohajerani, S. and Saeedi, P., 2019, July. Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 1029-1032). IEEE.

Moses, W.J., Gitelson, A.A., Berdnikov, S., Saprygin, V. and Povazhnyi, V., 2012. Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters—The Azov Sea case study. *Remote Sensing of Environment*, *121*, pp.118-124.

Moore, T.S., Campbell, J.W. and Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sensing of Environment, 113(11), pp.2424-2430.

Moore, T.S., Dowell, M.D., Bradt, S. and Verdu, A.R., 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote sensing of environment, 143, pp.97-111.

Morel, A. and Prieur, L., 1977. Analysis of variations in ocean color 1. Limnology and oceanography, 22(4), pp.709-722.

Moses, W.J., Gitelson, A.A., Perk, R.L., Gurlin, D., Rundquist, D.C., Leavitt, B.C., Barrow, T.M. and Brakhage, P., 2012. Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data. *Water research*, *46*(4), pp.993-1004.

Moses, W.J., Saprygin, V., Gerasyuk, V., Povazhnyy, V., Berdnikov, S. and Gitelson, A.A., 2019. OLCI-based NIR-red models for estimating chlorophyll-a concentration in productive coastal waters—a preliminary evaluation. *Environmental Research Communications*, *1*(1), p.011002.

Muller-Karger, F.E., Varela, R., Thunell, R., Luerssen, R., Hu, C. and Walsh, J.J., 2005. The importance of continental margins in the global carbon cycle. *Geophysical research letters*, *32*(1).

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.H., Devred, E. and Bouman, H., 2008. Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, *112*(8), pp.3366-3375.

Nalli, N.R., Minnett, P.J. and Van Delst, P., 2008a. Emissivity and reflection model for calculating unpolarized isotropic water surface-leaving radiance in the infrared. I: Theoretical development and calculations. *Applied optics*, *47*(21), pp.3701-3721.

Nalli, N.R., Minnett, P.J., Maddy, E., McMillan, W.W. and Goldberg, M.D., 2008b. Emissivity and reflection model for calculating unpolarized isotropic water surface-leaving radiance in the infrared. 2: Validation using Fourier transform spectrometers. *Applied optics*, *47*(25), pp.4649-4671.

Nechad, B., Ruddick, K.G. and Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, *114*(4), pp.854-866.

Nechad, B., Alvera-Azcaràte, A., Ruddick, K. and Greenwood, N., 2011. Reconstruction of MODIS total suspended matter time series maps by DINEOF and validation with autonomous platform data. *Ocean Dynamics*, *61*(8), pp.1205-1214.

Neil, C., Cunningham, A. and McKee, D., 2011. Relationships between suspended mineral concentrations and red-waveband reflectances in moderately turbid shelf seas. *Remote Sensing of Environment*, *115*(12), pp.3719-3730.

Neil, C., Spyrakos, E., Hunter, P.D. and Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. Remote Sensing of Environment, 229, pp.159-178.

Niang, A., Gross, L., Thiria, S., Badran, F. and Moulin, C., 2003. Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. Remote Sensing of Environment, 86(2), pp.257-271.

Nixon, S.W., 1995. Coastal marine eutrophication: a definition, social causes, and future concerns. Ophelia, 41(1), pp.199-219.

Nordberg, W., 1965. Geophysical Observation from Nimbus I: Cloud heights, sea surface, and soil temperatures are mapped from the satellite by infrared radiometry. Science, 150(3696), pp.559-572.

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

O'Reilly, J.E., Maritorena, S., Siegel, D.A., O'Brien, M.C., Toole, D., Mitchell, B.G., Kahru, M., Chavez, F.P., Strutton, P., Cota, G.F. and Hooker, S.B., 2000. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. SeaWiFS postlaunch calibration and validation analyses, Part, 3, pp.9-23.

O'Reilly, J.E. and Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6. Remote sensing of environment, 229, pp.32-47.

Pahlevan, N., Roger, J.C. and Ahmad, Z., 2017. Revisiting short-wave-infrared (SWIR) bands for atmospheric correction in coastal waters. Optics express, 25(6), pp.6015-6035.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N. and Matsushita, B., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment, 240, p.111604.

Pahlevan, N., Mangin, A., Balasubramanian, S.V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M. and Giardino, C., 2021. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. Remote Sensing of Environment, 258, p.112366.

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D. and Fernandez, V., 2022. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sensing of Environment, 270, p.112860.

Pan, S.J., Kwok, J.T. and Yang, Q., 2008, July. Transfer learning via dimensionality reduction. In *AAAI* (Vol. 8, pp. 677-682).

Parsons, T.R. and Strickland, J.D.H., 1963. Discussion of spectrophotometric determination of Marine-plant Pigments, with Revised Equations far Ascertaining Chlorophylls and Carotenoids. J. mar. Res, 21, pp.155-163.

Pastor, F., Valiente, J.A. and Palau, J.L., 2019. Sea surface temperature in the Mediterranean: Trends and spatial patterns (1982–2016). In Meteorology and climatology of the Mediterranean and Black Seas (pp. 297-309). Birkhäuser, Cham.

Philander, S.G.H., Yamagata, T. and Pacanowski, R.C., 1984. Unstable air-sea interactions in the tropics. *Journal of Atmospheric Sciences*, *41*(4), pp.604-613.

Planck, M., 1901. On the law of distribution of energy in the normal spectrum. Annalen der physik, 4(553), p.1.

Pitarch, J., Volpe, G., Colella, S., Krasemann, H. and Santoleri, R., 2016. Remote sensing of chlorophyll in the Baltic Sea at basin scale from 1997 to 2012 using merged multi-sensor data. Ocean Science, 12(2), pp.379-389.

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfündel, E.E., Moreno, J., Frankenberg, C. and Berry, J.A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. Journal of experimental botany, 65(15), pp.4065-4095.

Preisendorfer, R.W., 1986. Secchi disk science: Visual optics of natural waters 1. Limnology and oceanography, 31(5), pp.909-926.

Qian, Y., Zhao, Y., Wu, Q.L. and Yang, Y., 2018. Review of salinity measurement technology based on optical fiber sensor. *Sensors and Actuators B: Chemical*, *260*, pp.86-105.

Ramankutty, N. and Foley, J.A., 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), pp.997-1027.

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D. and Pande, V., 2015. Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*.

Ranganathan, A., 2004. The levenberg-marquardt algorithm. Tutoral on LM algorithm, 11(1), pp.101-110.

Rasconi, S., Winter, K. and Kainz, M.J., 2017. Temperature increase and fluctuation induce phytoplankton biodiversity loss–Evidence from a multi-seasonal mesocosm experiment. Ecology and Evolution, 7(9), pp.2936-2946.

Raven, J.A. and Geider, R.J., 19

88. Temperature and algal growth. New phytologist, 110(4), pp.441-461.

Rayleigh, J.W.S.B., 1871. On the scattering of light by small particles.

Reda, I. and Andreas, A., 2004. Solar position algorithm for solar radiation applications. *Solar energy*, *76*(5), pp.577-589.

Reise, K. and Kohlus, J., 2008. Seagrass recovery in the northern Wadden Sea?. *Helgoland Marine Research*, *62*(1), pp.77-84.

Ren, J., Zabalza, J., Marshall, S. and Zheng, J., 2014. Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]. *IEEE Signal Processing Magazine*, *31*(4), pp.149-154.

Renosh, P.R., Doxaran, D., Keukelaere, L.D. and Gossn, J.I., 2020. Evaluation of atmospheric correction algorithms for sentinel-2-MSI and sentinel-3-OLCI in highly turbid estuarine waters. Remote Sensing, 12(8), p.1285.

Retalis, D. and Retalis, A., 1998. Effects of air pollution and wind on the large-ion concentration in the air above Athens. *Journal of Geophysical Research: Atmospheres*, *103*(D12), pp.13927-13932.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. and Wang, W., 2002. An improved *in situ* and satellite SST analysis for climate. *Journal of climate*, *15*(13), pp.1609-1625.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. *Journal of climate*, *20*(22), pp.5473-5496.

Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C.E., Balch, W.M. and Guikema, S.D., 2015. Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO2. *Science*, *350*(6267), pp.1533-1537.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J. and Nykvist, B., 2009. A safe operating space for humanity. nature, 461(7263), pp.472-475.

Ropelewski, C.F. and Halpert, M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly weather review*, *115*(8), pp.1606-1626.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), p.386.

Röttgers, R., Schönfeld, W., Kipp, P.R. and Doerffer, R., 2005. Practical test of a point-source integrating cavity absorption meter: the performance of different collector assemblies. *Applied Optics*, *44*(26), pp.5549-5560.

Röttgers, R., Häse, C. and Doerffer, R., 2007. Determination of the particulate absorption of microalgae using a point-source integrating-cavity absorption meter: verification with a photometric technique, improvements for pigment bleaching, and correction for chlorophyll fluorescence. *Limnology and Oceanography: Methods*, *5*(1), pp.1-12.

Röttgers, R. and Doerffer, R., 2007. Measurements of optical absorption by chromophoric dissolved organic matter using a point-source integrating-cavity absorption meter. *Limnology and Oceanography: Methods*, *5*(5), pp.126-135.

Ruder, S., 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Ruddick, K.G., Ovidio, F. and Rijkeboer, M., 2000. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. Applied optics, 39(6), pp.897-912.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1985. Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science.

Ryther, J.H. and Yentsch, C.S., 1957. The estimation of phytoplankton production in the ocean from chlorophyll and light data 1. *Limnology and oceanography*, *2*(3), pp.281-286.

Ryu, J.H., Han, H.J., Cho, S., Park, Y.J. and Ahn, Y.H., 2012. Overview of geostationary ocean color imager (GOCI) and GOCI data processing system (GDPS). *Ocean Science Journal*, *47*(3), pp.223-233.

Salgado-Hernanz, P.M., Regaudie-De-Gioux, A., Antoine, D. and Basterretxea, G., 2022. Pelagic primary production in the coastal Mediterranean Sea: variability, trends, and contribution to basin-scale budgets. *Biogeosciences*, *19*(1), pp.47-69.

Sarmiento, J.L., 2013. Ocean biogeochemical dynamics. In Ocean Biogeochemical Dynamics. Princeton University Press.

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M.H., Maass, H. and Platt, T., 2009. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Marine Ecology Progress Series, 383, pp.73-84.

Schiller, H. and Doerffer, R., 1999. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. International journal of remote sensing, 20(9), pp.1735-1746.

Schiller, K., 2006. Derivation of photosynthetically available radiation from METEOSAT data in the German Bight with neural nets. Ocean Dynamics, 56(2), pp.79-85.

Schindler, D.W., 2006. Recent advances in the understanding and management of eutrophication. Limnology and oceanography, 51(1part2), pp.356-363.

Schippers, P., Lürling, M. and Scheffer, M., 2004. Increase of atmospheric CO2 promotes phytoplankton productivity. Ecology letters, 7(6), pp.446-451.

Schroeder, T., Schaale, M., Lovell, J. and Blondeau-Patissier, D., 2022. An ensemble neural network atmospheric correction for Sentinel-3 OLCI over coastal waters providing inherent model uncertainty estimation and sensor noise propagation. Remote Sensing of Environment, 270, p.112848.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A. and Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Optics express, 26(6), pp.7404-7422.

Seibel, B.A., 2011. Critical oxygen levels and metabolic suppression in oceanic oxygen minimum zones. Journal of Experimental Biology, 214(2), pp.326-336.

Shockley, W., 1949. The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors. Bell System Technical Journal, 28(3), pp.435-489.

Siegel, D.A., Wang, M., Maritorena, S. and Robinson, W., 2000. Atmospheric correction of satellite ocean color imagery: the black pixel assumption. *Applied optics*, *39*(21), pp.3582-3591.

Signorini, S.R., Franz, B.A. and McClain, C.R., 2015. Chlorophyll variability in the oligotrophic gyres: mechanisms, seasonality and trends. *Frontiers in Marine Science*, *2*, p.1.

Smith, R.C., 1981. Remote sensing and depth distribution of ocean chlorophyll. Marine Ecology Progress Series, pp.359-361.

Smith, R.C. and Baker, K.S., 1982. Oceanic chlorophyll concentrations as determined by satellite (Nimbus-7 Coastal Zone Color Scanner). Marine Biology, 66(3), pp.269-279.

Smith, V.H. and Schindler, D.W., 2009. Eutrophication science: where do we go from here?. *Trends in ecology & evolution*, *24*(4), pp.201-207.

Smyth, T.J., Fishwick, J.R., Al-Moosawi, L., Cummings, D.G., Harris, C., Kitidis, V., Rees, A., Martinez-Vicente, V. and Woodward, E.M., 2010. A broad spatio-temporal view of the Western English Channel observatory. Journal of Plankton Research, 32(5), pp.585-601.

Sommer, U., Peter, K.H., Genitsaris, S. and Moustaka-Gouni, M., 2017. Do marine phytoplankton follow B ergmann's rule sensu lato?. Biological Reviews, 92(2), pp.1011-1026.

Sørensen, K., Grung, M. and Röttgers, R., 2007. An intercomparison of in vitro chlorophyll a determinations for MERIS level 2 data validation. International Journal of Remote Sensing, 28(3-4), pp.537-554.

Spyrakos, E., Vilas, L.G., Palenzuela, J.M.T. and Barton, E.D., 2011. Remote sensing chlorophyll a of optically complex waters (rias Baixas, NW Spain): Application of a regionally specific chlorophyll a

algorithm for MERIS full resolution data during an upwelling cycle. Remote Sensing of Environment, 115(10), pp.2471-2485.

Spyrakos, E., O'donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G., Neil, C., Barbosa, C.C., Binding, C.E., Bradt, S. and Bresciani, M., 2018. Optical types of inland and coastal waters. Limnology and Oceanography, 63(2), pp.846-870.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), pp.1929-1958.

Srokosz, M.A. and Bryden, H.L., 2015. Observing the Atlantic Meridional Overturning Circulation yields a decade of inevitable surprises. *Science*, *348*(6241), p.1255575.

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

Stramska, M., Stramski, D., Kaczmarek, S., Allison, D.B. and Schwarz, J., 2006. Seasonal and regional differentiation of bio-optical properties within the north polar Atlantic. Journal of Geophysical Research: Oceans, 111(C8).

Stumpf, R.P., Arnone, R.A., Gould, R.W., Martinolich, P.M. and Ransibrahmanakul, V., 2003. A partially coupled ocean-atmosphere model for retrieval of water-leaving radiance from SeaWiFS in coastal waters. NASA Tech. Memo, 206892, pp.51-59.

Sun, H., Ren, J., Zhao, H., Sun, G., Liao, W., Fang, Z. and Zabalza, J., 2020. Adaptive distance-based band hierarchy (ADBH) for effective hyperspectral band selection. IEEE transactions on cybernetics.

Sun, H., Ren, J., Zhao, H., Yuen, P. and Tschannerl, J., 2021. Novel gumbel-softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection. *IEEE transactions on geoscience and remote sensing*, *60*, pp.1-13.

Sun, H., Zhang, L., Ren, J. and Huang, H., 2022. Novel Hyperbolic Clustering-based Band Hierarchy (HCBH) for Effective Unsupervised Band Selection of Hyperspectral Images. *Pattern Recognition*, p.108788.

Sunday, J.M., Bates, A.E., Kearney, M.R., Colwell, R.K., Dulvy, N.K., Longino, J.T. and Huey, R.B., 2014. Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences*, *111*(15), pp.5610-5615.

Tanaka, A., Kishino, M., Doerffer, R., Schiller, H., Oishi, T. and Kubota, T., 2004. Development of a neural network algorithm for retrieving concentrations of chlorophyll, suspended matter and yellow substance from radiance data of the ocean color and temperature scanner. *Journal of Oceanography*, *60*(3), pp.519-530.

Terseleer, N., Gypens, N. and Lancelot, C., 2010. Evaluating the river de-eutrophication gain on the magnitude of Phaeocystis blooms in the Southern North Sea between 1985 and 2005: a model study. KALLIOPI A. PAGOU, p.96.

Tilstone, G.H., Peters, S.W., van der Woerd, H.J., Eleveld, M.A., Ruddick, K., Schönfeld, W., Krasemann, H., Martinez-Vicente, V., Blondeau-Patissier, D., Röttgers, R. and Sørensen, K., 2012. Variability in specific-absorption properties and their use in a semi-analytical ocean colour algorithm for MERIS in North Sea and Western English Channel Coastal Waters. Remote Sensing of Environment, 118, pp.320-338.

Thomas, M.K., Kremer, C.T., Klausmeier, C.A. and Litchman, E., 2012. A global pattern of thermal adaptation in marine phytoplankton. *Science*, *338*(6110), pp.1085-1088.

Trenberth, K.E., 1997. The definition of el nino. *Bulletin of the American Meteorological Society*, *78*(12), pp.2771-2778.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W.M., Barker, K. and Barlow, R., 2016. A compilation of global bio-optical *in situ* data for ocean-colour satellite applications. Earth System Science Data, 8(1), pp.235-252.

Valentine, D., 2016. Atmosphere: Context, detachment, and the view from above Earth. American Ethnologist, 43(3), pp.511-524.

Van Bennekom, A.J. and Wetsteijn, F.J., 1990. The winter distribution of nutrients in the Southern Bight of the North Sea (1961–1978) and in the estuaries of the Scheldt and the Rhine/Meuse. *Netherlands Journal of Sea Research*, *25*(1-2), pp.75-87.

Van der Zande, D., Eleveld, M., Lavigne, H., Gohin, F., Pardo, S., Tilstone, G., Blauw, A., Markager, S. and Enserink, L., 2019. Joint Monitoring Program of the EUtrophication of the NOrthSea with SAtellite data user case. Journal Of Operational Oceanography , 12 (SI).

Vanhellemont, Q. and Ruddick, K., 2015. Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8. Remote Sensing of Environment, 161, pp.89-106.

Vanhellemont, Q. and Ruddick, K., 2021. Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters. Remote Sensing of Environment, 256, p.112284.

Vantrepotte, V., Loisel, H., Dessailly, D. and Mériaux, X., 2012. Optical classification of contrasted coastal waters. Remote Sensing of Environment, 123, pp.306-323.

Vepsäläinen, J., Pyhälahti, T., Rantajärvi, E., Kallio, K., Pertola, S., Stipa, T., Kiirikki, M., Pulliainen, J. and Seppälä, J., 2005. The combined use of optical remote sensing data and unattended flow-through fluorometer measurements in the Baltic Sea. International Journal of Remote Sensing, 26(2), pp.261-282.

Walton, C.C., Pichel, W.G., Sapper, J.F. and May, D.A., 1998. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research: Oceans*, *103*(C12), pp.27999-28012.

Wang, M. and Gordon, H.R., 1994. A simple, moderately accurate, atmospheric correction algorithm for SeaWiFS. Remote sensing of environment, 50(3), pp.231-239.

Wang, M. and Bailey, S.W., 2001. Correction of sun glint contamination on the SeaWiFS ocean and atmosphere products. Applied Optics, 40(27), pp.4790-4798.

Wang, M., 2002. The Rayleigh lookup tables for the SeaWiFS data processing: Accounting for the effects of ocean surface roughness. International Journal of Remote Sensing, 23(13), pp.2693-2702.

Wang, M. and Shi, W., 2007. The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. Optics express, 15(24), pp.15722-15733.

Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E. and Wolfe, R., 2012. Impact of sensor degradation on the MODIS NDVI time series. Remote Sensing of Environment, 119, pp.55-61.

Wasmund, N., Nausch, G., Gerth, M., Busch, S., Burmeister, C., Hansen, R. and Sadkowiak, B., 2019. Extension of the growing season of phytoplankton in the western Baltic Sea in response to climate change. *Marine Ecology Progress Series*, *622*, pp.1-16.

Watras, C.J., Garcon, V.C., Olson, R.J., Chisholm, S.W. and Anderson, D.M., 1985. The effect of zooplankton grazing on estuarine blooms of the toxic dinoflagellate Gonyaulax tamarensis. Journal of Plankton Research, 7(6), pp.891-908.

Werdell, P.J. and 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), pp.122-140.

Weiss, K., Khoshgoftaar, T.M. and Wang, D., 2016. A survey of transfer learning. *Journal of Big data*, *3*(1), pp.1-40.

Wirtz, K.W., 2011. Non-uniform scaling in phytoplankton growth rate due to intracellular light and CO2 decline. Journal of plankton research, 33(9), pp.1325-1341.

Wolters, E., Toté, C., Sterckx, S., Adriaensen, S., Henocq, C., Bruniquel, J., Scifoni, S. and Dransfeld, S., 2021. iCOR Atmospheric Correction on Sentinel-3/OLCI over Land: Intercomparison with AERONET, RadCalNet, and SYN Level-2. Remote Sensing, 13(4), p.654.

Wong, M.S., Lee, K.H., Kim, Y.J., Nichol, J.E., Li, Z. and Emerson, N., 2007. Modeling of suspended solids and sea surface salinity in Hong Kong using Aqua/MODIS satellite images. *Korean Journal of Remote Sensing*, *23*(3), pp.161-169.

Wu, X. and Smith, W.L., 1997. Emissivity of rough sea surface for 8–13 μm: modeling and verification. *Applied optics*, *36*(12), pp.2609-2619.

Wynne, T.T., Tomlinson, M.C., Briggs, T.O., Mishra, S., Meredith, A., Vogel, R.L. and Stumpf, R.P., 2022. Evaluating the Efficacy of Five Chlorophyll-a Algorithms in Chesapeake Bay (USA) for Operational Monitoring and Assessment. Journal of Marine Science and Engineering, 10(8), p.1104.

Wyrtki, K., 1962, January. The oxygen minima in relation to ocean circulation. In Deep Sea Research and Oceanographic Abstracts (Vol. 9, No. 1-2, pp. 11-23). Elsevier.

Xiong, X., Chiang, K., Esposito, J., Guenther, B. and Barnes, W., 2003. MODIS on-orbit calibration and characterization. Metrologia, 40(1), p.S89.

Xiong, X. and Barnes, W., 2006. An overview of MODIS radiometric calibration and characterization. Advances in Atmospheric Sciences, 23(1), pp.69-79.

Xiong, X. and Barnes, W., 2006. An overview of MODIS radiometric calibration and characterization. Advances in Atmospheric Sciences, 23(1), pp.69-79.

Xiong, X., Angal, A., Twedt, K.A., Chen, H., Link, D., Geng, X., Aldoretta, E. and Mu, Q., 2019. MODIS reflective solar bands on-orbit calibration and performance. IEEE Transactions on Geoscience and Remote Sensing, 57(9), pp.6355-6371.

Yamori, W., Hikosaka, K. and Way, D.A., 2014. Temperature response of photosynthesis in C3, C4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynthesis research*, *119*(1), pp.101-117.

Yan, Y., Zhang, Y., Ren, J., Hadjal, M., Mckee, D., Kao, F.J. and Durrani, T., 2022. Estimation of Chlorophyll Concentration for Environment Monitoring in Scottish Marine Water. In International Conference in Communications, Signal Processing, and Systems (pp. 582-587). Springer, Singapore.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J. and Gao, J., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment, 241, p.111716.

Zabalza, J., Ren, J., Wang, Z., Marshall, S. and Wang, J., 2014a. Singular spectrum analysis for effective feature extraction in hyperspectral imaging. *IEEE Geoscience and Remote Sensing Letters*, *11*(11), pp.1886-1890.

Zabalza, J., Ren, J., Yang, M., Zhang, Y., Wang, J., Marshall, S. and Han, J., 2014b. Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, pp.112-122.

Zabalza, J., Ren, J., Zheng, J., Han, J., Zhao, H., Li, S. and Marshall, S., 2015. Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. *IEEE transactions on geoscience and remote sensing*, *53*(8), pp.4418-4433.

Zaneveld, J.R.V., Kitchen, J.C. and Moore, C.C., 1994, October. Scattering error correction of reflection-tube absorption meters. In *Ocean Optics XII* (Vol. 2258, pp. 44-55). SPIE.

Zhang, Q., Wang, H., Dong, J., Zhong, G. and Sun, X., 2017. Prediction of sea surface temperature using long short-term memory. IEEE geoscience and remote sensing letters, 14(10), pp.1745-1749.

Zhao, W. and Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), pp.4544-4554.

Zhong, H. and Prentice, R.L., 2008. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics*, *9*(4), pp.621-634.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), pp.8-36.

Zibordi, G., Mélin, F., Berthon, J.F., Holben, B., Slutsker, I., Giles, D., D'Alimonte, D., Vandemark, D., Feng, H., Schuster, G. and Fabbri, B.E., 2009. AERONET-OC: a network for the validation of ocean color primary products. Journal of Atmospheric and Oceanic Technology, 26(8), pp.1634-1651.