TU Delft Delft University of Technology

Deltares Enabling Delta Life

Eutrophication prediction in the Dutch coastal waters using remote sensing data and machine learning

Master of Science

By Konstantinos Chatzopoulos Vouzoglanis

Cover: AQUA MODIS observation of a phytoplankton bloom in the North Sea on May 8, 2018. Source: https://earthobservatory.nasa.gov/images/92113/spring-color-in-the-north-sea, accessed on 24/07/2019.

Eutrophication prediction in the Dutch coastal waters using remote sensing data and machine learning

by

Konstantinos Chatzopoulos Vouzoglanis

in partial fulfilment of the requirements for the degree of

Master of Science
in Civil Engineering

at the Delft University of Technology,
to be defended publicly on July 29, 2019 at 13:30.

Student number:	4604717	
Thesis committee:	Dr. R. C. Lindenbergh,	TU Delft
	Dr. ir. G. El Serafy,	Deltares
	Dr. ir. I. V. Smal,	TU Delft
	Dr. S. Alfieri,	TU Delft
	Dr. M. A. Schleiss	TU Delft

This thesis is confidential and cannot be made public until July 29, 2019.

TU Delft

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727277.

This project has received funding from the European Commission's Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO) under grant agreement No ECHO/SUB/2016/740172/PREV18.

Abstract

Eutrophication processes in coastal waters are becoming more prominent as a result of high nutrient discharges from intensive agriculture and increased urban waste. These processes can be devastating for local ecosystems and lead to dissolved oxygen depletion, which applies considerable stress on aquatic organisms. For ecosystems to preserve their status, stop and reverse the negative effects of eutrophication, regular estimation of corresponding indicators has to take place. In this direction, mostly process-driven models have been used, but the presented project argues that freely available remote sensing data can also provide useful insights for the oxygen saturation of the water. The proposed methodology uses Sea Surface Temperature and Chlorophyll-a estimations from AQUA and ENVISAT satellite sensors for the period 2003-2011 to predict the dissolved oxygen content in the Dutch coastal waters. It does so by implementing various Machine Learning models, namely Random Forest, Artificial Neural Network and Gradient Boosting Regressors, with the latter demonstrating the best results. After extensive data pre-processing, the results show that dissolved oxygen can be predicted with an average Root-Mean-Squared error of 0.8 g/m^3. Important steps towards a lower error include the use of gap-filled variables and their decomposition into their temporal components as inputs for the model. Furthermore, the effect of the Sea Surface Temperature on the dissolved oxygen is documented through its contribution in the estimation of the latter's seasonal variability, while the estimation of the maximum dissolved oxygen values is attributed to Chlorophyll-a. Further feature engineering and model development can possibly improve the estimation of the minimum dissolved oxygen values in the coast and the overall prediction in more complex intertidal areas, like the Wadden Sea.

Acknowledgements

First, I would like to express my gratitude to my main TU Delft and Deltares supervisors Dr. Roderik Lindenbergh and Dr. Ghada El Serafy, for their help and support from the formulation of my research proposal up to the very end of the project. Secondly, I would like to thank Dr. Ihor Smal for his invaluable assistance in the topics of Data Science and Machine Learning, which were a huge part of my methodology. I would also like to thank Dr. Silvia Alfieri for offering her remote sensing expertise insights during the progress meetings we had and Dr. Marc Schleiss who, despite joining the committee at the Green-Light meeting, put time and effort in helping me improve the structure of my text. Moreover, I would like to thank PhD candidate Alex Ziemba that assisted me with the project and the text, and Deltares employee Sonja Wanke that offered to help me whenever I needed it. Furthermore, I am grateful for Deltares that provided me with a laptop, resources and an open and international workspace were I expanded my network and made lifelong friends. Finally, I would like to thank my parents for their continuous support and sacrifices to offer me an opportunity to study abroad.

Konstantinos Chatzopoulos Vouzoglanis
Delft, July 2019

Table of Contents

List of Figures

List of Tables

Acronyms

ANN	Artificial Neural Network
Chla	Chlorophyll-α
DO	Dissolved Oxygen
DL	Deep Learning
EU	European Union
GB	Gradient Boosting
HAB	Harmful Algal Bloom
ML	Machine Leaning
MSFD	Marine Strategy Framework Directive
N	Nitrogen
P	Phosphorus
RF	Random Forest
RWS	Rijkswaterstaat – Dutch Ministry of Infrastructure and Water Management
SST	Sea Surface Temperature
TSM	Total Suspended Matter
WFD	Water Framework Directive

1. Introduction

In this chapter the motivation behind this thesis, the problem statement and the research approach are presented.

1.1. Background

According to the European Commission (2011) seas and oceans are resources that provide great benefits. Their uses range from transport, fishing and tourism, to mineral extraction (e.g. deep-sea metals, gas and oil) and energy production (e.g. offshore windfarms). Meanwhile, the environmental importance of the habitats they include is high, and for this reason, their monitoring for preservation and protection is critical. Nevertheless, many regional seas and oceans at large are under threat due to their unsustainable utilization. Anthropogenic pollution of the coastal waters is amplifying the effect of certain processes in aquatic habitats which apply stress on the plants and animals they include. It further affects water quality by lowering the pH through a process of acidification. In addition, the ever growing and persistent impact of marine transports and activities have a multitude of effects on habitat quality. These include the introduction of alien species via ballast waters, direct pollution and oil spills, plastics, litter and lost fish nets, as well as the overfishing of fish stocks. The dredging required to maintain shipping corridors can also raise toxic compounds to the upper water layers. All the above along with the intensification of climate change have a great toll on the coastal and marine environment (European Commission, 2011).

One of the major stressors for many coastal ecosystems is eutrophication. Eutrophication is the result of excessive nutrient concentrations resulting in an imbalance of inland and marine waters causing excessive algal and water vegetation growth. This abundance of vegetation disturbs the dynamics of aquatic ecosystems and inflicts undesirable effects including dissolved oxygen depletion and choking of waterways (OSPAR, 2009).

In order to address these issues, the European Union (EU) formulated directives to promote fair and sustainable usage of water and marine resources, like the Water Framework Directive (WFD) (EU, 2000) and Marine Strategy Framework Directive (MSFD) (EU, 2008). The former includes specific policies for restriction of nutrient and other chemical discharges in the water through urban waste and agriculture, while the latter describes a holistic approach towards the preservation and sustainable development of the marine resources. Furthermore, the MSFD promotes the achievement of "good environmental status" by 2020, i.e. "environmental status of marine waters which are clean, healthy and productive", through ecosystem and biodiversity resilience to anthropogenic activities. In order to attain this, the current knowledge on marine environmental processes and how these are affected, must be enhanced. The importance of the preservation of such ecosystems is also underlined in UN's Agenda 2030 (United Nations, 2017) and several EU Horizon 2020[1] projects including "ECOPOTENTIAL", which focuses on monitoring and preserving Protected Areas (PAs), and "Green Infrastructure" promote their monitoring with the open-source satellite data that are available.

[1] Horizon 2020 or H2020 is the European Framework Program for Research and Innovation for the period 2014-2020 (European Commission, 2014).

1.2. Problem statement and research questions

The research context of this thesis resides predominantly in the fields of optical remote sensing and data science. It concerns the monitoring of water quality and associated eutrophication status of coastal ecosystems through the integration of multi-source data, including remote sensing observations and in-situ measurements. The area of interest is the Dutch Territorial Waters zone as it includes unique ecosystems with great economic and environmental benefits, while posing a significant challenge because of their optically complex Case 2 water[2] status (IOCCG, 2000).

The field of application and the study area were chosen based on the review of the scientific literature, the advances in the field of remote sensing for water resources management and the plethora of available in-situ data for the study area. The available data include various satellite missions dedicated to ocean monitoring (e.g. ENVISAT, AQUA, Sentinel-3) and the Rijkswaterstaat in-situ measurement database. In addition, Machine Learning can offer time efficient and robust tools to combine these data in a prediction scheme. Nonetheless, there is not a straightforward way to predict eutrophication; hence certain indicators and proxies (e.g. dissolved oxygen content), as well as their feasibility to be remotely measured, have to be identified and explored. All of the above, along with the need for better environmental monitoring, as reflected in the EU programs that promote the preservation of highly valuable ecosystems, assisted in the formulation of the main research question at the beginning of this thesis. This question followed by complementary sub-questions is presented below:

How adequate are freely available Earth Observation data for eutrophication monitoring and prediction in the Dutch coastal waters?

- Which remote sensing variables better describe the process of eutrophication?
- How do the remotely sensed variables relate to Dissolved Oxygen concentration, as an important eutrophication indicator?
- How does a data-driven statistical modelling approach perform in comparison with existing assimilative numerical models?

[2] Marine waters are classified as “Case 1” and “Case 2” based on their optical properties. The first refers to waters in which only organic matter affects those properties, while the latter describes the waters where inorganic components under suspension also affect their optical properties (IOCCG, 2000).

1.3. Research approach

The methodology that was designed to answer these research questions required a multidisciplinary approach, involving the understanding of ecological processes, biology, physics, optical remote sensing and in-situ data and the use of statistical tools for their analysis. First, a literature review on the topic of eutrophication was conducted and the main environmental indicators were identified. Then, the available remote sensing and in-situ data were examined, along with their potential to approximate the values of these eutrophication indicators. Finally, data science techniques were used for data exploration and processing and machine learning prediction models were implemented to estimate different indicators of eutrophication. A flow chart of the research approach can be seen in Figure 1.

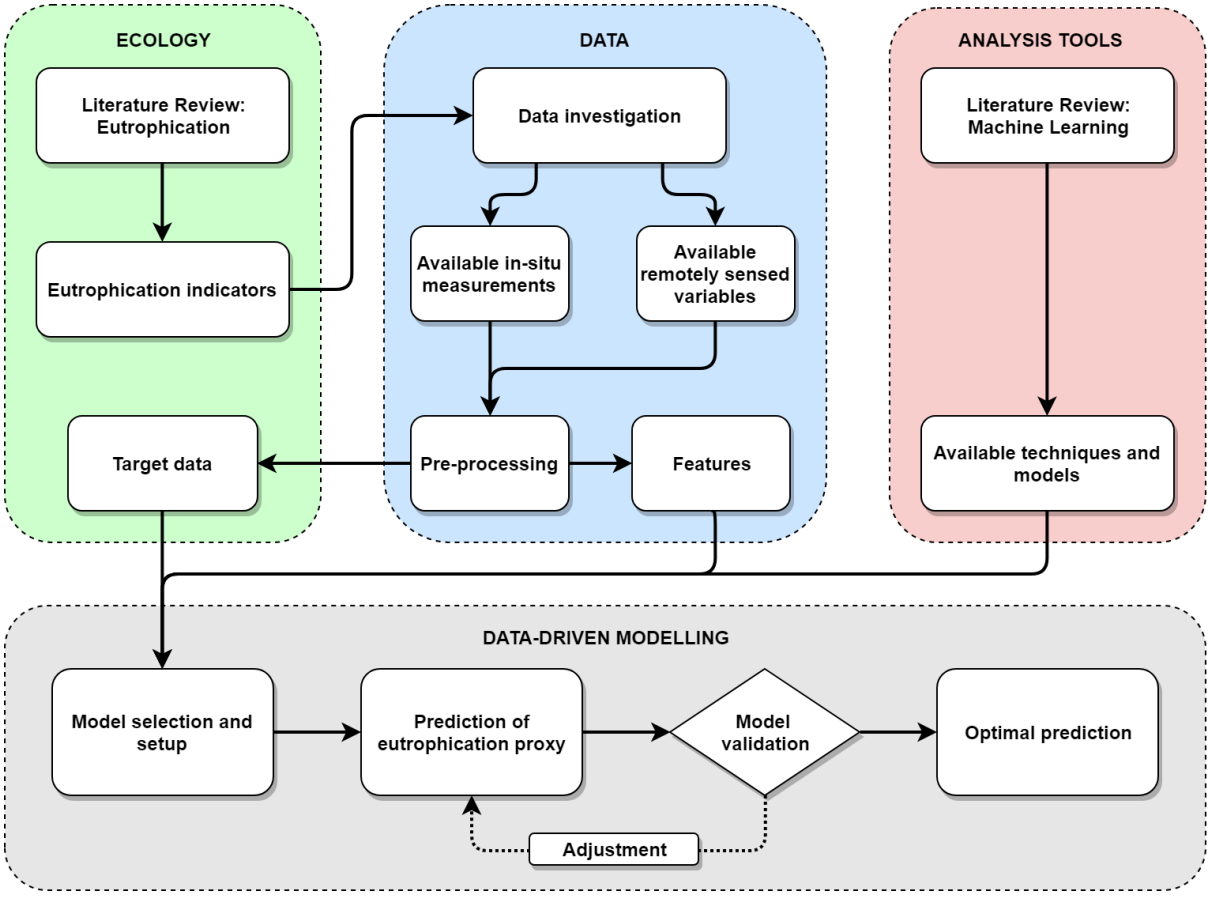


Figure 1 Flow chart of the research approach.

2. Eutrophication monitoring

In this chapter, certain fundamental concepts are introduced that are important in understanding the problem presented in this thesis, and in supporting its proposed solution. A more in-depth analysis of the physical and bio-chemical attributes of eutrophication, its effects on coastal ecosystems and their significance, as well as the common practices for its monitoring and assessment will be presented.

2.1. Eutrophication and coastal ecosystems

2.1.1. Coastal waters and water quality

Water is a fundamental resource for Earth's system, as the survival of all species and the thriving of the human society depend on it. Despite the many benefits derived from the technological developments in the last centuries, humans have damaged the environment and have affected its balance through their interventions. Water resources are threatened by these regimes, especially through pollution and changes in the geomorphology that affects water's route. Meanwhile, several standards have been set by legislation to perceive and protect the ecological status of various ecosystems, including the aquatic ones. Their quality, or chemical status, needs to be promoted and protected synergistically on a global scale (Funfak et al., 2014).

The water quality in coastal areas has been in decline around the world, as the absence of regulation concerning human activities led to eutrophication, increased turbidity and the loss of biodiversity. As mentioned before, oxygen depletion is an indicator of eutrophication and it affects mobile marine organisms (e.g. crabs, shrimps, fish etc.) that migrate away from the problematic areas. It also affects stationary organisms, e.g. oysters, clams and corals, on which it applies stress, lowering their growth and raising their mortality rates (Gobler, Depasquale, Griffith, & Baumann, 2014; Patterson, Boettcher, & Carmichael, 2014). Although oxygen depletion is a natural phenomenon that has been present in various marine ecosystems (e.g. deep basins, fjords etc.), its frequency of occurrence has increased in coastal waters (Wolanski *et al.*, 2013).

Coastal areas include rather diverse ecosystems, which are usually comprised of various habitats like salt marshes, estuaries, tidal flats and mangroves. Eutrophication threatens the balance of these habitats and it has been found to be a driver of salt marsh loss (Deegan *et al.*, 2012), zoobenthos increment (Bonsdorff *et al.*, 1997) and seasonal change of macrophyte (Bachelet *et al.*, 2000), among others. These ecosystems are among the most stressed natural systems in the world, because of their heavy human intervention and use. On a global scale half of the salt marshes and a third of mangroves, coral reefs and seagrass meadows have been lost or severely damaged (Barbier *et al.*, 2011).

2.1.2. Eutrophication

From the beginning of the 19th century until today, humanity's impact on the natural environment has increased rapidly. Urbanization, deforestation, air, soil and water pollution have shaped ecosystems and natural habitats to a great extent. Among others, the biochemical cycles of certain compounds such as nitrogen (N) and phosphorus (P) have been altered through anthropogenic activities from point sources (e.g. factories) and from diffusive sources (e.g. increased fertilizer usage in agriculture, incomplete household sewage systems, and atmospheric inputs). These compounds are fundamental for photosynthetic reactions and act as

catalysts for plant growth through their nourishing effects. Hence, when compounds like P and N are abnormally high within aquatic systems, primary producers (photosynthetic organisms), which are the basis of the food web as they produce organic matter by the transformation of inorganic compounds, may become over-abundant (OSPAR, 2009; Smith et al., 1998). This phenomenon is called eutrophication, which derives from the Greek word "eutrophos" ("εύτροφος") meaning "well-nourished". Ecosystems respond to eutrophication in many ways, with the most evident one being the "greening" of the water originating from the aftereffects of the nutrient enrichment to the algae and water vegetation (Diaz & Rosenberg, 2008). In addition, the WFD defines eutrophication as "*a process driven by enrichment of water by nutrients, especially compounds of nitrogen and/or phosphorus, leading to: increased growth, primary production and biomass of algae; changes in the balance of organisms; and water quality degradation. The consequences of eutrophication are undesirable if they appreciably degrade ecosystem health and/or the sustainable provision of goods and services*" (Ferreira *et al*., 2010).

Eutrophication of coastal waters has been an issue of increasing importance because of its rising frequency and its manifestation through marine phytoplankton blooms and Harmful Algal Blooms (HABs). These marine algal blooms imply the sudden proliferation of photosynthetic microalgae that, in some cases, are capable of producing toxins (Anderson, Cembella, & Hallegraeff, 2017). The diversity of the toxic species that compose these HABs is wide and can cause severe damage to the aquaculture products of the marine ecosystem (e.g. mass fish deaths, see Andersen, Hansen, & Engell-sørensen, 2015) and therefore negatively impact the respective economic sector. On the other hand, when eutrophication in marine ecosystems creates blooms of macro-algae in shallow estuarine waters, large areas of critically low dissolved oxygen in the water column are formed, while pH values rise (Smith et al., 1998).

According to MSFD's eutrophication quality descriptor, the effects of eutrophication are biodiversity loss, degradation of the ecosystem, harmful algal blooms and oxygen deficiency in the water (European Commission, 2010; Ferreira *et al.*, 2011). Furthermore, the assessment of the eutrophic state in the marine and coastal waters can be investigated through indicators that fall under three categories (European Commission, 2010):

1. Nutrient levels, concentration in the water column and ratios (silica, N, P);
2. Direct effects of nutrient enrichment, such as Chlorophyll-a (Chla) concentration, water transparency (an increase of suspended algae), an abundance of opportunistic macro-algae, species shift in floristic composition and anthropogenic HABs;
3. Indirect effects of nutrient enrichment, like impacts of decreased water transparency to the abundance of perennial seaweeds and seagrasses, as well as dissolved oxygen (DO) depletion.

More specifically, the direct effects of an increased primary production can be assessed through observations of Chla concentrations, an optically unique component associated with macro-algal presence (greenness of the water). These explosions of algae and macrophytes can also lead to subsequent mass die-offs resulting in high rates of decomposition in the base layer of the water column. The high decomposition rates include oxygen demanding reactions and are often correlated with decrease of submerged aquatic vegetation and HAB. Furthermore, as decomposition occurs nutrients are released which may dissolve back into the water column thereby providing a feedback mechanism. Increases of turbulence in shallow waters can release

even more phosphorus from the sediment, especially if the temperature is high (Ferreira *et al.*, 2011).

However, the presence of extensive nutrients does not directly lead to water eutrophication. For eutrophication to occur and persist, other favourable conditions are required in conjunction with nutrient levels such as adequate temperature, which supports accelerated bacterial growth. Another condition is low current velocity, which can limit the aeration of the water allowing the feedback mechanisms to deplete oxygen without additional introduction from outside sources (Yang *et al.*, 2008). Therefore, while current commonplace frameworks for determining eutrophication can give an indication based on one variable, more research needs to be conducted in order to map the underlying processes and the relationships between the different drivers of this phenomenon.

2.1.3. Hypoxia and Anoxia

The intensification of eutrophication leads to a profusion of organic matter which is decomposed by microbes that increase the respiratory demand. In this way, the DO content of the water is depleted leading to hypoxia (i.e. low DO), especially when there is stratification in the water column as this prevents oxygen dissolving in the upper layers from reaching the lower oxygen deprived layers . Then, with the continuous influx of nutrients, which is stored in the sediments, hypoxia becomes a seasonal phenomenon that periodically represses the life cycle of various species. In case this phenomenon persists in the local ecosystem and DO continues to fall, anoxia occurs. Anoxia limits the breakdown of organic matter by bacteria, resulting in a continual increase in the supply of organic carbon within the sediments and finally altering the normal nutrient cycle, furtherly reducing water quality. Anoxic conditions favour the formation of HABs, increase water's acidity and the microbial activity produces Hydrogen Sulfide (H_2S) which is toxic. Phenomena like these has been observed in various areas around the globe, like the Gulf of Mexico and the Danish waters (Diaz & Rosenberg, 2008; Howarth *et al.*, 2011; Wolanski *et al.*, 2013). Despite the mechanisms that deplete oxygen from the water though (e.g. water temperature and the respiratory demand of microbes), there exist processes that can add oxygen back in the system such as photosynthesis, stream flows and aeration (Gholizadeh, Melesse, & Reddi, 2016).

If the oxygen depletion conditions are allowed to continue, the eventual outcome is not just a deteriorated ecosystem, but rather dead zones. Engineers and ecologists have little to no power to reverse the ecological degradation as the management of these areas is challenging because of their scale and open boundaries (Wolanski *et al.*, 2013). More specifically, the water is described as hypoxic when the amount of DO drops under 2mg/L of water, resulting in abnormalities in the benthic fauna behaviour and shifts in its composition. Examples of such problematic areas have been documented in the Baltic Sea, Black Sea, Gulf of Mexico and the East China Sea, which are areas of important fishing-related economic activities. Furthermore, dead zone creation is more common in higher water temperatures, particularly during summer months, which incommode the solubility of oxygen and intensify water stratification (Diaz & Rosenberg, 2008; Wolanski *et al.*, 2013). Finally, a direct relation can be observed between the oxygen reduction in the coastal waters and the population explosion in the watersheds of the corresponding coastal areas that took place in the last 50 years, heavily increasing the local

nutrient accumulation. Figure 2 depicts the global human footprint map[3] along with known hypoxic systems (Diaz & Rosenberg, 2008). Here, it is evident that the areas where there is a clear correlation between human influence and hypoxia are the North-Western European coast, the East USA coast and the East Asia region.

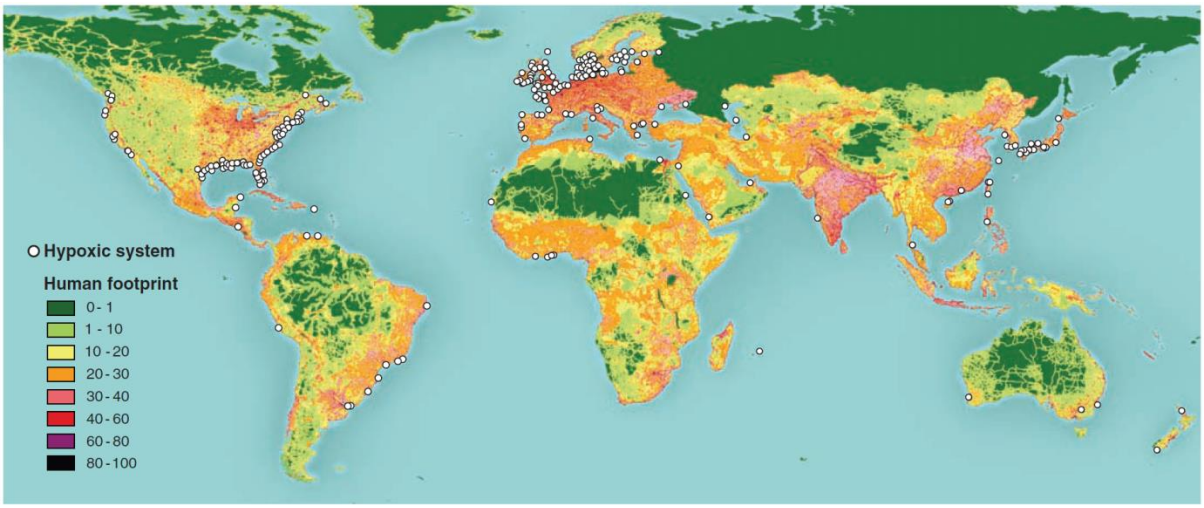


Figure 2 Reported eutrophication-induced dead zones in comparison to the global human footprint. The difference of reported hypoxic systems between the Northern and Southern hemisphere is due to the lack of research and reporting of such events in the Southern hemisphere (Diaz & Rosenberg, 2008).

2.2. Water quality monitoring

In order to assess the quality of marine water bodies the frequent measurement of causes (e.g. nutrients) and their effects on the ecosystem (e.g. eutrophication) is required. If this measurement frequency spans over a long period of time (e.g. years), then the term "monitoring" is attributed to the process. Monitoring programmes have multiple functions including enhancing system understanding, tracking key indicators of ecosystem, flora and fauna health, as well as providing necessary data to comply with national and supra-national reporting requirements. Additionally, they can aid in the development of models, projections, trend detections, and determination of possible causes for changes in the dynamics and composition of ecosystems (Karydis & Kitsiou, 2013).

The first practices of marine water monitoring were limited to toxic chemicals (e.g. heavy metal concentration). Emerging needs and interests, though, expanded the objectives of monitoring projects, which nowadays include a broader and more complex set of biological, chemical and physical variables. Meanwhile, the development of data analytics and technology helped in the incorporation of tools, such as statistical and spatial analysis methods and simulation models, which assisted in the direction of a better water quality assessment. Moreover, it is advised that monitoring projects should operate between pre-defined boundaries to be able to fulfil their objectives. Such boundaries are the geographical limits of the study area, but also the definition of a study period and the ecological limitations, i.e. exactly which processes are examined (Karydis & Kitsiou, 2013).

[3] The human footprint is a visualization of the global anthropogenic influence as a combination of individual ecological footprints proposed by Sanderson *et al*, (2002). It aims to present this influence as a continuous surface of patterns rather than single numbers.

According to the Eutrophication Monitoring Strategy of the United Nations Environment Programme (UNEP, 2007) every country is obliged to monitor certain water quality parameters that have been identified as relevant for the assessment of the degree of eutrophication or the prevention of entering a eutrophic state. These include temperature, pH, transparency, salinity, dissolved oxygen, chlorophyll-a, nitrates, silicates, nitrites, ammonium, orthophosphate and abundance of phytoplankton. In the past, water-related governmental authorities in various countries have established and funded broad-objective water quality programmes that failed to deliver, as they were lacking necessary preliminary research and tools to interpret their results (WMO, 2013). In order for these practices to become more specific, the World Meteorological Organization (WMO, 2013) underlines nine categories that can be used to better define a water quality monitoring program. Namely these include trend and background monitoring, basic, preliminary, operational and emergency surveys, modelling and impact surveys, as well as early warning systems. For instance, in non-problematic surface water monitoring programmes it is suggested that a background analysis and interpretation of the spatial and temporal trends can help in the definition of natural processes that can be excluded in case of emergency events (WMO, 2013).

Furthermore, the planning of such a programme should be comprised of three phases. The first phase includes the establishment of its necessity and its targets. The second phase concerns its design (e.g. monitoring locations for in-situ measurements and their frequency), the examined physical, chemical and biological variables, the choice of measurement techniques (e.g. in-situ and/or remote sensing) and the availability of data. The third and final phase consists of the integration of the above choices into an application that manages the data, implements and validates the defined methodology, presents results and constructs a product (WMO, 2013).

Traditionally, water quality status was assessed by laboratory analysis of in situ samples collected at the field. This process offers accurate estimations of indicators but requires a lot of time and hard work, especially when there is need for coverage of large areas. Additionally, an assessment of the spatial and temporal patterns of the various indicators is often a requirement for water quality studies, which is difficult to derive from point measurements (Gholizadeh et al., 2016). In recent years, advantages in remote sensing technology have broadened the spectrum of observation capabilities for various land covers and resources, including surface water. This enabled the formulation of low-cost tools that can assist decision makers, as well as help with the calibration and validation of existing environmental models. Nevertheless, despite the fact that many applications have been developed for the estimation of hydrological variables (e.g. precipitation, evapotranspiration), water quality variables remain critically unrepresented (Chang, Imen, & Vannah, 2015).

2.3. Remote sensing for water resources

The evaluation of water quality requires frequent and costly in-situ measurements, specifically in highly dynamic and large marine and coastal water bodiess. Remote sensing can assist in that direction, as it is a monitoring alternative with the capability to inspect large areas with high frequency. Remote sensing satellites carry sensors that measure the reflected electromagnetic radiation (EMR) from the earth's surface. If the radiation that they measure originates from another source (e.g. reflected sunlight) then the term passive sensor is attributed to them (e.g. imaging radiometer). Passive sensors measure in the visible and near-infrared part of the electromagnetic spectrum (Figure 3) and their acquisition of the earth's surface reflectance is

often affected by the presence of clouds and aerosols as they completely block it or corrupt it. On the other hand, if the satellite emits its own radiation and has a dedicated sensor that measures the same reflected radiation, then these sensors are called active (e.g. radar). The EMR produced by the active sensor platforms belongs in the microwave part of the spectrum (Figure 3) that usually can penetrate the clouds, the aerosols and sometimes the vegetation, while not being heavily affected by them (Chang et al., 2015).

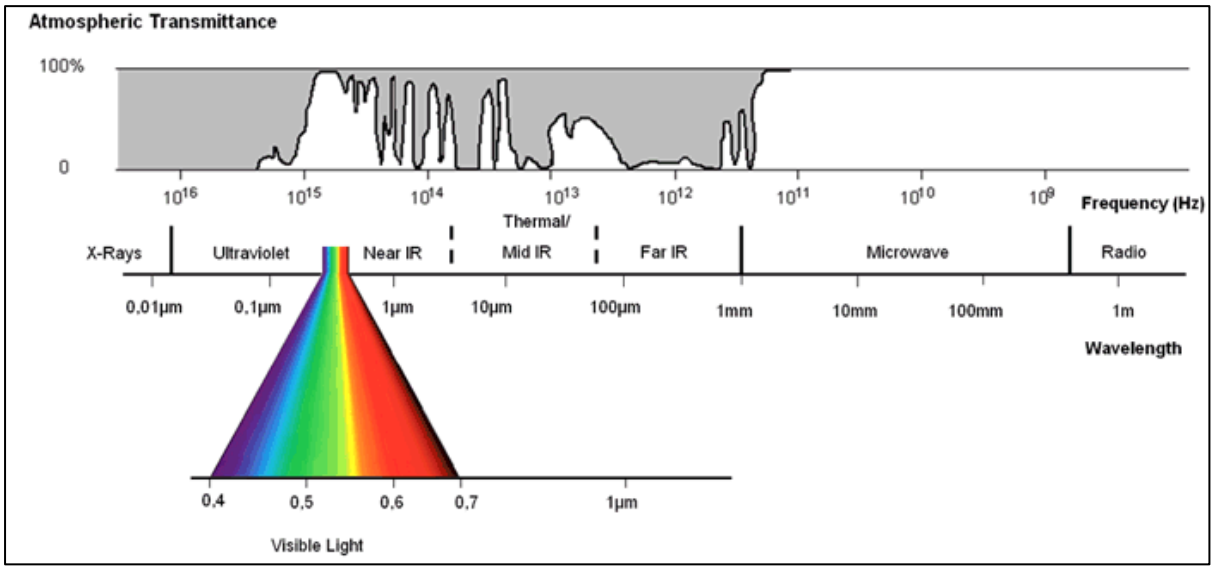


Figure 3 Electromagnetic spectrum and percentage of its atmospheric transmittance. Source: SEOS project (https://seos-project.eu/)

In passive remote sensing, the spectral signature of water depends on the combination of its compounds, i.e. the amount of the suspended organic and inorganic particles, which affect its scattering and absorption properties. For example, phytoplankton presence alters the spectral signature of the water substantially, making it feasible to distinguish between clear and trophic water in the visible part of the spectrum (Figure 4) and especially in the blue and red colour parts (Figure 5). Water emits a part of the absorbed radiation in the thermal infrared (TIR) part of the spectrum and passive microwave sensors measuring in that bandwidth can be used to determine the water's surface temperature and salinity. Active remote sensing measures in the part of the spectrum that the water is not reflecting (only absorbing) the incoming radiation, leaving it unable to retrieve much information from water covered surfaces. Nevertheless, it can still be used to measure the roughness of the water's surface and approximate the wave height, the wind speed above the surface of the water or the extent of oil spills (Solimini, 2016).

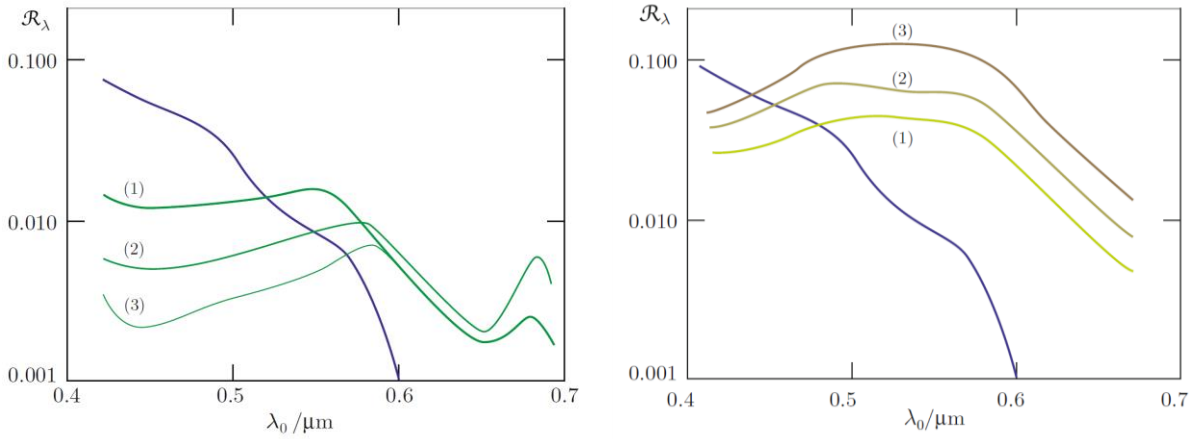


Figure 4 Reflectance (R_λ) in the visible part of the spectrum for different water types. Blue line indicates clear water, while (1), (2) and (3) lines correspond to water with increasing phytoplankton (left) and inorganic sediment (right) content. Source: (Solimini, 2016). Note: λ_0 = wavelength.

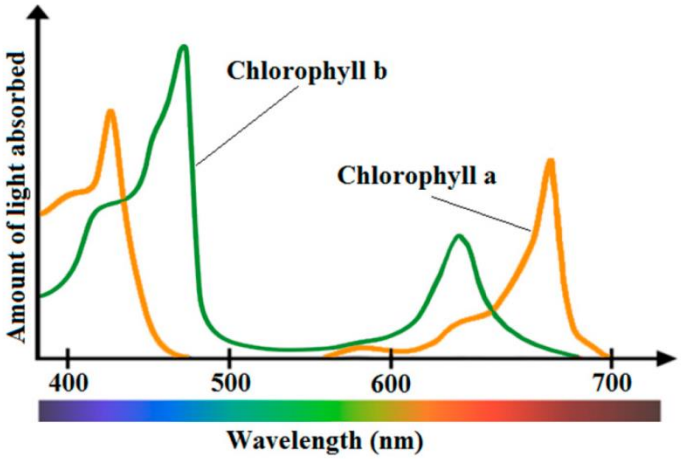


Figure 5 Spectral signatures of different algal pigments. Source: (Gholizadeh et al., 2016).

Among the variables that can be estimated from the satellite data are the Sea Surface Temperature (SST), which is one of the most important and common studied parameters, Chla concentration, which is used as a proxy for the estimation of phytoplankton biomass, Total Suspended Matter content (TSM), which is used for water transparency estimation and water current movement observation (Kim *et al.*, 2017).

Detection and tracking of the water resources is supported by the efforts of ESA[4] and NASA[5] who are providing free satellite data from missions dedicated to the ocean (e.g. ENVISAT MERIS, Sentinel 3 OLCI, AQUA-MODIS, SNPP VIIRS), as well as general monitoring (e.g. Sentinel 2, Landsat 8). These data range from raw top-of-the-atmosphere (TOA) reflectance (Level 0) to per grid cell water quality or geophysical parameter value (Level 2/3, e.g. Chla, TSM, SST, Coloured Dissolved Organic Matter (CDOM), Diffused Light Attenuation Coefficient (Kd)). Nevertheless, even satellite data of higher levels of processing contain errors and uncertainties and therefore need to be explored and corrected before their use for a specific application or

[4] European Space Agency (ESA)
[5] National Aeronautics and Space Administration (NASA)

study area. Normally, quality information for each pixel of the provided images is available, like an error metric or a flag that is raised if conditions that degrade the data are present.

Other important factors of these datasets are their spatial and temporal resolution. The former regards the level of detail with which the measurements are depicted on the images, while the latter concerns the revisit time of the satellite over the same location, or in other words the frequency of the measurements. Thus, depending on the spatial and temporal scales of the phenomena that are being studied, a first step towards choosing a useful remote sensing dataset would be its capabilities to capture the explored phenomena. For example, in case coastal phenomena are being studied, measurements of higher frequency can be beneficial in capturing the highly varying spatial extent of processes affected by the dynamics of the area.

2.4. Remote sensing for eutrophication

Several studies regarding eutrophication monitoring using remote sensing have been conducted. Most focus on the assessment of the phytoplankton biomass, monitoring of HABs and available nutrients, mainly through the estimation of Chla content, while only a few of them have attempted to explore the impact of the eutrophication process on the DO content.

More specifically, the report on eutrophication monitoring guidelines using remote sensing for the North-West Pacific Region (CEARAC, 2007) underlines the importance of all the available remotely sensed variables (e.g. Chla, SST, TSM, Kd, CDOM etc.) and in-situ measurements (e.g. nutrients, Chemical Oxygen Demand (COD[6]), etc.). In the four case studies it includes though, while various in-situ measurements of Chla, COD and nutrients are used, remotely sensed Chla products are used only for red algae detection, investigation of correlations or for identification of spatio-temporal Chla patterns. In a report issued by Deltares (Blaas, 2013) concerning eutrophication assessment capabilities the use of remotely sensed and in-situ Chla data is promoted as an important tool for the monitoring of the process, with the spatial and temporal coverage being invaluable. Ibrahim *et al.* (Ibrahim, Mabuchi, & Murakami, 2009) developed algorithms for the detection of five different pigments in hyperspectral data, concerning suspended matter and chlorophyll-a variables. In the preprint version of a paper by Hess *et al.* (Hess, Duvvur, & Srinivasan, 2019), the authors make use of Convolutional Neural Networks (CNN) in an attempt to predict HABs and their future patterns based on Chla satellite estimations. All the aforementioned studies, even though they approached the topic from a different angle, have had remarkable results regarding the primary eutrophication impacts.

On the other hand, studies that attempted to address some indirect effects of eutrophication, e.g. by exploring different oxygen-related variables, have been rarer and less promising. A review paper published in 2016 (Gholizadeh et al., 2016) sited twelve different studies that have attempted to remotely sense DO, COD and Biochemical Oxygen Demand (BOD) using various sensors (e.g. Landsat 5, WorldView 2, MODIS, MERIS). In these studies, connections between observed reflectances or band ratios and oxygen-related water variables were examined using linear, exponential and logarithmic regressions. The authors argue that there is no single sensor or methodology that offers a solid insight for the prediction of these variables as their relationship with the spectral information of the images is not yet fully understood. They conclude that there are little opportunities in the reviewed papers unless extensive in-situ

[6] Chemical Oxygen Demand is an indicator for the oxygen that can be consumed in water based on the available organic compounds.

measurements of both water variables and ground reflectances for atmospheric correction are being used. It should also be mentioned that out of these studies, only two dealt with DO as a part of a set of pollution variables that they examined through a land cover change inspection around rivers (He, Oki, Wang, & Oki, 2009; Somvanshi, Kunwar, Singh, Shukla, & Pathak, 2012).

In a paper published in 2015 (Muchlisin, 2015), the author attempts to predict DO from Landsat water temperature estimations. A positive correlation between the 28 in-situ observations of DO and the in-situ measured water temperature is reported (Figure 6), which is the opposite of what was expected. The author attributes that behaviour to calibration inconsistencies of the used instrument and to the unique natural conditions of the study area. Also, he suggests that the Landsat water temperature estimation is loosely correlated with the in-situ measurements (Figure 7), which he attributes to the different acquisition time of the two sets of measurements. Concluding, the author argues about the possibilities in the estimation of DO and for the necessity of better measurement instruments and methods. Furthermore, more robust results have been given by researchers that combined an assimilative ocean circulation model (US Naval Research Laboratory Ocean Forecast System) and MODIS SST estimation along with river discharge data to account for salinity fluxes in the Louisiana Continental Shelf region. In addition, light attenuation and salinity remotely sensed variables assisted in the improvement of their result (Ko, Gould, Penta, & Lehrter, 2016). All in all, the research topic of this thesis that was set in chapter 1 can be considered relatively unexplored after taking into account the findings of the literature presented above.

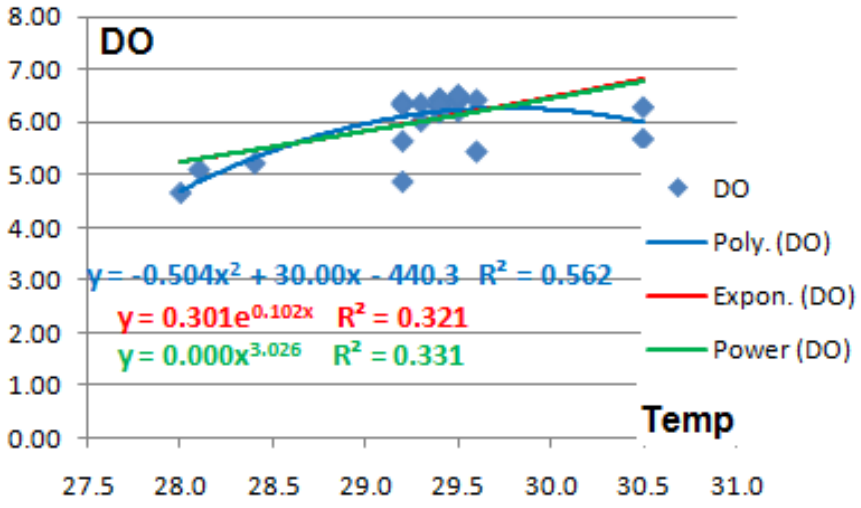


Figure 6 Proposed models by Muchlisin (2015) for the relation of DO with in-situ water temperature.

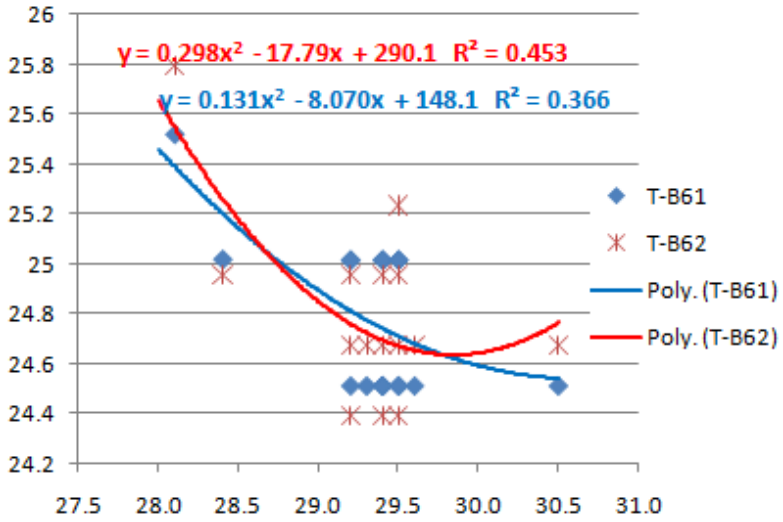


Figure 7 Proposed models by Muchlisin (2015) for the relation of in-situ water temperature with Landsat derived SST.

2.5. Chapter conclusions

The literature review on eutrophication process monitoring revealed which indicators need to be measured in order to draw conclusions about the environmental status of a marine ecosystem. Nutrient levels, algal blooms, marine vegetation and dissolved oxygen content in the water were identified as the most important indicators. From these, only algae and vegetation alter the optical properties of the water and can therefore be measured by satellite sensors. On these grounds, prior research has mainly focused on algal bloom measurement and prediction through the estimation of Chlorophyll-a content, which is directly related to their abundance. Meanwhile, it was argued that there is not a straightforward way to estimate dissolved oxygen content, as there is lack of understanding of its relationship with the spectral information of the images. This limitation reveals a knowledge gap and an opportunity for this thesis to expand the toolset of ecologists with the indirect estimation of dissolved oxygen using variables that are known to affect its presence, according to the literature review. For this reason, the thesis focuses on dissolved oxygen prediction.

3. Satellite missions, in-situ and model data

In this chapter the available satellite data with a focus on this research's objectives are documented. This means that certain datasets and products regarding land or atmospheric products are intentionally not mentioned. Furthermore, based on the literature review presented in chapter 2, it was decided that only passive sensors and their corresponding variable estimations are going to be considered, as they are assumed to be more relevant to the studied topic. Moreover, in-situ and modelled data for several locations along the Dutch and European coasts are available through various organizations (e.g. Rijkswaterstaat, EMODnet, COPERNICUS). Remote sensing data are often used to validate these models and assist in their calibration. The combination of this wealth of available data for the aquatic domain could assist towards an improved monitoring of eutrophic waters.

3.1. AQUA

The AQUA satellite was launched in 2002 and still remains operational providing several data products (Table 1) for ocean monitoring (OBPG-NASA, 2018). It carries six different sensors dedicated for different applications, namely the Atmospheric Infrared Sounder (AIRS), the Advanced Microwave Scanning Radiometer (AMSR-E), the Advanced Microwave Sounding Unit (AMSU-A), the Clouds and the Earth's Radiant Energy System (CERES), the Humidity Sounder for Brazil (HSB) and the Moderate Resolution Imaging Spectroradiometer (MODIS). The objectives of the mission are to support the understanding of mainly water resources, but also of other components that react with the earth's climate system, as well as to improve weather forecasts (NASA, 2006). For this thesis project only the MODIS sensor is be considered as it provides adequate ocean colours and temperature products.

3.1.1. MODIS

The MODerate resolution Imaging Spectroradiometer (MODIS) on NASA's AQUA satellite provides Ocean Colour and Sea Surface Temperature data with a temporal resolution of one to two days depending on the location. The spatial resolution for Level 1 and 2 data in the spectral region that concerns ocean monitoring (Table 13 – Appendix) is 1km at nadir but Level 3 (global data) are available at 4km and 9km spatial resolutions as well. Its data are available, among other ocean datasets, on NASA's Ocean Biology Processing Group website along with their description and related documents. Furthermore, the quality flags of the dataset are provided and can be seen in Table 14 (Appendix). By default, the flags for land, cloud and saturated pixels are on; hence they do not need to be applied when the datasets are being processed. An example of MODIS global Chla content for spring of 2014 can be seen in Figure 8.

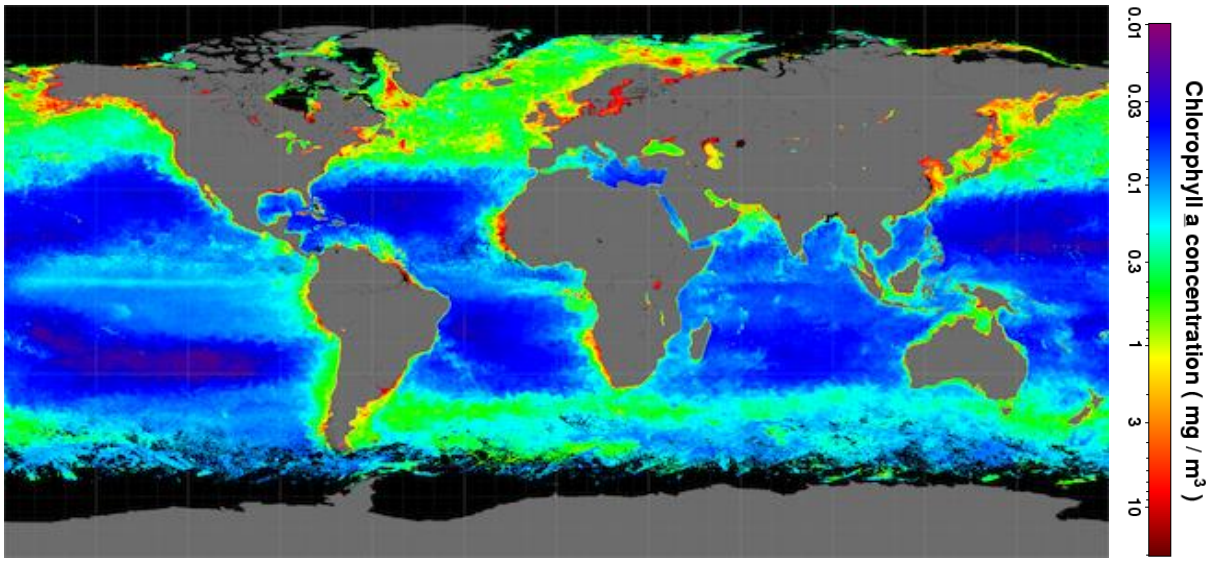


Figure 8 AQUA MODIS average Chla product for the spring of 2014 with spatial resolution 9km. Source: https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/[7].

Table 1 MODIS Ocean Colour Level 2 data product (NASA, n.d.-b, 2010).

Param.	Description	Units	Bands Used	Algorithms
Chla	Chlorophyll-a concentration	$\boldsymbol{mg/m^3}$	440-670nm	OCx and CI (Hu *et al.*, 2012)
K490	Diffuse attenuation coefficient at 490nm	$\boldsymbol{m^{-1}}$	490nm 547-565 nm	
PIC	Calcite Concentration (Inorganic Carbon)	$\boldsymbol{mol/m^3}$	443 & 555nm 670, 750, 870 nm	Balch *et al.*, 2005; Gordon *et al.*, 2001
POC	Particulate Organic Carbon	$\boldsymbol{mg/m^3}$	443nm, 547-556nm	
SST	Sea Surface Temperature	Degrees Celsius	SW: 3959 & 4959nm, LW: 11000-12000nm	Walton *et al.*, 1998

[7] Accessed on 22/07/2019.

3.2. ENVISAT

3.2.1. MERIS

ENVISAT's MEdium Resolution Imaging Spectrometer (MERIS) was a push-broom imaging spectrometer that measured Earth-leaving solar radiation in 15 spectral bands in the Visible and Near-Infrared (NIR) part of the spectrum. Its products are available in 3 pre-processing levels (0, 1, 2), in 3 spatial resolutions and have a temporal resolution of 1 to 2 days depending on the geographical location (Table 2) (ESA, 2006).

Table 2 MERIS different resolution products (ESA, 2006).

Product name	Spatial resolution [m]	Temporal resolution [days]
Full-Resolution (FR)	260 x 290	
Reduced-Resolution (RR)	1040 x 1160	1-2
Low Resolution (LR)	4160 x 4640	

MERIS's ocean mission objectives were to estimate the potential of phytoplankton to photosynthesize, to detect dissolved organic matter in the water as well as suspended matter originating from riverine sediment discharged in the sea. Meanwhile, detection of planktonic blooms, assessment of the water quality and coastal evolution observation could be attributed as secondary objectives (ESA, 2006). MERIS Level 2 marine data products can be seen in Table 3. Furthermore, the data are provided with their corresponding standard errors and an auxiliary layer of quality flags, which can be used to assess the uncertainty affecting the selected pixel values. In Table 16 (Appendix), these flags along with their descriptions are available, and the relevant ones to this project are highlighted with light green colour. An example of its Chla products can be seen in Figure 9.

Table 3 MERIS Level 2 marine products (ESA, 2006).

Name	Description
Chla	Chlorophyll-a content in $\boldsymbol{mg/m^3}$
TSM	Suspended particulate matter content in $\boldsymbol{g/m^3}$
CDOM	Attenuation of radiative flux due to dissolved organic matter in the sea water in $\boldsymbol{m^{-1}}$.
Kd	Spectral extinction coefficient in $\boldsymbol{m^{-1}}$

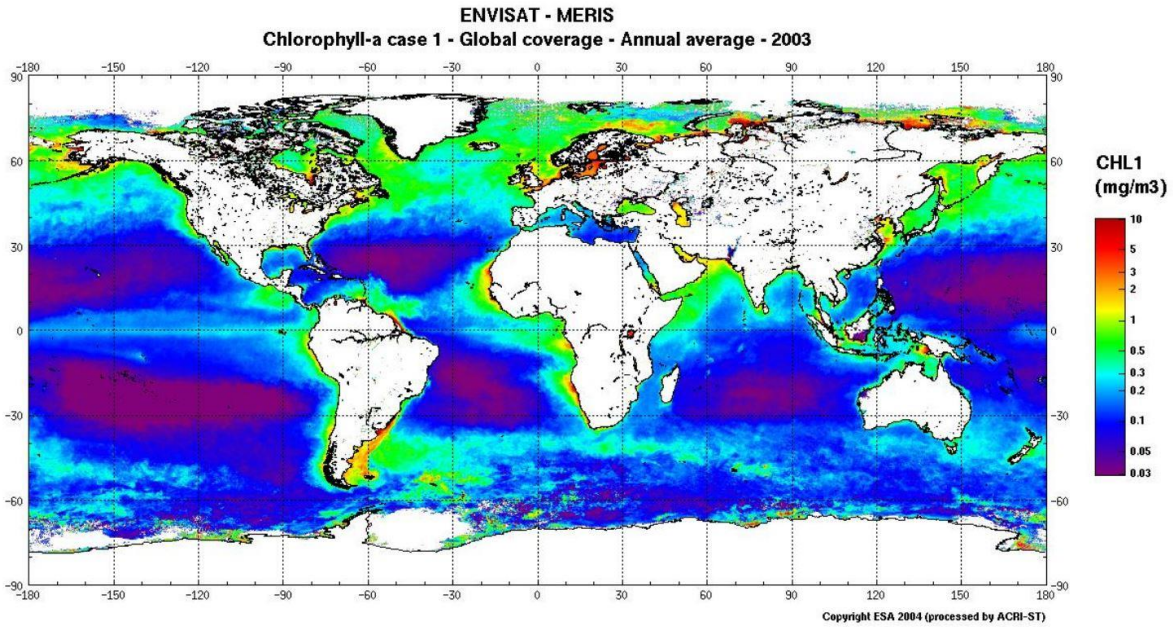


Figure 9 ENVISAT MERIS Chla product of annual average for 2003. Source: (ESA, 2006).

3.2.2. HYDROPT Algorithm for MERIS

Most Level 2 and 3 products use different algorithms to estimate variables (especially chlorophyll-a) in coastal waters (Case 2) that are more turbid, and in the open ocean (Case 1). For MERIS, two different algal pigment products are available for Case-1 and Case-2 waters, similar to MODIS (Table 1). These two products are freely available but a third algorithm called HYDROPT was developed specifically for complex coastal waters like the ones of the North Sea (Woerd & Pasterkamp, 2007). HYDROPT has been proven to perform better than pre-existing methodologies that integrated absorption models with satellite measurements (Blondeau-Patissier *et al.*, 2004; Gohin *et al.*, 2005). Initially, the HYDROLIGHT radiation transfer model (Mobley, 1994) was used to train a neural network which, along with a set of optical properties, was implemented to derive the variable estimates from MERIS data. However, the need for faster computation in order for the estimation to be used in real-time and the lack of flexibility in the previous model led to the development of HYDROPT, which is reliable in the range of 1 to 20 mg/m^3 of chlorophyll-a. Additionally, standard errors are provided per pixel corresponding to the uncertainty of the observation based on the optical model used (Woerd & Pasterkamp, 2007).

3.2.3. AATSR

The Advanced Along Track Scanning Radiometer (AATSR) was designed to measure the Sea Surface Temperature with great precision and it was one of the sensors on ENVISAT. It had a 500km swath, its spatial resolution was 1 km at nadir and it had a relative radiometric accuracy better than 0.1 K. Furthermore, it utilized three visible and four infrared spectral bands (Table 4) (Cardaci, 2013).

Table 4 AATSR Spectral Bands (Cardaci, 2013).

Band	1 Visible	2 Visible	3 Thermal	4 near-IR	5 Infrared	6 Infrared	7 Infrared
Bandwidth [μm]	0.555	0.659	0.865	1.61	3.7	10.85	12

3.3. In-situ measurements

The Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat – RWS) offers in-situ data for the Dutch inland and marine waters. The data acquisition interval can range from five minutes (water temperature) to one month (Chla, DO) and among others concern water level, eutrophication parameters and nutrients (e.g. Chla, Dissolved P and N), general water quality (DO, acidity, TSM) and water surface temperature. These measurements are collected from point locations and from around 0 to 3 meters of water depth. Depending on the variable, they are measured continuously using dedicated permanent platforms (e.g. water temperature) that provide near real-time information, or by collecting samples and analysing them later in the lab (e.g. Chla). Despite the very dense network of measuring locations found in RWS's data distribution portal, many of them seem to have been abandoned in 1978 and 1995, while others that continue until today include big gaps. Additionally, the variables that require lab analysis take substantially more time to be available online, e.g. at the time of writing this thesis (mid-2019) the latest available in-situ Chla and DO data for the remaining stations are from December 2017.

According to the Ministry's website (www.helpdeskwater.nl), each variable is being measured using different tools and techniques. For the calculation of water temperature a platinum-coated electrode that is attached to a galvanometer is put in the water. Then an electric current is send through the electrode and the changes in the resistance of the wire are used to determine the water's temperature. Meanwhile, for the computation of DO a water sample is taken to the lab, where a golden and a silver electrode are put separately in the water. The different electrochemical reactions produced by the golden and silver wires are used to determine the amount of DO in the sample. All in all, these in-situ measurements are going to be considered as ground-truth data despite the uncertainties in their measuring methods, as they are direct estimations of the variables in question. An example of a buoy collecting water quality data, as well as a Rosette water sampler that collects water samples along the water column can be seen in Figure 10.

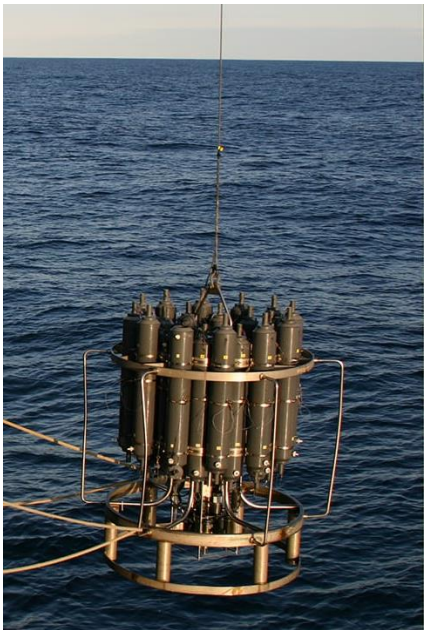


Figure 10 A buoy station and a Rosette water sampler. Sources: http://eurogoos.eu/member-product/real-time-data-oceanographic-buoy-vida/[8], https://www.whoi.edu/what-we-do/explore/instruments/instruments-sensors-samplers/rosette-sampler/[9]

3.4. Process-driven models

Copernicus[9] offers via its Copernicus Marine Environmental Monitoring Service (CMEMS) various ocean-related model data as well as in-situ and satellite data. These models can be purely numerical, i.e. rely on known physical laws and mathematical equations, or they can assimilate in-situ, satellite and other data. For the study area and period of this project, CMEMS provides data derived from a coupled hydrodynamic-ecosystem model, the Forecasting Ocean Assimilation Model for the European North-west continental shelf at 7km spatial resolution (FOAM AMM7). This model incorporates data from more general, sometimes global, models for a better representation of its boundaries and the exchange with the greater North Atlantic. It takes into account the tides, salinity, air and sea temperature, sea surface height and current information for all depths at its open boundaries (CMEMS, 2018).

More specifically, FOAM AMM7 uses a physical ocean model called *Nucleus for European Modelling of the Ocean* (NEMO v3.6), coupled to the *European Regional Sea Ecosystem Model* (ERSEM v15.06). Furthermore, its exchange with the Baltic Sea in terms of temperature and salinity is modelled by the CMEMS Baltic MFC system for freshwater river discharges, accounting for 279 rivers in the area. Weather data are provided by a Met Office's numerical model incorporating heat and moisture fluxes every three hours and hourly wind speed and pressure information. Additionally, diffuse light attenuation in the water column is modelled by POLCOMS (Proudman Oceanographic Laboratory Coastal-Ocean Modelling System) and used in FOAM. As a final step of the FOAM modelling procedure, in-situ measurements and satellite data (SST estimations, altimetry data etc.) are used to assimilate the model, while the 3DVar NEMOVAR system to assimilate remote sensing data regarding Chla concentration (CMEMS, 2018). In conclusion, FOAM AMM7 is a fully integrated hydro-bio-geo-chemical model that can offer complex insights about the overall situation of the area. On the other hand, since the model

[8, 9] Accessed on 22/07/2019.
[9] COPERNICUS is EU's Earth Observation programme (www.copernicus.eu).

is calibrated to have the best representation of the entire Northwest European shelf, it is expected to not prioritize on the correction of errors in local systems (e.g. Dutch coastal waters) during its assimilative process. Below, a map of the DO content prediction that FOAM offers for an example date can be seen (Figure 11).

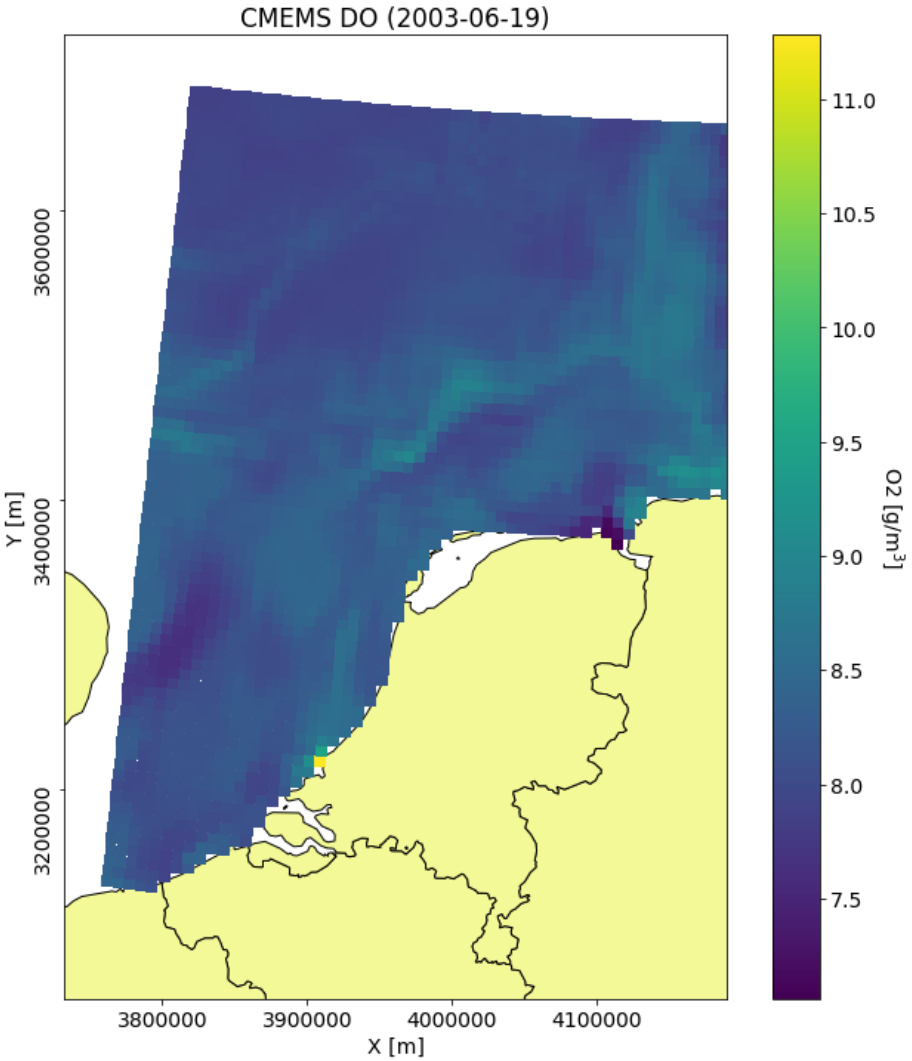


Figure 11 CMEMS FOAM AMM7 Dissolved Oxygen estimation example for 19th of June 2003.

3.5. Chapter conclusions

An extensive data investigation was performed for the available datasets that can help answer the research questions of this thesis. AQUA and ENVISAT satellites were chosen and showcased above as their long time series and overlap with ground truth data provided by Rijkswaterstaat fulfils the criteria for the implementation and validation of the methodology. More specifically, the standard ocean colour and SST products of MODIS, the SST from AATSR and the optimized MERIS ocean colour products for the complex waters of the North Sea (HYDROPT) were found to be the more adequate. In regards with the available process-driven model outputs, the FOAM AMM7 model from Copernicus provides a prediction for DO for the study area. Its prediction is based on multiple model outputs, satellite and in-situ data and is freely available online, thus it was considered a high-quality product to use in this thesis and to answer the third research sub-question in particular.

4. Machine learning

With such a wide variety of data from multiple sources on key indicators of eutrophication, manual methods of generating relationships between the datasets would be labour intensive and time consuming. Therefore, machine learning techniques have been identified as key features that can aid in the modelling of the relationships between the variables. With the definition of these relationships, the development of predictive models for DO and thereby eutrophication status can become possible. Based on the literature review, Random Forests, Gradient Boosting machines and Artificial Neural Networks are some of the most commonly used and well-performing models in the machine learning field. In the following chapter, these models are presented, as they can assist in the estimation of the examined variables and phenomena from remote sensing observations.

4.1. Background

Artificial Intelligence (AI) is the scientific field that aims at automating computers to perform processes and tasks that are normally performed by humans. This can be achieved by explicitly programming a computer to act on a defined set of rules that humans have handcrafted for it (e.g. early chess programs) or by "teaching" it to create its own rules or approximate a set of rules based on the data that it is presented with. This teaching process that gives the computer the ability to adapt to new unseen input data is what is often called machine learning (ML) and it is only a subset of the field of AI (Chollet, 2017).

In order to perform a ML task to estimate a target, two sets of data along with an assessment method are needed. The required datasets are the input data (also known as features) that the computer will use to make the prediction and the target or expected data (or labels) are the ones that it will try to approximate. The ML model applies various transformations on the input data so that they resemble the target data. These transformations need to be assessed and adjusted accordingly based on their ability to make a successful approximation (Chollet, 2017).

A subfield of ML is called Deep Learning (DL) (Figure 12) where the learning process is being realised by using "*successive layers of increasingly meaningful representations*" of the data. These layers are put in a sequence and they create models that are called Artificial Neural Networks (ANN) (Chollet, 2017).

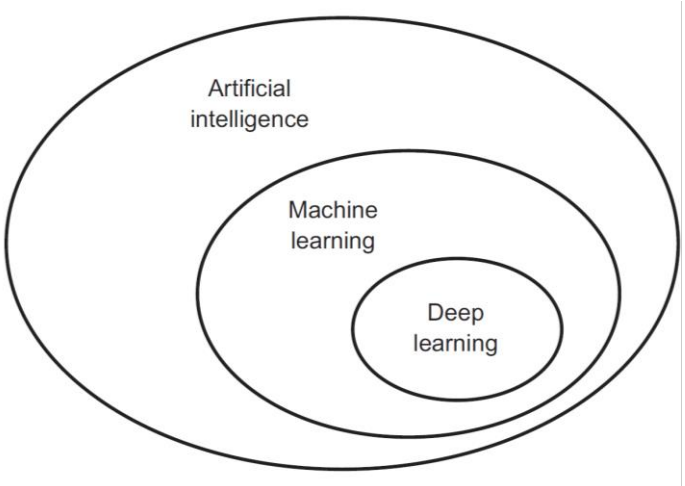


Figure 12 Relations between AI, ML and DL. Source: (Chollet, 2017).

4.2. Important concepts

ML can be used for Classification or for Regression problems. In classification problems, the target data are discrete or categorical (e.g. water or land), they have a finite number of classes and the input data are assigned to one of them. On the other hand, in regression problems input data are used to approximate target data that are normally continuous (e.g. Chla content between 0 and 100 mg/m^3) (Mohri, Rostamizadeh, & Talwalkar, 2012). In other words, Classification is used for qualitative prediction, while Regression is used for quantitative prediction (Hastie, Tibshirani, & Friedman, 2009). Another categorization of the ML models can be made based on the presence or not of training labels. If training labels are available, the ML model has a priori knowledge on what it tries to predict and this type of learning is called *supervised*. In case that there is a lack of training labels, the ML model can still find patterns in the input features and group them in similar clusters. This is called *unsupervised* learning (Theodoridis & Koutroumbas, 2009). Since the rough research goal is to determine the DO content, from now on the text will focus on supervised regression problems. The reason for this is that DO is a continuous variable and other remotely sensed continuous variables (e.g. SST) will be used to approximate it.

4.3. Ensemble learning

A lot of commonly used and well-performing models for regression belong to the so-called Ensemble Learning methods (Chollet, 2017). Ensemble learning is a subspace of machine learning that uses multiple simple models in order to increase the confidence of the proposed solution for the problem at hand (Figure 13). In other words, an ensemble model consults many models with a similar bias and then it combines their individual decisions, reducing the overall variance of the final decision. Thus, while each individual model will carry its own bias that is different from the others, there will be an agreement for the correct decisions, reinforcing their power in the ensemble and reducing the average bias (Polikar, 2012).

Typically, an ensemble learning model comprises a method to select subsets of the dataset that makes the model more diverse, a predictive model type to use on each subset and a mechanism to combine these models. For the first component, the dataset is normally either split randomly (e.g. bagging) or the selected subset is updated iteratively in order to fulfil certain criteria (e.g. boosting). Then, a simple model, also known as base learner, (e.g. decision tree) is often chosen. Finally, for the combination mechanism, usually a simple or weighted majority voting (classification) or averaging (regression) scheme is set (Polikar, 2012) .

Furthermore, ensemble methods are usually favoured over single models in case there are limited target data to represent the target space, as by sub-setting and averaging the different predictions, they can increase the confidence on the choice of the model. Meanwhile, it is argued that single models can fall into the pitfall of getting stuck in local optima and hence under-represent the target space, while ensemble models by combining many predictors that have varying starting points can better represent the underlying unknown process (Dietterich, 2000). All these attributes of ensemble learning can be crucial within the framework of this thesis, as the variety of data sources that are required to overlap can limit the generalization potential of a DO predictive model. Moreover, the unknown function that connects the chosen features (remote sensing variables) to the target data (in-situ DO) encapsulates complex physical and bio-chemical processes that would be less accurate to model with a single predictor.

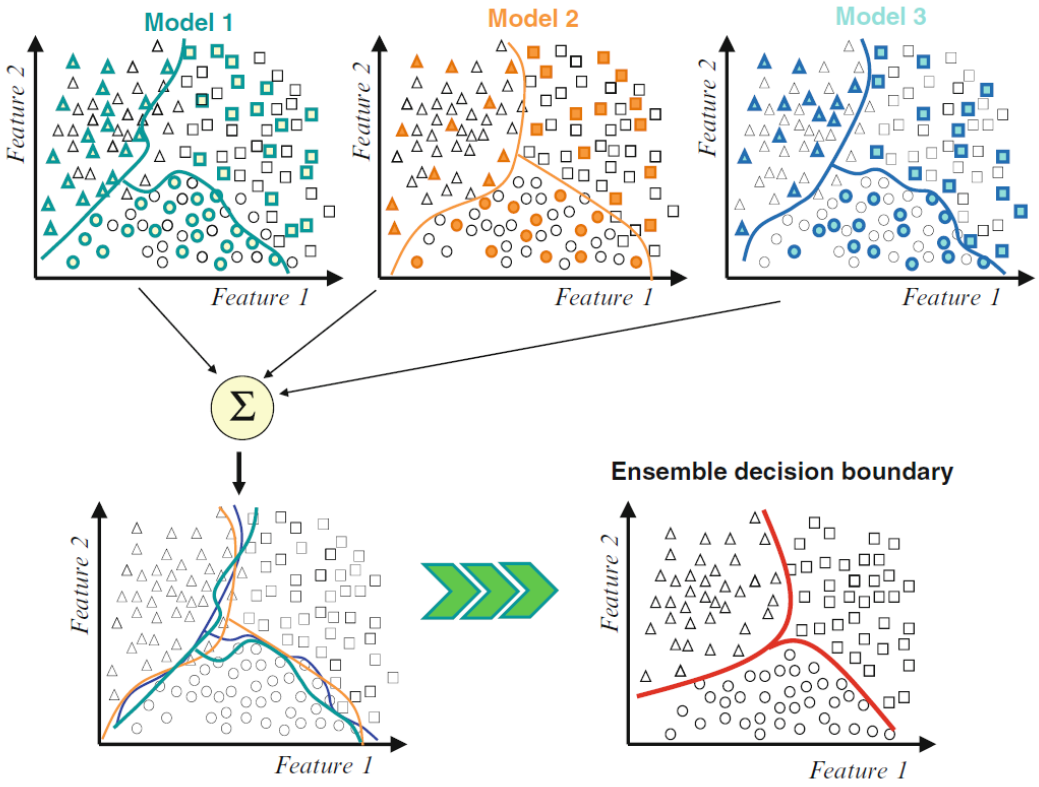


Figure 13 Ensemble learning scheme. Source: (Polikar, 2012).

4.3.1. Tree-based models

The cornerstone of ensemble methods are the individual models they use. The simplest models are the tree-based ones and the most commonly used is the Classification and Regression Tree (CART) (Breiman, Friedman, Olshen, & Stone, 1984). Regression trees split the feature space using decision boundaries (or hyperplanes) that are perpendicular to one of the feature axes. Furthermore, trees are composed of three types of nodes, namely the root node, the internal nodes and the leaf nodes (Figure 14). The root node contains all the data samples as well as the first splitting criterion. The data samples that satisfy the criterion are split from the rest of the samples and both parts form two separate internal nodes. For regression trees this split criterion can be minimization of a loss function, e.g. sum of squared differences. The same procedure continues recursively until a desired tree depth is reached or another stopping criterion is fulfilled. The terminal nodes that finished splitting are called leaf nodes and they contain the decision of the prediction for each group of the data samples (Hastie *et al.*, 2009; Maimon & Rokach, 2010).

The way a tree is designed can greatly affect its performance. For instance, very large trees are known for failing to generalize and predict unseen data, while small trees are known to lack the needed complexity. Thus, growing rather large trees and then pruning them with various restrictions, such as maximum allowed nodes, is often a good practice (Hastie et al., 2009). Various models, including CART trees and ensemble methods, along with many options for pruning and fine-tuning are available in Python's *Scikit-learn* package (Pedregosa, Weiss, & Brucher, 2011), which is the tool of choice for ensemble method implementation in this project.

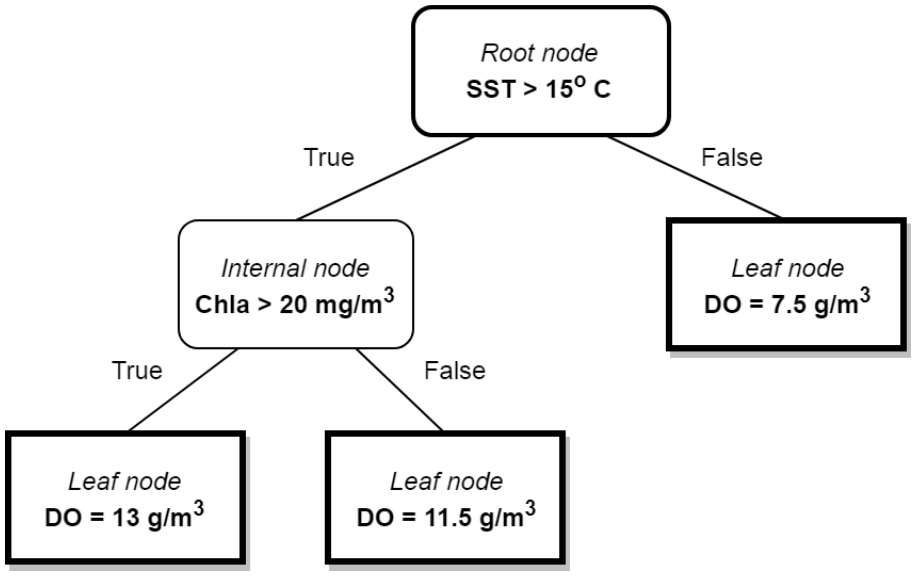


Figure 14 Example of a simple regression tree.

4.3.2. Random forest

Random forest (RF) is an ensemble learning method that creates a big number of simple decision trees, which are implemented to random subsets of the feature space. Then, in classification problems the prediction is being made using a majority voting scheme between trees, while for regression problems the prediction equals to the mean of all individual regressions averages them (Cutler, Cutler, & Stevens, 2012; Hastie et al., 2009). This method incorporates the technique of *bootstrap aggregation* (bagging) and the logic behind it is that by averaging many noisy and "weak" trees that are independent of each other, a lower prediction variance can be achieved. Bagging has been proven to reduce the errors in classification of unstable procedures (Breiman, 1996).

RF models work well without much tuning, but there are still certain parameters that can be altered in order to improve the overall performance. More specifically, the number of decision trees used, their size in terms of minimum node size for splitting and the number of data records used at each node. The latter has the significant influence on the performance of the model and for a dataset of length N the default values used are $\sqrt{N}$ for classification problems and $N/3$ for regression problems. Furthermore, the number of used trees should not be too big as well as their depth should be small in order to avoid overfitting (Cutler *et al.*, 2012). Figure 15 depicts how a random forest models works, where the feature space is split into subsets and a tree is trained for each of them. Then the results of each tree are averaged and the final prediction is given.

4.3.3. Gradient boosting

While bagging methods like Random Forest use an ensemble of models and they train them in parallel, boosting methods train them in a sequence (Figure 15). In detail, they sequentially apply a predictor to the data and the next predictor is trying to improve the output of its predecessor. Then, the result of each step is assessed based on its contribution to the direction of successfully predicting the target and a weight is assigned to it and in the end a weighted vote indicates the final prediction (Hastie *et al.*, 2009).

In the first step of the sequential process, equal weights are assigned to the training samples. Following the first iteration, the weight corresponding to each training sample is modified according to its miss-classification status. If it is misclassified then it receives a higher weight for the next iteration and vice versa. In this manner, the model will put more effort to predict the data points that were harder to predict during the previous iteration. This is how the AdaBoost algorithm, one of the most popular boosting algorithms that was introduced by Freund & Schapire (1997), works. In the following years, more boosting algorithms were developed like Gradient Boosting (GB) which is similar to AdaBoost but it tries to minimize the defined differentiable loss function *L* based on the *L*'s gradient descent. In other words, the weights that are chosen are the ones that decrease the loss by the highest gradient (slope) and thus the method is characterized as "greedy" (Hastie *et al.*, 2009). Recently, the XGBoost (Extreme Gradient Boosting) algorithm, which is actually an optimised version of the GB algorithm in terms of performance and computational resource usage, has proven to be very effective for regression problems in machine learning competitions like Kaggle (http://kaggle.com) and is favoured over bagging methods (e.g. RF) (Chollet, 2017).

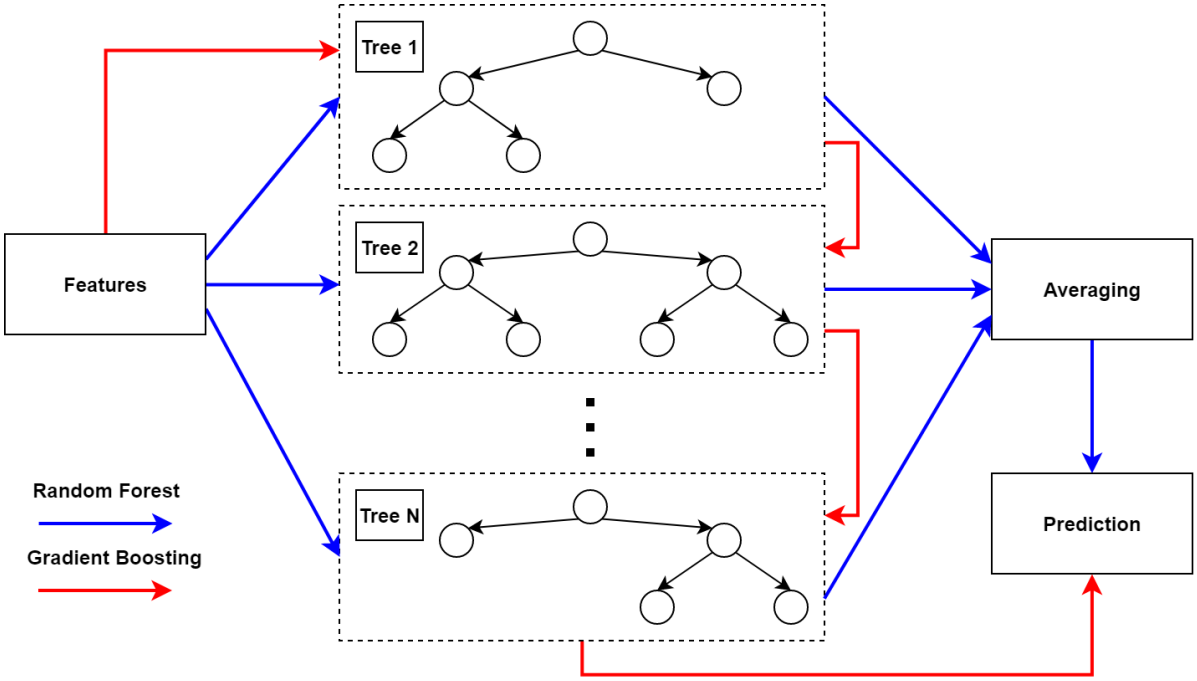


Figure 15 Random Forest and Gradient Boosting schematic representation.

4.4. Artificial Neural Networks

ANNs belong to the deep learning subspace of ML and therefore follow a different logic from the ensemble methods. Basically, they use successive layers of simple transformations that aim to map the input data (features) to the expected data (targets). The input data are inserted through the input layer and the target data are inserted in the output layer. In between these layers, a number of hidden layers can be defined based on the user's needs. These hidden layers perform the transformation according to a set of weights that are defined arbitrarily. In order to assess the outcome of the first set of transformations in terms of how close it is to the target data, a distance metric is computed which is called the loss function. Finally, the goal is to optimize (minimize) the loss function by adjusting the weights in a backward movement called backpropagation (Chollet, 2017). A graph depicting this process can be seen in Figure 16.

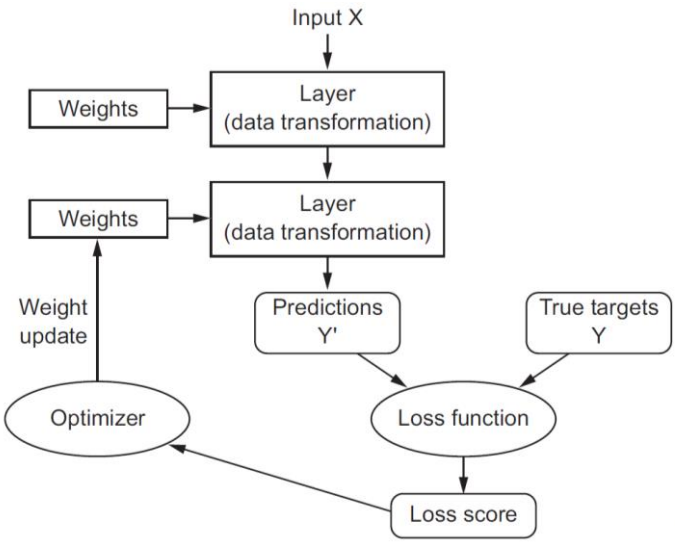


Figure 16 Neural Network workflow. Source: (Chollet, 2017).

4.5. Model validation

Every ML model faces certain challenges during the different stages of its design and implementation. The main problem that needs to be answered is if the model is overfitting or underfitting the train data. By overfitting it is meant that the model is able to explain the known data adequately but it fails to generalize to new data. On the other hand, if the model is not trained sufficiently then it fails to predict sufficiently both the train and the target data; this is called underfitting (Géron, 2017).

In order to design a model that does not fall into these pitfalls a validation process has to take place. During this process, the dataset needs to be split into the training and testing sets, where the model is fitted on the first only and then applied in both. As a result, two predictions are made, which when compared to the corresponding target sets, give the training and testing errors. Finally, a good model is the one that gives relatively low training and testing errors without them being significantly different from each other. Furthermore, if this process is repeated multiple times using different splits of the dataset then an optimal selection of model hyper-parameters can be defined. This is called cross-validation and once it is completed then the optimized model can be applied to the whole dataset (Géron, 2017). To calculate these errors that represent how different is the predicted from the true value, usually the Root-Mean-Squared error (RMSE) metric is used (Eq. 1).

$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \qquad \textbf{Eq. 1}$$

4.6. Chapter conclusions

Random Forest, Gradient Boosting and Artificial Neural Networks are the machine learning models of choice and can assist in the prediction of the target variable using the remotely sensed feature variables. In this manner, they are the fundamental tool to answer the research questions, depending on their performance and other analyses. Nevertheless, their time efficiency and validation will indicate the best one to be used in this study.

5. Study area and period

5.1. Area selection

The chosen study area consists of the 12 nautical mile zone that corresponds to the Dutch Territorial Waters (Figure 17), as it is defined by the Convention on the Law of the Sea (United Nations, 1982). The choice was based on multiple criteria with the most important being the availability of in-situ measurements for training the ML models, as well as the coverage of process-driven model outputs for comparison purposes (e.g. CMEMS FOAM AMM7). Fortunately, RWS operates in this area and measures the DO content in 11 locations after 2003, out of which eight have been found to contain adequate data. The study area, along with the RWS station locations that had adequate data for the extent of the study period, can be seen in Figure 17.

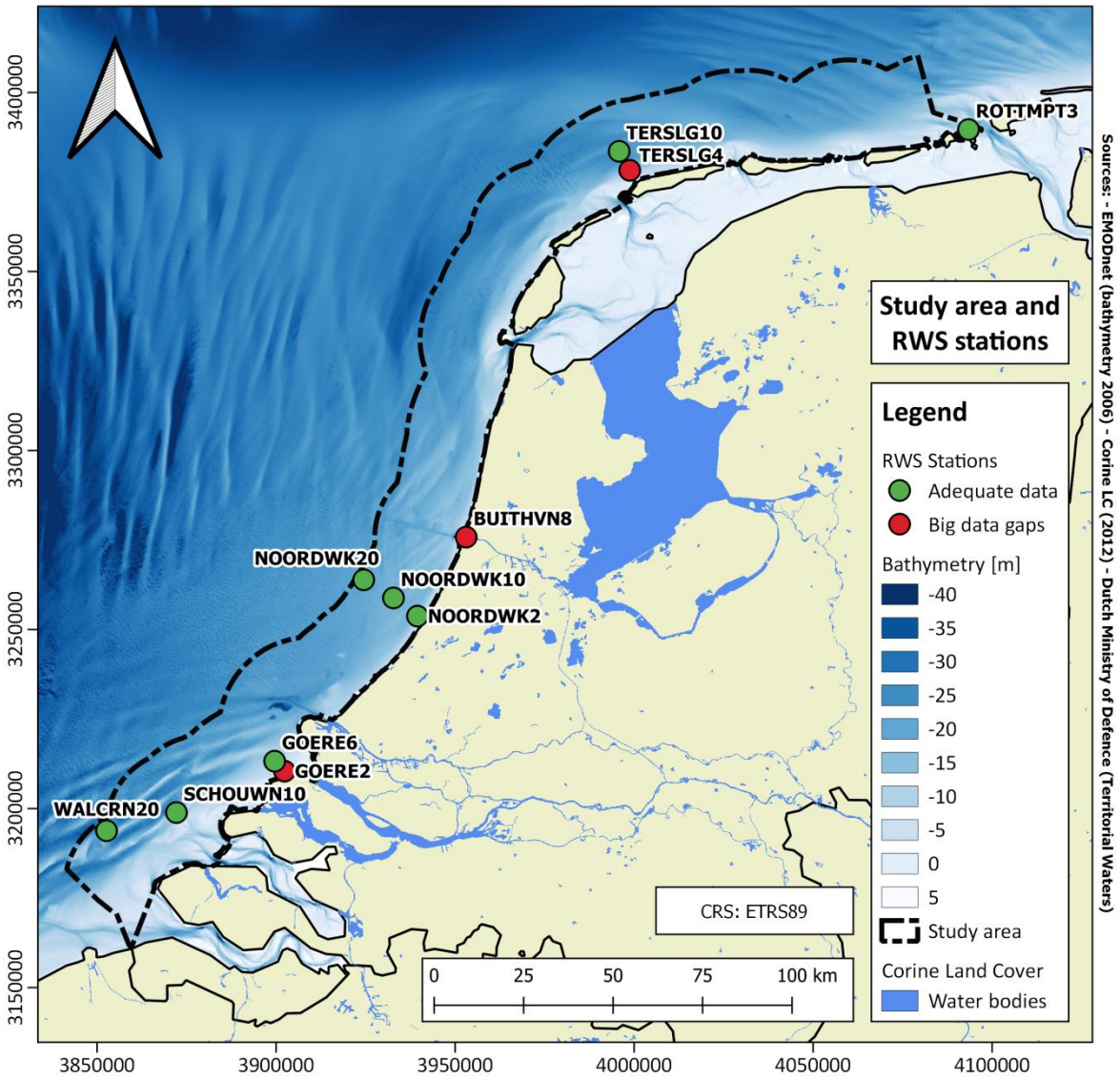


Figure 17 Study area including the Dutch Territorial Waters and the available RWS stations. The Territorial Sea borders were acquired from the Dutch Ministry of Defence website[10]. Created with QGIS 3.2.

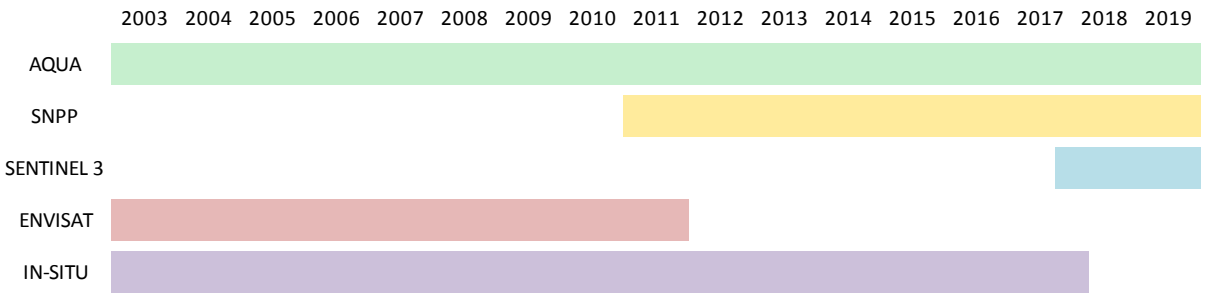
[10] Dutch Ministry of Defence, https://english.defensie.nl/topics/hydrography/documents/applications/2019/04/17/territorial-sea

In the map of the study area (Figure 17), the bathymetry that was available from the European Marine Observation and Data Network (EMODnet) can be seen. The Dutch Territorial Waters seem to be rather shallow with an average depth of 17.7m and a maximum depth of 47.5m. This means that the study area in all of its extent is a euphotic zone, i.e. the sunlight can illuminate every part of it and assist in photosynthesis, given that the turbidity of the water is not blocking all the incoming radiation. Furthermore, from the Corine Land Cover layer for 2012[11] and its inland waters layer, it is evident that the southern part of the study area has a strong influence of river discharges as it includes the Rhine–Meuse–Scheldt delta. According to the Dutch Ministry of Infrastructure and Water Managements (Rijkswaterstaat, 2016), the nutrient loads in these inland water bodies need to be furtherly reduced as in many cases their concentration exceeds the already set thresholds. Additionally, the Ministry mentions that especially the Scheldt River and the port of Rotterdam area are highly problematic in terms pollutant presence that deteriorate the overall water quality. Therefore, there are documented concerns and issues about the trophic status of the chosen study area, which makes it relevant and interesting for the application of the proposed methodology.

5.2. Temporal restrictions

The availability of the different satellite data and their temporal overlap with the in-situ measurements led to the choice of a study period that spans between 2003 and 2011. This period corresponds to the data provided by the ENVISAT satellite that is the predecessor of the ESA's new satellite Sentinel-3. Although it could be considered more interesting to use the latest Sentinel-3 satellite, its data overlap with the respective in-situ only for the period November 2017 to February 2018 at the time of this project (Table 5). This makes its data coverage minimal and inadequate, since the explored phenomena have strong annual cycles (van Aken, 2010), like the SST cycle that can have a direct physical effect on the DO content as was mentioned in 2.1.3. Furthermore, as the design of the optical instrument that Sentinel 3 (OLCI) is carrying is based on the specifications of MERIS (ENVISAT) the methodology and findings of the current project can plausibly be useful for future projects that incorporate Sentinel-3 data. After all, one of the key goals of the Sentinel-3 mission is to continue the observation capabilities of MERIS (Donlon et al., 2012).

Table 5 Gantt chart of temporal data coverage.



[11] The COPERNICUS website (http://land.copernicus.eu/pan-european/corine-land-cover) offers Corine Land Cover for years 1990, 2000, 2006, 2012 and 2018.

6. Materials and Methods

In this chapter the data used are presented, along with their acquisition, exploration and processing methods. Finally, the prediction model optimization and implementation are showcased. Figure 18 depicts a flow chart corresponding to the presented methodology. The input data are highlighted with the blue colour and regard to the remote sensing images, the in-situ data from RWS and a resampling regular grid. Meanwhile, the outputs are the DO maps and the validated and fine-tuned ML model that can be seen at the bottom of the chart.

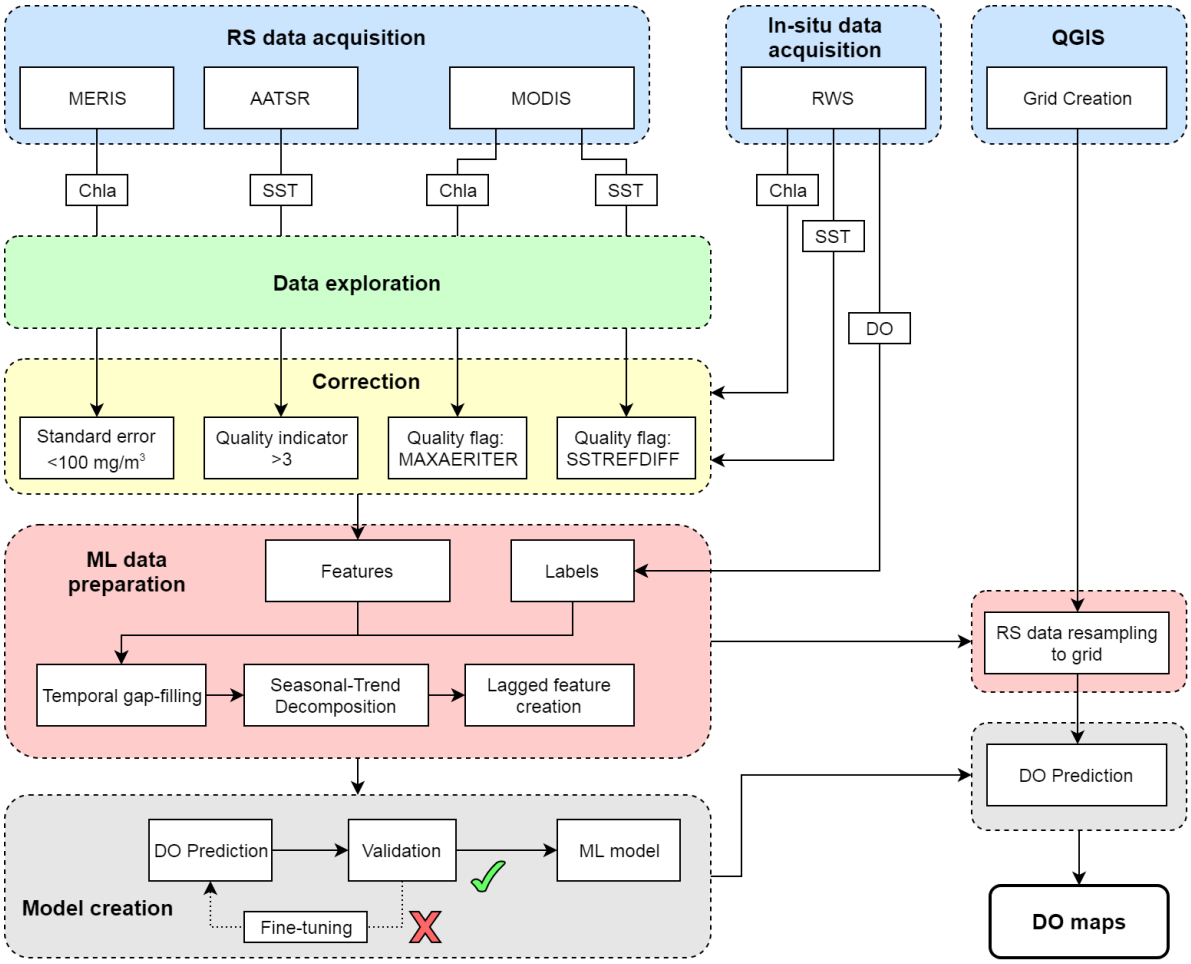


Figure 18 Methodology flow chart.

6.1. Data acquisition and pre-processing

Using a Python script, MODIS (OBPG-NASA, 2014) Level 2 data for the predefined spatial and temporal range (chapter 5) were downloaded automatically from NASA's Ocean Ecology Laboratory. The datasets provide close to daily coverage of the study area and there was a separate netCDF[12] file for each observed variable set (ocean colour and sea temperature) and date, accounting for 7975 files. Likewise, 5729 netCDF file of MERIS Level 2 data were also

[12] Network Common Data Form (netCDF) file format that is commonly used for the creation, access and distribution of matrix-like scientific data (https://www.unidata.ucar.edu/software/netcdf/docs/index.html#learn-more).

downloaded from the Deltares Remote Sensing repository (OPENDAP). Additionally, in-situ data for validation and training of the ML models were downloaded again from Deltares' OpenDAP service, using a Python script. This resulted in a large amount of data that was in excess of 1 TB and had to be processed and saved in an efficient way, removing unnecessary auxiliary data. Thus, the netCDFs were first cropped to the extent of the study area using the Territorial Waters shapefile (5.1), secondly validated for information utility (e.g. excessive cloud cover resulting in no pixel information - flagged for deletion) and then filtered, with those images not containing useful data deleted. Furthermore, geographic transformations were applied to the images as the coordinate system of the original files was the World Geodetic System spherical format (WGS84). In order to correlate the images to the in-situ data, as well as to create less distorted visualizations of the study area, a transformation to the European Terrestrial Reference System 1989 (ETRS89), which is a Cartesian system with meter units, was realised. An overview of the acquired datasets along with information on their spatial and temporal resolution and their number can be seen in Table 6.

Table 6 Data and acquisition information

Dataset	Source	Period	Spatial Resolution [km]	Temporal resolution [days]	# of files
MODIS	oceandata.sci.gsfc.nasa.gov/[13]	2003-2011	1	1-2	7975
ENVISAT MERIS	Deltares Remote Sensing data repository	2003-2011	1.2	1-2	1734
ENVISAT AATSR	ESA's EO portal	2003-2011	1	1-2	2992
RWS In-situ data	Deltares OpenDAP Service	2003-2011	Point measurements	DO: ~30 SST: 1 Chla: ~7	1233

6.2. Data preparation

6.2.1. Data formatting

Once the data were retrieved, they had to be inspected for accuracy and sensibility. This was done in a visual manner and explored to determine useful and required corrections which could be applied to the data. Firstly, 11 locations corresponding to RWS stations of the study area were chosen and are listed in Table 7 and can also be seen in the map of Figure 17. These locations will be used for training and validation of the methodology, as they are keystone points, containing both in-situ and satellite data. Secondly, the in-situ data for each of the 11 stations were transformed into a data cube (i.e. matrix), where corresponding dates of the period of interest between 01/01/2003 and 31/12/2011 were logged for each specific station.

[13] NASA's Ocean Biology Processing Group webpage

Then, for each remote sensing dataset the nearest neighbouring grid cell to each of the stations was found and its different variables were saved in separate matrices, similar to the in-situ one. However, since data from different sources were not available for each location on a daily basis, these data matrices included a lot of data gaps that were filled with Not-a-Number (NaN) data types. This choice was made to have all different datasets in the same structure, in order for all the comparisons to be implemented in a faster and more efficient way. Furthermore, the study period included 3287 days (2003-2011) corresponding to the same amount of data records for the 11 locations, resulting in 36157 allocated data records per matrix. This is a relatively small number and thus no memory problems were encountered nor a need to make it more efficient.

Table 7 Rijkswaterstaat In-situ measurement Stations with an indication of their location in regards to the study area.

Name	Location	Name	Location
BUITHVN8	Central part, 8km offshore	**ROTTMPT3**	North part, 3km offshore
GOERE2	South part, 2km offshore	**SCHOUWN10**	South part, 10km offshore
GOERE6	South part, 6km offshore	**TERSLG4**	North part, 4km offshore
NOORDWK2	Central part, 2km offshore	**TERSLG10**	North part, 10km offshore
NOORDWK10	Central part, 10km offshore	**WALCRN20**	South part, 20km offshore
NOORDWK20	Central part, 20km offshore		

6.2.2. Data exploration and quality control

Next, the datasets were plotted and inspected for existing errors or correlations with each other, both visually and by using statistical metrics. For the Chla data, both MERIS and MODIS products were examined spatially, by plotting the individual scenes and visually checking for implausible patterns. Their temporal behaviour was also visually examined for the 11 stations by plotting the time series of satellite observations at station locations against the corresponding in-situ time series. Moreover, scatterplots between the remotely sensed products and the in-situ measurements of the same variable were used, along with a linear model fit and correlation metrics to assess the accuracy of the estimations. In case both datasets corresponded to the same environmental variable, a very strong positive correlation would be the optimal case, while in case different variables were compared a less clear relation would be expected.

In order to correct the identified issues, the different quality indicators, flags and metrics that were provided for each product were used, while trying to preserve as many data records as possible and avoid extensive data thinning. In addition to the quality flags, MERIS HYDROPT Chla product had a standard error metric per pixel (chapter 3.2.2), while AATSR SST and MODIS SST products had a quality indicator per pixel. All these flags and indicators were applied one at a time to each dataset and their effect was assessed in the direction of co-aligning the remotely sensed variables with the corresponding in-situ measurements. A more in depth description of the data exploration process along with the final choices on the corrections is presented later in chapter 7.1. After applying the necessary corrections, the relations between the remotely sensed variables and the target variable, i.e. DO, were explored using the same scatterplots as

above. Since the absence of data was an issue, variable-size window averages were used to further expand the available overlapping data. Finally, the time series were divided into annual subsets and plotted together for each variable for a visual identification of temporal patterns and seasonal behaviour without the need of direct overlap.

6.3. Statistical learning and prediction

In this section, alternative approaches on feature selection for different prediction methodologies are presented. In order to counter the data restrictions coming from the gaps and the irregular sampling of the different variables, various experiments were performed.

6.3.1. Time-independent prediction

The simplest and most straightforward approach for predicting DO using the outlined methodology is to identify the dates were in-situ measurements have been taken and see whether satellite measurements were also available for that date. For the dates that this criterion was fulfilled, the satellite data were saved as features and the in-situ measurement was saved as the target. Although, the rarity of simultaneous overlap of all the datasets decreased the availability of data pairs for feature construction on the specific days that target data were present. Hence, the inputs for the predictive model were inadequate in number to support a robust model that was trained on a dense set with broad variety of seasonal events and variable value ranges and therefore generalize well. In order to counter this problem, a moving average filter was applied and the effect of different window sizes on the data overlap and the improvement of the prediction was documented.

6.3.2. Temporal gap-filling

The variables studied in this thesis have naturally strong temporal components, which were disregarded in the approach taken in 6.3.1. These components are the natural annual variation of the SST and a seasonal variation of Chla that coincides with a spring and autumn algal bloom in the Dutch coastal waters. Meanwhile, the irregularity of spatial and temporal data gaps and the small density of the in-situ data records made their overlap with the satellite measurements rarer. For this reason, an alternative approach was implemented, where the irregular data sequences (both in-situ and satellite) would be gap-filled to become regularly spaced, equidistant signals (constant interval) and analysed in components of different frequencies. This way, the strong seasonal behaviour of the environmental variables would be easier to introduce into the modelling process.

On these grounds, different techniques were tested and compared for temporal gap-filling, namely a linear and an Akima interpolation (Akima, 1970). The first approach was rather straightforward as the variable sequences were linearly interpolated by connecting the adjacent data records with a straight line. The second approach uses multiple up to third degree local polynomial fits to estimate a smooth and natural curve that passes through all the data points (exact method). It takes into account the positions of two adjacent points with respect to the query point, as well as the slopes at those points, making it more complex than the linear interpolation. Moreover its results are more realistic than the linear interpolation, which tends to produce spikes and sudden changes in the variable values. The implementation of these two methods was realised using the Pandas Python library and was straightforward and fast. Here, it

should be mentioned that other available techniques were also tried out (e.g. polynomial, spline, cubic), but the produced estimations had extreme fluctuations with non-realistic values (either negative or very big) and were discarded. Overall, the choice of Akima was based upon its higher flexibility compared to the linear interpolation, as well as on the fact that its results were realistic (i.e. between the expected ranges) and followed a visually natural curve and thus no further exploration was needed.

6.3.3. Seasonal-Trend Decomposition

The continuous time-series of the variables where decomposed into their trend, seasonal and residual components using a Seasonal-Trend Decomposition tool (STL) in Python (stldecompose package), which is based on the filtering methodology proposed by Cleveland, Cleveland, McRae, & Terpenning (1990). It is a straight-forward procedure where different local regressions are applied to the original time-series. Except for the time-series, the periodicity of the data has to be defined. For this project, this was set to 365 days, as the presence of annual periodicity is physically expected and has been visually identified during the exploration procedure. Furthermore, through trial and error it was found that the annual periodicity resulted in the lowest residuals and therefore it was the most suitable.

6.3.4. Feature engineering

The gap-filling of the variable time-series and their decomposition created opportunities to further improve the prediction of the target variable, i.e. dissolved oxygen. Hence, the results from the decomposition process for each variable were treated as separate features that were to be experimented with in the prediction process. This choice was supported by the fact that certain components, e.g. the seasonal component, are expected to have a higher correlation with the documented seasonality of the target variable, while the contribution of others, e.g. the residual component, needed to be assessed. Furthermore, with this separation of the components the ML models will be able to indicate the importance of each feature separately, and assist with the interpretation of the results. Additionally, experiments were run using separate satellite sensor data, i.e. MERIS and MODIS, but also combining them all together in one model as different features (Figure 19).

Moreover, three temporal features were created corresponding to each observation, namely Day in the year [1,365], Day in the time-series [1, 3287] and Month [1, 12]. However, the usage of these features was soon abandoned as they proved to just force the model to predict the average value of the training data. Despite the fact that the prediction using these features was better, the effect was deemed artificial as it was heavily influenced by the proximity of the station to each other.

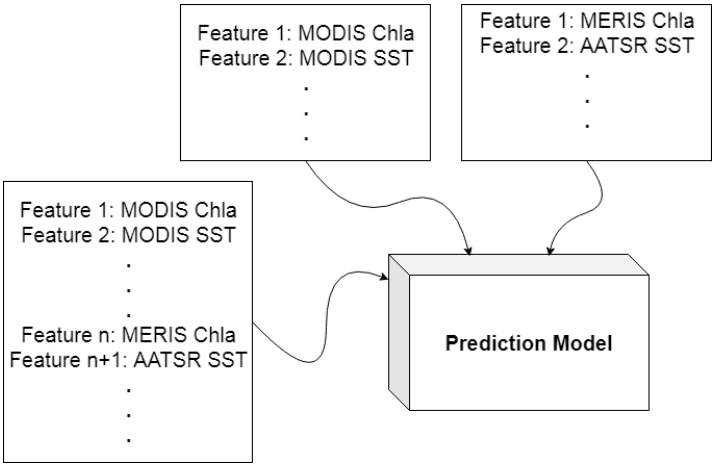


Figure 19 Graphical representation of the feature selection approaches

Additionally, the existing features were complemented by *N* lagged versions of themselves. This was done to introduce the temporal dependency between the data records of the same feature in the prediction. In other words, for the prediction of a DO record at time *t*, the features corresponding to time *t, t-1, t-2, … , t-N* were used. This way, whether the previous state of the feature variables affects the DO in the future can be assessed. Finally, trials for the optimal value for *N* were carried out in terms of RMSE reduction (Figure 20).

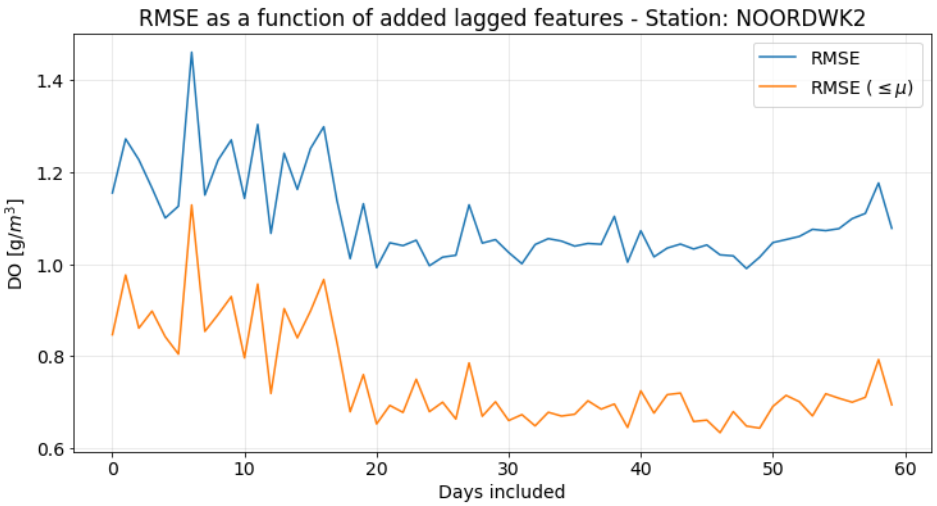


Figure 20 RMSE progression using different sets of lagged features for the prediction. Example for station Noordwijk 2km. RMSE ≤μ denotes the error of the prediction for the true values that are equal to or smaller than the average in-situ DO.

6.3.5. Prediction, fine-tuning and validation

Many experiments and trials were carried out for the prediction of the target variable. Namely, GB, RF and ANN (chapter 4) regression models were implemented, providing different predictions errors and running times. The hyperparameters of each model, i.e. the parameters that are set before training the model and define how the model is designed, were chosen carefully after iteratively testing different combinations of them in an attempt to optimize the Root-Mean-Square error and to assess whether the model was overfitting to the training labels. The accurate detection of the DO minima was also under investigation, as these values correspond to higher stress and indicate a direction towards the formulation of hypoxic

conditions. For the two ensemble methods, this "fine-tuning" was relatively simple as the hyperparameters have a clear description and their impact on the result can be identified and interpreted. Meanwhile for the ANN the effect of the number of hidden layers and neurons per layer is a bit harder to understand and need an extensive sensitivity analysis that goes beyond the scope of this project. On these grounds, it was chosen to present mostly results from the GB model as they were the optimal in terms of runtime and accuracy.

The ANN was designed using the Keras library through the TensorFlow package in Python (Agarwal et al., 2015). Meanwhile, the RF and GB models were designed using the Scikit-learn Python package (Pedregosa *et al.*, 2011). For these two methods the most important hyperparameters are documented in Table 8, while others that are not so relevant with the proposed methodology, e.g. sample weights, have been left out. For the fine-tuning of these hyperparameters, certain tools are available in the Scikit-learn package where the user can choose a gridded, i.e. brute-force search, or a random search for their optimal combination. This procedure is time consuming and only the final results are usually available. Nevertheless, an example has been plotted to demonstrate the sensitivity of the prediction when changing the number of estimators (Figure 21). Furthermore, these methods offer certain tools that can provide an insight into the performance of the model, like feature importance metrics. The computation of these metrics is based on the Gini criterion, i.e. the highest reduction in impurity from one node to the next of the decision tree, and describes the amount of contribution each feature has on the prediction expressed as a percentage. Also, according to the documentation of the package it is mentioned that in case of multiple highly correlated features (collinear) only one of them is given the corresponding importance value and the rest of the features are given a remarkably low one. This implies that features with remarkably low importance are redundant and could possibly be left out of the prediction as their contribution is minimal.

Table 8 Most important hyperparameters of Scikit-learn RF and GB Regressors.

Hyperparameter	RF	GB	Description
Number of estimators	X	X	Number of weak learners (trees) that are going to be used in the ensemble
Criterion	X	X	The function with which the quality of each split is assessed. Normally the Mean-Squared-Error (MSE) or the Mean-Absolute-Error (MAE)
Learning Rate	X	X	Sets a contribution metric for each learner
Minimum Samples Split	X	X	The minimum number of data points needed for a split of an internal node
Minimum Samples Leaf	X	X	The minimum number of data points needed to have a separate leaf node. A node can be split only if it is ensured that the two created nodes are going to have samples equal to the minimum samples leaf number. This hyperparameter affects the smoothness of the prediction
Maximum Features	X	X	Maximum number of features involved in each node split. Choosing a subset of the features may lead to an increase in

		bias, while the variance is expected to drop
Loss	X	Loss function that the model fitting is trying to minimize. Options are 'Least Square Regression', 'Least Absolute Deviation' and 'Huber'[14]
Bootstrap	X	Defines whether the whole or a subset of the dataset is used to design each tree

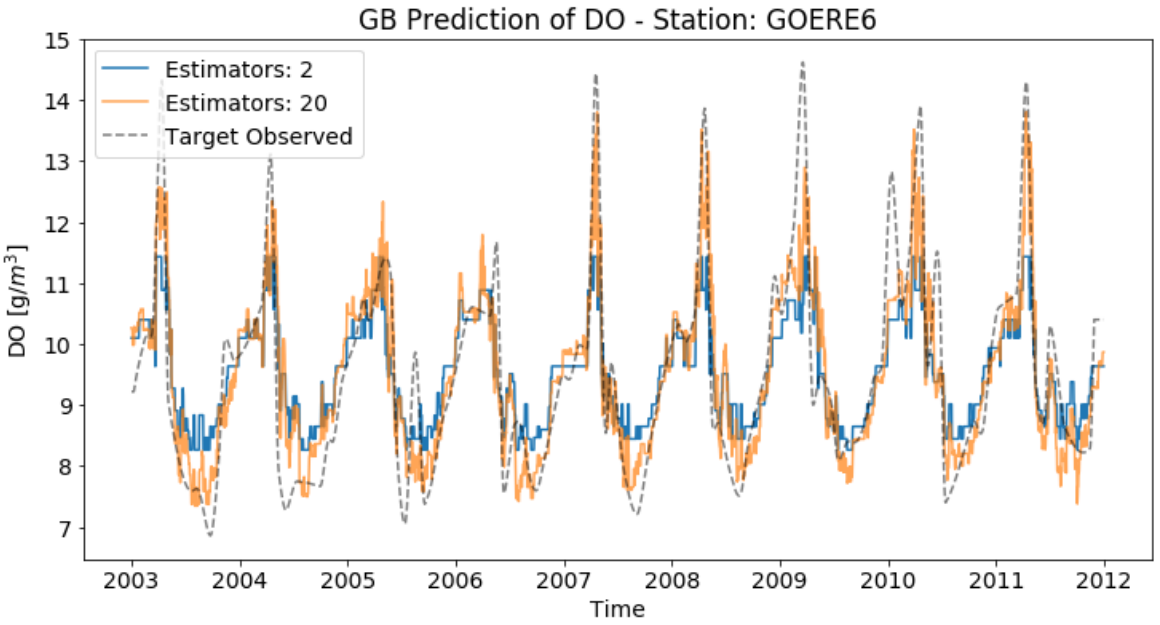


Figure 21 Difference in GB prediction using different numbers of estimators in the ensemble.

Finally, in order to assess the performance of the prediction model, a leave-one-out cross validation scheme was used. More specifically, the data of one station was taken out of the original dataset, which was comprised of eight stations with nine years of daily measurements of the different features and target variables. Then, a model was fitted to the remaining seven stations' data. Following the training, the model was applied to the seventh "unseen" station and its prediction was assessed with the RMSE metric by comparing it with the known target data that were put aside. This procedure was implemented iteratively for each station and the average RMSE of all the runs was computed and used to assess the methodology. As a new model is produced in each iteration, the mean RMSE does not assess the model itself, but the potential a model to predict certain targets given the chosen hyperparameters and specific train data. Meanwhile, the standard deviation of the individual RMSE values can indicate whether the procedure is robust or not, as a low spread of values means a more or less uniform performance of the prediction.

[14] Huber loss makes the regression robust, i.e. less affected by outliers (Hastie *et al.*, 2009).

6.4. Data resampling

The various datasets were sampled in different spatial resolutions and configurations, thus they had to be resampled to a common grid-space in order for the selected processing and comparison methods to be applied. At first, a common grid had to be defined for which all datasets would be resampled to. Such a grid was determined through a comparison of the original data grids with attention paid to the effective final resolution. Its spatial resolution had to be sufficient to provide detailed information but at the same time not require excessive downscaling, which would introduce uncertainties and errors, or upscaling, which would result in loss of information. The target grid was defined via a 1.2-by-1.2 km resolution regular grid corresponding to the grid of MERIS, since the HYDROPT Chla estimations were found to contribute more in the predictions in terms of feature importance compared to the respective MODIS estimation. Additionally, the choice of the grid resolution was based on the fact that SST's variation is more temporal than spatial, i.e. it is rather homogeneous at each date, and therefore the resolution of the grid should not affect the result much. The resampling technique that was implemented was the Inverse Distance Weighting (IDW) interpolation (Shepard, 1968). It is a simple implementation of a weighted average using the formula shown below:

$$\hat{\boldsymbol{Z}}_{\boldsymbol{new}} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{w_i z_{sat,i}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{w_i}}, \boldsymbol{w_i} = \frac{\mathbf{1}}{\boldsymbol{distance^p}} \qquad \textbf{Eq. 2}$$

Here, w_i are the weights, corresponding to the inverse distance with a power of $p = 1$, using the Euclidean distance metric. $\boldsymbol{Z}$ denotes the variable value of the satellite (***sat***) and interpolation point (***new***). Not all the satellite grid points where taken into account for this computation but only the n nearest neighbours to the query point were selected using a $1.2 \times \sqrt{2}\ km$ radius threshold, representing the diagonal of the square grid cell.

The choice of the interpolation method was based upon the fact that this technique is often used for continuous data, as are the variables in this project, and the final result was physically meaningful. It was also easy and fast to implement. A comparison between the original and resampled data can be seen in Figure 22. Here, the spread of the values on the two surfaces seems very close but as expected the borders between values have lost their sharpness and some gaps of the original image have been partially filled as a result of the spatial interpolation. These minimal alterations between the two images were deemed acceptable and therefore other more complex methods were not considered.

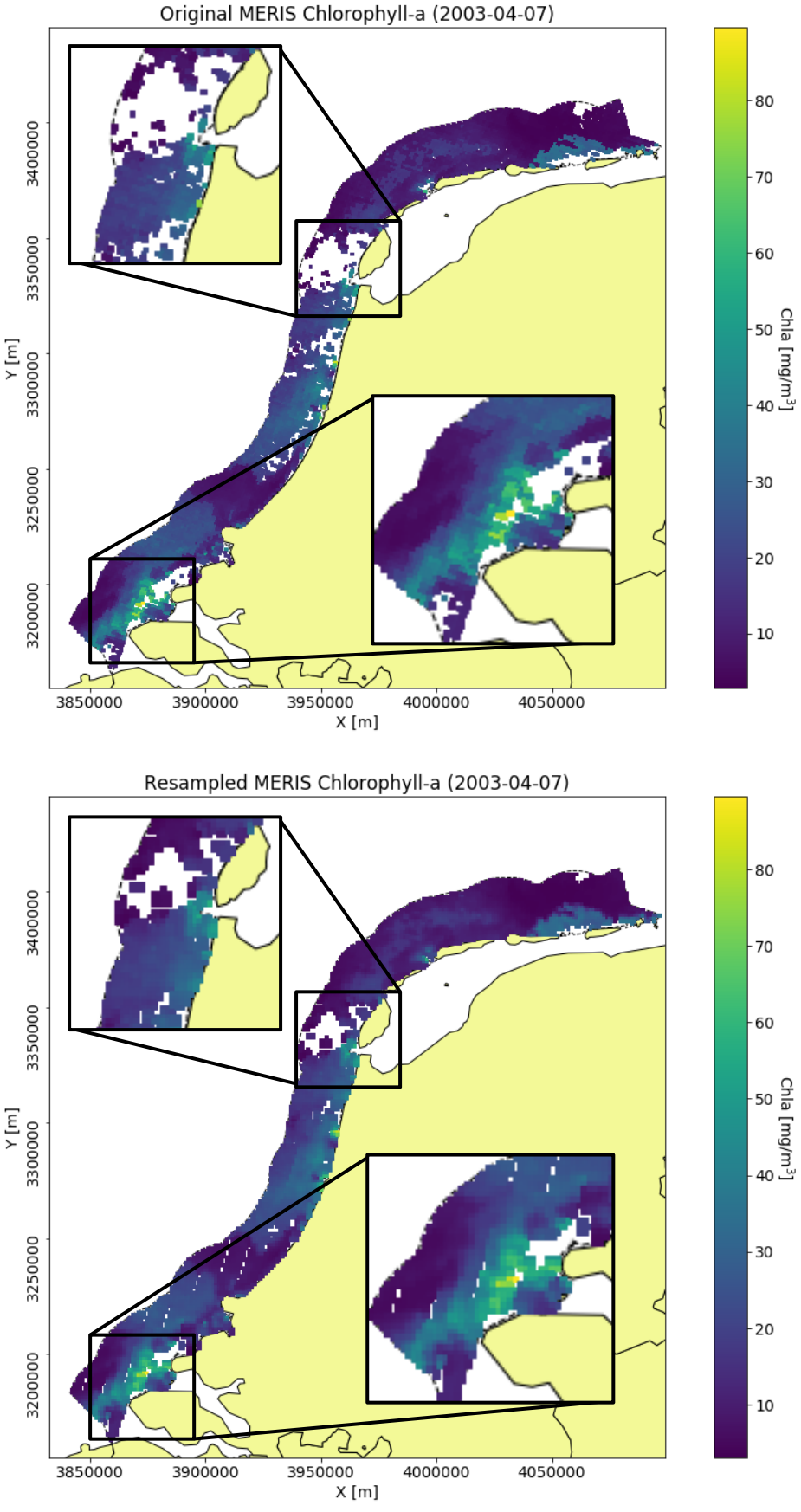


Figure 22 Original (up) and resampled (down) MERIS chlorophyll-a estimation for a sample date.

6.5. Chapter conclusions

The large volume of data acquired by different sources was explored and corrected using the ground truth data. The datasets when then saved and used for the prediction of the target variable. Experimentation with different feature sets, gap-filling techniques and feature engineering was carried out to investigate the effects on the prediction accuracy using the three chosen ML models. Finally, the gap-filled feature variables were resampled in a common regular grid to spatially expand the prediction in the study area and identify possible spatio-temporal patterns.

7. Results

In this chapter, examples of the most important results of the methodology are presented in order to justify the choices of the proposed workflow. Different approaches on data exploration and correction, as well as in prediction methodologies and models, reveal how each choice affects the results and the accuracy of the prediction. The majority of the presented results are derived from a GB model, as it delivered the best performance. Finally, the DO prediction for the whole spatial extent of the study area is showcased.

7.1. Data exploration and quality control

The data exploration section presents the actions needed to be performed when dealing with datasets of unknown uncertainty. It will help assess the quality of the selected products and correct possible errors they include. This way it is ensured that the quality of the used data is known and the possible limitations are identified. It should be noted that the direct comparison of *1km x 1km* satellite pixels with point in-situ measurements adds to the uncertainty of the conclusions that can be derived, as it is hard to account for possible sampling issues. Nonetheless, the extent to which satellite data converge to the in-situ data of the same variable is going to be documented. Then, the correlation of the corrected variables to the DO variable will be investigated.

7.1.1. MERIS estimation of Chlorophyll-a content

Starting from Chla datasets, the in-situ and MERIS time-series were plotted for each station location (Figure 23). Initially, the original data were explored and the corresponding plots gave strong indication for presence of errors. More specifically, two screening effects were evident around 14 and 25 mg/m^3 that indicated a possible error in the processed data (by HYDROPT), as these errors are consistent throughout the study period and no known MERIS error has been documented with such behaviour. Furthermore, for every in-situ measurement it was checked whether MERIS Chla data have also been observed and, if so, they were plotted in a scatter plot to assess their correlation (Figure 24). The screening effect was visible in this representation of the data as well, so corrective actions had to take place. For this reason, the effects of masking the data using the available quality flags and the theoretical standard deviation error from the HYDROLIGHT model were explored.

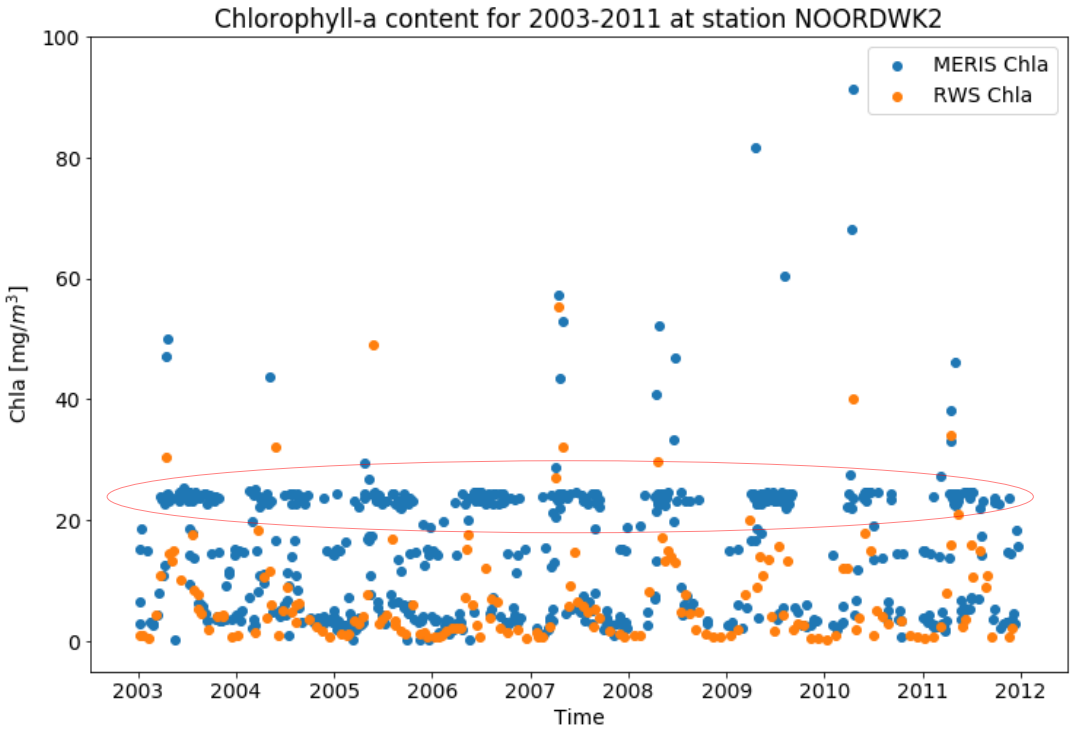


Figure 23 Chlorophyll-a content at the location of the Noordwijk2 station that is 2km offshore on the Noordwijk transect.

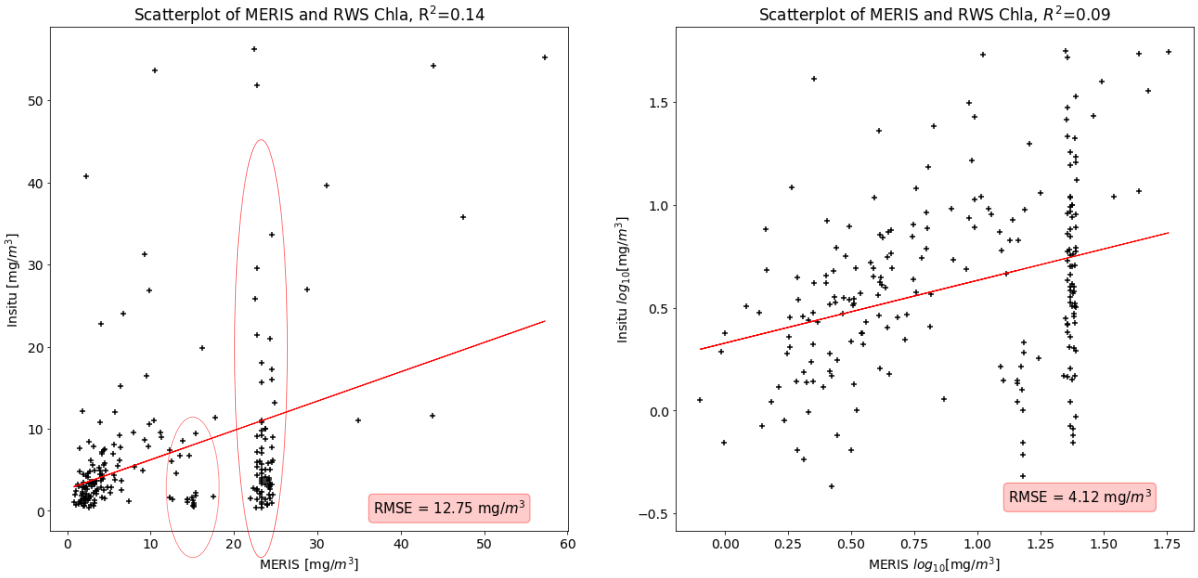


Figure 24 Scatter (left) and log-log plots (right) of data record pairs between in-situ and MERIS data.

Quality flag exploration

Because of the aforementioned systematic errors, the quality flags that were included in the MERIS dataset were turned on and off successively and their effect on the results was documented. Through trial and error, it was found that the flag HIGH_GLINT[15] (Table 16) was

[15] HIGH_GLINT flag indicates a pixel with high sun-glint, which has not been corrected. It may indicate that sun-glint reflectance calculated from viewing geometry and wind speed exceeds the medium glint threshold, and is too high for glint correction.

able to mask out a big number of the corrupted data records while maintaining more data records in the expected value range. In a first glance at Figure 26, the screening effect seems to be minimized. Nevertheless, significant thinning (roughly by 65%) of the data record pairs can be seen in Figure 25, while the screening effect is not completely eradicated.

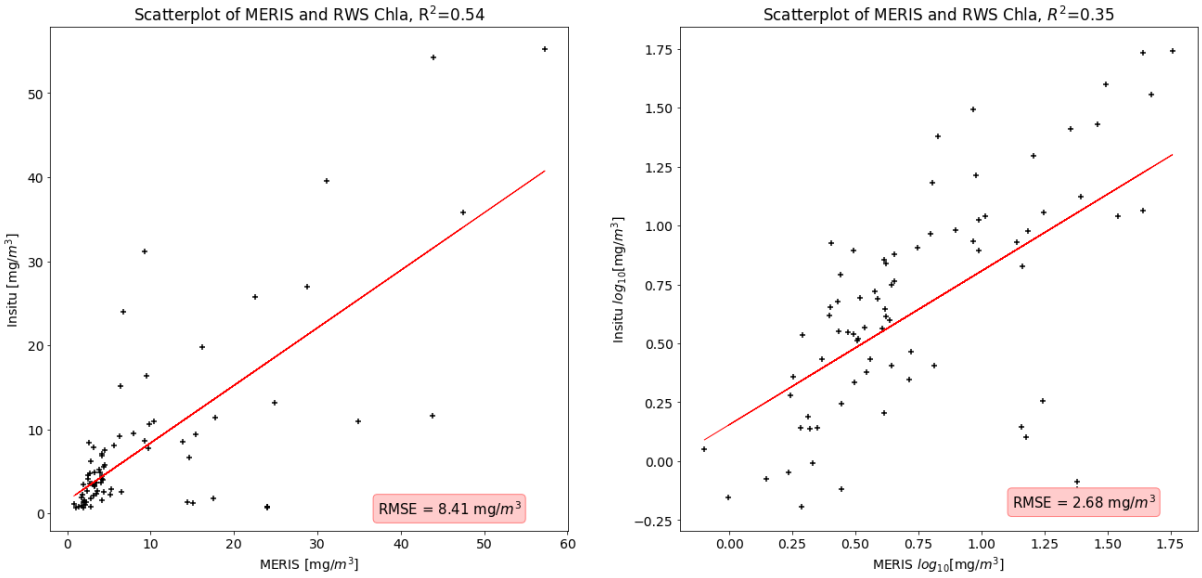


Figure 25 Scatter plot (left) and log-log plot (right) of pairs of data records between in-situ and MERIS data, having flag HIGH_GLINT on.

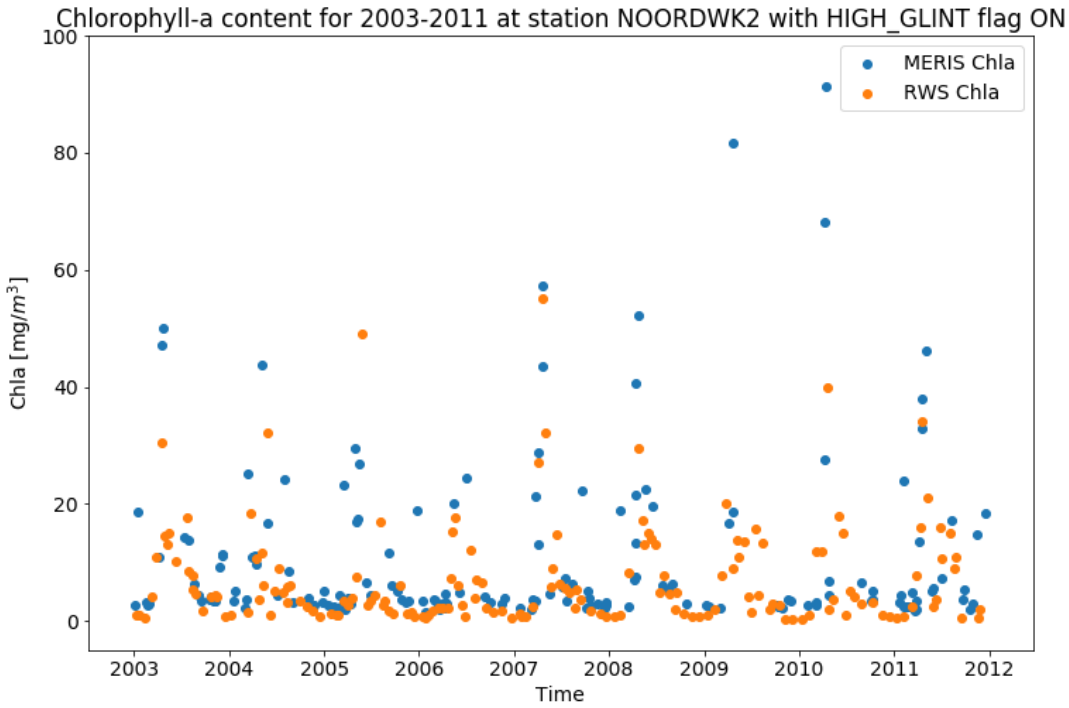


Figure 26 In-situ and MERIS Chlorophyll-a content at the location of the Noordwijk2 station that is 2km offshore on the Noordwijk transect, having flag HIGH_GLINT on.

Thus, the quality flag approach may have improved the dataset but not in an adequate degree. Another indicator of the thinning of the data is the histogram of the data record pairs, where the effect of turning the flag on can be seen (Figure 27). In more detail, from the two histograms it is evident that mostly the values between 0 and 20 are affected and especially the ones in the range 0-5 mg/m^3, which are halved.

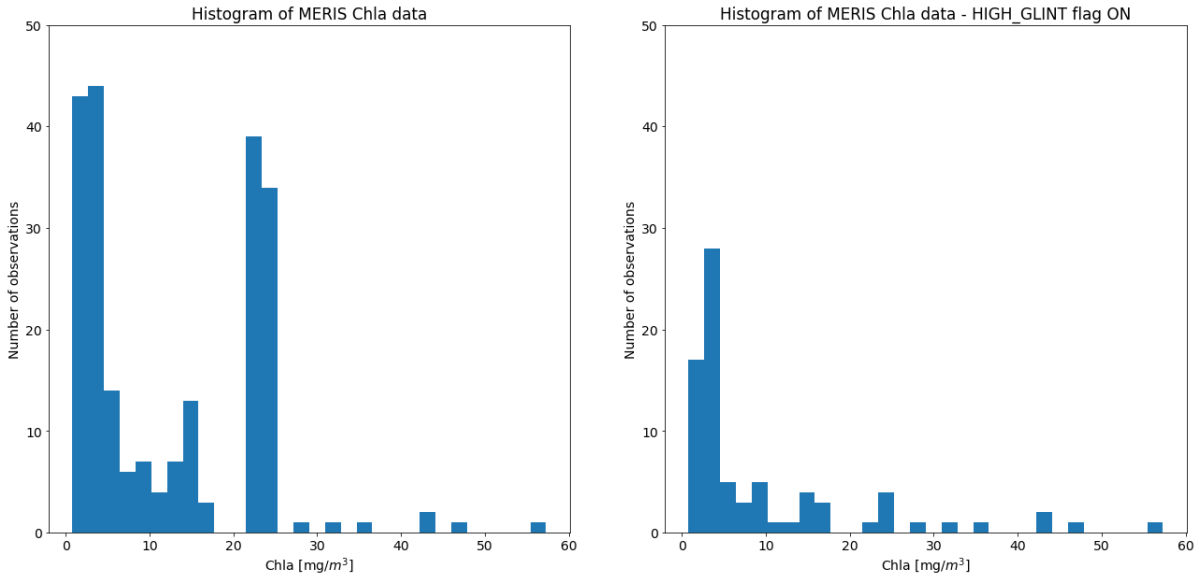


Figure 27 Histograms of MERIS Chla data with all flags are turned off (left) and with HIGH_GLINT turned on (right).

HYDROPT standard deviation error exploration

MERIS datasets were provided with per pixel standard deviation error which is given by the HYDROPT algorithm described in chapter 0. The histogram of these errors was plotted (Figure 28) and it was discovered that while most errors are relatively low, certain pixels have errors around 200 and 500 mg/m^3, which made no physical sense for a variable whose expected range is between 0 and 100 mg/m^3. Hence, the pixels with an error greater than 100 mg/m^3 were masked out and the corresponding histograms from before and after the filtering can be seen in Figure 29. Here, it is evident that this correction is more efficient in terms of data thinning (data reduced by ~35%) and RMSE thus the dataset was corrected using only this standard error.

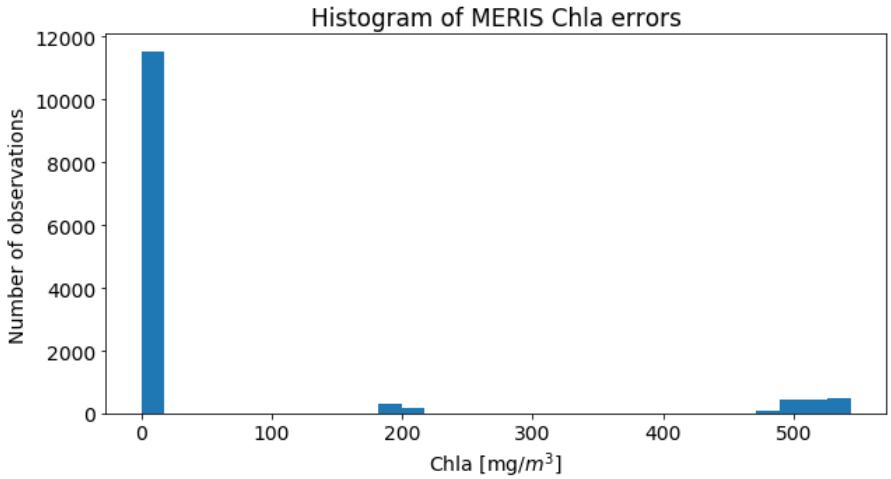


Figure 28 Standard deviation errors per pixel for MERIS Chla dataset.

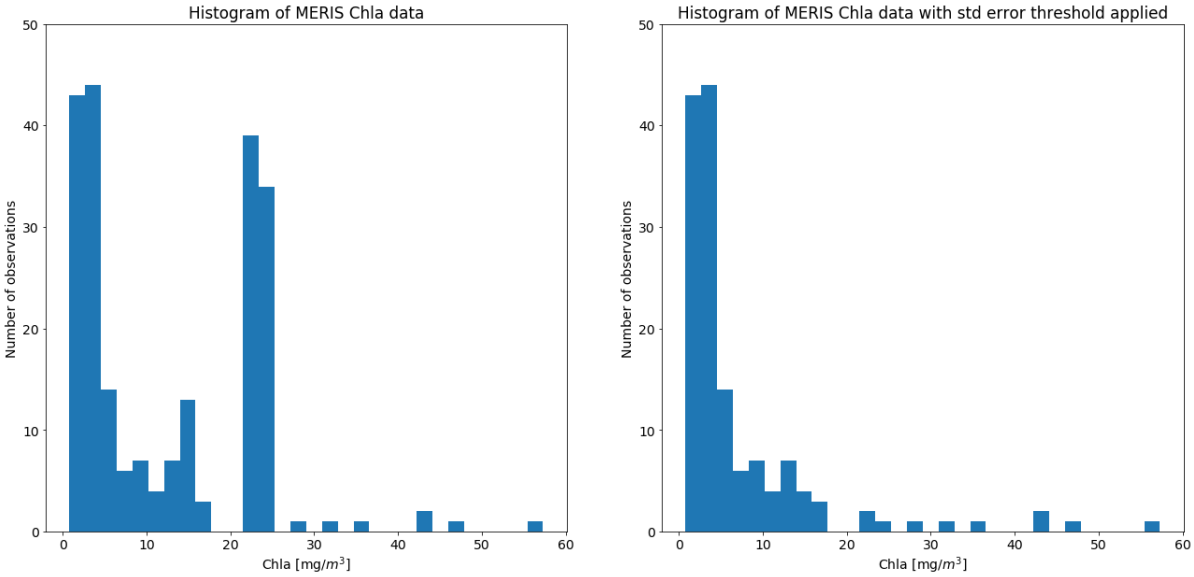


Figure 29 Histograms of MERIS Chla original data (left) and Standard error corrected (right).

Additionally, Figure 30 shows that the MERIS and RWS datasets of chlorophyll-a are positively correlated, which is visually more clear in the log space. This was expected as they refer to the same variable. This final setup was chosen as the best one because of the low data thinning and adequate RMSE and correlation metrics.

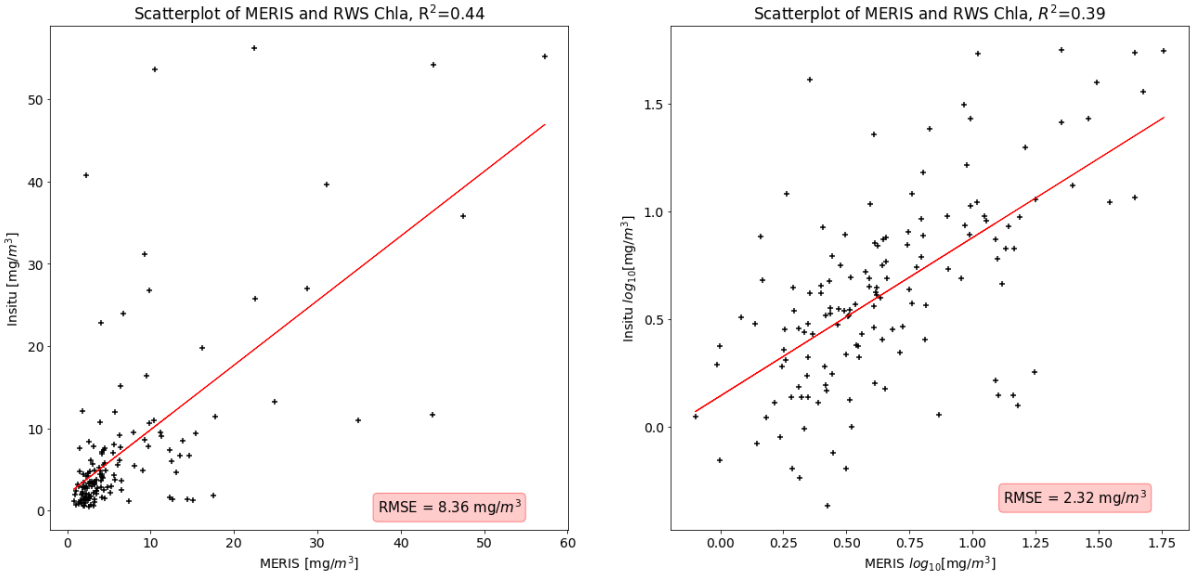


Figure 30 Scatter plot (left) and log-log plot (right) of pairs of data records between in-situ and MERIS data, for error <100mg.

For a better understanding of the extent of these errors, the original and corrected images for the whole study area were plotted (Figure 31). Here, it is evident that in the original image there are implausible values on the north-eastern part as well as in the Zeeland delta and the Wadden Sea. These areas show an unnatural transition in values from their neighbouring pixels, and they are completely erased when the correction is applied. Moreover, as it was denoted in the previous sub-chapter, the implausible values at the north-eastern part can be caused by the high glint flag. In such a case, the angle of the sun's reflection causes erroneous measurements in the optical values of the surface water and returns a constant value when the algorithm is applied. Similarly, in the Wadden Sea there are several values which have been removed. These could be a result of the intertidal regime of the area, as the location of the flagged and removed pixels corresponds to areas known for exposed intertidal mud flats.

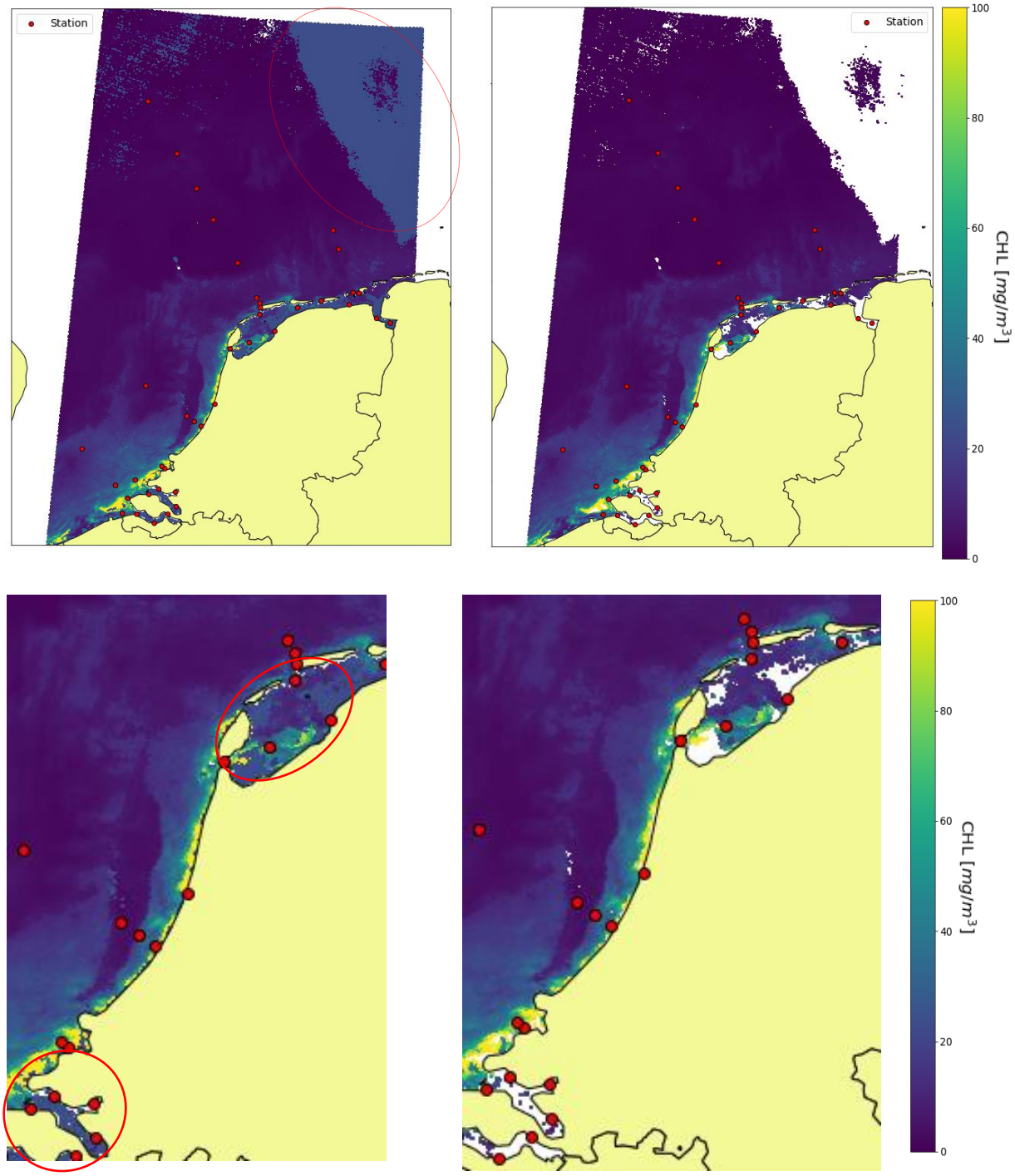


Figure 31 MERIS chlorophyll-a observations for 16/04/2003. On the left side are the original images and on the right side the corrected ones, where the saturated values have been eliminated.

7.1.2. MODIS estimation of Chlorophyll-a content

The AQUA MODIS product for Case-2 waters regarding chlorophyll-a content is produced with the OCx algorithm, which uses two bands at 490 and 555nm (Reilly *et al.*, 2000). The OCx is a simpler algorithm compared to HYDROPT (chapter 3.2.2), so it is expected to have a lower performance. The figures below show that the data pairs are less linearly correlated in the log space based both on R^2 value and RMSE, as well as on their overall distribution, in comparison to Figure 30. Meanwhile, in the regular space, MODIS has lower RMSE and higher R^2 values,

although that can be attributed to the many outliers in the high values that MERIS demonstrated. In Figure 34, the time series of the MODIS data seems to follow the seasonal trend of the in-situ measurements but also to overestimate the chlorophyll-a content. Furthermore, after checking out one-by-one all the available flags for MODIS, it was concluded that by excluding pixels which had the MAXAERITER[16] flag on, some outliers were removed and the correlation between the in-situ data and MODIS data, as well as their RMSE, were improving (Figure 33).

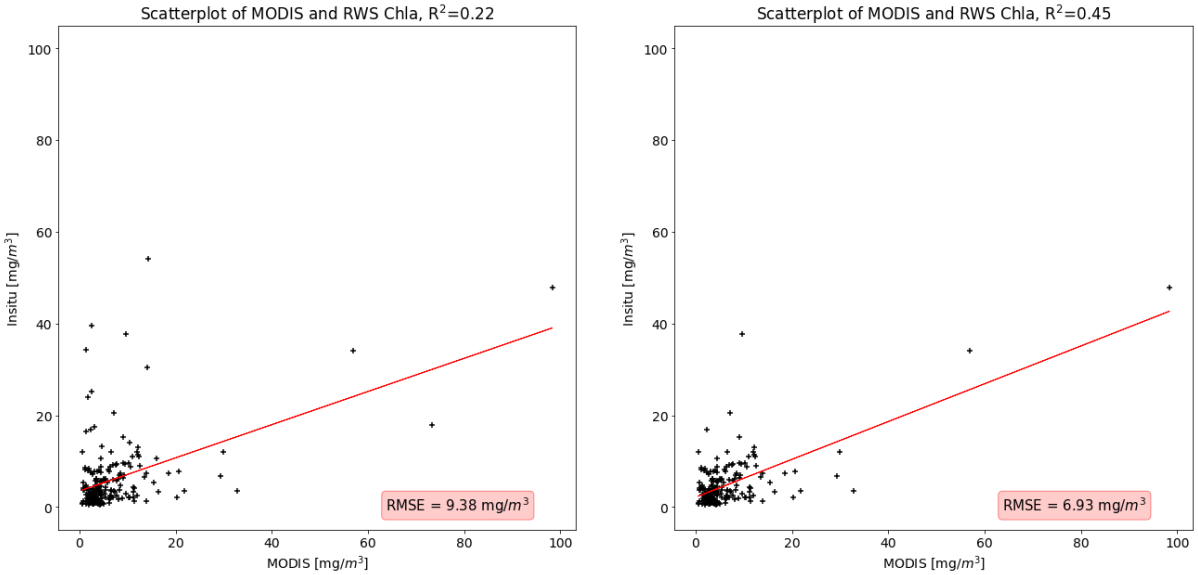


Figure 32 Scatter plot of matched pairs of data records between in-situ and MODIS data, original (left) and having flag MAXAERITER turned on (right).

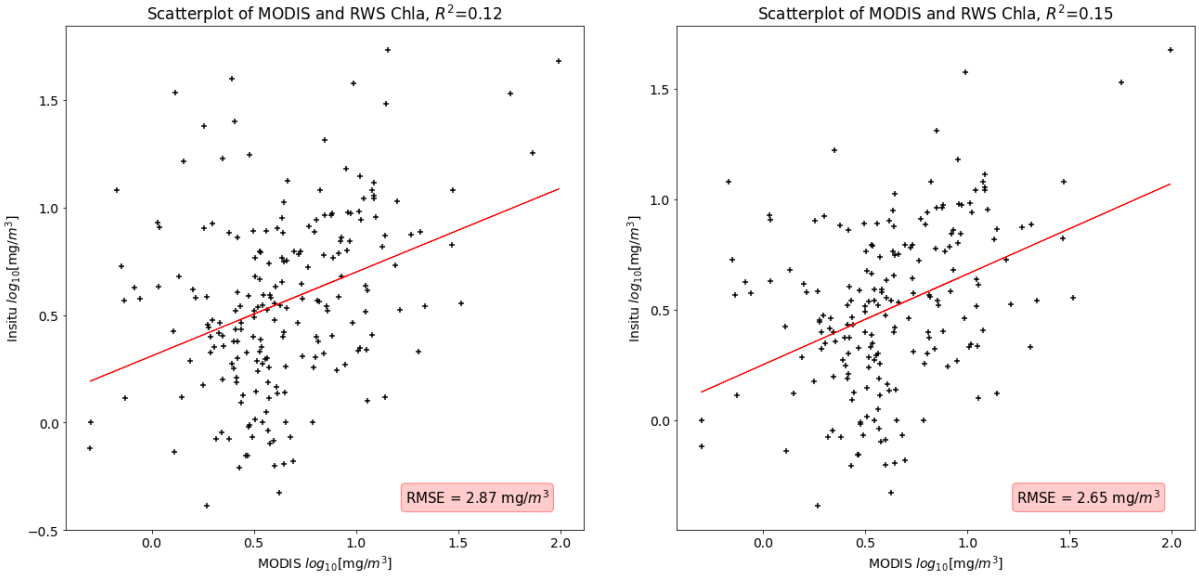


Figure 33 Scatter log-log plot of matched pairs of data records between in-situ and MODIS data, original (left) and having flag MAXAERITER turned on (right).

[16] MAXAERITER flag is turned on when a maximum number of iterations is reached for NIR band without it converging in a predefined range. It normally implies saturation of the pixel because of presence of clouds (NASA, n.d.-a)

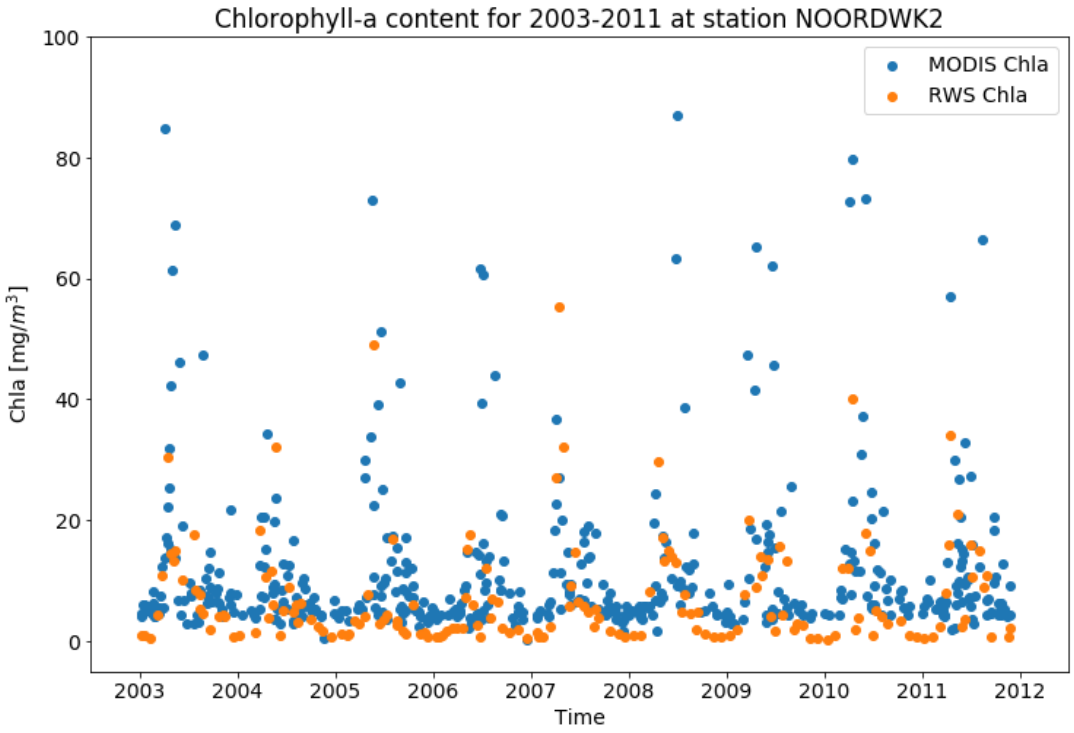


Figure 34 In-situ and MODIS Chla content at the location of the Noordwijk2 station that is 2km offshore on the Noordwijk transect, in an area that in this study is considered as Coastal Zone.

Similarly as before, the original and corrected image for a date with almost full coverage of the study area were plotted (Figure 35). On the left image, there is clear saturation in the values of the coastal area, especially in the south-west part, with very low values being observed while surrounded by much higher values with a sudden change. This error is eradicated after the correction on the right image. As denoted by the used flag, it is possible that these errors imply the presence of clouds that were not masked out in during the pre-processing of the images.

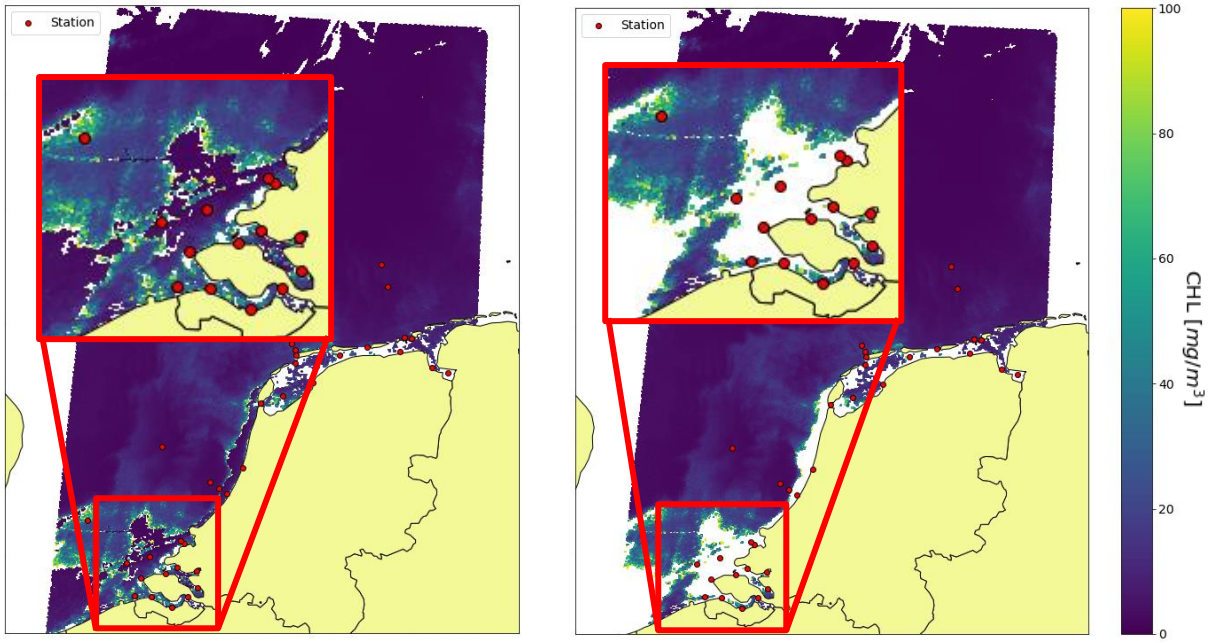


Figure 35 MODIS chlorophyll-a observations for 16/04/2003. On the left side is the original image and on the right side is the corrected one.

7.1.3. AATSR estimation of Sea Surface Temperature

The Sea Surface Temperature data taken by ENVISAT's AATSR were compared with the in-situ water temperature measurements taken by RWS. The original satellite data did not correlate adequately with the in-situ measurements (Figure 36) so they needed to be corrected. The dataset included an overall quality indicator divided into 6 classes by ESA, namely "no data", "bad data", "worst quality", "low quality", "acceptable quality" and "best quality". Thus, through trial and error it was found that by using only the pixels that were labelled as "acceptable "or "best" quality, the correlation and RMSE between the observed and the in-situ data were greatly improved. In detail, the R^2 was 0.37 before the correction and 0.99 after, while the RMSE dropped from 5.87°C to 0.5°C. As it is also evident in Figure 36, the chosen correction resulted in a strong positive correlation between the two datasets accompanied with a low error, hence it was concluded that this methodology resulted in an adequately accurate estimation of the variable.

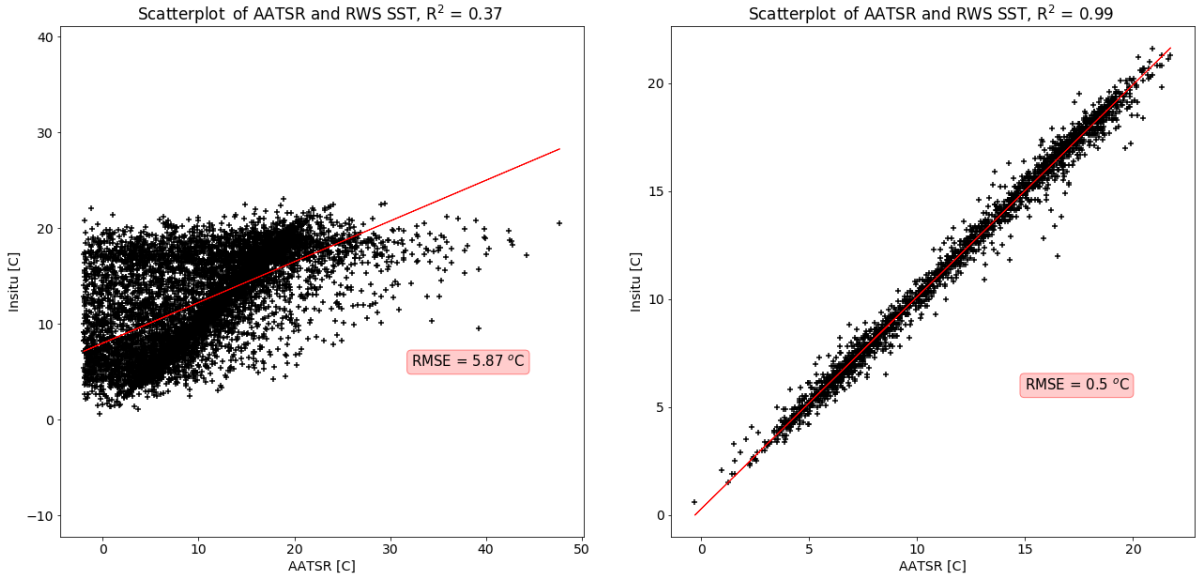


Figure 36 Scatter plot of matched pairs of data records between in-situ and AATSR data, original (left) and after excluding measurements of low quality (right).

7.1.4. MODIS estimation of Sea Surface Temperature

MODIS's SST estimations were also compared with the in-situ water temperature measurements taken by RWS. The original data included various errors and the estimation of the in-situ was poor (Figure 37). The provided per-pixel quality indicator did not improve the result as all pixels were labelled the same. Once again through trial and error but now using the quality flags the correlation and RMSE between the observed and the in-situ data improved. More specifically, R^2 was improved from 0.33 to 0.95, indicating a much stronger positive correlation, while the RMSE dropped from 7.39°C to 1.38°C. Figure 37 shows the final result after identifying the flag that was resulting in the highest number of errors. This flag is 'SSTREFDIFF' (Table 15) and it checks whether the observed SST is significantly higher than the reference SST using a threshold of greater than -1.25 degrees (OBPG-NASA, n.d.-b).

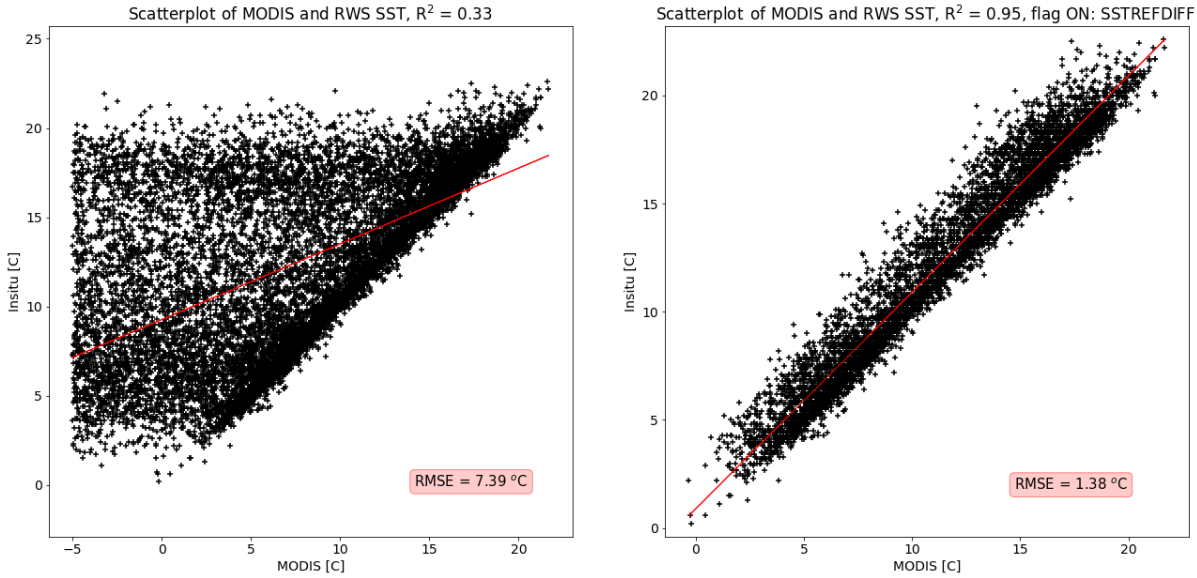


Figure 37 Scatter plot of matched pairs of data records between in-situ and MODIS SST data, after excluding measurements of a specific flag.

7.1.5. Variable relation exploration: Chlorophyll-a and Dissolved oxygen

The satellite Chla estimations and the in-situ measured DO do not seem to be linearly correlated for space-time pairs as it is evident in Figure 38. On the contrary, the pairs of the regular space scatterplot seem to follow a loose power law pattern. Furthermore, using a time lag of up to 120 days as well as rolling windows of different sizes (e.g. 7, 10, 20 days), statistics like the minimum, the average and the maximum value of the data records in the window as well as the slope of a fitted linear model to the existing data points inside the window were calculated. Nevertheless, Figure 39 shows that there was still no strong linear correlation between the two variables so there is not a simple and straightforward way to predict the dissolved oxygen using only the MERIS estimation of chlorophyll-a content.

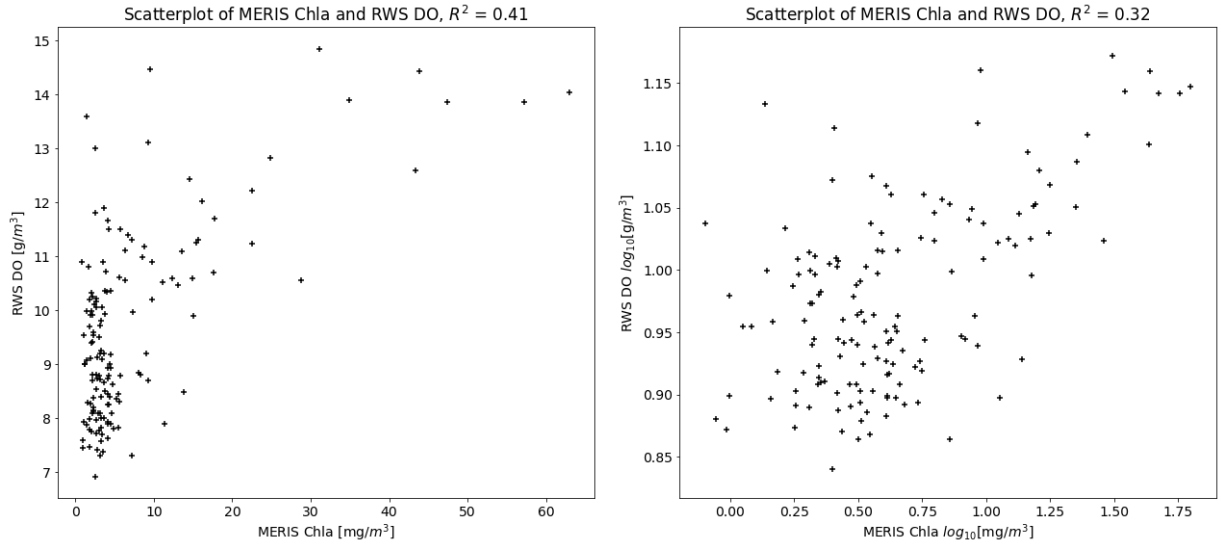


Figure 38 Scatter plot (left) and log-log plot (right) of pairs of data records between in-situ DO and MERIS chlorophyll-a data.

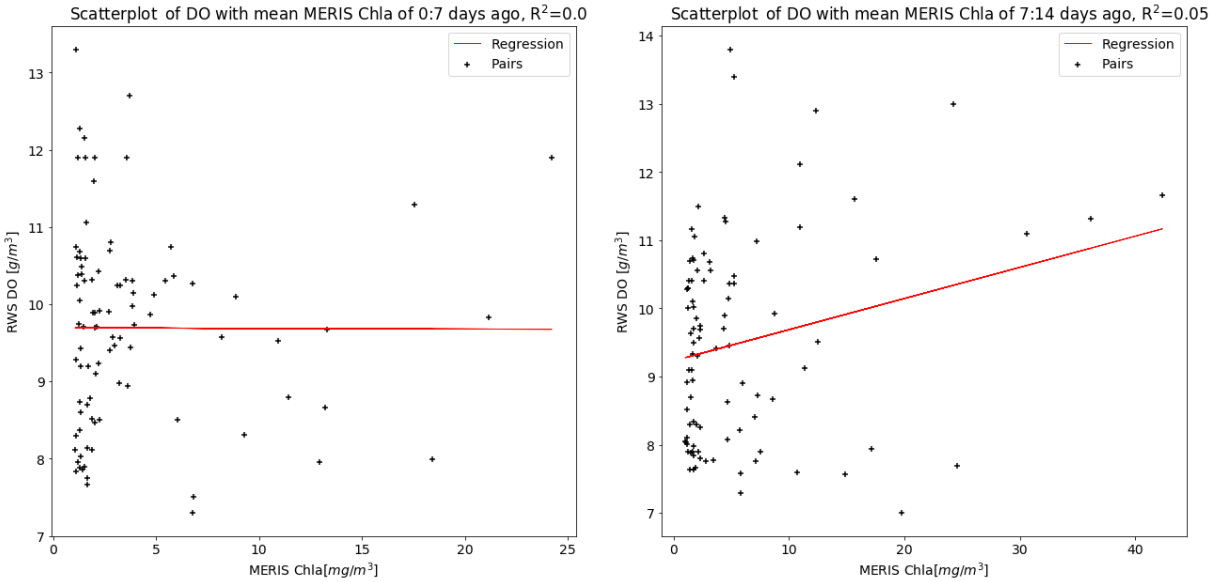


Figure 39 Correlation of Mean value of Chlorophyll-a over a 7 day window computed for two periods before the in-situ dissolved oxygen measurement

7.1.6. Variable relation exploration: Sea Surface Temperature and Dissolved oxygen

Similarly, the same relations as in the previous chapter were explored, this time using the Sea Surface Temperature estimation from the AATSR sensor of ENVISAT. Here, a strong negative correlation between SST and DO was documented with an r-squared value of 0.6-0.65 (Figure 40). This was the highest r-squared value using any of the available window sizes and statistics.

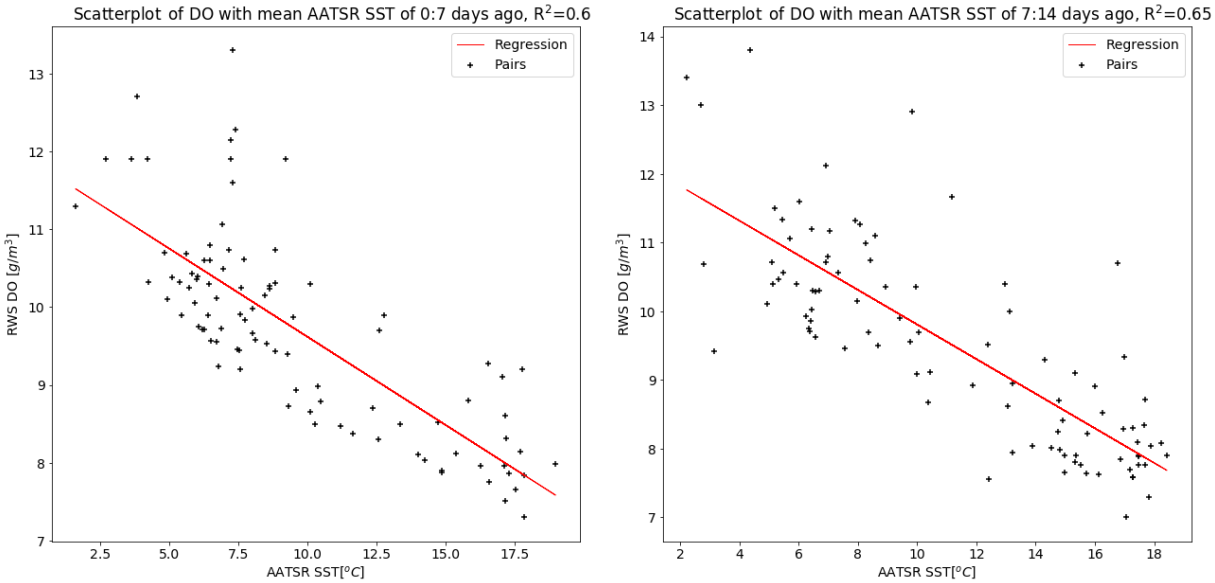


Figure 40 Correlation of Mean value of SST over a 7 day window computed for two periods before the in-situ dissolved oxygen measurement

7.1.7. Temporal patterns and seasonality

Next, the datasets were split into annual subsets and plotted together for each variable (Figure 41). In this way, the seasonality of each variable becomes more evident and the temporal relations between them are easier to read. From the figure below, the negative correlation documented in 7.1.6 can be observed as SST's peak occurring around the 230th day of each year corresponds to the period of DO's annual valley. However, Chla and DO seem to bloom together around the 100th day of each year, while for the rest of the year there is not major evidence that they are correlated.

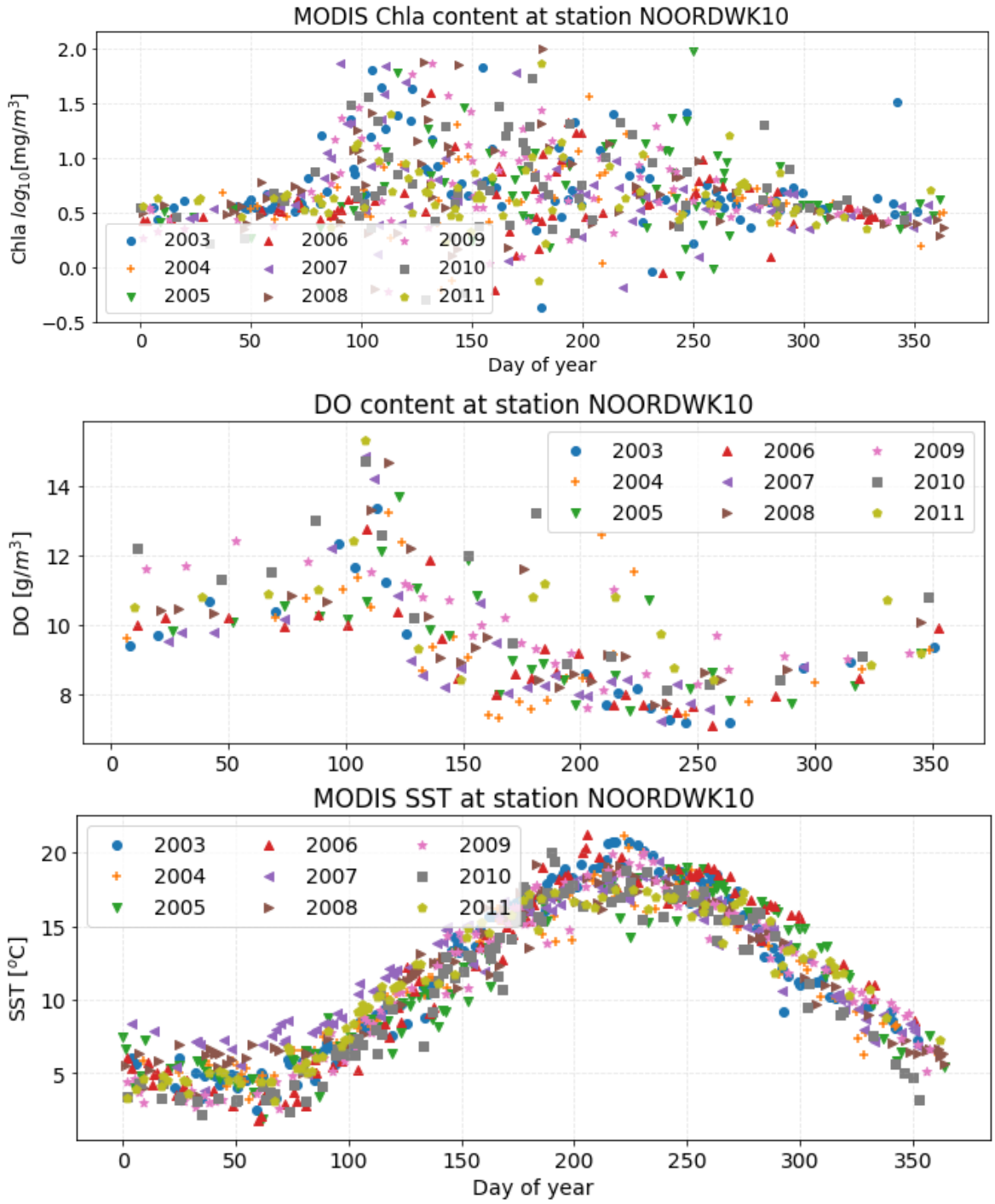


Figure 41 Plots of the annual data records for different variables.

Additionally, Figure 41 reveals that DO occasionally blooms around mid to late summer (in six years out of nine), despite the repeating SST peak around the same time that should force it in the opposite direction (chapter 2.1.3). Likewise, Chla shows a smaller but still present secondary peak in late summer – early autumn.

7.2. Time-independent prediction

By taking n-day averages, the satellite observations were extended in time and more common ground was available to create feature-target data pairs (chapter 6.3.1). The progression of the Root-Mean-Squared Error and R^2 as a function of the window size used for the moving average can be seen in Figure 42 and Figure 43. Here, with window sizes zero and one, the number of data pairs before the division into train and test sets is 13 and 114 respectively, which is too small to take the RMSE provided by the prediction under consideration. Then, with window size equal or greater than three the trade-off between RMSE and R^2 becomes better. Although, keeping the window size relatively small can avoid smoothing the measurements too much. It can also prevent the formulation of dangerous assumptions of stable conditions throughout several days in a dynamic coastal system.

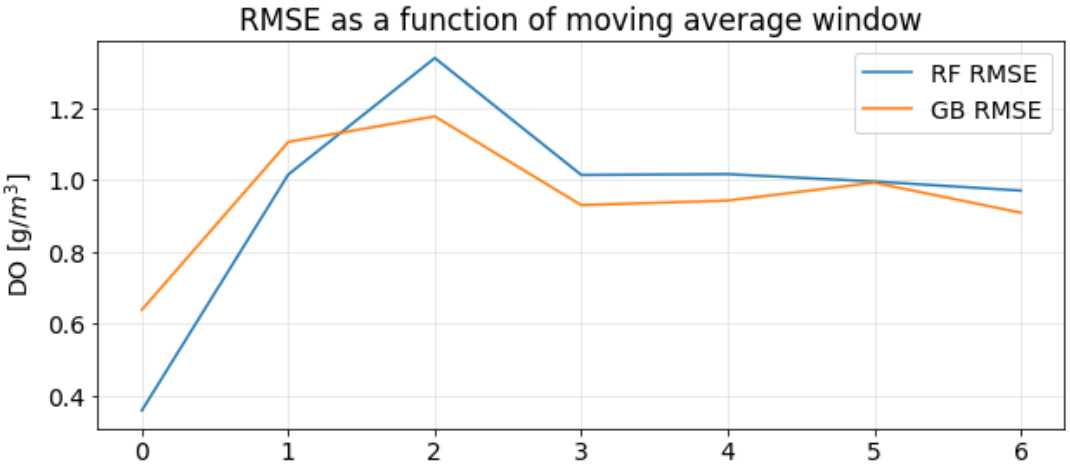


Figure 42 RMSE of the DO prediction as a function of window size of a moving average filter.

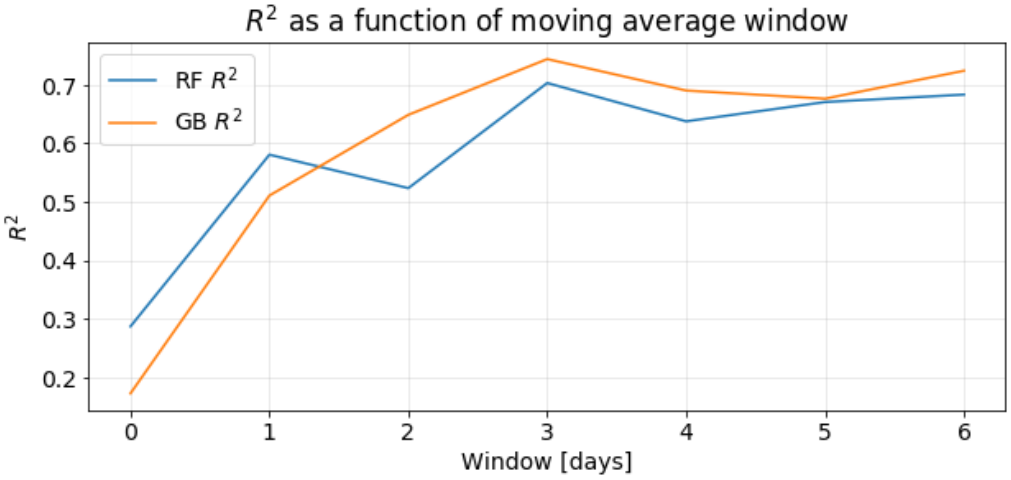


Figure 43 R^2 derived from the linear regression between the true and predicted data as a function of window size of a moving average filter.

In Figure 44 and Figure 45, the prediction results using the original time-independent features are presented. The first figure corresponds to the 310 data pairs after using a three-day window for the moving average, while the second corresponds to seven-day window using 2376 data pairs. Furthermore, in each figure the first column corresponds to the RF Regression result and the second column to the GB Regression. The best result out of four in terms of RMSE and R^2 is achieved using the seven-day window and the GB Regression, with RMSE = 0.93 g/m^3 and R^2 = 0.71.

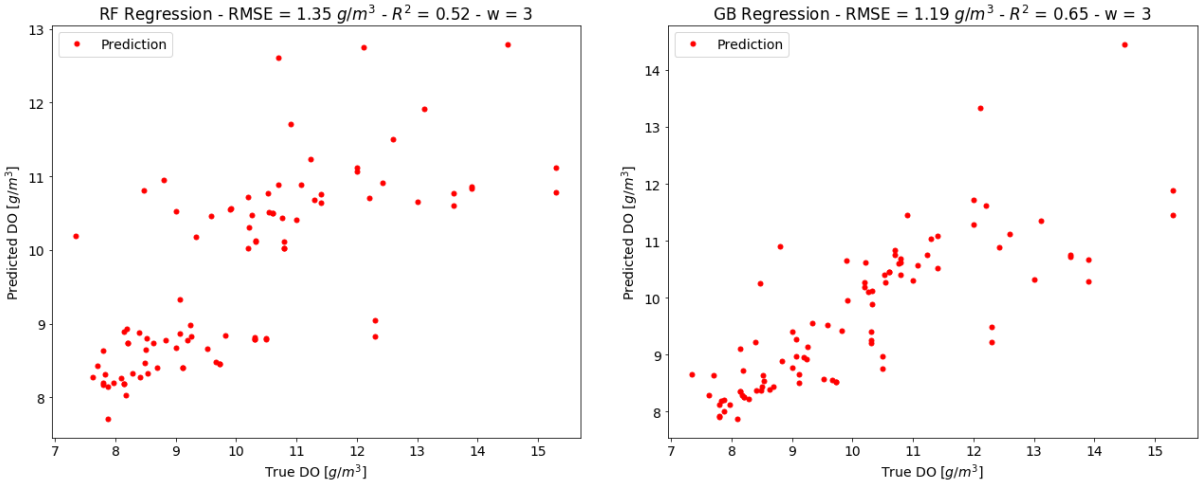


Figure 44 Results of DO prediction using time-independent features and an averaging window equal to 3.

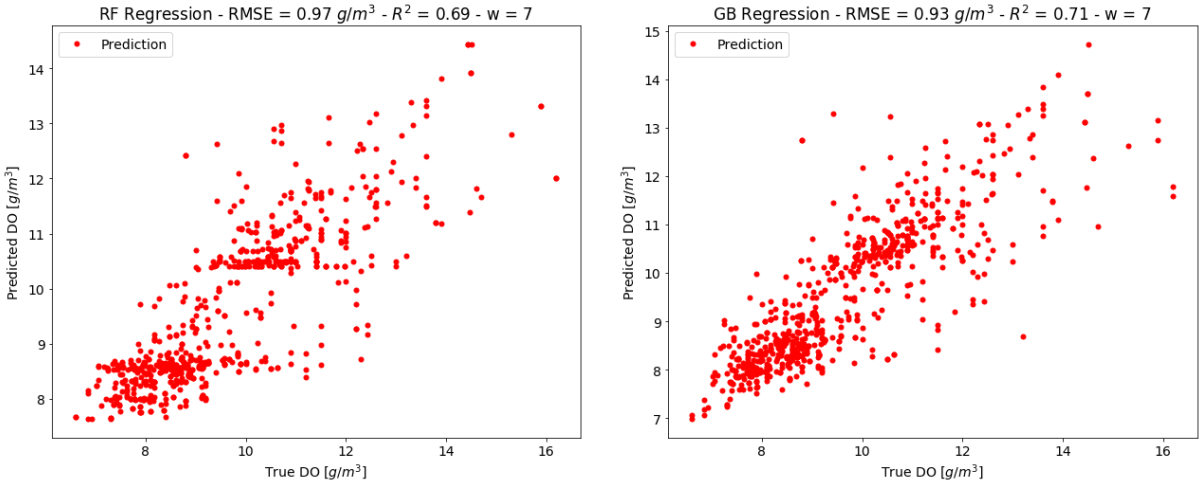


Figure 45 Results of DO prediction using time-independent features and an averaging window equal to 7.

Moreover, the feature importances in the different regressions were calculated. Figure 46 shows a feature importance example corresponding to the GB regression with the seven-day window. The features used correspond to the Chla estimations from MODIS and MERIS and the SST estimations of AATSR and MODIS. Here, it is evident that SST is the most important feature, while Chla contributes less but is still useful in the prediction of DO. Furthermore, the documented collinearity of the SST estimations from the different sensors (see chapters 7.1.3 and 7.1.4) greatly reduces the importance of one of them (in this case of AATSR). Nevertheless,

AATSR's importance is around 15%, which means that it still contributes to the prediction to some extent.

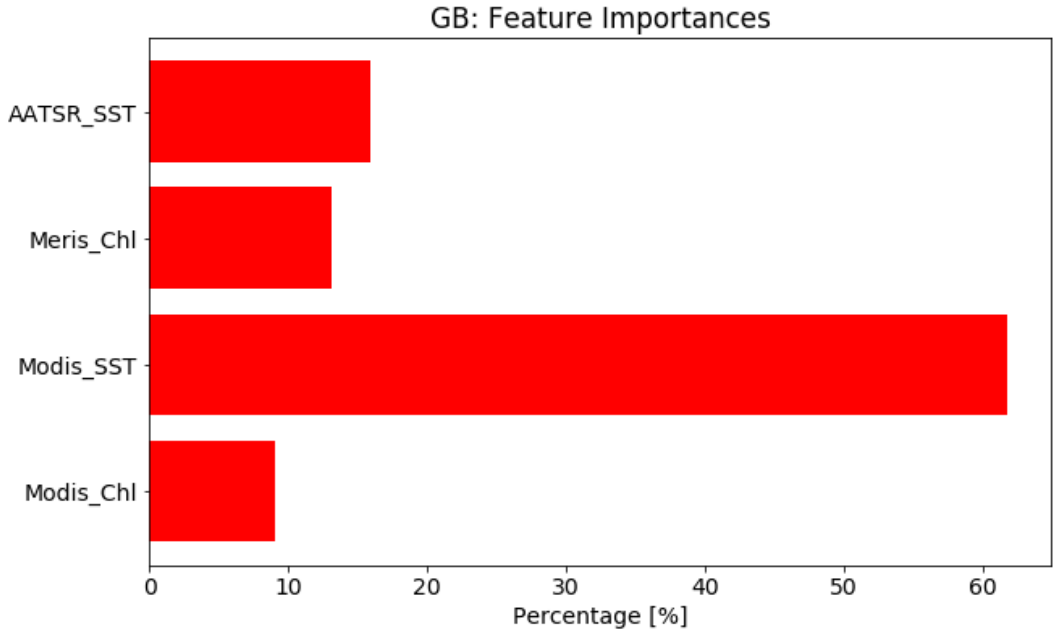


Figure 46 Sample of feature importances for the applied regression models.

7.3. Time-dependent prediction

7.3.1. Temporal gap-filling

As described in chapter 6.3.2 different interpolation techniques were tested out in order to perform gap-filling and create an equidistant time series for every variable. Figure 47 and Figure 48 show two of the most meaningful results for DO and Chla using linear and Akima interpolation. Even though the two gap-filled time-series do not seem to differ significantly from each other, the sharpness of the linear interpolation resulted in an unrealistic estimation. Therefore the Akima method was chosen to be used for the rest of the project as its accuracy on the representation of the underlying process was assumed to be the best among the two.

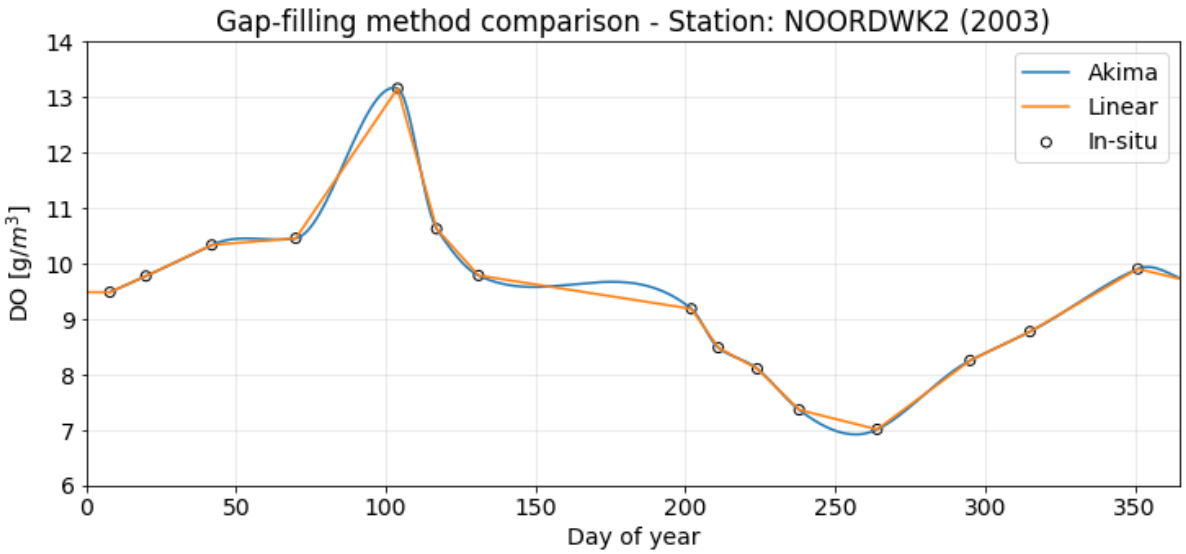


Figure 47 Different results for DO gap-filling at Noordwijk 2km location.

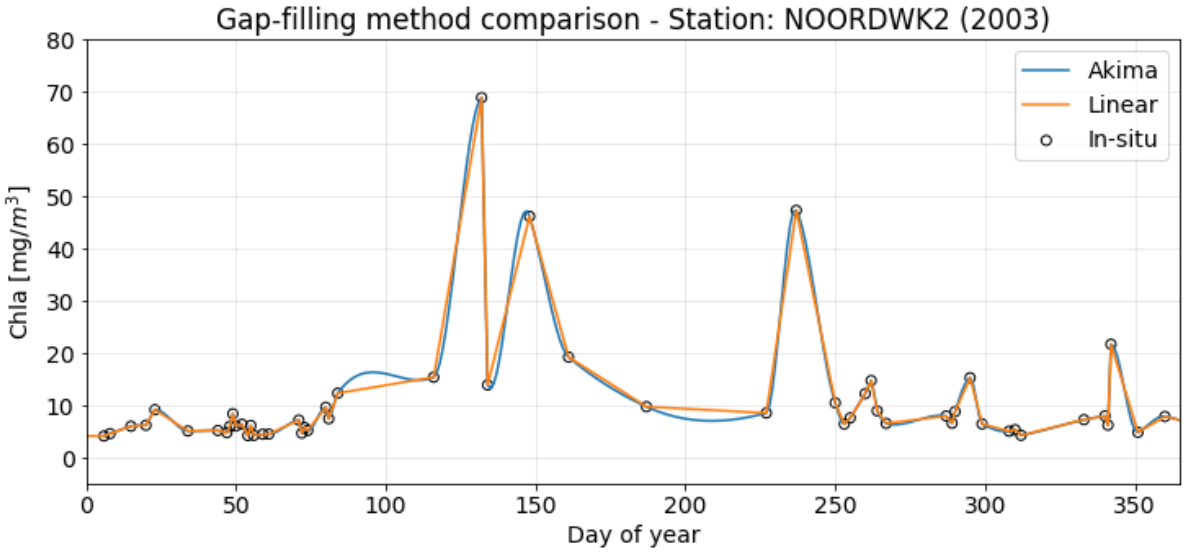


Figure 48 Different results for Chla gap-filling at Noordwijk 2km location.

7.3.2. Seasonal-Trend Decomposition

In Figure 49 the result of the STL decomposition (chapter 6.3.3) for DO can be seen. Here, the DO shows an upward trend in the final years of the study period, while the relative low residuals show that the seasonal component describes the seasonal variability sufficiently. Nonetheless, the drop in the residuals from 2 to -2 g/m³ in 2009, reveals that the DO peak that year happened later than expected. Moreover, the double peaks occurring in the residual signal in 2005, 2006, 2007, 2010 and 2011 underline the inability of the seasonal decomposition to account for the second late summer/early autumn peak, happening with different orders of magnitude in some years.

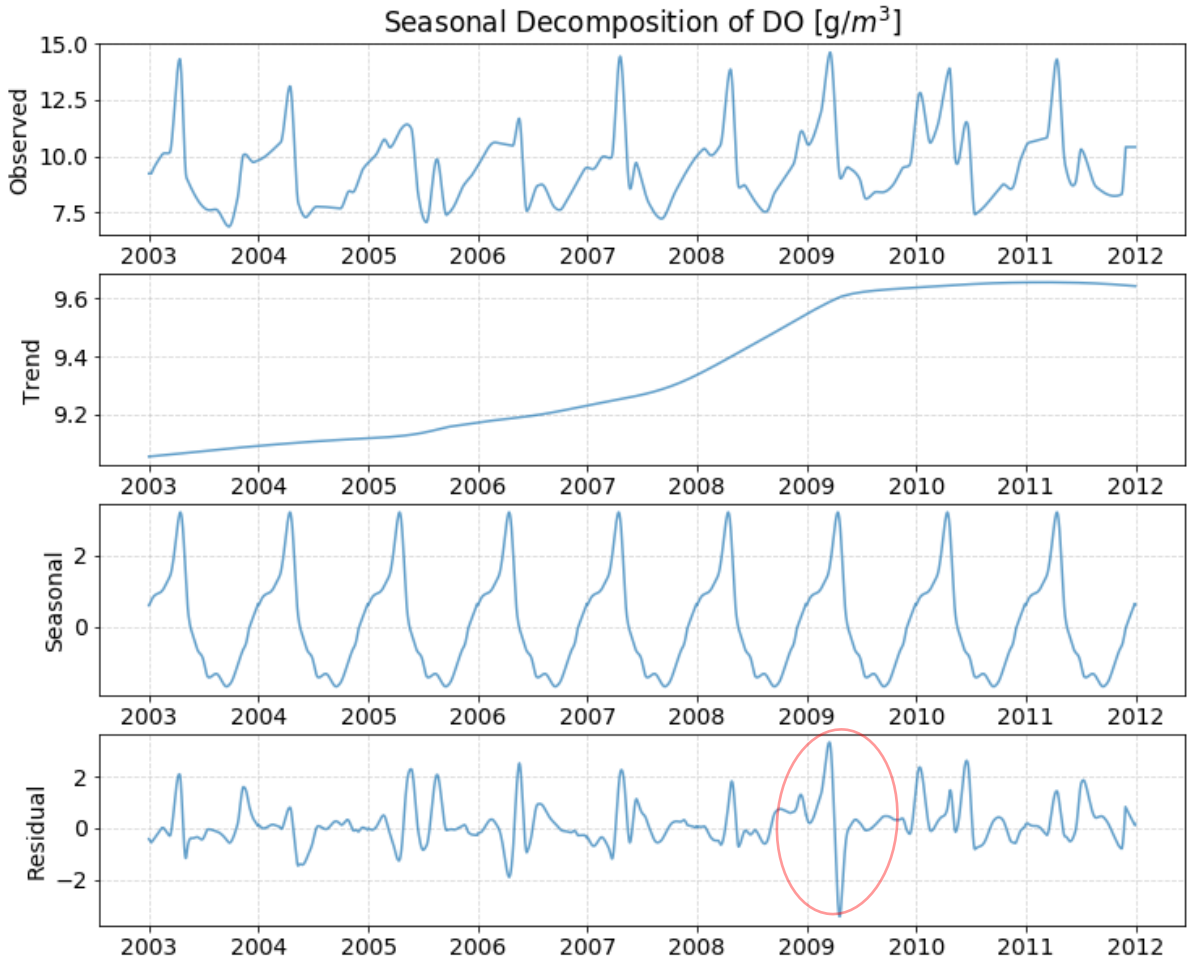


Figure 49 STL decomposition of the Dissolved Oxygen time-series of Goere 6km station.

Furthermore, Figure 50 and Figure 51 show the de-noised time-series for DO, Chla and SST obtained by adding together the corresponding seasonal and trend components and plotting the results normalized in the range [0, 1] for better visual comparison. Here, it is observed that SST had a steadily rising trend in the nine-year study period, following the one of DO. It is also evident that the two variables have a strong negative correlation, as the local maxima of the SST are happening at the same time as the local minima of DO. On the other hand, the trend of Chla decreases until 2006 and then levels off, following an opposite progression than the one of DO and plausibly revealing a decadal or greater cycle. This behaviour of Chla can probably be explained by the documented decrease in winter nutrient loads in the southern coastal part of the North Sea in the period 1990-2014, as well as the absence of a trend in the period 2006-2014. Additionally, the annual nitrogen and phosphorus inputs have also decreased significantly in the Greater North Sea area in the last decades (OSPAR, 2017).

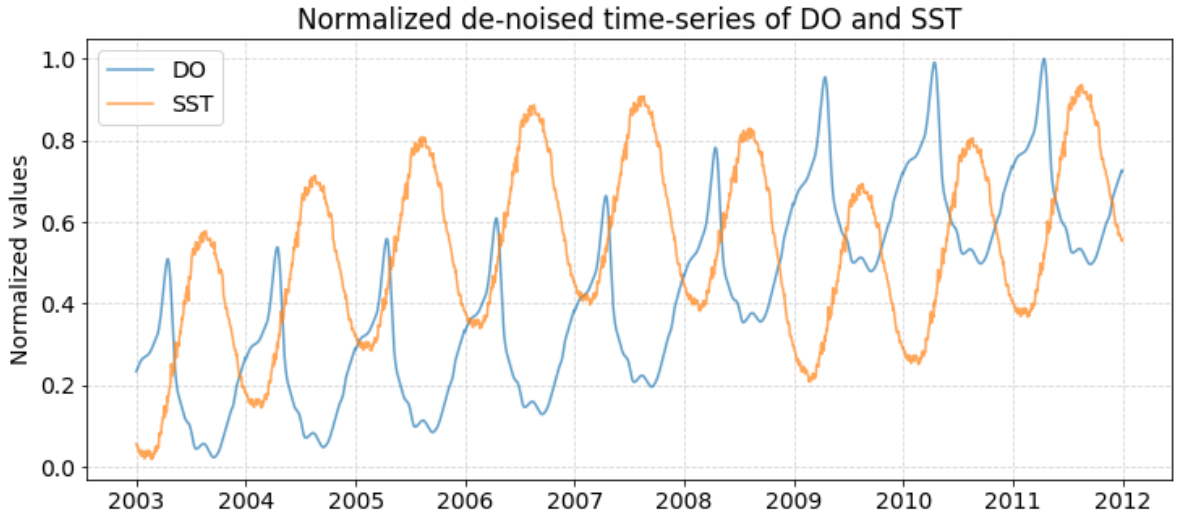


Figure 50 Seasonal and Trend components added together for Dissolved Oxygen comparison with SST at the location of the Goere 6km station.

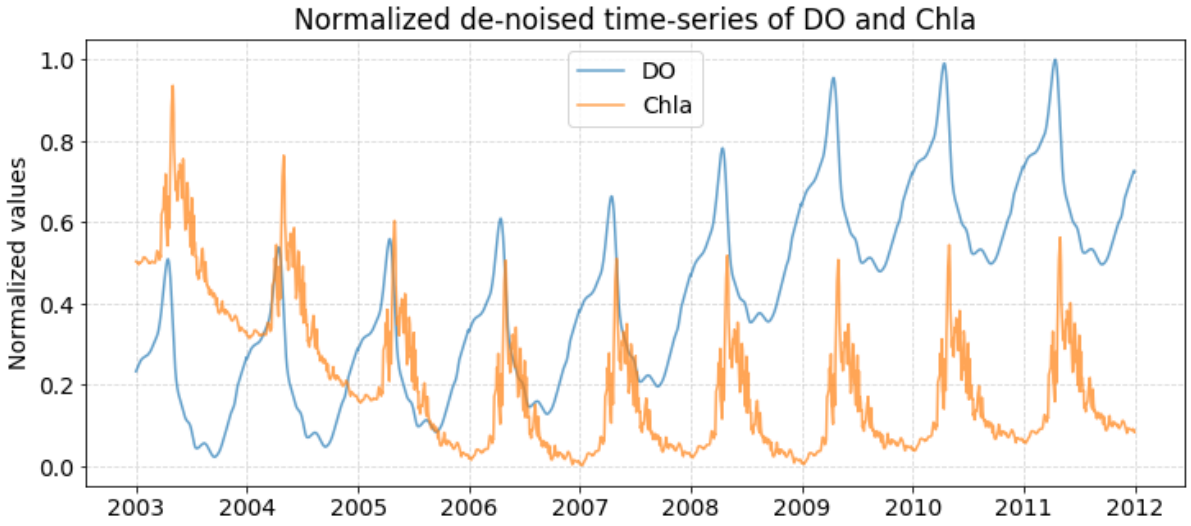


Figure 51 Seasonal and Trend components added together for Dissolved Oxygen comparison with Chlorophyll-a contented at the location of the Goere 6km station.

7.3.3. Prediction using gap-filled data

Many different results have been produced using the approach discussed after chapter 6.3.2 using different combinations of features and conducting a sensitivity analysis. In Figure 52 the result of a GB prediction using the four original signals of SST and Chla from the two satellites can be seen. Here, the overall seasonal variability of the DO is captured along with most of the annual peaks and valleys. The overall RMSE is 0.90 g/m^3, while the RMSE corresponding to the values that are equal or smaller than the average DO value at the specific location is 0.89 g/m^3. The latter RMSE can be characterized as more interesting for this project as it corresponds to the detection of the lower values of DO, which are the ones that induce stress in the ecosystem and can potentially lead to hypoxic conditions (see chapter 2.1.3).

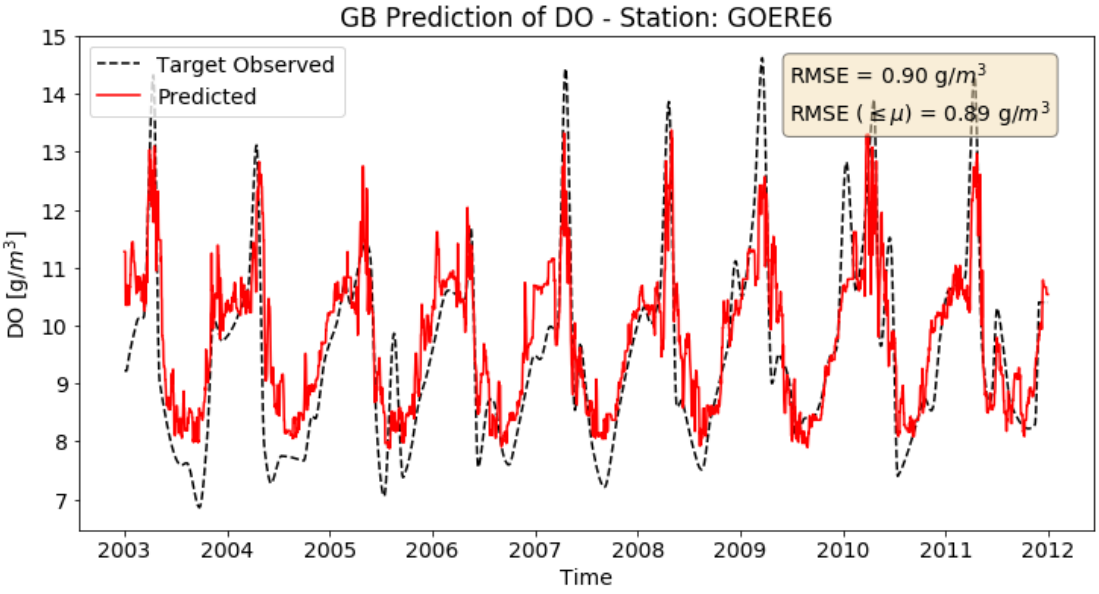


Figure 52 GB Regression using Chla and SST original signals from both MERIS/AATSR and MODIS.

Meanwhile, complementing the original signals with their components produced by the STL decomposition (see 6.3.3) seems to greatly improve the accuracy of the prediction (Figure 53). The predicted signal in this case also seems less noisy than the previous one, which is one of the reasons it demonstrates a higher accuracy in terms of RMSE. In regards with the importance of the decomposed features, the seasonal component of the MODIS SST importance is 60%, while the seasonal component of MERIS Chla has 10% importance. The rest of the features range around 3-7% and the trend component of the MODIS SST is the least important feature with a score close to 0%. Since variables from different sensors contribute to the prediction with high enough feature importances, the use of multiple sensors proves to be beneficial.

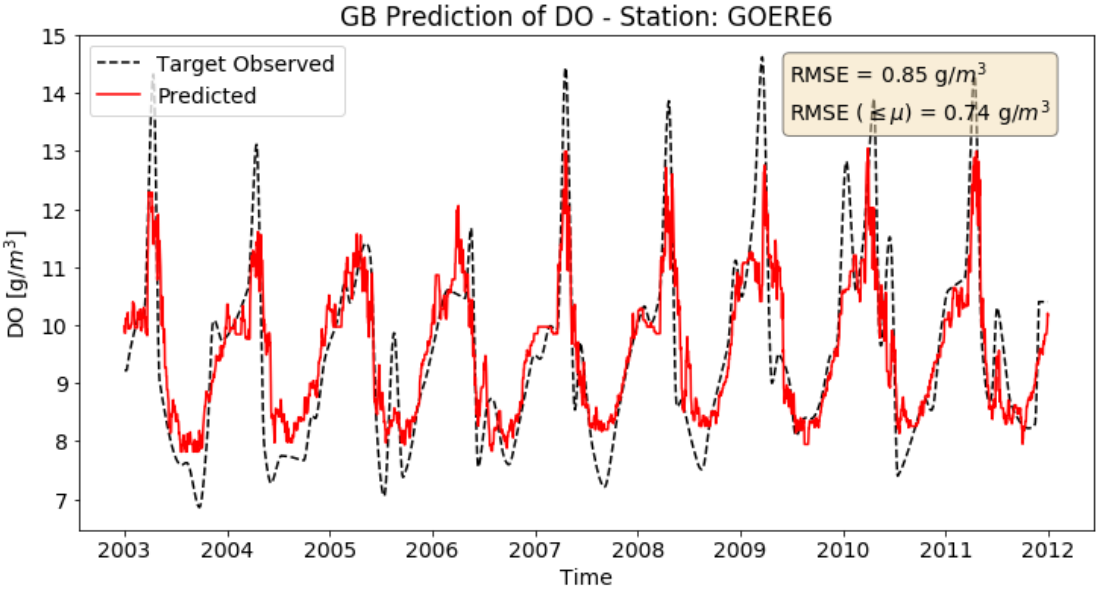


Figure 53 GB Regression using Chla and SST signal components as separate features from both MERIS/AATSR and MODIS.

Finally the best prediction is achieved when both the original and the decomposed signals are combined along with the lagged features (Figure 54). Here, the dominant seasonal variability is predicted reliably as the width of the annual period and the values between the extrema of the predicted data are overlapping with the true data. On the other hand, the extrema, i.e. the peaks and valleys, of the in-situ data are predicted with less accuracy. For example, the maximum value of the peak in 2009 is around 14.5 g/m^3 but the prediction shows a value around 13 g/m^3, while the valley of 2003 is overestimated by 1 g/m^3. Nevertheless, the model approximates the minima better, as the RMSE corresponding to the DO values that are below the average value is remarkably lower than the overall RMSE by roughly 25%. The average errors of the cross-validation procedures for the different setups presented above were computed and can be seen in Table 9. From this table it is inferred that the use of decomposed components as additional features improves the prediction substantially, while the use of lagged features offers only a smaller improvement on the output. Finally, when the results are compared to the simpler prediction scheme explained in chapters 6.3.1 and 7.2., it can be concluded that by performing temporal gap-filling and feature engineering there is great potential for the improvement of the prediction.

Table 9 Average RMSE of the cross-validation procedure using GB

	Time-independent w=7	Original	Original + STL	+ With lags
RMSE [g/m^3]	0.97	0.93	0.84	0.82
RMSE for values ≤ μ_{DO} [g/m^3]	0.73	0.66	0.62	0.60

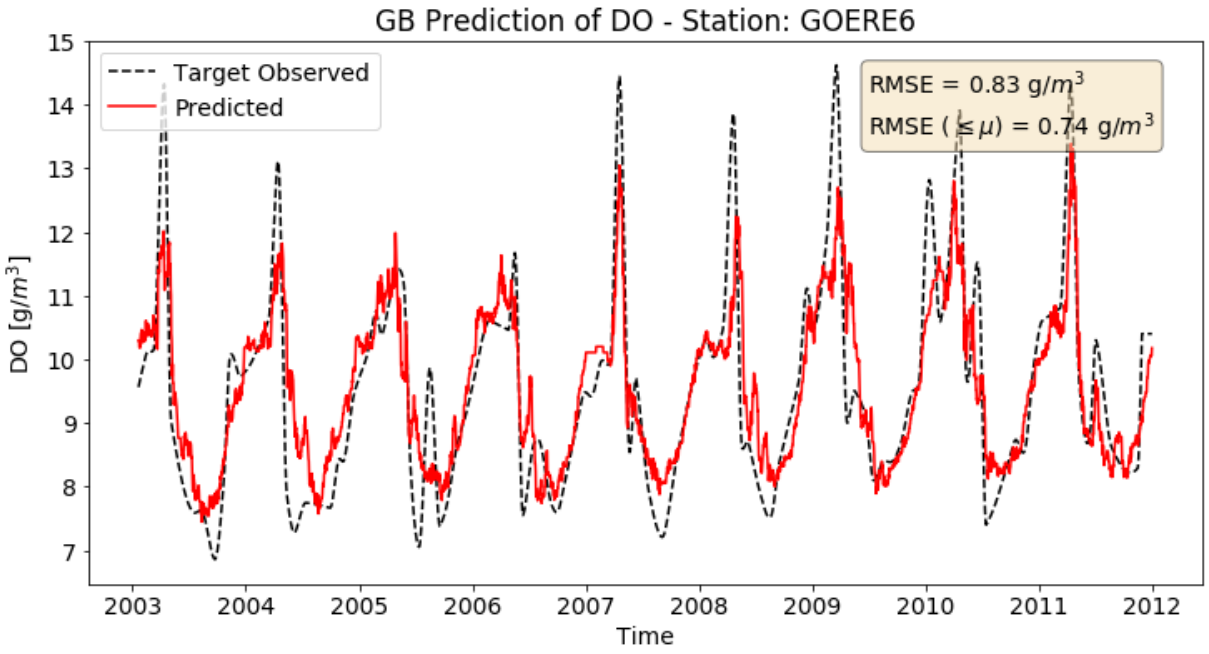


Figure 54 GB Regression using Chla and SST original and decomposed signals.

7.3.4. Sensitivity analysis for feature selection

Another important concept that needed to be explored was the effect of the different variables on the modelling of the DO. On these grounds, a GB model was applied using only SST (Figure 55) and another one using only Chla (Figure 57). From these plots it can be inferred that SST plays a major role in the seasonal layout of the DO but it fails to provide enough information for the prediction of the peaks and valleys. The result is comparable to the CMEMS FOAM AMM7 model prediction of the DO and this is clearer in Figure 56, where the two predictions are subtracted by their respective mean values and plotted together. FOAM only needs a bias correction to overlap with the GB's result.

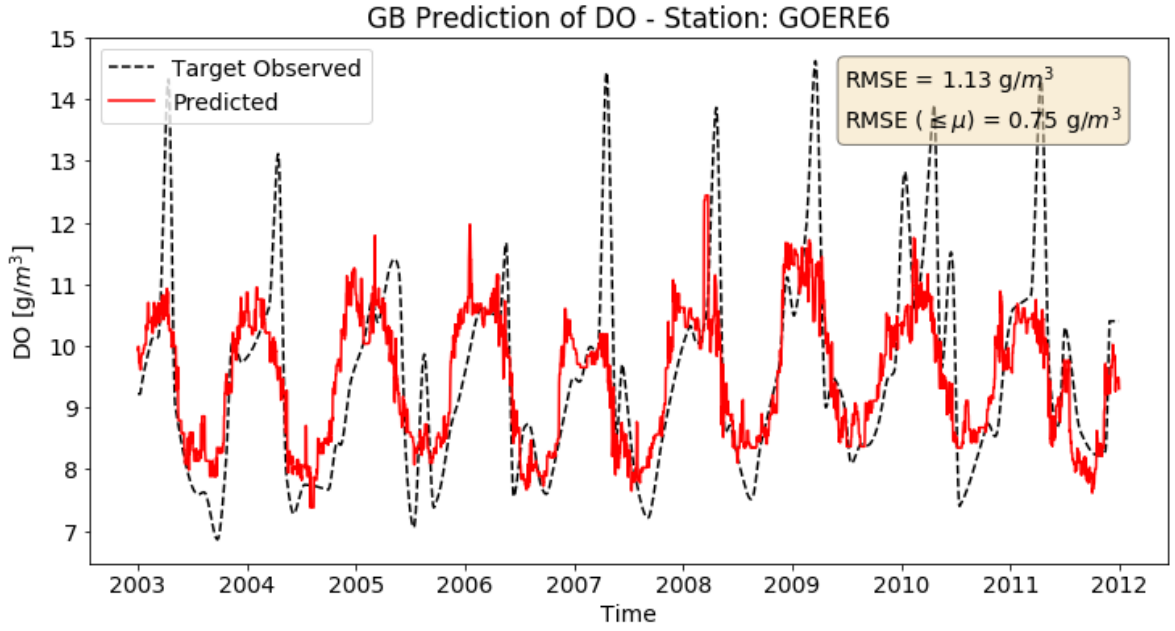


Figure 55 GB prediction using only the SST-related features.

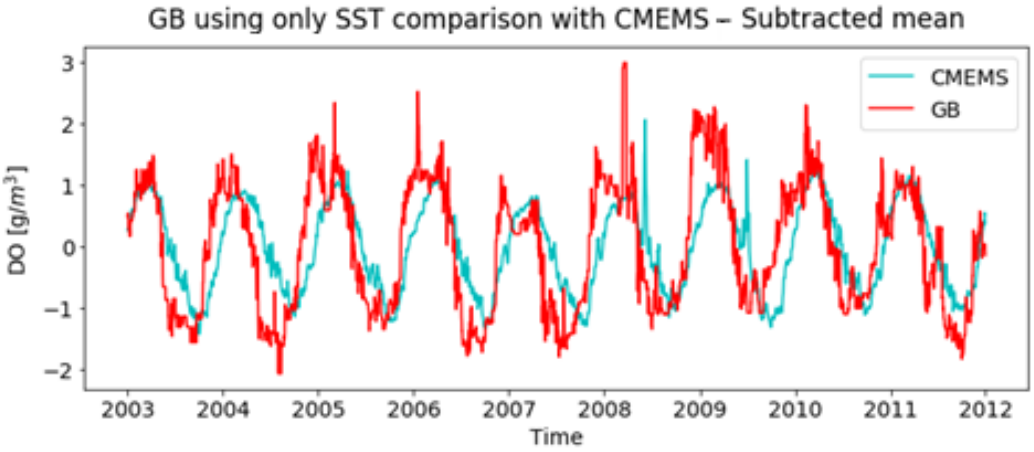


Figure 56 Comparison between GB prediction of DO using only SST features and CMEMS FOAM.

On the other hand, Chla helps the model to predict the DO peaks occurring in the first half of each year but for the rest of the year it seems to introduce noise. This can probably be linked to the correlation between the variables described in chapter 7.1.7. There it is obvious that Chla

values show some variability around the period that DO is relatively high. Meanwhile, Chla remains close to zero for the rest of the year and does not show much correlation with DO. Hence, it is possible that the sensitivity analysis of Chla and its impact on DO reveals the process of photosynthesis, where phytoplankton absorbs the available sunlight and carbon dioxide and produces oxygen.

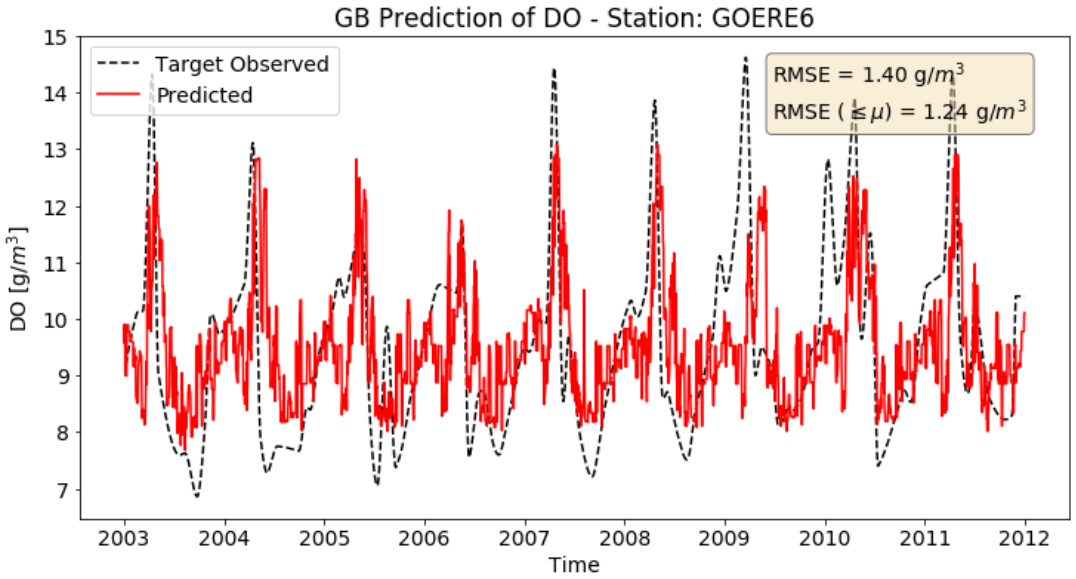


Figure 57 GB prediction using only the Chla-related features.

In combination, both variables provide the best prediction of DO without the need of bias correction that was required for CMEMS FOAM. The results of this variable combination have been shown in the chapter 7.3.3. Moreover, the effect of variables like MERIS's estimation of the TSM was explored. The results seen in Figure 58 indicate that TSM and its components (e.g. seasonal cycle) do not contribute in the prediction of DO as the output of the model is not approximating the true data at all. This was expected as in experiments run during the data exploration, no correlation between TSM and DO was found, as can be seen in the scatterplot in Figure 59.

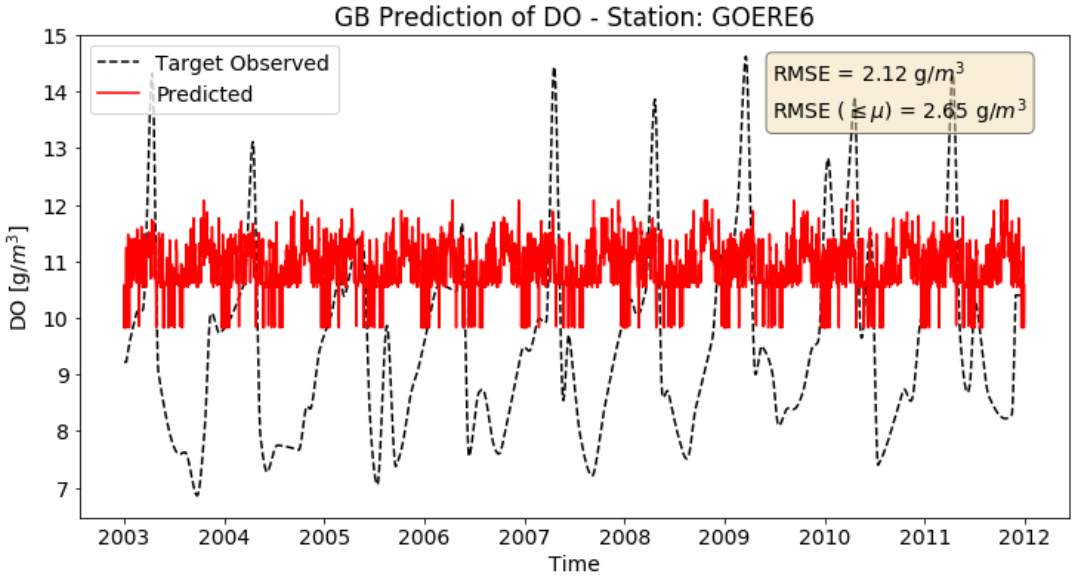


Figure 58 GB prediction using only the TSM-related features.

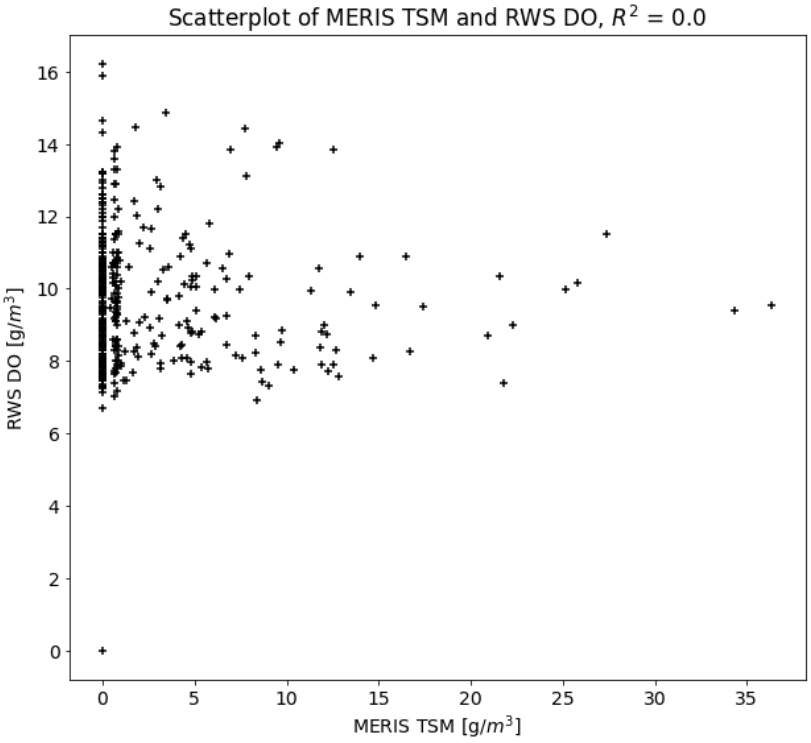


Figure 59 Scatter plot of pairs of data records between in-situ DO and MERIS TSM data.

7.3.5. Model selection and design

An extensive fine-tuning procedure using hyperparameter exploration and design experimentation with RF and GB was realised. Meanwhile, shallow ANN models were tried out and explored to a smaller extent. The most important attributes of each model are shown in Table 10. As it was mentioned in the chapter 6, GB was the model of choice, based on its performance and speed. Nevertheless, a comparison of the different methods in terms of training and test RMS errors was plotted to further justify this choice. Figure 60 shows the individual errors for each station and each technique. The GB demonstrates a small gap between the training and test errors that is comparable to the one of ANN, while RF seems to overfit to the training data more. Moreover, GB outperforms the other models in four out of eight locations and is almost identical to them in another three locations. Finally, it clearly predicts the smaller values better at six out of eight locations, as can be seen in Figure 61, and therefore it is justified as the model of choice.

Table 10 Prediction model general design parameters

	GB	RF	ANN
Number of estimators	20	15	-
Loss function	Huber	MSE	MSE
Max depth of trees/Number of hidden layers	7	7	3
Number of nodes per tree/neurons per layer	10	-	1/10 of the length of the features

Figure 60 Machine learning model comparison in terms of overall test and train RMSE.

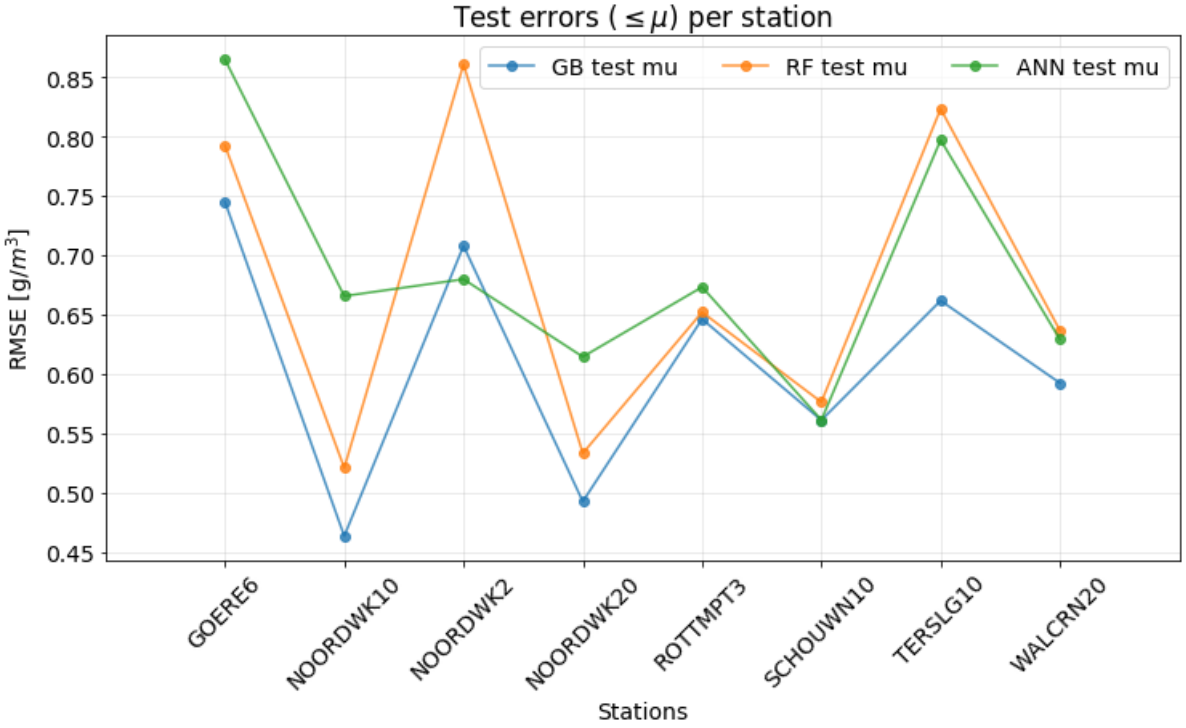


Figure 61 Machine learning method comparison in terms of test RMSE corresponding to the smaller half of the values.

7.4. DO maps for study area

The final step of the methodology included the application of the fine-tuned and validated model to the gap-filled remote sensing data and the production of a DO estimation map for each date for the study area. In Figure 62 an example of these products can be seen. Here, the mid-southern part of the study area shows higher DO content (around 10 g/m^3) and especially in the area of the port of Rotterdam. On the other hand, offshore of the barrier island system of the Wadden Sea DO content estimates are as low as 7.5 g/m^3.

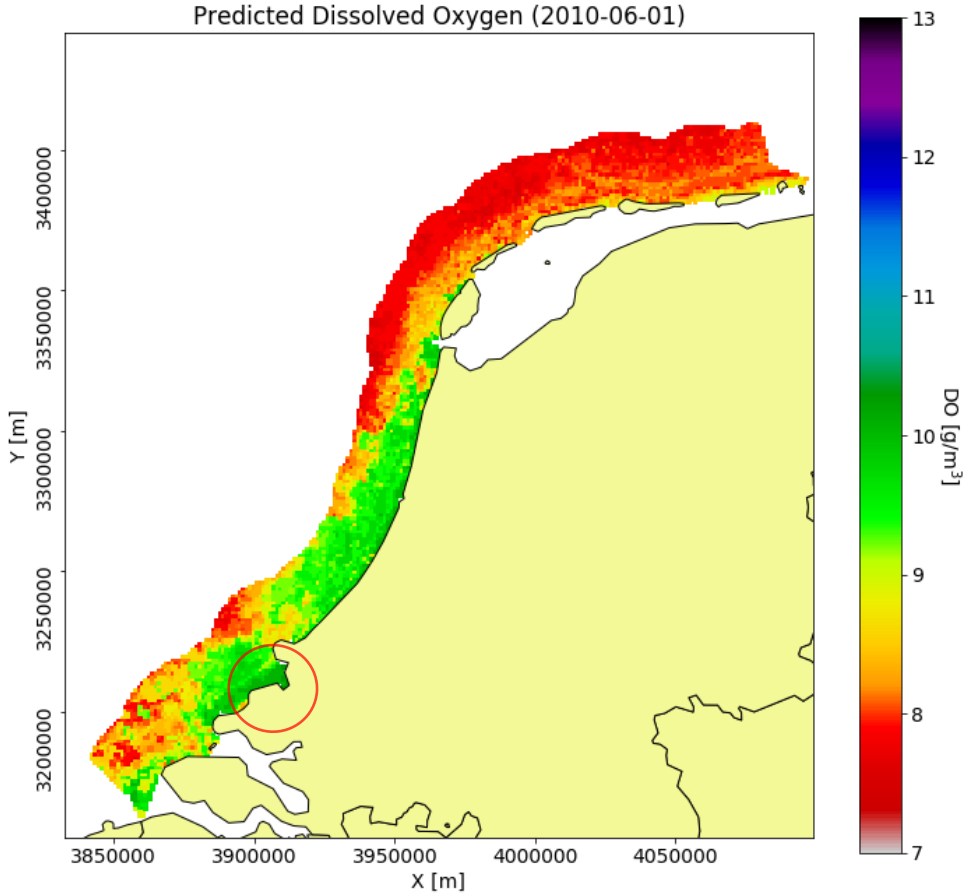


Figure 62 DO map example for 01/06/2010. The ref circle indicates the area of the port of Rotterdam.

Nonetheless, checking each individual map provides limited information on the general spatio-temporal patterns. Using simple statistics for the time-series of each grid point can reveals more information on the dynamics of the area. In detail, Figure 63 includes a map depicting the average DO content estimation, as well as a map depicting the standard deviation of each grid point's time-series. In regards with the first map, average DO content seems to be higher along the coast with the port of Rotterdam, river Scheldt's estuary and east part of the Dutch Wadden Sea barrier islands being the highest, while in general the average DO seems to decrease as a function of distance from shore. Meanwhile, the standard deviation map shows that there are two main directions of variability. On one hand, there is a zone that corresponds to the first few kilometres from the shore that is more variable than the next zone corresponding to the remaining area. On the other hand, the southern part that is heavily affected by the Zeeland Delta estuary seems to be more variable than the middle and northern part of the study area.

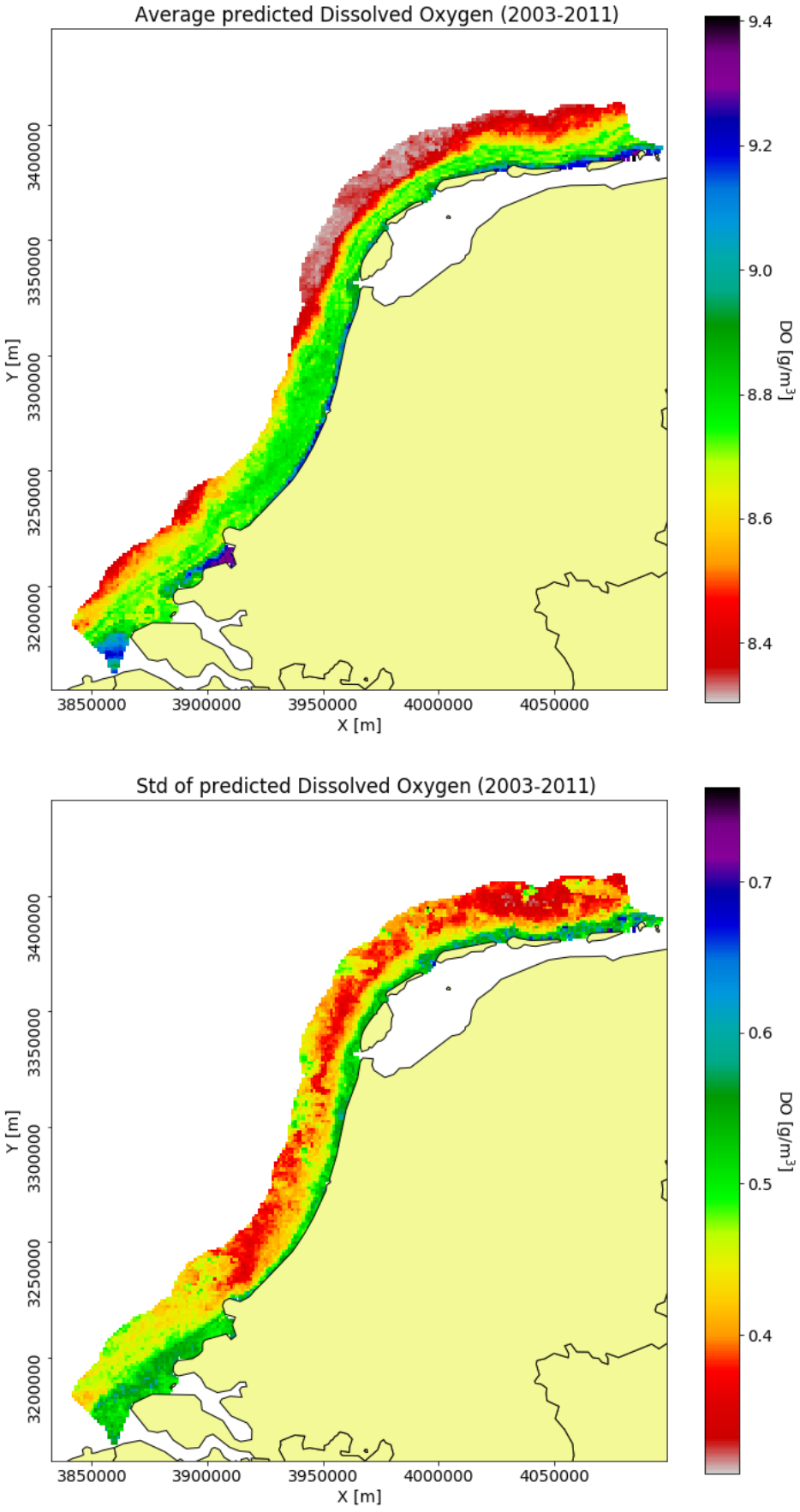


Figure 63 Average value (above) and standard deviation (below) of the predicted time-series of DO content at each grid location for the period 2003-2011.

7.5. Chapter conclusions

In this section, ways to correct the different datasets in terms of systematic (e.g. screening effect) and other errors of the satellite observed variables were found. Furthermore, the presence or absence of correlations between the satellite variables and target in-situ variable (Dissolved Oxygen) was documented and provided insights for the next steps. Based on these it was inferred that SST is negatively correlated with the DO content in the water, supported also by the known effect of water temperature on the solubility of oxygen mentioned in chapter 2.1.3. Meanwhile, Chla as a proxy of phytoplankton biomass does not have a linear relation with DO, but non-linear relations were observed that can be modelled with ML. The extensive presence of errors in the acquired data underlines the importance of an extensive data exploration with various visualizations and statistical metrics before proceeding to later steps of a research.

Different experiments were showcased, namely the time-independent and time-dependent predictions, along with a brief description of their benefits and limitations. The effects of the feature engineering with the decomposed signals and lagged features revealed ways to improve the DO estimation. Additionally, the GB, RF and ANN model validation results were presented, supporting the choice on focusing on the GB model, which proved to be the best among the three in terms of accuracy and over-fitting potential. The chapter concludes with the DO maps after the resampling of the feature variables and the application of the fine-tuned GB models to them.

8. Discussion

The following chapter discusses some important points that have arisen from the research and the results presented before. These include the main reasoning behind the choices made in the methodology, a comparison of the developed data-driven approach with a process-driven model and their documented benefits and limitations. Furthermore, the transferability of the model to a different area is examined through an experiment in Wadden Sea. Finally, the ways that this thesis is relevant today and how its findings can assist stakeholders are presented.

8.1. ML model settings and performance

Experiments have been carried out with various ML models to assess their differences in the DO prediction. RF's performance was characterised by fast and relatively accurate prediction without much tuning. Nevertheless, it was found to overfit more than the other two models (chapter 7.3.5) and it showed the lowest potential in predicting the low DO values. GB proved to be the best in terms of overall accuracy, runtime and minima prediction, while it generalized better to unseen data. This can possibly be attributed to its architecture that trains decisions trees sequentially and pushes them to better predict the target data that the previous tree missed. For instance, the seasonal DO component that attributed for the biggest part of DO's variability was modelled in the first steps of the process using the high importance SST features. Then, the model was pushed to predict the remaining fluctuations by putting less weight in the estimation of the seasonal component. All in all, both ensemble methods were fine-tuned through trial and error, visual inspection and an intensive optimisation scheme described in chapter 6.3.5. The most important hyperparameters that greatly affected the results were the number of estimators

In regards with the ANN, a simple multilayer perceptron was used, where the individual hyperparameters, e.g. number of hidden layers, neurons per layer, activation function, were chosen through trial and error. Furthermore, its runtime was 30-60 times slower than the other two models depending on the number of features and the settings and thus less effort was put on its fine-tuning. Nonetheless, it managed to perform comparably to the GB in terms of accuracy and overfitting; hence, with more extensive experimentation on its parameters and design there is possibly potential to make better predictions in case runtime is not a significant issue.

8.2. Process-driven versus Data-driven approach

Despite the in-situ measurements, certain process-driven models (e.g. numerical, assimilative etc.) provide predictions of DO for the study area, like CMEMS FOAM AMM7 as described in chapter 3.4. Below, Figure 64 and Figure 65 show the in-situ measurements taken by RWS at stations Goeree-6km and Noordwijk-20km, their gap-filled time-series, the prediction of FOAM and the prediction of the proposed methodology of this project using a GB model. The prediction of the GB follows the (unseen) in-situ time-series quite well, capturing the seasonal DO cycle of every year in terms of duration accurately and representing most peaks and valleys adequately. On the other hand, FOAM is not performing equally well and seems to be underestimating DO systematically, which results in significantly higher errors as can be seen in Table 11 for the five locations where it was available. In addition, there are very few fluctuations between the different years as it looks like a similar sinusoidal function is repeated. Furthermore, FOAM presents a lag in its estimation of the valleys while it fails to capture the peaks. Finally, the FOAM-modelled DO in both examples seems identical, in contrast to the GB where differences between the in-situ dataset of the two stations are reflected on the different predictions as well (e.g. positive slope before the peaks of 2004 and 2011).

According to the Quality Information Document accompanied by the model data (Kay, Mcewan, Ford, & Mcewan, 2019), FOAM underestimates DO in comparison to observation-based estimations from World Ocean Atlas (Garcia *et al.*, 2013), but it still follows the same temporal patterns. A smaller but still present negative bias is observed when compared to the in-situ data from ICES Oceanographic Database (ICES, 2014) as well. The authors of the report account that behaviour to the overestimation of the water temperature by FOAM, which leads to lower solubility of oxygen in the water and hence resulting in the biased outcome.

Table 11 Error metrics for CMEMS estimation of in-situ.

Station	RMSE $[g/m^3]$
GOERE6	1.71
NOORDWK10	1.77
NOORDWK20	1.77
SCHOUWN10	1.89
WALCRN20	1.65
Average	1.76

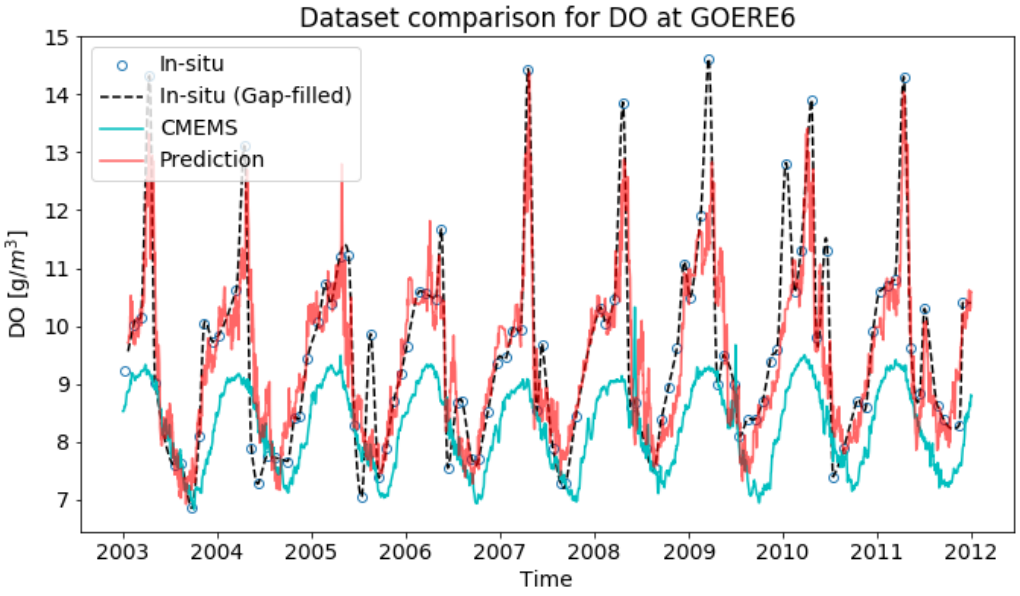


Figure 64 Visual comparison of datasets from different sources for DO, station Goeree 6km.

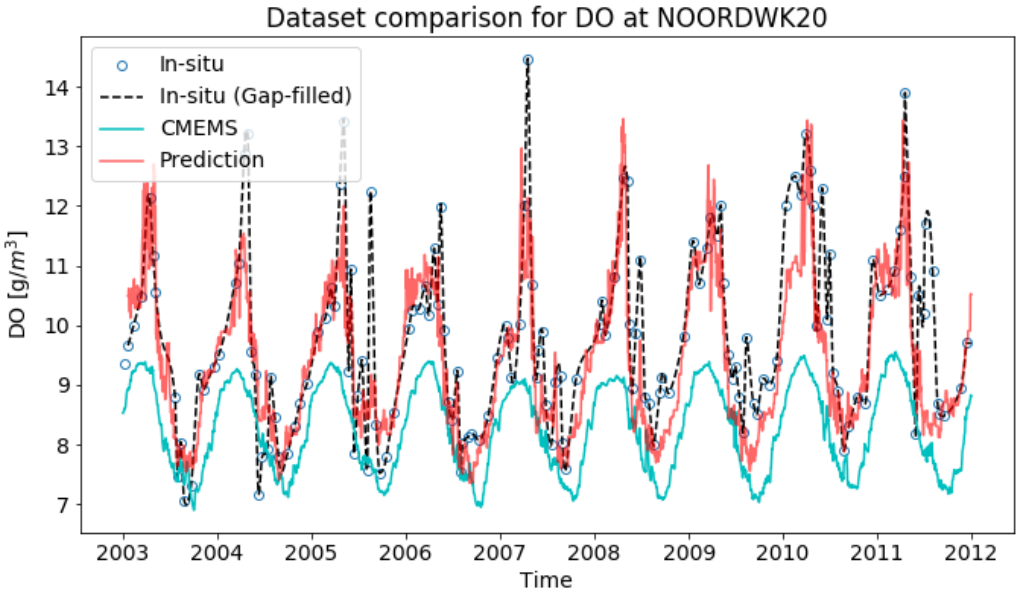


Figure 65 Visual comparison of datasets from different sources for DO, station Noordwijk 20km.

8.3. Model transferability to other study areas

The transferability of the Coast-trained model was assessed by applying it to the stations that are situated in the Dutch Wadden Sea (WS – area description in Appendix 0). Furthermore, a new model was trained and tested using only data from the WS and the comparative results can be seen in Table 12. In Figure 67, the locally WS-trained model seems to perform better than the coast-trained one but they are both failing to capture the extreme values. WS is a highly dynamic shallow sea affected by tides and river discharges. On these grounds, it is possible that the daily varying water depth can affect the variable estimation and therefore ML models using only two water variables may not suffice. For this reason it would be interesting to document other variables, e.g. the water level at the acquisition time of the satellite data, and integrate it in the analysis, e.g. as a separate feature.

Table 12 Error metrics after the application of different models to the Wadden Sea stations

Station	WS-trained		Coast-trained	
	RMSE $[g/m^3]$	RMSE $\leq \mu_{DO}$ $[g/m^3]$	RMSE $[g/m^3]$	RMSE $\leq \mu_{DO}$ $[g/m^3]$
DANTZGT	0.96	0.95	1.29	1.39
DOOVBWT	1.01	0.74	1.08	0.78
GROOTGND	1.04	1.14	1.81	2.17
HUIBGOT	0.89	0.64	1.16	1.10
VLIESM	0.75	0.65	0.88	0.86
Average	0.92	0.80	1.18	1.16

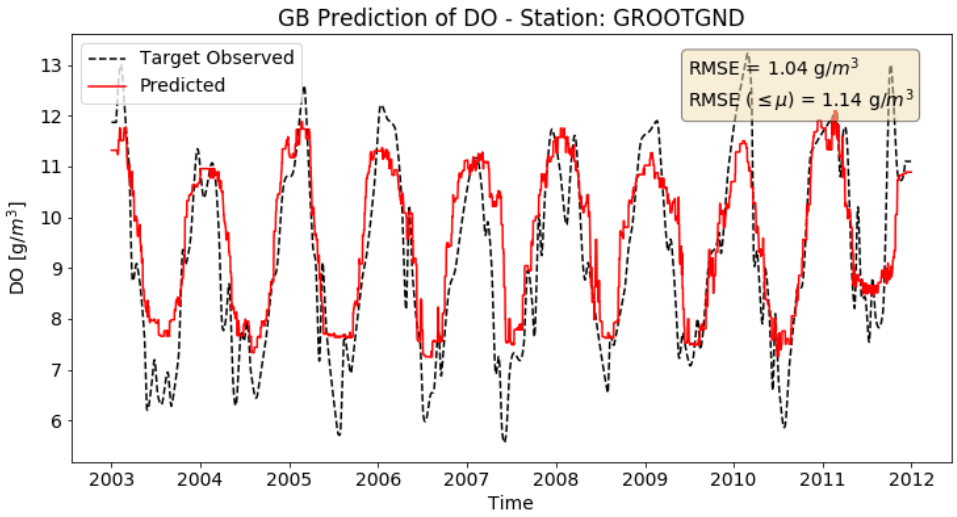


Figure 66 Prediction of DO at station GROOTGND in Wadden Sea using the locally trained model.

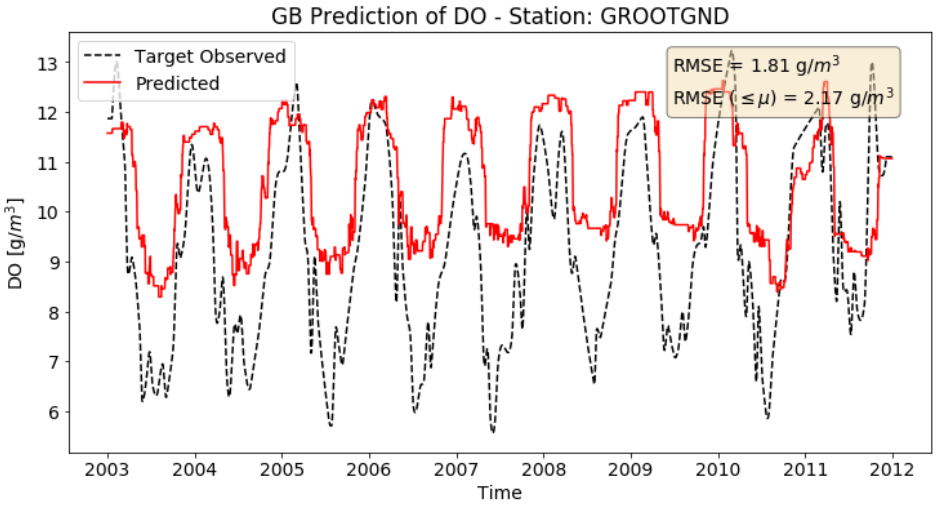


Figure 67 Prediction of DO at station GROOTGND in Wadden Sea using the Coastal trained model.

Normally, the acquisition time of satellite images is known, and for instance the MODIS acquisition time window for the study area spans between 11:30 and 13:30 GMT. Meanwhile, in-situ water level measurements with 10-minute intervals are available by Rijkswaterstaat at five stations in the Wadden Sea, namely Den Helder, Oudeschild, Vlieland Haven, Nes and Schiermonnikoog. The first three are taking these measurements on coast while the last two in areas that are classified as intertidal flats according to Corine Land Cover (2012), as can be seen in Figure 68. In addition, the clear differences in land cover, i.e. permanent water and intertidal flats, pose new challenges for the spatial expansion using a resampling grid as shown in chapter 7.4.

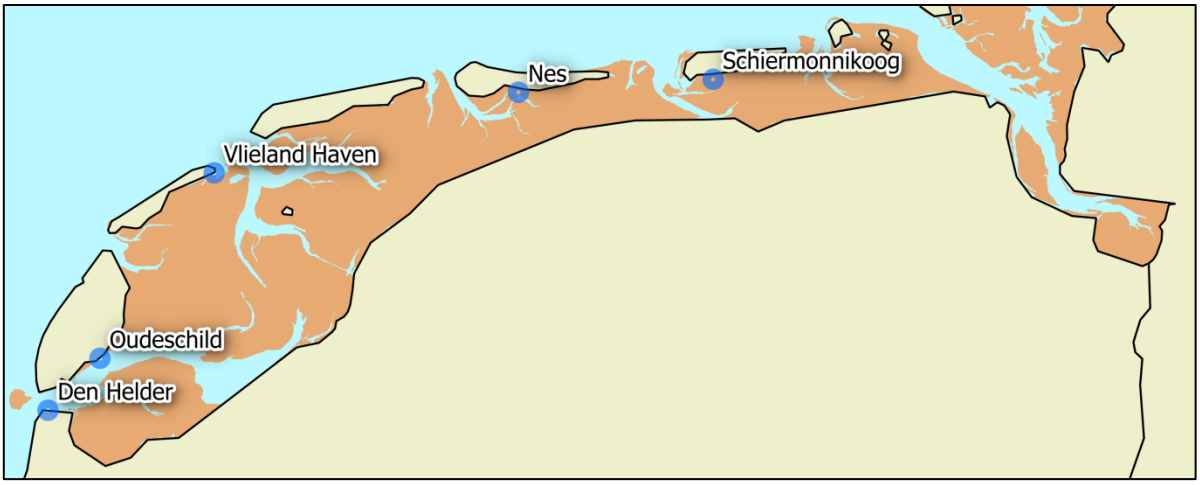


Figure 68 RWS stations measuring water level in the Wadden Sea.

Furthermore, the water level was plotted together with the rough acquisition time window (Figure 69), and from these plots it is evident that it greatly varies across the stations. Additionally, at its lowest values, the water level is around 50 cm, which reveals the infeasibility of the proposed model to work, since it is trained to predict DO content at 1-3 meters of depth. Nevertheless, auxiliary information on the dynamics of the intertidal system, exclusion of areas that may be exposed or very shallow during the low tide and a more ML complex model could possibly improve the prediction.

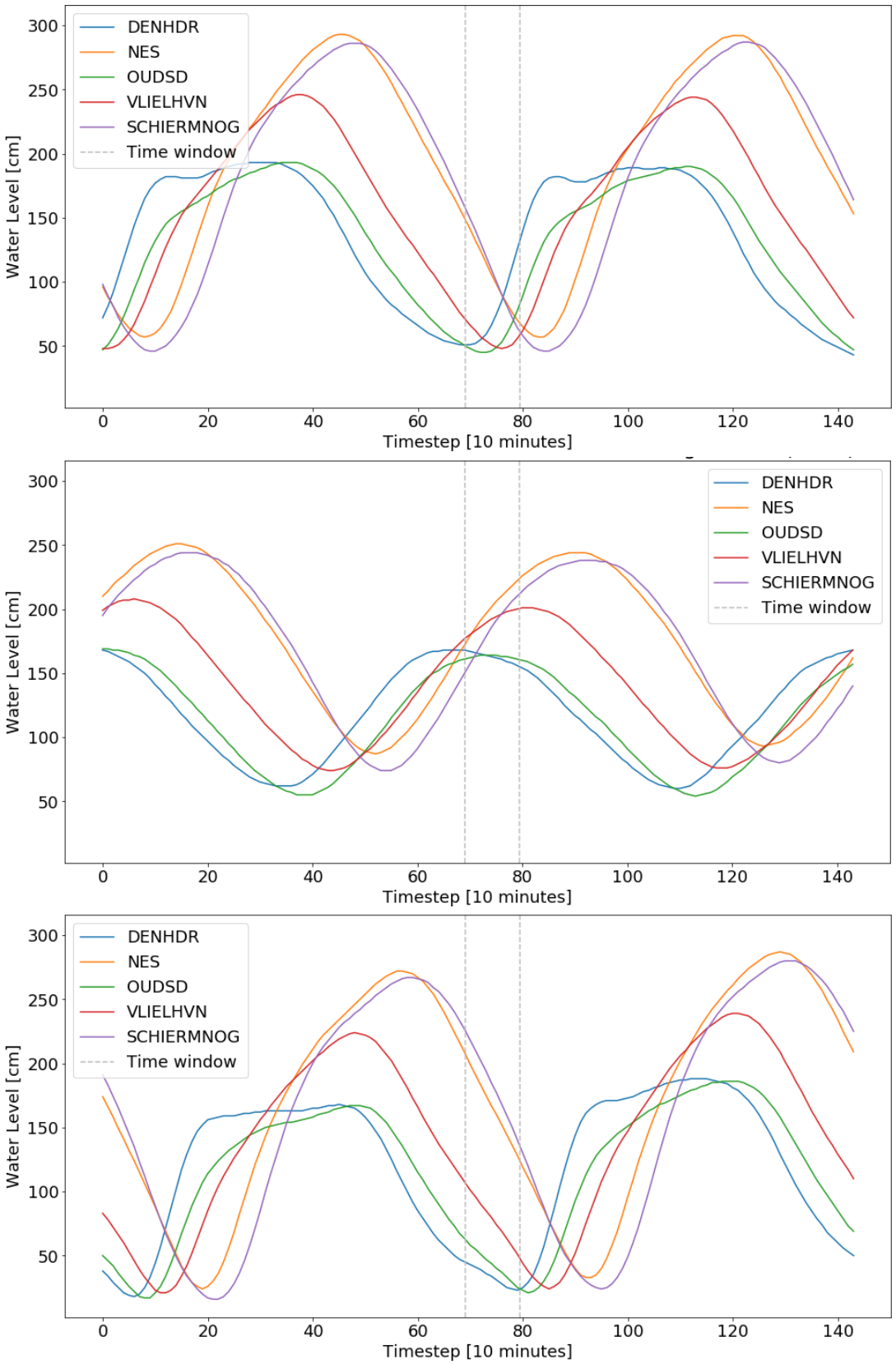


Figure 69 Water level at the five stations and MODIS acquisition window at day 10 (above), day 50 (middle) and day 100 (below) of 2018.

8.4. Benefits and limitations

When compared to the existing process-driven models, the developed model can offer many benefits. These include faster implementation, less need for computer power and higher spatial resolution outputs, while it also does not require in-depth coastal system and physics understanding. It can also be calibrated and applied locally and does not require variables from other influencing systems, e.g. river discharges and large scale oceanic phenomena. Therefore, it can be useful for areas that lack localized complex physical model outputs or areas that need rapid response systems as, with the revisit time of modern satellites, the data can be updated in a close-to-daily basis. Moreover, models like FOAM tend to provide good predictions for the broader area they correspond to (e.g. North-West European Shelf), but lose accuracy closer to the coast. The proposed methodology can assist with filling that gap and even provide data to correct and assimilate the process-driven models.

Despite the benefits, certain limitations of the data-driven model can also be identified. One limitation derived from the application of the model to the Dutch coast and the Dutch Wadden Sea, is the error progression for locations closer to the shore. More specifically, for the prediction in the coastal part, a clear gradient of error increase in the south part can be seen in Figure 70. At the same time, in the Wadden Sea the errors are generally higher and the stations that are further inside from the barrier island have the highest errors. These two stations are the DOOVBWT and GROOTGND and they are in positions that are heavily influenced by river discharges, namely the Rhine through the Ijsselmeer dike and the Ems River respectively. The analysis of the spatial distribution of errors reveals the higher complexity that characterizes the near-shore locations and the locations inside the intertidal zone of the Wadden Sea. In order to model this complexity, further development of the proposed methodology is needed, possibly with additional features and more complex models overall.

Other limitations of the methodology can be identified when compared to existing process-driven models. The proposed model requires an adequate amount of in situ data for the area it is applied to, which can be proven expensive and labour intensive. Meanwhile, it can only be used for “now-casting” and short term forecasting, unless it undergoes more development to integrate long term trends, i.e. need of more data. On the other hand, process-driven models can make predictions and long-term forecasts based only on known physical models and system behaviour, with little to no input data. Unlike the proposed model, process-driven models can also be used to explore different scenarios, test and document the ecosystem’s response and finally assist in long term planning in the direction of preservation and resilience of the local environment. In summary, both methodologies have strengths and weaknesses but with their integration, opportunities for better prediction of environmental variables can arise and facilitate more efficient coastal management.

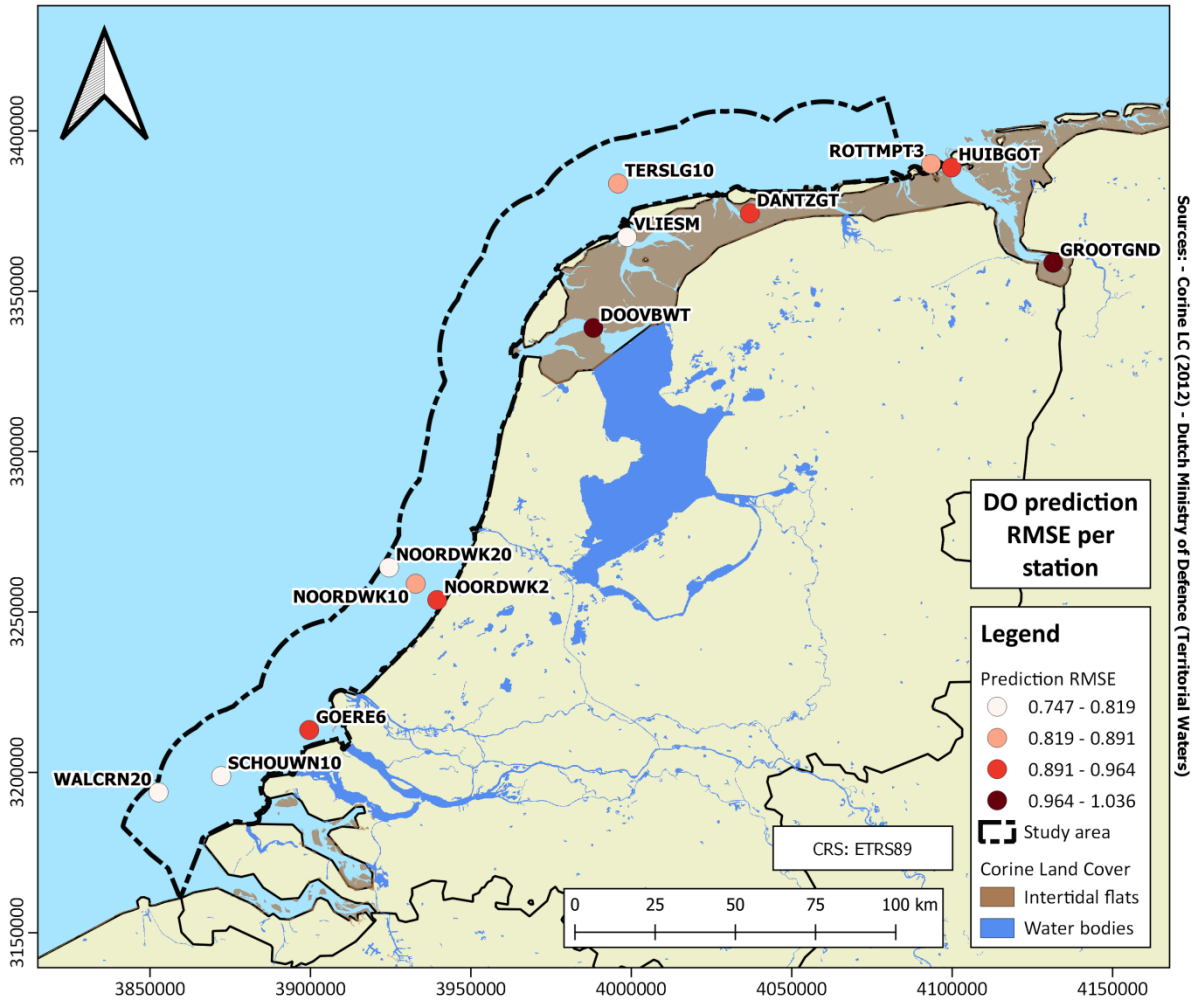


Figure 70 Map showing the spatial distribution of the RMS errors for the DO prediction at each station. The errors are derived using the locally fitted models in each area (study area and Wadden Sea).

8.5. Relevance of research

The MSFD in its report about eutrophication and the achievement of "Good Environmental Status" underlines the importance of anthropogenic eutrophication prevention. This is promoted through monitoring of marine waters for the minimization of ecosystem degradation, diversity loss, HAB formation and oxygen deficiency. Meanwhile, as presented in chapter 2.4, prior research in the field of eutrophication prediction using remote sensing data has been focused mainly in chlorophyll-a estimation and HAB observation. Few publications have experimented with oxygen-related variables, while even fewer attempted their estimation without the assistance of process-driven model outputs. In this framework, the methodology proposed here attempted to bridge the gap between eutrophication prediction studies and DO content estimation using remote sensing data. This was achieved using time-series of ENVISAT data and integrating them with other long time-series datasets like MODIS. While the use of data from satellites that are no longer operational (ENVISAT) may seem counterintuitive, their long time-series can provide valuable insights on how to use similar datasets, as well as expand the predicted variable datasets across heritage sensors. With better quality open data being available nowadays from ESA's Sentinel-3, the continuation of MODIS as well as other sensors dedicated to ocean monitoring like the SNPP VIIRS, the opportunities for better monitoring and

assessment of the environmental status of national and transnational marine waters are expanding. Therefore, the development of open source tools and algorithms can help scientists, engineers and policy makers to effectively monitor marine waters and take counter-measures in a time-efficient way in order to achieve the national and European goals.

9. Conclusions and recommendations

9.1. Answering the research questions

The motivation of this thesis was to make a first step towards addressing the impact of eutrophication on coastal waters by answering a number of research questions (chapter 1.2). These questions are listed below and the proposed answers based on the argument that was being built throughout the chapters are given.

Three complementary research sub-questions were set in chapter 1.2. These were "*Which remote sensing variables better describe the process of eutrophication?*", "*How do the observed variables relate to Dissolved Oxygen concentration, as an important eutrophication indicator?*" and "*How does a data-driven statistical modelling approach perform in comparison with existing assimilative numerical models?".* The answer to the first and second sub-questions can be drawn from the results of the thesis and the feature sensitivity analysis where SST emerged as the most important feature, clearly describing the seasonal variability of DO. This was attributed on the known effect of temperature on the solubility of the oxygen in the water. Furthermore, Chla was also proven to be important for the detection of DO peaks, which was attributed to the process of photosynthesis. Finally, in regards with the third sub-question, the answer has been given in chapter 8.1, where the performance of the proposed methodology against the CMEMS FOAM prediction is assessed. In particular, the proposed methodology outperformed FOAM in the study area, managing to deliver a more accurate prediction in terms of extrema detection. Nevertheless, the fact that FOAM is a model that corresponds to a much larger study area (North-West European shelf) makes it less capable to perform well in a small coastal area it contains, especially when compared to a model that was trained specifically for that area. Additionally, known model biases on the estimation of the SST resulted in a negative bias in the DO prediction, which was not an issue for the data-driven approach. In this sense, a data-driven approach can have advantages over process-driven models.

The main research question was *"How adequate are freely available Earth Observation data for eutrophication monitoring and prediction in the Dutch coastal waters?"*. In general, remote sensing data have been used to approximate the Chla content in the water but when combined with ML models and ground truth data for training and validation they can be used to predict oxygen-related variables as well. The results of this thesis show that open remote sensing data are capable of providing important information for the predictions of DO content, an important indicator of eutrophication that has been studied in the past using mostly in-situ measurements and process-driven models. More specifically, using the proposed GB model the seasonal variability of the DO can be predicted with high accuracy, while its minima that indicate a tendency towards hypoxic conditions and ecosystem stress can also be approximated. Overall, the RMSE of the prediction fluctuates around 0.8 g/m^3, while for the values below the average DO content the RMSE drops around 0.6 g/m^3. Moreover, the methodology can be used to spatially extend the DO prediction, and along with the required pre-processing, provide a continuous spatial and temporal approximation of the variable. This way more general conclusions concerning the environmental status of the study area can be drawn through the identification of the dominant DO spatial patterns, which would be invisible in case only a few point measurements were used.

The results of this thesis reveal the feasibility of Dissolved Oxygen content prediction in coastal waters using remotely sensed water quality variables. Sea surface temperature and chlorophyll-a content have arisen as the most useful variables to achieve this, while others like the suspended matter content did not assist the prediction in the study area. Moreover, the temporal gap-filling of the variables and the decomposition of their time-series to temporal components seemed to further improve the prediction. Although there is still a lot of room for improvement, this thesis can assist with the monitoring of the Dissolved Oxygen content in marine areas for more efficient mitigation of its negative environmental impacts. It can also assist in the assimilation of general process-driven numerical models like FOAM, as well as for rapid response systems in areas that lack outputs from localized models.

9.2. Recommendations

As this field of application is relatively unexplored, there are many opportunities for future research. A first step would be to implement the methodology using the latest Sentinel-3 data once in-situ measurements giving at least a year of overlapping observations become available. Then, taking into account the demonstrated benefits of multiple sensors integration, opportunities may arise in the sensor fusion field. AQUA MODIS, SNPP VIIRS and possibly other satellite data, can be used together to create denser variable datasets with lower uncertainty that can result in more robust Chla estimation and DO prediction.

Furthermore, additional data and features can be explored for the improvement of the model for more complex areas, like the Wadden Sea. The proposed methodology can be expanded with additional feature engineering and higher complexity ANN models, e.g. 1D Convolutional Neural Networks. Moreover, time-series forecasting methods can be explored for short-term forecasting either using a different setup of features in the current models (using the already created lagged features to predict for a few days in the future), or by implementing more sophisticated models like Recurrent Neural Networks. Finally, the implementation of the methodology in a coastal area with observed hypoxic or anoxic conditions could offer great insights as to whether these conditions can be captured and predicted.

References

Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., … Zheng, X. (2015). TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the Assosiation for Computing Machinery*, *17*(4), 589–602.

Andersen, N. G., Hansen, P. J., & Engell-sørensen, K. (2015). Ichthyotoxicity of the microalga Pseudochattonella farcimen under laboratory and field conditions in Danish waters. *Diseases of Aquation Organisms*, *116* (October), 165–172. https://doi.org/10.3354/dao02916

Anderson, D. M., Cembella, A. D., & Hallegraeff, G. M. (2017). Progress in understanding harmful algal blooms (HABs): Paradigm shifts and new technologies for research, monitoring and management. *Annual Review of Marine Science*, *4*, 143–176. https://doi.org/10.1146/annurev-marine-120308-081121.Progress

Bachelet, G., De Montaudouin, X., Auby, I., & Labourg, P. J. (2000). Seasonal changes in macrophyte and macrozoobenthos assemblages in three coastal lagoons under varying degrees of eutrophication. *ICES Journal of Marine Science*, *57*(5), 1495–1506. https://doi.org/10.1006/jmsc.2000.0902

Balch, W. M., Gordon, H. R., Bowler, B. C., Drapeau, D. T., & Booth, E. S. (2005). Calcium carbonate measurements in the surface global ocean based on Moderate-Resolution Imaging Spectroradiometer data. *Journal of Geophysical Research C: Oceans*, *110*(7), 1–21. https://doi.org/10.1029/2004JC002560

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., & Silliman, B. R. (2011). The value of estuarine and coastal ecosystem services. *Ecological Monographs*, *81*(2), 169–193. https://doi.org/10.1890/10-1510.1

Blaas, M. (2013). *Eutrophication Assessment Using Remotely Sensed and In Situ Chlorophyll-a Data*. Delft, The Netherlands.

Blondeau-Patissier, D., Tilstone, G. H., Martinez-Vicente, V., & FMoore, G. (2004). Comparison of bio-physical marine products from SeaWiFS , MODIS and a bio-optical model with in situ measurements from Northern European waters. *Journal of Optics A: Pure and Applied Optics*, *6*, 875–889. https://doi.org/10.1088/1464-4258/6/9/010

Bonsdorff, E., Blomqvist, E. M., Mattila, J., & Norkko, A. (1997). Coastal eutrophication: Causes, consequences and perspectives in the archipelago areas of the Northern Baltic Sea. *Estuarine, Coastal and Shelf Science*, *44*(SUPPL. A), 63–72. https://doi.org/10.1016/S0272-7714(97)80008-X

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*, 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees* (1st ed.). Wadsworth, New York: Chapman and Hall.

Cardaci, M. (2013). ENVISAT-1 products specifications - AATSR products specifications. ESA.

CEARAC. (2007). *Eutrophication Monitoring Guidelines by Remote Sensing for the NOWPAP Region*. Toyama, Japan.

Chang, N., Imen, S., & Vannah, B. (2015). Remote Sensing for Monitoring Surface Water Quality

Status and Ecosystem State in Relation to the Nutrient Cycle : A 40-Year Perspective. *Critical Reviews in Environmental Science and Technology © Taylor & Francis Group, LLC*, *45*, 101–166. https://doi.org/10.1080/10643389.2013.829981

Chollet, F. (2017). *Deep Learning with Python*. Manning.

Cleveland, R. B., Cleveland, Wi. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, *6*(1), 3–73.

CMEMS. (2018). *North-West Shelf Biogeochemical Reanalysis Product NORTHWESTSHELF REANALYSIS BIO _004_011*.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 157–175). © Springer Science+Business Media, LLC 2012. https://doi.org/10.1007/978-1-4419-9326-7

CWSS. (2017). *"Introduction. In: Wadden Sea Quality Status Report." Wadden Sea Ecosystem* (Vol. 9). Common Wadden Sea Secretariat, Wilhelmshaven, Germany. Retrieved from http://cwss.www.de/TMAP/Qsr99/Qsr99.html

Deegan, L. A., Johnson, D. S., Warren, R. S., Peterson, B. J., Fleeger, J. W., Fagherazzi, S., & Wollheim, W. M. (2012). Coastal eutrophication as a driver of salt marsh loss. *Nature*, *490*(7420), 388–392. https://doi.org/10.1038/nature11533

Diaz, R. J., & Rosenberg, R. (2008). Spreading Dead Zones and Consequences for Marine Ecosystems. *Science*, *321*(5891), 926–929. https://doi.org/10.1126/science.1156401

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In J. Kittler & F. Roli (Eds.), *Multiple Classifier Systems* (pp. 1–15). Calgiari, Italy: Springer.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M., Féménias, P., Frerick, J., … Sciarra, R. (2012). Remote Sensing of Environment The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment*, *120*, 37–57. https://doi.org/10.1016/j.rse.2011.07.024

ESA. (2006). MERIS Product Handbook, (2).

ESA. (2017). Sentinel-3 User Handbook, (January), 1–157. https://doi.org/GMES-S1OP-EOPG-TN-13-0001

EU. (2000). DIRECTIVE 2000/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2000 establishing a framework for Community action in the field of water policy (Water Framework Directive). *Official Journal of the European Communities*, (327).

EU. (2008). Directive 2008/56/EC of the European Parliament and of the Council (Marine Strategy Framework Directive). *Official Journal of the European Union*, *164*, 19–40. https://doi.org/10.1016/j.biocon.2008.10.006

European Commission. (2010). Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (notified under document C(2010) 5956)Text with EEA relevance, 11.

European Commission. (2011). *Seas for life*. European Union. https://doi.org/10.2779/18719

European Commission. (2014). *HORIZON 2020 in brief. The EU Framework Programme for Research & Innovation*. *Directorate-General for Research and Innovation*. https://doi.org/10.2777/3719

Ferreira, J. G., Andersen, J. H., Borja, A., Bricker, S. B., Camp, J., Cardoso da Silva, M., … Claussen, U. (2010). *MSFD Task Group 5 Report Eutrophication*. *JRC Scientific and Technical Reports*. https://doi.org/10.2788/86830

Ferreira, J. G., Andersen, J. H., Borja, A., Bricker, S. B., Camp, J., Cardoso da Silva, M., … Claussen, U. (2011). Overview of eutrophication indicators to assess environmental status within the European Marine Strategy Framework Directive. *Estuarine, Coastal and Shelf Science*, *93*(2), 117–131. https://doi.org/10.1016/j.ecss.2011.03.014

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting *, *139*, 119–139.

Funfak, A., Fischlechner, M., Donath, E., & Köhler, J. M. (2014). Fluorescence-labelled polymer beads for chemical monitoring of nanoliter fluid segments. *Topical Conference on Applications of Micro-Reactor Engineering 2007, Held at the 2007 AIChE Spring National Meeting*, 153–165. https://doi.org/10.1007/978-1-4020-4410-6_217

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Mishonov, A. V., Baranova, O. K., … Johnson, D. R. (2013). *WORLD OCEAN ATLAS 2013 Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation* (Vol. 3). Maryland, USA. Retrieved from https://www.nodc.noaa.gov/OC5/indprod.html

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn* (1st ed.). O'Reilly Media, Inc.

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, *16*. https://doi.org/10.3390/s16081298

Gobler, C. J., Depasquale, E. L., Griffith, A. W., & Baumann, H. (2014). Hypoxia and Acidification Have Additive and Synergistic Negative Effects on the Growth , Survival , and Metamorphosis of Early Life Stage Bivalves. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0083648

Gohin, F., Loyer, S., Lunven, M., Labry, C., Froidefond, J., Delmas, D., … Herbland, A. (2005). Satellite-derived parameters for biological modelling in coastal waters : Illustration over the eastern continental shelf of the Bay of Biscay, *95*, 29–46. https://doi.org/10.1016/j.rse.2004.11.007

Gordon, H. R., Boynton, G. C., Balch, W. M., Groom, S. B., Harbour, D. S., & Smyth, T. J. (2001). Retrieval of coccolithophore calcite concentration from sea WiFS imagery. *Geophysical Research Letters*, *28*(8), 1587–1590. https://doi.org/10.1029/2000GL012025

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. *Math. Intell.* (Vol. 27). https://doi.org/10.1007/b94608

He, B., Oki, K., Wang, Y., & Oki, T. (2009). Using remotely sensed imagery to estimate potential annual pollutant loads in river basins. *Water Science & Technology*, *60*(8). https://doi.org/10.2166/wst.2009.596

Hess, D., Duvvur, V., & Srinivasan, K. (2019). Prediction of Algal Blooms in the Great Lakes through a Convolution Neural Network of Remote Sensing Data. *Preprint*. https://doi.org/10.1101/450551

Howarth, R., Chan, F., Conley, D. J., Garnier, J., Doney, S. C., Marino, R., & Billen, G. (2011). Coupled biogeochemical cycles: Eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. *Frontiers in Ecology and the Environment*, *9*(1), 18–26. https://doi.org/10.1890/100008

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans*, *117*(1), 1–25. https://doi.org/10.1029/2011JC007395

Ibrahim, A. N., Mabuchi, Y., & Murakami, M. (2009). Remote sensing algorithms for monitoring eutrophication in Ishizuchi storm water reservoir in Kochi Prefecture, Japan, *6667*. https://doi.org/10.1623/hysj.50.3.525.65024

ICES. (2014). *Dataset on Ocean Hydrography*. Copenhagen. Retrieved from http://ocean.ices.dk/HydChem/

IOCCG. (2000). *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters*. (S. Sathyendranath, Ed.) (Reports of). Dartmouth, Canada.

Karydis, M., & Kitsiou, D. (2013). Marine water quality monitoring: A review. *Marine Pollution Bulletin*. Elsevier Ltd. https://doi.org/10.1016/j.marpolbul.2013.09.012

Kay, C. S., Mcewan, R., Ford, D., & Mcewan, R. (2019). *North West European Shelf Production Centre - NORTHWESTSHELF_REANALYSIS_BIO_004_011*.

Kim, H. C., Son, S., Kim, Y. H., Khim, J. S., Nam, J., Chang, W. K., … Ryu, J. (2017). Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Marine Pollution Bulletin*, *121*(1–2), 425–434. https://doi.org/10.1016/j.marpolbul.2017.05.026

Ko, D. S., Gould, R. W., Penta, B., & Lehrter, J. C. (2016). Impact of satellite remote sensing data on simulations of coastal circulation and hypoxia on the Louisiana continental shelf. *Remote Sensing*, *8*(5). https://doi.org/10.3390/rs8050435

Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (2nd ed.). Springer New York Dordrecht Heidelberg London. https://doi.org/10.1007/978-0-387-09823-4

Mobley, C. D. (1994). *Light and water; Radiative transfer in natural waters*. London: AcademicPress.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning*. Cambridge, Massachusetts: MIT press.

Muchlisin, A. (2015). Developement of dissolved oxygen concetration extraction model using Landsat data. Case study: Ringgung Coastal Waters, *12*(1), 1–12.

NASA. (n.d.-a). MODIS NASA website. Retrieved November 15, 2018, from https://modis.gsfc.nasa.gov/

NASA. (n.d.-b). NASA OcealColor Web. Retrieved from https://oceancolor.gsfc.nasa.gov/

NASA. (2006). *Earth Science Reference Handbook*. Washington D.C.

NASA. (2010). Ocean Level-3 mapped image Products, 1–7.

OBPG-NASA. (n.d.-a). MODIS Level 2 Ocean Color Flags. Retrieved March 18, 2019, from https://oceancolor.gsfc.nasa.gov/atbd/ocl2flags/

OBPG-NASA. (n.d.-b). SST Quality Flags. Retrieved March 29, 2019, from https://oceancolor.gsfc.nasa.gov/atbd/sst/flag/

OBPG-NASA. (2014). Sea-viewing Wide Field-of-view Sensor (MODIS) Ocean Color Data, NASA

OB.DAAC, Greenbelt, MD, USA. https://doi.org/10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0

OBPG-NASA. (2018). Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Ocean Color Data. Retrieved March 5, 2019, from https://oceancolor.gsfc.nasa.gov/data/10.5067/AQUA/MODIS/L2/OC/2018/

OSPAR. (2009). *Eutrophication Status of the OSPAR Maritime Area Second OSPAR Integrated Report Eutrophication Series*.

OSPAR. (2017). *Eutrophication Status of the OSPAR Maritime Area Third Integrated Report on the Eutrophication Status of the OSPAR Maritime Area Eutrophication Series*.

Patterson, H. K., Boettcher, A., & Carmichael, R. H. (2014). Biomarkers of Dissolved Oxygen Stress in Oysters : A Tool for Restoration and Management Efforts. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0104440

Pedregosa, F., Weiss, R., & Brucher, M. (2011). Scikit-learn : Machine Learning in Python, *12*, 2825–2830.

Polikar, R. (2012). Ensemble Learning. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 1–34). © Springer Science+Business Media, LLC 2012. https://doi.org/10.1007/978-1-4419-9326-7

Reilly, J. E. O., Brien, M. C. O., Siegel, D. A., Toole, D., Menzies, D., Smith, R. C., … Aiken, J. (2000). *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3* (Vol. 11). Retrieved from http://oceancolor.gsfc.nasa.gov/REPROCESSING/R2009/ocv6/

Rijkswaterstaat. (2016). *River basin management plans 2016-2021 of the Netherlands*. The Hague.

Sanderson, E. W., Jaiteh, M., Levy, M. A., Redford, K. H., Wannebo, A. V, & Woolmer, G. (2002). The Human Footprint and the Last of the Wild. *BioScience*, *52*(10), 891–904. Retrieved from https://academic.oup.com/bioscience/article/52/10/891/354831

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *ACM National Conference* (p. 8). New York, USA.

Smith, V. H., Tilman, G. D., & Nekola, J. C. (1998). Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, *100*(1–3), 179–196. https://doi.org/10.1016/S0269-7491(99)00091-3

Solimini, D. (2016). *Understanding Earth Observation - The Electromagnetic Foundation of Remote Sensing*. (F. D. van der Meer, Ed.). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-25633-7

Somvanshi, S., Kunwar, P., Singh, N. B., Shukla, S. P., & Pathak, V. (2012). Integrated remote sensing and GIS approach for water quality analysis of Integrated remote sensing and GIS approach for water quality analysis of Gomti river , Uttar Pradesh. *International Journal of Environmental Sciences*, *3*(1), 707–713. https://doi.org/10.6088/ijes.2012030131008

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern Recognition*. Elsevier Inc.

UNEP. (2007). Eutrophication Monitoring Strategy for the MED POL (REVISION). *English*, (December), 12–14.

UNESCO-Bilko. (2009). Meris Level 2 Flags, 9–11. Retrieved from http://www.bilko.org/, ftp://ftp.noc.soton.ac.uk/devcocast/training_materials/References/MERIS_level2flags.pdf

United Nations. (1982). Convention on the Law of the Sea, 202. Retrieved from https://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf

United Nations. (2017). Transforming Our World: The 2030 Agenda for Sustainable Development. In *A New Era in Global Health*. New York, NY: Springer Publishing Company.

van Aken, H. M. (2010). Meteorological forcing of long-term temperature variations of the Dutch coastal waters. *Journal of Sea Research*, *63*(2), 143–151. https://doi.org/10.1016/j.seares.2009.11.005

van Beusekom, J. E. E. (2005). A historic perspective on Wadden Sea eutrophication. *Helgoland Marine Research*, *59*(1), 45–54. https://doi.org/10.1007/s10152-004-0206-2

van Beusekom, J. E. E., Bot, P., Carstensen, J., Grage, A., Kolbe, K., Lenhart, H.-J., … Rick, J. (2017). *Eutrophication. In: Wadden Sea Quality Status Report 2017*. Wilhelmshaven, Germany. Retrieved from qsr.waddensea-worldheritage.org/reports/eutrophication

van der Veer, H. W., Raaphorst, W. van, & Bergman, M. J. N. (1989). Eutrophication of the Dutch Wadden Sea: external nutrient loadings of the Marsdiep and Vliestroom basin, *515*(18), 501–515. https://doi.org/10.1152/jn.00698.2010

Walton, C. C., Pichel, W. G., Sapper, J. F., & 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*, *103*(NO. C12), 27999–28012.

WHC. (2009). *Convention concerning the protection of the World Cultural and natural Heritage*. Quebec, Canada.

WMO. (2013). *Planning of water-quality monitoring systems. Technical report series No.3*.

Woerd, H. J. Van Der, & Pasterkamp, R. (2007). HYDROPT : A fast and flexible method to retrieve chlorophyll-a from multispectral satellite observations of optically complex coastal waters, (April 2008). https://doi.org/10.1016/j.rse.2007.09.001

Wolanski, E., Newton, A., Rabalais, N., & Legrand, C. (2013). Coastal Zone Management. *Reference Module in Earth Systems and Environmental Sciences*, (March), 1–7. https://doi.org/10.1016/B978-0-12-409548-9.00537-6

Yang, X., Wu, X., Hao, H., & He, Z. (2008). Mechanisms and assessment of water eutrophication. *Journal of Zhejiang University SCIENCE B*, *9*(3), 197–209. https://doi.org/10.1631/jzus.B0710626

Appendices

In this chapter, various auxiliary information are presented, including additional sensor details for MODIS and MERIS, a Sentinel-3 mission description, some useful Python libraries and a more detailed description of the Wadden Sea that was used as a case study in chapter 8.3. Finally, the results of all the stations of the study area as well as a comparison of the three model outputs for one station are shown.

Appendix: MODIS and MERIS sensor auxiliary information

Table 13 MODIS spectral bands (NASA, n.d.-a)

Band	Bandwidth [nm]	Usage
1	620 – 670	Land/Cloud/Aerosol – Boundaries
2	841 – 876	
3	459 – 479	Land/Cloud/Aerosol – Properties
4	545 – 565	
5	1230 – 1250	
6	1628 – 1652	
7	2105 – 2155	
8	405 – 420	Ocean Colour/Phytoplankton/Biochemistry
9	438 – 448	
10	483 – 493	
11	526 –536	
12	546 –556	
13	662 –672	
14	673 –683	
15	743 –753	
16	862 –877	
17	890 –920	Atmospheric Water Vapour
18	931 – 941	
19	915 – 965	
20	3660 –3840	Surface/Cloud Temperature
21	3929 – 3.989	
22	3929 – 3989	

23	4020 – 4080	
24	4433 – 4498	Atmospheric Temperature
25	4482 – 4549	
26	1360 – 1390	Cirrus Clouds Water Vapour
27	6535 – 6895	
28	7175 – 7475	
29	8400 – 8700	Cloud Properties
30	9580 – 9880	Ozone
31	10780 – 11280	Surface/Cloud Temperature
32	11770 – 12270	
33	13185 – 13485	Cloud Top Altitude
34	13485 – 13785	
35	13785 – 14085	
36	14085 – 14385	

Table 14 MODIS Ocean Colour Level 2 flag description (OBPG-NASA, n.d.-a)

Bit	Name	Description	L2 Default
00	ATMFAIL	Atmospheric correction failure	OFF
01	LAND	Pixel is over land	ON
02	PRODWARN	One or more product algorithms generated a warning	OFF
03	HIGLINT	Sunglint: reflectance exceeds threshold	OFF
04	HILT	Observed radiance very high or saturated	ON
05	HISATZEN	Sensor view zenith angle exceeds threshold	OFF
06	COASTZ	Pixel is in shallow water	OFF
07	spare		
08	STRAYLIGHT	Probable stray light contamination	ON
09	CLDICE	Probable cloud or ice contamination	ON
10	COCCOLITH	Coccolithophores detected	OFF

11	TURBIDW	Turbid water detected	OFF
12	HISOLZEN	Solar zenith exceeds threshold	OFF
13	spare		
14	LOWLW	Very low water-leaving radiance	OFF
15	CHLFAIL	Chlorophyll algorithm failure	OFF
16	NAVWARN	Navigation quality is suspect	OFF
17	ABSAER	Absorbing Aerosols determined	OFF
18	spare		
19	MAXAERITER	Maximum iterations reached for NIR iteration	OFF
20	MODGLINT	Moderate sun glint contamination	OFF
21	CHLWARN	Chlorophyll out-of-bounds	OFF
22	ATMWARN	Atmospheric correction is suspect	OFF
23	spare		
24	SEAICE	Probable sea ice contamination	OFF
25	NAVFAIL	Navigation failure	OFF
26	FILTER	Pixel rejected by user-defined filter OR Insufficient data for smoothing filter	OFF
27	spare		OFF
28	BOWTIEDEL	Deleted off-nadir, overlapping pixels (VIIRS only)	OFF
29	HIPOL	High degree of polarization determined	OFF
30	PRODFAIL	Failure in any product	OFF
31	spare		

Table 15 MODIS Sea Surface Temperature flag description (OBPG-NASA, n.d.-b)

Bit	Name	Description
00	ISMASKED	Pixel was already masked.
01	BTBAD	Brightness temperatures are bad outside radiance to brightness table conversion or saturated
02	BTRANGE	Brightness temperatures are out-of-range for top of the atmosphere

		realistic ocean surface values, -4 to 37 °C
03	BTDIFF	Brightness temperatures spectral differences between channels are outside of expected valid ranges, 0 – 1.6 °C
04	SSTRANGE	SST outside valid range -1.8 to 45 °C
05	SSTREFDIFF	Retrieved SST is too different from Reynolds reference field. Threshold is > -3 °C non dust regions and a more stringent > -1.25 °C in known dust region defined as 10S to 30N latitude and 105W and 105E longitude.
06	SST_triple_DIFF	Longwave NSST is different from SST_triple
07	SST_triple_VDIFF	Longwave NSST is very different from shortwave SST_triple
08	BTNONUNIF	Brightness temperatures are spatially non-uniform > 0.7 °C
09	BTVNONUNIF	Brightness temperatures are very spatially non-uniform ° 1.2 °C
10	spare	spare
11	REDNONUNIF	Red-band reflectance spatial non-uniformity or saturation > 0.01. Test not applied in sun glint region.
12	HISENZ	Sensor zenith angle high > 55 degrees
13	VHISENZ	Sensor zenith angle very high > 65 degrees
14	SSTREFVDIFF	SST is too different from reference > 5 °C
15	SST_CLOUD	Pixel failed the cloud Alternating decision tree

Table 16 MERIS Level 2 flag description (ESA, 2006; UNESCO-Bilko, 2009).

Flag	Description
LAND	Pixel classified as Land in L1B, adjusted radiometrically during L2 pixel classification to allow for geocorrection errors and tidal changes.
CLOUD	Pixel classified as cloud by the L2 cloud screening algorithm. Sub-pixel, scattered cloud not included.
WATER	Pixel classified as Ocean in L1B, adjusted radiometrically during L2 pixel classification to allow for geocorrection errors and tidal changes in land/water boundary. Includes inland water.
PCD_1_13	Confidence flag for MDS 1 to 13 (reflectances). Raised at low sun angles, when atmospheric correction fails or there are difficulties with aerosol correction. Also for pixels with whitecaps or uncorrected glint, when reflectances in any band are negative, or when reflectance at 510nm exceeds a threshold without the CASE2_S flag having been raised.
PCD_14	Confidence flag for MDS 14 (water_vapour).

PCD_15	Confidence flag for MDS 15 (algal_1). Raised when atmospheric correction fails or there are difficulties with aerosol correction. Also for pixels with uncorrected glint or whitecaps, and for pixels with high turbidity.
PCD_16	Confidence flag for MDS 16 (yellow_subs / total_susp). Raised when PCD_13 is raised, or when the algorithm input or output is outside the expected range.
PCD_17	Confidence flag for MDS 17 (algal_2) Raised when PCD_13 is raised, or when the algorithm input or output is outside the expected range.
PCD_18	Confidence flag for MDS 18 (photosyn_rad) Raised when atmospheric correction fails or there are difficulties with aerosol correction. Also for pixels with uncorrected glint, whitecaps, and high yellow substance
PCD_19	Confidence flag for MDS 19 (aero_epsilon, aero_opt_thick). Raised when atmospheric correction fails, when there is glint, whitecaps, or high yellow substance. Also when the retrieved aerosol model does not match the aerosol climatology.
COASTLINE	Coastline flag.
COSMETIC	Cosmetic flag (from level-1b): Missing data filled in by interpolation.
SUSPECT	Suspect flag (from level-1b): Transmission errors means measurements may be unreliable.
ABSOA_CONT	Continental absorbing aerosol.
ABSOA_DUST	Dust-like absorbing aerosol.
CASE2_S	Turbid water, nominally TSM > 0.3gm-3. Indicative of sediment laden coastal water or coccolithophore blooms, sometimes other algal blooms. Raised after correction for atmospheric Rayleigh scattering and checks for glint and whitecaps, when near infrared (NIR) radiances are high.
CASE2_ANOM	Bright water pixel, anomalous scattering water. Flags the presence of Case 2 water, based on.
TOAVI_BRIGHT	Bright land pixel, suspect vegetation index.
CASE2_Y	Yellow substance loaded water.
TOAVI_BAD	Top of atmosphere vegetation index (TOAVI) is bad.
ICE_HAZE	Pixel with high radiance in the blue, likely to be caused by ice or high aerosol load.
TOAVI_CSI	
MEDIUM_GLINT	Pixel corrected for medium level glint. Sun-glint reflectance calculated from viewing geometry and wind speed exceeds the low glint threshold, but is low enough for glint correction to take place.
TOAVI_WS	

DDV	Dark, dense vegetation
HIGH_GLINT	Pixel with high sun-glint, which has NOT been corrected. Sun-glint reflectance calculated from viewing geometry and wind speed exceeds the medium glint threshold, and is too high for glint correction.
TOAVI_INVAL_REC	
P_CONFIDENCE	The two atmospheric pressure estimates (from the data and from ECMWF) do not compare successfully.
LOW_PRESSURE	Pressure computed from data lower than ECMWF pressure.

Appendix: Sentinel 3 description

ESA's Sentinel 3 is a satellite mission that primarily aims to support the Global Monitoring for Environment and Security (GMES) program and to continue its predecessors' (ERS, ENVISAT-MERIS, SPOT) time series data. The mission's scope is to accurately measure the Sea Surface Topography, the Sea and Land Surface Temperature and the Ocean and Land surface Colour and aid the ocean forecasting, environmental and climate monitoring scientific communities. It consists of two different satellites, Sentinel 3A and 3B that offer a temporal resolution of 1 to 2 days depending on the location. On board the satellites there are four instruments, namely the Ocean and Land Colour Instrument (OLCI), the Sea and Land Surface Temperature Radiometer (SLSTR), the Altimetry Instrument and the MicroWave Radiometer (MWR) (ESA, 2017). For this project, only the first two instruments are considered as relevant.

The services and data of Sentinel 3 are provided by a cooperation of ESA and EUMETSAT[17] aiming to contribute to the work already being done in the fields of "Numerical Ocean Prediction", "Maritime Safety and Security", "Coastal Zone Monitoring", "Open Ocean and Ice Monitoring", "Atmospheric Services", "Global Land Monitoring Applications", "Environmental Policy and Law", "Climate Change Monitoring" and "Support of European Security, Humanitarian and Emergency Services" (ESA, 2017).

Ocean and Land Colour Instrument (OLCI)

The OLCI is an along-track optical sensor (push-broom) that measures the Earth's reflectance of the sunlight in 21 spectral channels with 300m spatial resolution. These spectral channels or bands are situated in the part of the electromagnetic spectrum that spans from 400nm to 1020nm (Table 17). It is a more developed version of its predecessor, MERIS, and it has six additional bands (ESA, 2017).

The provided products are divided into two processing levels, namely 1B and 2, while there is an additional division for the latter in land and water products. Level 1 products offer per-pixel reflectance values in the grid of the sensor for each spectral band. Meanwhile, Level 2 products offer land (or water) and atmospheric parameter estimates in the original (300m) but also in reduced resolution (1.2km) (ESA, 2017). For this research, Level 2 data for water are going to be considered and the corresponding estimated parameters can be seen in Table 18.

[17] The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is a European organization that offers weather and climate data.

Table 17 Sentinel 3 OLCI Spectral bands (ESA, 2017). The highlighted cells correspond to the additional bands compared to MERIS.

Band	Band Centre [nm]	Bandwidth [nm]	Usage
Oa1	400	15	Aerosol correction, improved water constituent retrieval
Oa2	412.5	10	Yellow substance and detrital pigments (turbidity)
Oa3	442.5	10	Chla absorption max., biogeochemistry, vegetation
Oa4	490	10	High Chla, other pigments
Oa5	510	10	Chla, sediment, turbidity, red tide
Oa6	560	10	Chlorophyll reference (Chla minimum)
Oa7	620	10	Sediment loading
Oa8	665	10	Chla, sediment, yellow substance/vegetation
Oa9	673.75	7.5	For improved fluorescence retrieval and to better account for smile together with the bands 665 and 680 nm
Oa10	681.25	7.5	Chl fluorescence peak, red edge
Oa11	708.75	10	Chl fluorescence baseline, red edge transition
Oa12	753.75	7.5	O2 absorption/clouds, vegetation
Oa13	761.25	2.5	O2 absorption band/aerosol corr.
Oa14	764.375	3.75	Atmospheric correction
Oa15	767.5	2.5	O2A used for cloud top pressure, fluorescence over land
Oa16	778.75	15	Atmospheric correction/aerosol correction
Oa17	865	20	Atmos. corr./aerosol corr., clouds, pixel co-registration
Oa18	885	10	Water vapour absorption reference band. Common reference band with SLSTR instrument. Vegetation monitoring
Oa19	900	10	Water vapour absorption/vegetation monitoring (max. reflectance)
Oa20	940	20	Water vapour absorption, atmos./aerosol corr.
Oa21	1 020	40	Atmospheric correction/aerosol correction

Table 18 Sentinel 3 OLCI Level 2 water and atmospheric parameter products (ESA, 2017).

Product	Description	Units	Used bands
Rxxx	Surface directional reflectance	-	Oa1-Oa12, Oa16-Oa18, Oa21

chl_oc4me and chl_NN	Chla concentration, computed using "OC4Me" and Neural Network algorithms	mg/m^3	- Oa3 and Oa6 - Oa1-Oa12, Oa16, Oa17 and Oa21
TSM_NN	Total suspended matter concentration	g/m^3	Oa1-Oa12, Oa16, Oa17 and Oa21
KD490_M07	Diffuse attenuation coefficient for down-welling irradiance, at 490 nm	m^{-1}	Oa4 and Oa6
ADG_443_NN	Absorption of coloured detrital and dissolved material at 443 nm.	m^{-1}	Oa1, Oa12, Oa16, Oa17, Oa21
PAR	Quantum energy flux from the sun in the spectral range 400-700 nm.	$\mu E/m^2 s$	-
T865 and A865	Aerosol load, expressed in optical depth at a given wavelength (865 nm) and spectral dependency of the aerosol optical depth, between 779 and 865 nm.	-	Oa5, Oa16 and Oa17
IWV	Integrated Water Vapour column	kg/m^2	Oa18, Oa19

Product Grid

Level 0 and Level 1 product pixels are distributed in a varying resolution grid since sensors are measuring the curved surface of the Earth and not a 2D surface. Level 1B products are resampled from Level 1 products into an evenly spaced grid (300m). In the Reduced Resolution product (RR), groups of 4-by-4 pixels (16 in total) of the Full Resolution (FR) dataset are averaged into a bigger pixel (1.2km) (ESA, 2017).

Sea and Land Surface Temperature Radiometer (SLSTR)

The SLSTR instrument uses three infrared bands, centred at 3740nm, 10850nm and 12000nm respectively, to estimate the Sea Surface Temperature with an accuracy of 0.3 K. It does so by correcting the initial observation with the water vapour atmospheric absorption estimates, as well as by observing the same pixel from two different orbital positions. This way it combines two observations of the same location where the reflected solar radiation follows a different atmospheric path, thus correcting for aerosol effects (ESA, 2017). Additionally, SLSTR provides products from another eight Visual (VIS), Short-Wave Infrared (SWIR) and Thermal Infrared (TIR) bands that can be used to monitor the Land Surface Temperature (LST), fire location and fire radiative power and the vegetation index, but these are irrelevant to this study and therefore will not be discussed further (Table 19). Finally, the products are available in 300m and 1km spatial resolutions (ESA, 2017).

Because of the limited ability of the infrared radiation to get past the “skin” of the water, the SST estimates correspond only to the top layer of the ocean, i.e. the first few tens of μm. This is important to be noted as oceanographers normally refer to the SST as the temperature of the

ocean at about 10cm of depth, which can differ from the aforementioned one by a few tenths of a Kelvin degree (ESA, 2017).

Moreover, the SST products are produced in two different ways. The first one computes the SST as a weighted combination of the thermal bands with the brightness observed at nadir and at an oblique view. The second one (L2P) computes the SST according to the specifications defined by the Group for High Resolution Sea Surface Temperature (GHRSST) that smooths the result of the first method using the ATSR model (ESA, 2017).

Table 19 Sentinel 3 SLSTR bands (ESA, 2017).

Band	Band Centre [nm]	Bandwidth [nm]	Usage	Type of sensor	Res. [m]
S1	555	20	Cloud screening, vegetation monitoring, aerosol		
S2	659	20	NDVI, veg. monitoring, aerosol	VIS/NIR	
S3	865	20	NDVI, cloud flagging, pixel co-registration		500
S4	1375	15	Cirrus detection over land		
S5	1610	60	Cloud cleaning, ice, snow, veg. monitoring	SWIR	
S6	2250	50	Vegetation state and cloud clearing		
S7	3740	380	SST, LST, Active fire		
S8	10850	900	SST, LST, Active fire		
S9	12000	1000	SST, LST	TIR	1000
F1	3740	380	Active fire		
F2	10850	900	Active fire		

Appendix: Useful Python libraries

Table 20 Python libraries used for data acquisition and pre-processing

Libraries	Usage
os	File explorer and managing operations
bs4 and urllib	File retrieval from open ftp (NASA's OBPG) and https sites (Deltares Repository)
ftplib	File retrieval from ftp servers that required logging in, e.g. to download AATSR data
netCDF4	Opening, processing and creating netCDF files

pyproj	Coordinate transformation
Scikit-learn	ML models
TensorFlow	DL models

Appendix: Dutch Wadden Sea area description

The Dutch Wadden Sea (Figure 71) is an important intertidal ecosystem whose biodiversity and quality should be preserved. It has very specific characteristics and it has been characterized as an UNESCO world heritage site in 2009 (CWSS, 2017; WHC, 2009).

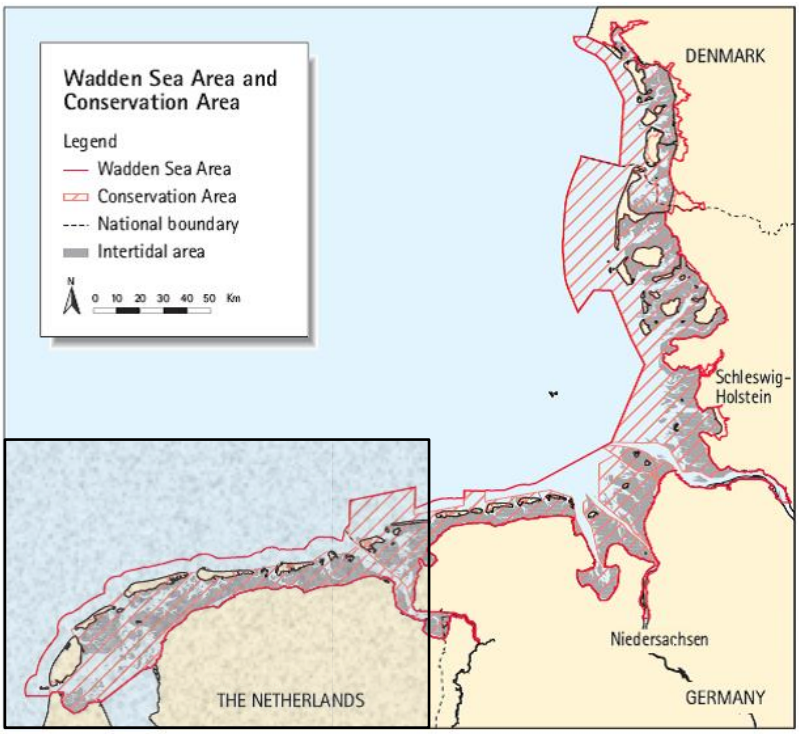


Figure 71 The Wadden Sea. The Dutch portion is identified by the black box seen on the lower left corner. Source: (CWSS, 2017).

Area description

The Wadden Sea is a geologically young shallow sea, as it has been formed during the last 8000 years. It is 500 km long covering various portions of the coast of three countries, namely The Netherlands, Germany and Denmark, and it separates from the North Sea though a sandy barrier island system (CWSS, 2017). Tides play a major role in the way it functions, as the differences in water level can stretch from 1.2m to 3.5m in some areas in a matter of hours. Meanwhile, about half of its area is an intertidal zone with mainly sandy tidal flats, meaning they are not covered with water when the tide is low (van Beusekom, 2005). Its barrier islands accommodate large salt marshes and along with its tidal flats, which comprise 60% of all European and North African tidal areas, create a very diverse and unique ecosystem of global importance (CWSS, 2017).

The Rhine River's influence on the Dutch part of the Wadden Sea is significant, as one of its branches on Dutch territory, the Ijssel, ends-up in the Ijssel Lake which is the main supplier of fresh water for the Wadden Sea. At the same time, the rest of Rhine's branches flow through the Dutch coast into the North Sea and their influence on the coastal waters propagates to the Wadden Sea through the tidal inlet (van Beusekom, 2005; van der Veer, Raaphorst, & Bergman, 1989). Similarly, the Ems River delivers freshwater to the Dutch Wadden Sea on its border with Germany, while Meuse and Scheldt along with the Rhine supply the Dutch coastal waters with riverine water through the Zeeland Delta (van Beusekom *et al.*, 2017).

Ecology of the Wadden Sea

Beyond its complex geomorphology, the Wadden Sea also contains numerous and diverse habitats with rich varieties of flora and fauna. More specifically, there are approximately 10,000 species living in those habitats, ranging from single-cell organisms to animals, while half of them are distributed between the salt marshes and the aquatic areas (marine or brackish waters) (CWSS, 2017).

Different species thrive in different habitats and different seasons allowing space for the accommodation of many migratory species. In particular, 41 migratory bird species that originate from Canada to Siberia pass through the Wadden Sea to pass the winter or moult. It is evident that the ecological status of the Wadden Sea is important in a transcontinental sense, and this is why the area is also protected under the RAMSAR Convention[18] (CWSS, 2017).

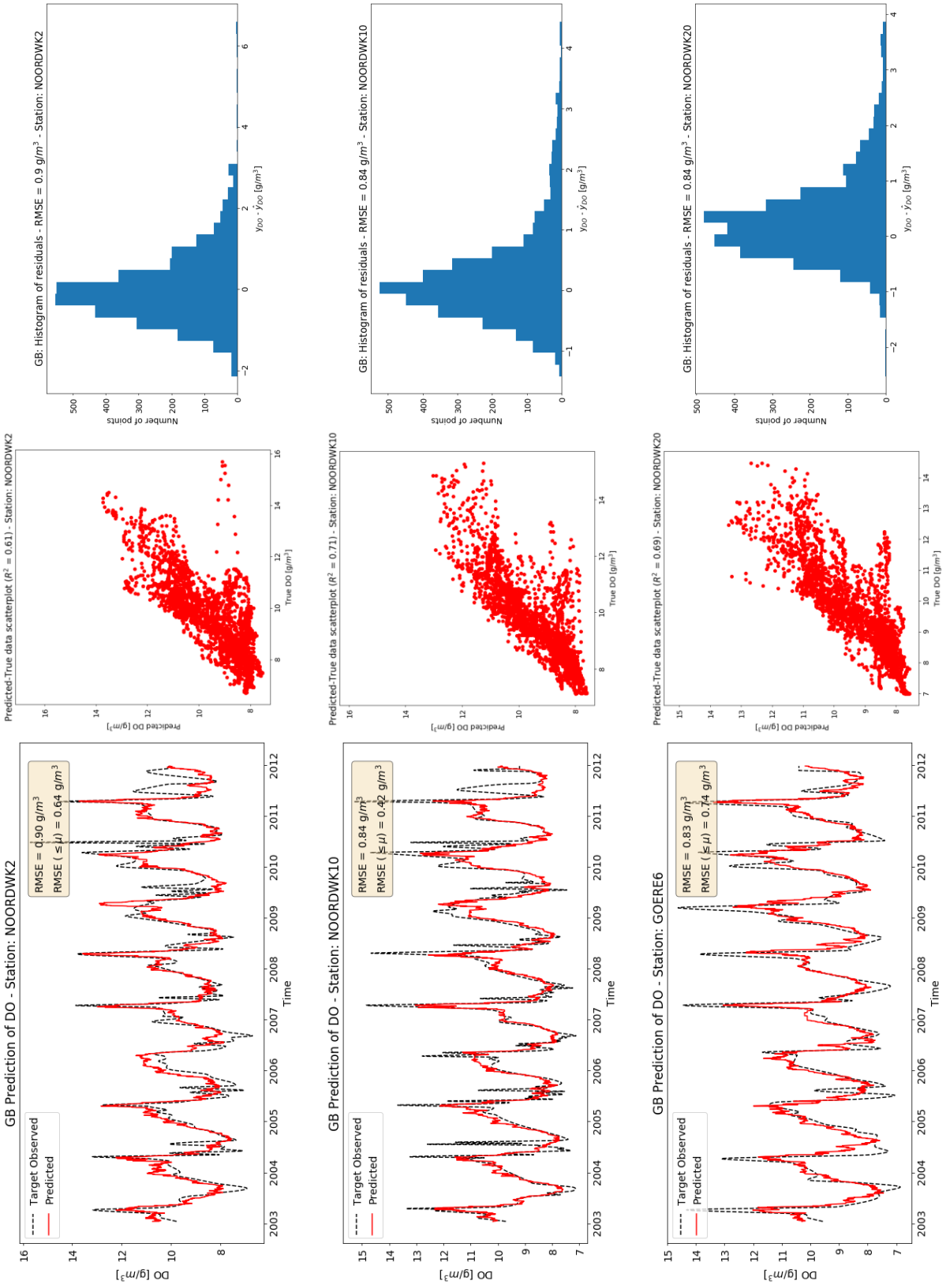
Eutrophication of the Wadden Sea

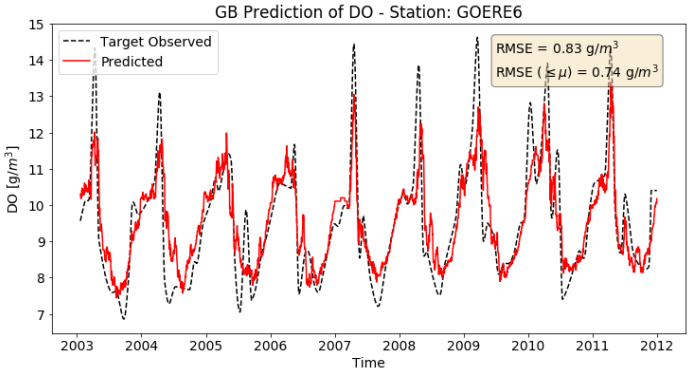
Since the 1950s, rivers have been delivering great amounts of nutrients into the North Sea, affecting the primary production in the Wadden Sea as well. In fact, phytoplankton biomass and benthic vegetation has increased by a factor of 2 to 3 during this period reaching a peak in 1994. Fortunately, certain EU regulations have been set though the WFS and MSFD limiting the anthropogenic contribution of nutrients in the riverine waters, resulting to a negative trend of eutrophication indicators. Nevertheless, as of 2017 it has been stated that the OSPAR goal of transforming the Wadden Sea into an "eutrophication non-problem area" has yet to be achieved (van Beusekom *et al.*, 2017; van Beusekom, 2005).

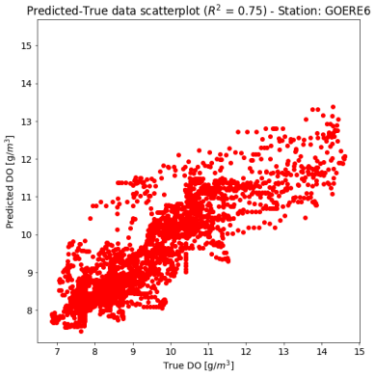
[18] Convention on Wetlands of International Importance especially as Waterfowl Habitat. Ramsar (Iran), 2 February 1971. UN Treaty Series No. 14583. As amended by the Paris Protocol, 3 December 1982, and Regina Amendments, 28 May 1987

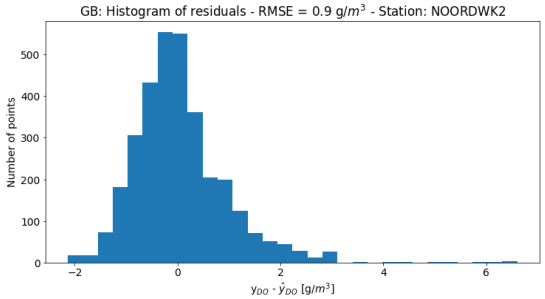
Appendix: Additional results

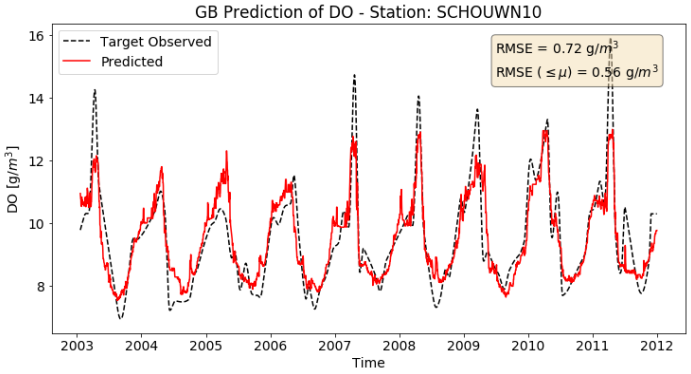
Below some additional results are shown for all the stations of the study area.

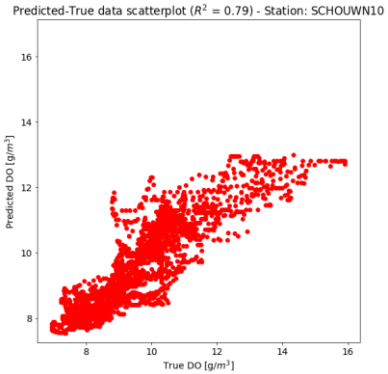


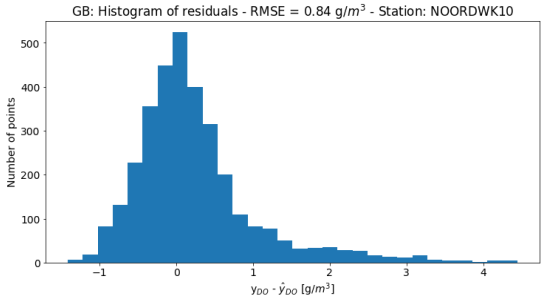

GB Prediction of DO - Station: GOERE6
Target Observed
Predicted
RMSE = 0.83 g/m^3
RMSE ($\leq \mu$) = 0.74 g/m^3
DO [g/m^3]
Time

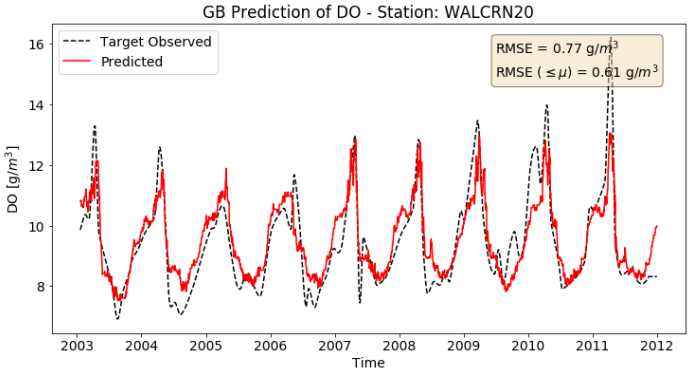

Predicted-True data scatterplot (R^2 = 0.75) - Station: GOERE6
Predicted DO [g/m^3]
True DO [g/m^3]

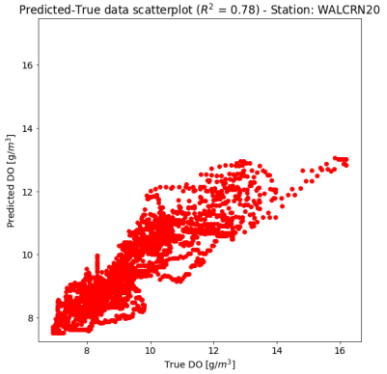

GB: Histogram of residuals - RMSE = 0.9 g/m^3 - Station: NOORDWK2
Number of points
y_{DO} - $\hat{y}_{DO}$ [g/m^3]

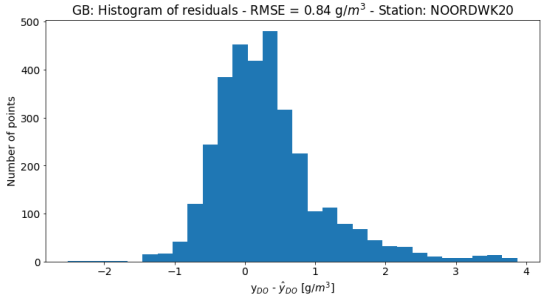

GB Prediction of DO - Station: SCHOUWN10
Target Observed
Predicted
RMSE = 0.72 g/m^3
RMSE ($\leq \mu$) = 0.56 g/m^3
DO [g/m^3]
Time

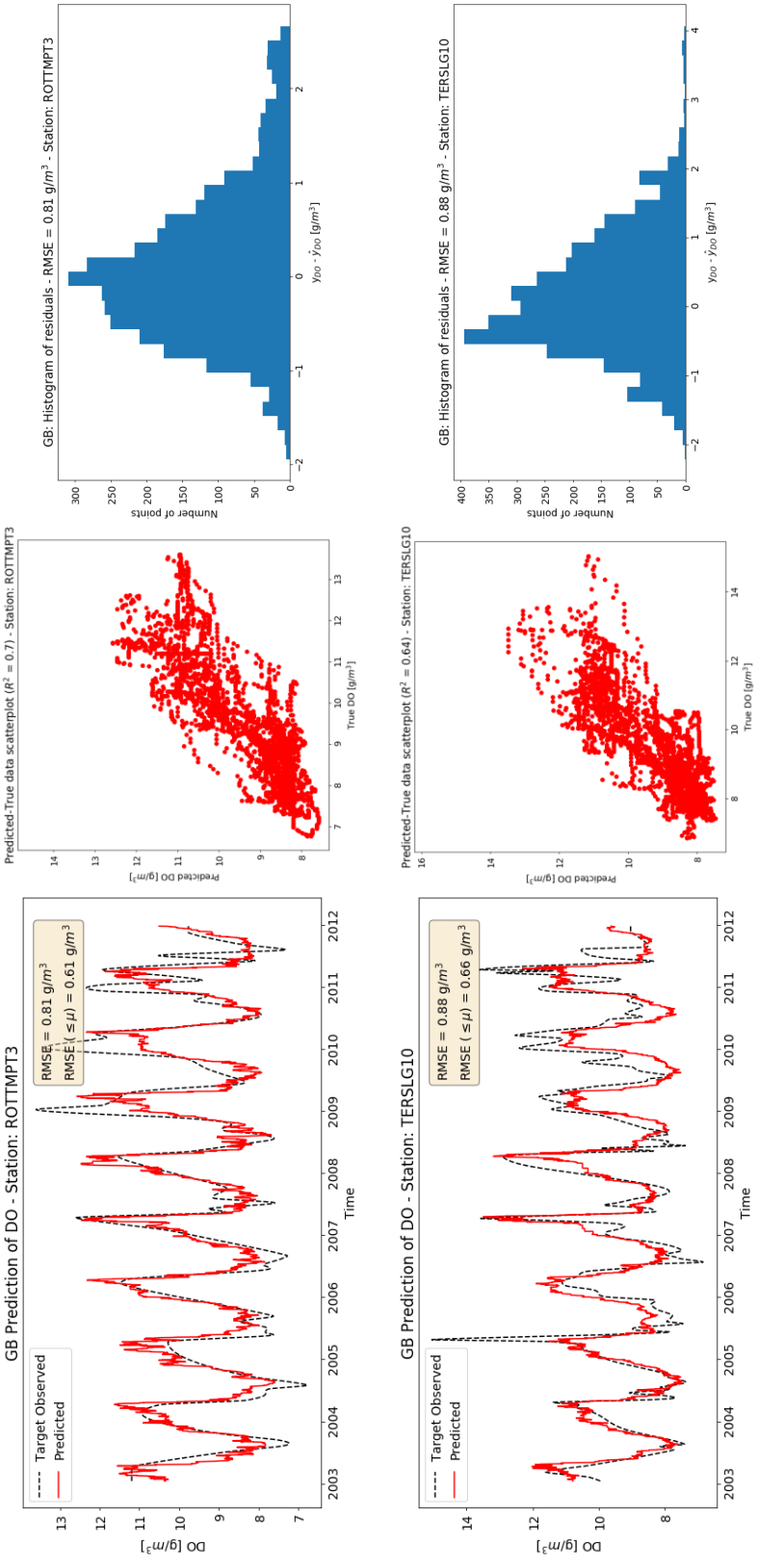

Predicted-True data scatterplot (R^2 = 0.79) - Station: SCHOUWN10
Predicted DO [g/m^3]
True DO [g/m^3]


GB: Histogram of residuals - RMSE = 0.84 g/m^3 - Station: NOORDWK10
Number of points
y_{DO} - $\hat{y}_{DO}$ [g/m^3]


GB Prediction of DO - Station: WALCRN20
Target Observed
Predicted
RMSE = 0.77 g/m^3
RMSE ($\leq \mu$) = 0.61 g/m^3
DO [g/m^3]
Time


Predicted-True data scatterplot (R^2 = 0.78) - Station: WALCRN20
Predicted DO [g/m^3]
True DO [g/m^3]


GB: Histogram of residuals - RMSE = 0.84 g/m^3 - Station: NOORDWK20
Number of points
y_{DO} - $\hat{y}_{DO}$ [g/m^3]


GB Prediction of DO - Station: ROTTMPT3
Target Observed
Predicted
RMSE = 0.81 g/m3
RMSE (≤ μ) = 0.61 g/m3
DO [g/m3]
Time
Predicted-True data scatterplot (R2 = 0.7) - Station: ROTTMPT3
Predicted DO [g/m3]
True DO [g/m3]
GB: Histogram of residuals - RMSE = 0.81 g/m3 - Station: ROTTMPT3
Number of points
yDO - ŷDO [g/m3]
GB Prediction of DO - Station: TERSLG10
Target Observed
Predicted
RMSE = 0.88 g/m3
RMSE (≤ μ) = 0.66 g/m3
DO [g/m3]
Time
Predicted-True data scatterplot (R2 = 0.64) - Station: TERSLG10
Predicted DO [g/m3]
True DO [g/m3]
GB: Histogram of residuals - RMSE = 0.88 g/m3 - Station: TERSLG10
Number of points
yDO - ŷDO [g/m3]

Below, predictions from ANN, GB and RF predictions are presented for visual comparison for station Noordwikj-10km.