Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

Ole Martin Borge

Atmospheric Correction over Coastal Waters Based on Machine Learning Models

Master's thesis in Applied Physics and Mathematics
Supervisor: Patrick Joseph Espy
June 2020

NTNU
Norwegian University of
Science and Technology

Ole Martin Borge

Atmospheric Correction over Coastal Waters Based on Machine Learning Models

Master’s thesis in Applied Physics and Mathematics
Supervisor: Patrick Joseph Espy
June 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

Summary

Atmospheric correction (AC) is often the first step in ocean-color remote sensing algorithms for satellites and would ideally remove atmospheric and ocean surface radiances to produce water-specific properties like the water-leaving radiance (L_w) or remote sensing reflectance (R_{rs}). These properties can again be used to derive ocean color products, such as the essential climate variable chlorophyll concentration (CHL), used to study phytoplankton changes in the ocean ecosystems to understand ecosystem responses to climate changes. AC in ocean color is critical as the atmosphere may contribute as much as $90\,\%$ of the measured top of the atmosphere (TOA) radiance in the blue wavelength regions over clear waters, and even higher in coastal areas. Standard AC algorithms work well for open oceans, but not for turbid coastal waters, where empirical assumptions done by the algorithms tend to fail.

In this master study, new AC and Inherent Optical Properties (IOP) retrieval algorithms for hyperspectral imaging over coastal waters based on the Machine Learning (ML) models Neural Network (NN), Partial Least Squares Regression (PLSR), Support Vector Regression (SVR) and Stochastic Gradient Descent Regression (SGDR) were tested. The different models were validated against each other with different metrics: the Pearson correlation coefficient (R), the mean percentage bias (Bias), the average percentage difference (APD), the root mean squared difference (RMSD) and the normalized RMSD (NRMSD), and were compared to determine the ML approach best suited for on-board processing. For this study, the coupled atmosphere-ocean Radiative Transfer Model AccuRT was used to simulate the interaction of solar radiation with particles and molecules in the atmosphere and ocean. Hyperspectral TOA radiance was generated based on various inputs of sensor geometries, AOD(869), and global in-field measurements of water IOPs from the NOMAD and CCRR datasets. The simulated data were used as training and validation data for the different ML models, where the goal was to predict R_{rs}.

When predicting R_{rs} from TOA radiance corrected for Rayleigh and absorption (L_{rac}), all ML models resulted in $\mathrm{R}^2 > 0.968$, indicating that they were able to predict the spectral relationship between L_{rac} and R_{rs}. The best results were obtained with the NN algorithm (R^2=0.999), especially compared to the linear models PLSR (R^2=0.974) and SGDR (R^2=0.968). On the other hand, the linear models provided interpretable coefficients. Also, unlike many standard AC algorithms, these models were capable of doing AC without the extra short-wave infrared (SWIR) bands, as they were trained on hyperspectral data in the wavelength region $400\text{-}800\,\mathrm{nm}$. Finally, the NN approach could also be used for water IOP retrieval, and provided $\mathrm{R}^2 > 0.9998$ when predicting chlorophyll concentration from R_{rs}. In further work, the ML models should be tested on $in\ situ$ data and be validated against standard AC algorithms.

The different AC algorithms based on ML became very fast once they had been trained and would therefore suit operational use in satellites and be part of the on-board data processing framework.

Sammendrag

Atmosfærisk korreksjon (AC) er ofte det først steget i "ocean color" algoritmer for satellitter og vil ideelt fjerne radians fra atmosfæren og havoverflaten, for å produsere vannspesifikke egenskaper som vannforlatende radians og -reflektans (L_w og R_{rs}). Disse egenskapene kan videre brukes til å predikere konsentrasjon av klorofyll i havet, som betegnes som en viktig klima-variabel og som benyttes for å studere endringer i planteplankton i havets økosystemer for å forstå mer om responser på klimaendringer. Atmosfærisk korreksjon er viktig da atmosfæren bidrar med opptil $90\,\%$ av det som en satellitt måler ved toppen av atmosfæren (TOA) i det blå bølgelengdeområdet, i åpne havområder. Det kan være enda høyere bidrag ved kystområder. Standard AC-algoritmer fungerer bra over åpne hav, men ikke for kystnære områder med mye grumsete vann, der empiriske antakelser gjort av modellen ikke er gyldige.

I denne masterstudien har nye algoritmer for AC og iboende optiske egenskaper (IOP) for hyperspektral avbildning over kystvann basert på nevrale nett (NN), Partial Least Squares Regression (PLSR), Support Vector Regression (SVR) og Stochastic Gradient Descent Regression (SGDR), blitt testet ut. De forskjellige modellene ble validert mot hverandre med forskjellige statistiske beregninger: Pearson-korrelasjonskoeffisienten (R), den gjennomsnittlige prosentvise biasen (Bias), den gjennomsnittlige prosentvise forskjellen (APD), root mean squared difference (RMSD) og den normaliserte RMSD (NRMSD). Disse ble sammenlignet for å bestemme hvilken ML model som var best egnet for bruk ombord i en sattelitt. I dette studiet ble en koblet atmosfære-hav stråleoverføringsmodell (AccuRT) brukt til å simulere hvordan solstråling ville interagere med partikler og molekyler i atmosfæren og havet. Hyperspektral TOA-radians, TOA-radians korrigert for absorpsjon og Rayleigh-radians, og vannforlatende radians og reflektans ble generert med AccuRT for forskjellige verdier av sensorgeometrier, AOD(869), klorofyll og mineralkonsentrasjoner, samt absopsjonskoeffisienten til farget oppløst organisk materiale ved 443nm. Disse spektrale dataene ble videre brukt som trenings- og valideringsdata for de forskjellige maskinlæringsmodellene, der målet var å predikere vannforlatende reflektans. De ulike inputverdiene for å generere havspesifikk radians ble hentet fra globale feltmålinger som NOMAD-datasettet, for å gjøre dataene så representative som mulig til faktiske situasjoner.

Ved predikering av vannforlatende reflektans fra TOA-radians korrigert for Rayleigh og absorpsjon (L_{rac}), så resulterte alle ML-modellene i R^2 > 0.968. De beste resultatene ble oppnådd med NN algoritmen (R^2=0.999), spesielt sammenliknet med de lineære modellene PLSR (R^2=0.974) og SGDR (R^2=0.968). Men de lineære modellene ga tolkbare koeffisienter. I motsetningen til standard AC-algoritmer så var disse ML-modellene i stand til å utføre AC uten ekstra-korte-infrarøde bølgelengder (SWIR). Avslutningsvis så viste bruk av NN for å predikere klorofyllkonsentrasjon fra R_{rs}. I videre arbeid bør ML-modellene også testet på in situ data og valideres mot standard AC algoritmer.

De forskjellige atmosfærisk korreksjonsalgoritmene basert på ML ble veldig raske når de først var ferdig trent, og algoritmene vil derfor egne seg godt som en del av dataprossesering rammeverket om bord i satelitter.

Preface

This thesis is submitted as part of the requirements for the Master of Science degree at the Norwegian University of Science and Technology (NTNU) and is the conclusion of a five-year integrated ”Master in Science”-programme in applied physics and mathematics. The extent of the thesis is 30 ECTS, suggesting an estimated workload of 40 hours per week. The work presented in this thesis is a part of the HYPSO project mainly concerned with remote sensing, carried out by The Center for Autonomous Marine Operations and Systems (NTNU-AMOS) at the Department of Engineering Cybernetics together with the Department of Electronic Systems.

The interest and introduction to the remote sensing field started with the very interesting and educational internship at Science and Technology AS in Jan-Feb 2019. Thanks to the S[&]T team and special thanks to Michael Soukup for interesting work, great mentoring and ideas for this master project, and Andreas Thorvaldsen for being co-supervisor from S[&]T, always available for advice and discussions. Without the internship, I would most probably not have ended up with this project, which turned out to be my most interesting academic year.

Thanks to my supervisor Professor Patrick Joseph Espy for your academic guidance and strong theoretical advice. I am also truly grateful for the supervision provided by PhD student Sivert Bakken who continuously has provided information and guidance to all the different parts of the master project. Thanks for using your valuable time for helpful discussions and giving me space to explore and chose a desired path within this project.

The friendship and good discussions with my fellow MSc. students deserve of course a special thanks. You all made a supportive environment and lots of fun, very needed for long days at Fysikkland. Also, a special thanks to my good friends Harald and Shahzeb for excellent IT support and valuable help for my progress with this project.

With the Covid-19 pandemic coming up, the work environment for my master suddenly changed from NTNU and Fysikkland, to be “isolated” in my rented room, to be evacuated to my home Ås from March. Thanks to my beloved parents for taking good care of me during my final student weeks.

Ås, June 14, 2020
Ole Martin Borge

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

6S Second Simulation of a Satellite Signal in the Solar Spectrum
AC Atmospheric Correction
AccuRT Accurate Radiative Transfer
AOD Aaerosol Optical Depth
AOP Apparent Optical Property
APD Averaged Percentage Difference
CCRR CoastColor Round Robin
CDOM Colored Dissolved Organic Matter
CHL Chlorophyll concentration $[\mathrm{g}/\mathrm{m}^3]$
CV Cross Validation
HAB Harmful Algal Bloom
HRS Hyperspectral Remote Sensing
HYPSO Hyper Spectral Imager for Oceanographic Applications
IOP Inherent Optical Property
MIN Mineral concentration $[\mathrm{mg}/\mathrm{m}^3]$
ML Machine Learning
MODTRAN MODerate resolution atmospheric TRANsmission
MSE Mean Squared Error
NASA National Aeronautics and Space Administration
NIR Near InfraRed
NN Neural Network
NOMAD NASA bio-Optical Marine Algorithm Data set
PCA Principle Component Analysis
PCR Principle Component Regression
PLSR Partial Least Squares Regression
RMSD Root Mean Squared Difference
RT Radiative Transfer
RTM Radiative Transfer Model
Savgol Savitzky Golay filter
SeaDAS SeaWiFS Data Analysis System
SGDR Stochastic Gradient Descent Regression
SVM Support Vector Machine
SVR Support Vector Regression
TOA top of the atmosphere
TSM Total Suspended Matter
VIS Visible

Chapter 1

Introduction

1.1 Motivation

About 70 % of the Earth is covered with water and although the oceans play a critical role in our climate, they remain the least explored of the Earth's environments. An understanding of the biogeochemistry, ecology, and hazards of our oceans in a changing climate is critical to sustaining Earth as a habitable planet [2]. Satellite remote sensing of the spectral albedo of the ocean is an effective tool to characterize and monitor our ocean environment on a global scale. However, this requires a knowledge of the biotic signatures of the different ecosystems as well as the modification of those signals by the atmospheric absorption typically associated with those ecosystems. Here we will utilize radiative transfer models to characterize and separate the atmospheric signals, and explore Artificial Intelligence (AI) solutions to identify near-shore ecosystems from their spectral albedo, discriminating against the atmospheric transmission scenarios typically present.

The difference between the ecosystem on land and in water is that the green plants in the ocean that contribute to pull carbon dioxide from the atmosphere, thus lower the volume of greenhouse gas, are predominantly microscopic, also known as phytoplankton. Phytoplankton uses the pigment molecule chlorophyll for photosynthesis, a molecule that preferentially absorbs blue and red portions of the light spectrum and reflects green light [3]. Most of the total scattering and absorption in waters is due to phytoplankton compared to other substances. This is the case for roughly 98 % of the world's open ocean and coastal waters, commonly also known as Case 1 waters [4]. The other category (Case 2 waters) is the nearshore waters, where the total absorption is dominated by sediment and yellow substances, often from land drainage, and are usually found in urban areas or close to areas with industrial development. Case 2 waters are typically more complex and varying than Case 1 waters, and are important for recreation, fisheries and military operations [4]. The impact of the land and river runoffs may impact key habits (e.g. mangroves, salt marshes) and fish recruitment. These regions are also in danger of eutrophication, which is a phenomenon of enrichment of natural systems through increased loads of nutrients [5].

Eutrophication causes an increase in the growth of plant and algae which can lead to harmful algal blooms (HAB) and can be of particular damage for mariculture. HAB can cause dead zones which refers to areas with a reduced level of oxygen (hypoxia) and release of toxins that can cause mass mortality of caged fish and long closures of shellfish farms. Further, HAB can also force closures of other wild fisheries, endanger human health, and can result in closure of recreational beaches and loss of tourism. The economic impacts of all these effects are substantial [5].

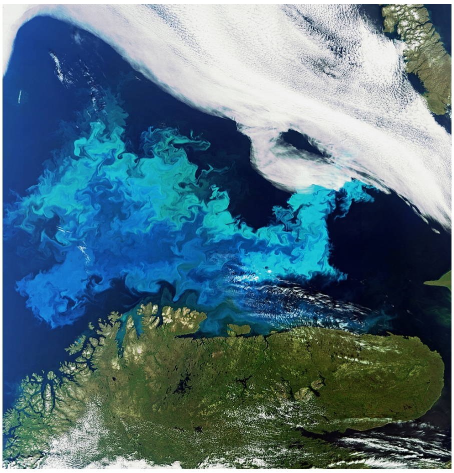


Figure 1.1: Satellite image of an algal bloom outside the coast of Northern Norway.

To understand the changes in the waters, a deeper understanding of the biotic signals and the constituents of the water is necessary. Most of the sunlight that penetrates the sea surface, especially red, orange, yellow and, green wavelengths, are absorbed by the water itself. The remaining light which is scattered back out of the water is a mainly shorter wavelength of blue and violets. Areas with high concentrations of phytoplankton can, due to their absorption and reflection properties, appear greener than areas with lower concentration. Observations and measurements of this can, therefore, be used to predict the amount of phytoplankton. However, the remoteness and vastness of oceans, in combination with phytoplankton that can change rapidly in response to environmental changes, require that the observations are on a synoptic scale. The only window into the marine ecosystem on these scales is with satellites, which provide a more global picture of the Earth as they cover spatially larger areas and therefore can capture data for larger areas at the same time better than traditional observing platforms such as moorings, free-drifting floats and ships [6].

Earth-orbiting spacecraft have been used the last thirty years to observe the oceans on larger scales and have with sufficient accuracy and precision showed to be useful for physical oceanography, ocean-system modeling, coastal management, biogeochemistry, and fisheries [2]. The remote sensing cameras can capture hundreds of images at a time over spatially large areas, where each image corresponds to a specific wavelength range in the electromagnetic spectrum. These wavelength ranges can also be outside the visible range (VIS) and capture a broader range of wavelengths than the human eye can see, giving valuable information useful for solving a variety of problems. "Ocean Color" in remote sensing refers to the capturing of the color of the ocean surface which depends on the chemicals and particles floating in it. One of the most important constituents of the water is the phytoplankton which also serves as a major indicator for eutrophication. The ocean color is the key to understand and monitor phytoplankton and can help scientists to study and predict environmental changes.

Traditional Earth Observation Satellites (EOS) are often very expensive and can take several years to develop and launch. The raw instrument data measured by the EOS is known as level 0 data (L0) and will normally go through some data processing before it is usable for other applications. The L0 data captured by the satellite is usually sent to Earth for data processing, where geometric and radiometric processing, data reshaping, and projection are done. [7]. Due to improvements in sensor technology, it has been a strong trend the recent years towards smaller satellites with smarter instruments. However, these satellites usually do not have the bandwidth to downlink all the L0 data down to Earth, but On-board Data Processing Software (ODPS) can solve this [7]. Satellites as small as 1-10 kg can use the raw data together with corresponding metadata to process and produce only the relevant data substantially reducing the downlink requirements. The small satellites offer a low-cost, flexible option to the traditional relative larger satellites and can be used to provide images of small areas of interest.

The Center for Autonomous Marine Operations and Systems (NTNU-AMOS) at the Department of Engineering Cybernetics together with the Department of Electronic Systems are developing and preparing a series of small satellites for launch. These satellites are specially made for maritime observation and surveillance [8]. As a result of this, The HYPer-spectral Smallsat for ocean Observation (HYPSO) a project by NTNU Small Satellite Lab that aims to develop NTNUs first SmallSat. The SmallSat in the HYPSO project will observe oceanographic phenomena with a hyperspectral camera, intelligent on-board processing, and robots. Many operational satellites use multispectral cameras, but hyperspectral sensors are being considered more frequently [9, 7]. Hyperspectral sensors can enhance target detection in rapidly varying signals where there is a high degree of spatial correlation at specific wavelengths and it is been expected that hyperspectral sensors can enhance the possibilities to detect targets of interest in comparison to multispectral cameras [10]. With today’s technology, it would also be possible to do specific data analysis on the image on-board the SmallSat. The HyperScout project is an example of this, where the aim is to detect fires, flooding or irrigation needs [7] based on small satellites with OBPS. Ocean-color sensors, like the HYPSO SmallSat, would for instance be capable of detecting changes in the chlorophyll levels well before it is obvious to the naked eye [5].

To extract useful information from the satellite data, like chlorophyll estimations, the measurements to be analyzed must consist of pure ocean signals. The satellite measurements contain not only ocean contributions but are perturbed by atmospheric effects. The removal of these atmospheric effects to retrieve surface-specific properties is known as Atmospheric Correction (AC).

1.2 Atmospheric Correction

Atmospheric Correction in remote sensing plays an important role, as the view of Earth's surface from satellites is degraded by the presence of the atmosphere. The satellites normally measure the upward radiant flux received at the top of the atmosphere (TOA). However, they will not only measure the useful contribution from the water but will also measure different atmospheric radiance contributions from light scattered by air molecules (Rayleigh), aerosols, and different kinds of surface reflections. Ideally, an AC algorithm for ocean water would remove these parts and retrieve the water-leaving radiance ($L_w(\lambda)$) or reflectance ($\rho_w(\lambda)$).

Several different AC algorithms that try to remove the surface and atmospheric contributions in various ways have been developed [11]. After retrieving the water-leaving radiance or reflectance from the spectral measurements of the TOA radiance, it would then be possible to extract inherent optical properties (IOPs) from it. IOP is linked with the absorption and scattering characteristics of the medium and its constituents (such as chlorophyll) [12] and would give valuable information about the medium. To extract these properties, it has also been developed several water IOP retrieval algorithms, often based on some semi-empirical assumptions [13].

The different atmospheric and surface contributions of the TOA radiance must be considered differently and studied carefully to produce an accurate AC algorithm. For radiance in the ultraviolet-visible range (UV-VIS), the atmospheric effects are crucial, as it is the major proportion of total measured TOA radiance. This term can be computed accurately, and it can be computed with a Radiative Transfer Model (RTM) with an uncertainly lower than $0.5\,\%$ [14]. RTMs build on the physical phenomenon of electromagnetic radiation propagation through different mediums and can model the path of photons and estimate the path radiance contribution by simulations of a constructed atmosphere when enough input and boundary conditions are given.

However, AC algorithms that build on Radiative Transfer (RT) can be computationally expensive and the need for proper corresponding metadata to the location where the satellite image was captured is important for accurate results. More empirical AC algorithms build on different assumptions to abstract themselves from RT models. One of the most used assumptions is the assumption that due to the highly absorbing properties of the water, no radiance is backscattered out of the water in the near-infrared region (NIR). This assumption has shown good results for open ocean, but tends to fail in coastal areas where both aerosol and ocean effects are much more complex [11, 13, 5] and the algorithms can be inaccurate. As already mentioned, monitoring the water in coastal and inland areas are of high importance. Therefore new AC methods are needed, capable of correcting for the complexity found in the coastal regions [5]. This study will make use of different ML models to investigate AC over coastal waters where the black ocean assumptions tend to fail.

1.3 Objective and Tasks

The main objective of this master project is to produce ML based AC algorithms capable of handle strong aerosol containment and Case 2 waters. The main objective will be achieved through four subgoals:

- Utilize RT model(s) to simulate different spectral radiance/irradiance data representative for a wide range of atmospheric and coastal oceanic environments.
- Produce different AC models capable of handle strong aerosol containment and Case 2 waters by training on the generated data. Test and validate the ML models against each other with respect to accuracy, time complexity and interpretation capability, to study which would be preferable to use in an on-board processing framework for the HYPSO project.
- Study ML hyperparameters and different pre-processing steps to improve the ML models. Test and validate the ML models with and without the pre-processing steps.
- Produce different ML based water IOP retrieval models that can be used to predict properties like chlorophyll concentration from AC end products ($R_{rs}(\lambda)$). Validate and compare these models against each other.

1.4 Structure of the Thesis

The master thesis is divided into 6 chapters:

- **Chapter 1** will give a short background description and present the motivation for this study.
- **Chapter 2** will present the problem around AC and briefly present the relevant background theory relevant for AC.
- **Chapter 3** will discuss how to utilize RT models to do AC, and relevant, operational AC algorithms for this thesis will be presented. The desired RT used in this thesis, namely AccuRT, will be presented and discussed.
- **Chapter 4** details the data generated with AccuRT and the final method for AC. Data pre-processing will be discussed and the ML models desired to use will be presented. Hyperparameter optimization will also be discussed and some results will be presented.
- **Chapter 5** presents and discusses the final results for the various ML models, both when it comes to AC and IOP retrieval.
- **Chapter 6** concludes and summarizes the problems discussed in this thesis based on obtained results. Further work will also be discussed.

Chapter 2

Background Theory

This chapter presents the basic theory that further progress is based upon, as well as complementary theory needed for a general understanding of the problem around AC and the remote sensing field. A general description of the most fundamental and important aspects of hyperspectral remote sensing, the theory regarding ocean optics and, AC algorithm description will be covered. In addition, a physical description of the different oceanic and atmospheric effects important to consider for an AC algorithm will be presented.

2.1 Hyperspectral Remote Sensing

In Hyperspectral Remote Sensing (HRS), the satellite combines imaging and spectrometry to capture information across the electromagnetic spectrum [15]. The satellite would ideally capture an image of a geographical scene where each pixel would have information over a spectrum. The data is stored in a three-dimensional dataset (x, y, λ) often called a hypercube, where the first two dimensions represent the spatial x and spatial y direction of the geographical scene, and the third dimensions represent the spectral information (often called bands). This is illustrated in Fig. 2.1 which shows a hypercube with 10×10 pixels and 14 different bands and the reflectance spectra for one pixel.

Each band corresponds to a narrow wavelength range where the satellite would capture information related to the chemical composition of materials by measuring the variation in power with the wavelength of the frequency of light. The variation of how reflectance or emissivity of the materials within an image pixel varies with wavelength is provided in the different bands and can often be enough to characterize the material observed [15]. A plot of the reflectance from four different materials as a function of wavelength is showed in Fig. 2.2 to illustrate how each material has a unique spectrum.

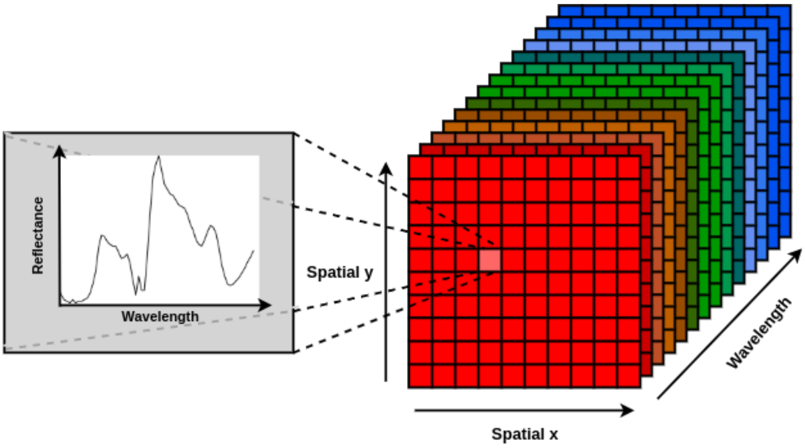


Figure 2.1: Illustration of the hypercube data structure. The data is captured in spatial x- and y-direction for sensor specific number of different bands, which each represents a narrow wavelength range of the electromagnetic spectrum (spectral band).

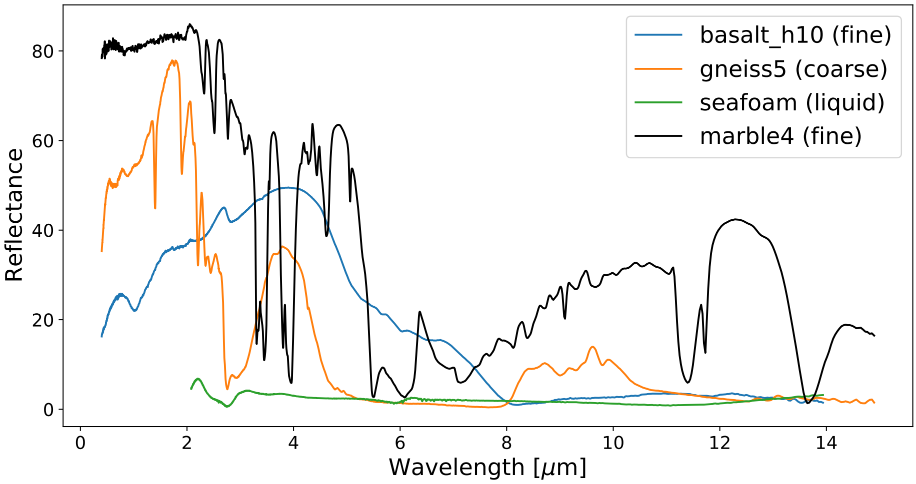


Figure 2.2: A plot of reflectance as a function of wavelength for three rocks (Basalt, Gneiss, and Marble) and water foam. The data is received from ECOSTRESS spectral library (formerly ASTER spectral library). The numbers in the label refer to the sample number, which can be used to find further information about the chemical composition of the material [16, 17].

A typical hyperspectral remote sensing hypercube consists of about 100-200 spectral bands with bandwidths of around $5\,\mathrm{nm}$. Another well-used imaging technique used for remote sensing is multispectral imaging. This differ from HRS by having fever bands (normally 5-10) and wider bandwidth ($70\text{-}400\,\mathrm{nm}$) [18]. In contrary to imaging systems that often captures reflected and/or emitted electromagnetic radiation integrated over the visible band, HRS also measures other wavelength regions not visible for humans. All these regions can further be divided into the visible (VIS), near-infrared (NIR), shortwave-(SWIR), midwave-(MWIR) and longwave (LWIR) infrared. Contributions measured by the remote sensing systems in these regions ($0.4\text{-}14\,\mu\mathrm{m}$) can be divided into reflected sunlight and thermal emission from an object in the scene. The different wavelength areas, their abbreviations, range, and contribution is specified in Table 2.1. Ocean color remote sensing normally utilize the VIS region, i.e. wavelengths between $400\,\mathrm{nm}$ and $700\,\mathrm{nm}$, and NIR light, i.e. wavelengths from $700\,\mathrm{nm}$ to just under $2000\,\mathrm{nm}$ [13].

Table 2.1: Name of different wavelength regions, with abbreviation, regions and what contribution the region is dominated by.

Wavelength Region	Abbreviation	Wavelength	Dominated by
Visible	VIS	0.4 - 0.8 µm	reflected sunlight
Near Infrared	NIR	0.7 - 1.1 µm	reflected sunlight
Visible + Near Infrared	VNIR	0.4 - 1.1 µm	reflected sunlight
Shortwave Infrared	SWIR	1.1 - 3 µm	reflected sunlight
Midwave Infrared	MWIR	3 - 5 µm	thermal emission
Longwave Infrared	LWIR	5 - 14 µm	thermal emission

Two ways to describe the precision of the sensors is typically by spectral and spatial resolution. The spectral resolution depends on two factors, sampling interval, and bandwidth. The sampling interval is the spectral distance between the centers or peaks of spectral channels along a spectrum and bandwidth which is the full width at half maximum (FWHM) of a spectral channel. This is illustrated in Fig. 2.3. The blue curves are typical for what one would find in multispectral sensors where the sampling interval is larger than the bandwidth. Hyperspectral sensors look more like the orange curves, where the bandwidth is narrow and less than the sampling interval. A large number of narrow bands would make it possible to do material analysis with few pixels, as it would consist of more information, as previously illustrated in Fig. 2.2. The spatial resolution represents the size of each pixel, which can be as small as 1 m for airborne systems to more than 1000 meters for satellite systems [13]. Large pixels could result in pixels containing multiple objects, making it difficult to identify the object. On the other hand, if the pixels are too small, the reliability of the measured features could be reduced due to decreased signal-to-noise ratio [15].

The goal of remote sensing is to convert the sensor measurements into useful and desired information [13]. It is of interest that the sensors have some standard measurements which can be compared to other sensors and reference values measured in the field for validation. Therefore, the measurements must be independent of the sensor.

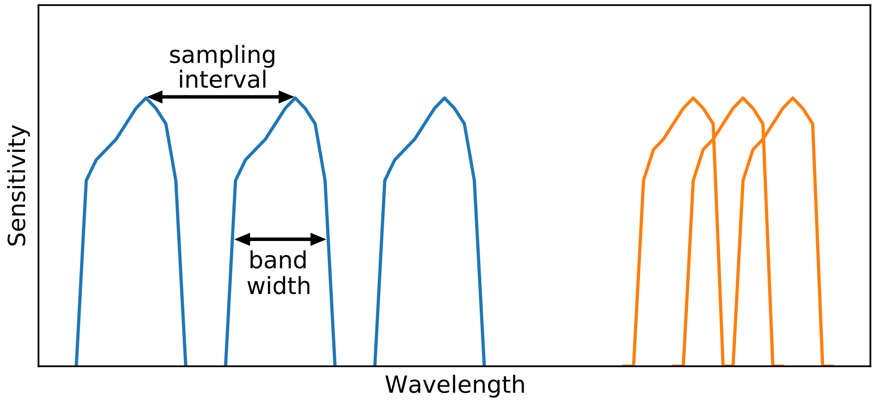


Figure 2.3: Illustration of sensitivity, sampling interval and band width. The blue curves represent the bands and spacing typically used in multi-spectral imaging, whereas the orange bands are more typical of hyperspectral imaging

2.1.1 Radiometry

Radiometry is the science of measuring electromagnetic energy. However, what quantity of electromagnetic energy to measure is an important aspect to decide in remote sensing. The detectors of optical instruments in the satellite often register the light energy they receive during an observation period. In some cases, the energy received is not only dependent on the properties of the observed scene, but also on the instrument itself. Therefore, it is important to measure physical quantities that are not dependent on the sensor so that comparisons of measurements between sensors are possible [19]. The total spectral energy measured by the sensor, ϵ, during a time, t, is provided by photons of different wavelengths and is expressed in joules. The sensor would receive energy during the integration time given as [19]:

$$\epsilon = \int_{\lambda_1}^{\lambda_2} S(\lambda)\epsilon(\lambda)d\lambda \tag{2.1}$$

where S(λ) is the unit-less, instrument spectral sensitivity between λ_1 and λ_2, which is the relative efficiency of detecting light as a function of the wavelength of the light. To remove the time dependency, the quantity spectral flux is introduced and is expressed in watts per unit wavelength as:

$$\phi(\lambda) = \frac{d\epsilon(\lambda)}{dt} \tag{2.2}$$

To get rid of the detector surface dependency, irradiance is introduced. This quantity is the spectral flux reaching the detector per surface and given as:

$$E(\lambda) = \frac{d\phi(\lambda)}{dA} \tag{2.3}$$

tustna hvdan sove where dA is the surface unit. However, this quantity will depend on the orientation of the detector as well as the field of view of the detector. The next improved quantity is called the radiance and is the spectral flux that reaches the instrument per unit area and per unit of solid angle (dΩ), perpendicular to the surface of the detector. The perpendicular solid angle is defined as dΩ_s = d$A_s$cos$(\theta_s)/r^2$ and is expressed in steradians [sr]. Radiance is given in $\mathrm{W/m^2/sr}$ and is given by [19]:

$$L(\lambda) = \frac{d^2\phi(\lambda)}{dA\, d\Omega\, \cos(\theta_s)} \tag{2.4}$$

where θ_s represents the angle between the surface normal and the direction of which L is measured. A $\times$ cos(θ_s) can, therefore, be interpreted as the area perpendicular to the direction L(λ) is measured. Radiance is widely used property in remote sensing because it is independent of the characteristics of the instrument and since it further can be used to derive other useful radiometric properties [13]. With this fundamental radiometric quantity in remote sensing defined, it is possible to address the problem around AC.

2.2 Atmospheric Correction Problem Formulation

When photons from the sun reach the earth, its radiance spectrum will be perturbed by the constituents in the atmosphere and ocean. If the photons would arrive on an Earth without atmosphere, the incoming solar radiation would eventually hit the surface and partially be absorbed by the surface and partially reflected back towards space. A remote sensing satellite capturing the backscattered radiation would then easily be able to predict surface properties based on the measured radiance, the sun-surface-remote sensor angles, and total solar irradiance. However, the photons would in an actual case be perturbed by two atmospheric processes, scattering by molecules and aerosols and gaseous absorption [20], in addition to ocean effects. These contributions are critical to understand and investigate in order to do accurate AC.

As most of the measured TOA radiance over waters is due to atmospheric contributions, retrieving useful properties from the water-leaving radiance could only be done properly if the atmospheric correction algorithms are accurate. This is because a relatively small portion of the incoming sunlight is backscattered from below the ocean surface in comparison with the sunlight backscattered from the atmosphere and specular reflection from the surface [21]. The total measured TOA radiance, $L_t(\lambda)$, for ocean-atmosphere systems can be expressed as a partitioned linearly equation given by [22, 23]:

$$L_t(\lambda) = L_{path}(\lambda) + t(\lambda)L_{wc}(\lambda) + t(\lambda)L_{sky}(\lambda) + T(\lambda)L_{sun}(\lambda) + t(\lambda)L_w(\lambda) \tag{2.5}$$

where $L_{path}(\lambda)$ is the radiance contribution from different atmospheric scattering, $L_{wc}(\lambda)$ is the radiance contribution from whitecap on the sea surface, $L_{sun}(\lambda)$ is the sun glitter radiance (specular reflection of direct sunlight off the sea surface), $L_{sky}(\lambda)$ is the radiance contribution from surface-reflected background atmospheric radiance and $L_w(\lambda)$ is water-leaving radiance due to photons that penetrate the sea surface and are backscattered out of the water. The water-leaving radiance is the contribution that consists of valuable information about the medium and is what the ocean color AC algorithms aim to retrieve. This signal is often only about $10\,\%$ of $L_t(\lambda)$, which gives an impression of how small this desired property is compared to the total measured radiance [13].

The contribution to the path radiance is the scattering by atmospheric gases and aerosols, where the part that solely comes from the molecules is called Rayleigh radiance, denoted as L_r [13]. Further, $L_{path}(\lambda)$ can be divided into three contributions given as:

$$L_{path}(\lambda) = L_r(\lambda) + L_a(\lambda) + L_{ra}(\lambda) \tag{2.6}$$

where $L_r(\lambda)$ is the radiance due to scattering by air molecules (Rayleigh scattering in absence of aerosols), $L_a(\lambda)$ is the aerosol scattering (in the absence of air molecules) and $L_{ra}(\lambda)$ is the multiple interaction terms between molecules and aerosols. $L_a(\lambda)$ + $L_{ra}(\lambda)$ is often put together and denoted as L_A [24], and referred to as the aerosol contribution.

The different radiance terms contributing to the total measured TOA radiance $L_t(\lambda)$ are illustrated in Fig. 2.4. Here, the arrows marked with A, B, C, D, E, F refer to Rayleigh (L_r), aerosol (L_a), water-leaving (L_w), direct sun glint (L_{sun}), whitecap (L_{wc}) and sky glint (L_{sky}) radiance, respectively.

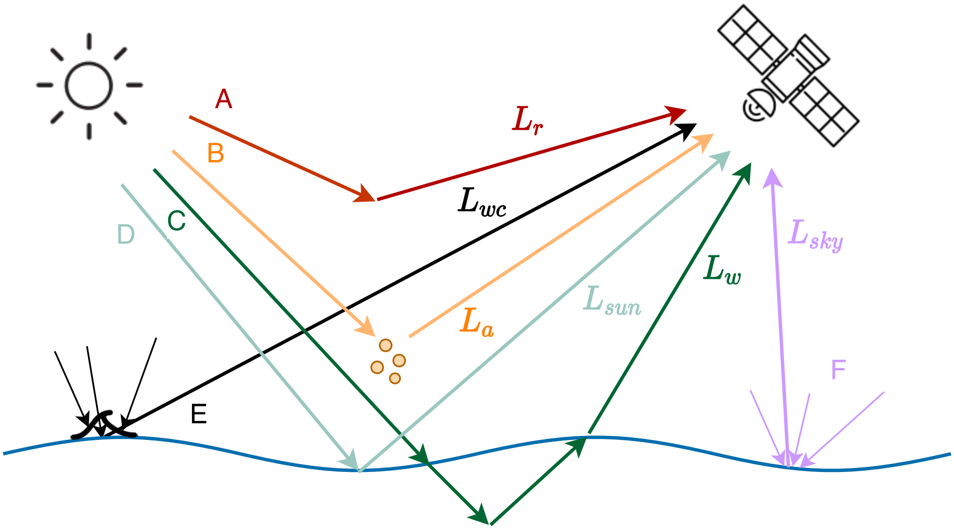


Figure 2.4: Illustration of different contributions to the sensor-measured radiance. A, B, C, D, E, F refer to Rayleigh (L_r), aerosol (L_a), water-leaving (L_w), direct sun glint (L_{sun}), whitecap (L_{wc}) and sky glint (L_{sky}) radiance, respectively.

All contributions in L_{path} are considered to be at the TOA, whereas the L_w, L_{sun}, L_{sky} and L_{wc} radiances refer to radiance measured just above the sea surface. The latter radiances will experience transmittance caused by atmospheric absorption and attenuation by atmospheric constituents [13] when propagating upwards toward a sensor, and would therefore not give the same measurements just above the surface as at TOA. Common notation is to define the radiance measured at TOA as L^{TOA}, where the subscript TOA is used to specify that this value is at TOA. L_w^{TOA} would, therefore, refer to how much of the water-leaving radiance just above the surface that reaches the sensor at TOA. In Eq. 2.5, these contributions are included as the diffuse (t(λ)) and direct transmittances (T(λ)).

Direct transmittance is used when only one particular path, or a narrow bundle of nearly colinear paths, connects the rays and the sensor [13], which is the case for the specular sun glint reflection. For the upward propagation of L_w, L_{sky} and L_{wc} towards a sensor, diffuse transmittance is used because radiance from all locations and various directions can be scattered into their direction of interest via only one scattering. $t(\lambda)$ is the diffuse transmittance of the atmosphere from the surface to the satellite sensor and $t_0(\lambda)$ is the corresponding diffuse transmittance from the sun to the surface. The direct transmittance from the surface to the sensor and from the sun to the surface is given as $T(\lambda)$ and $T_0(\lambda)$, respectively. The diffuse transmittance, found in Eq. 2.5 can therefore be defined as [13]:

$$t(\lambda) \equiv \frac{L_w^{TOA}(\lambda)}{L_w(\lambda)} \tag{2.7}$$

and is the atmospheric transmittance that accounts for the propagating of L_{wc}, L_{sky} and L_w from the sea surface to the top of the atmosphere [25]. A further explanation of the different radiance terms noted in Eq. 2.5 including other radiance terms relevant for this study, are listed in Tab. 2.2.

Table 2.2: Radiance and irradiance notations and definitions. The SI units for the spectral radiance L and irradiance E are $\mathrm{Wm}^{-2}\mathrm{nm}^{-1}\mathrm{sr}^{-1}$ and Wm^{-2}, respectively [13].

Symbol	Definition
L_t	total upwelling radiance measured at TOA
L_{rac}	TOA radiance without Rayleigh radiance and absorption.
L_{path}	total radiance coming from atmospheric scattering to the TOA radiance. L_{path} = L_r + L_a + L_{ra}
L_r	total Rayleigh radiance at TOA
L_a	total radiance at TOA coming from aerosol scattering only
L_{ra}	total radiance due to aerosol-molecule scattering
L_w	water-leaving radiance just above sea surface
L_w^{TOA}	total contribution of L_w reaching TOA
L_{wc}	radiance from whitecaps and foam just above sea surface
L_{wc}^{TOA}	radiance from whitecaps and foam reaching TOA
L_{sky}	surface-reflected radiance background sky radiance at the sea surface
L_{sky}^{TOA}	the part of L_{sky} that reaches TOA
L_{sun}	direct Sun glint radiance just above the sea surface
L_{sun}^{TOA}	the part of the Sun glint radiance L_{sun} that reaches TOA
L_g	total glint radiance. L_g = L_{sky} + L_{sun}.
$E_d(0^+)$	Spectral downward plane irradiance just above surface
R_{rs}	Remote sensing reflectance give as L_w / $E_d(0^+)$ [sr^{-1}].

For simplicity, all the radiance terms in Eq. 2.5 has so far been written as a function of wavelength only. However, the radiance distribution is a function of four other variables, namely depth z, solar zenith angle (θ_0), viewing zenith angle (θ), and relative azimuth angle ($\Delta\phi$). In order to describe these angles, one important term must be specified, namely zenith. Zenith is illustrated in Fig. 2.5, and refers to a vertical vector at a location pointing in the direct opposite direction as an apparent gravitational force at that location. The solar zenith angle (θ_0) is defined as the angle between a downwelling sunray and zenith, whereas the view zenith angle (θ) is defined as the angle between an upwelling sun ray propagating towards a sensor and the zenith. These angles are also illustrated in Fig. 2.5, together with the view (ϕ) and solar (ϕ_0) azimuth angles. The absolute value of the difference between the view and solar azimuth angles defines the relative azimuth angle, given as $\Delta\phi$. The amount of Rayleigh scattering is heavily dependent one these angles, as they would give an indication of how long the photons would propagate thought the atmosphere, which again would result in more/less scattering.

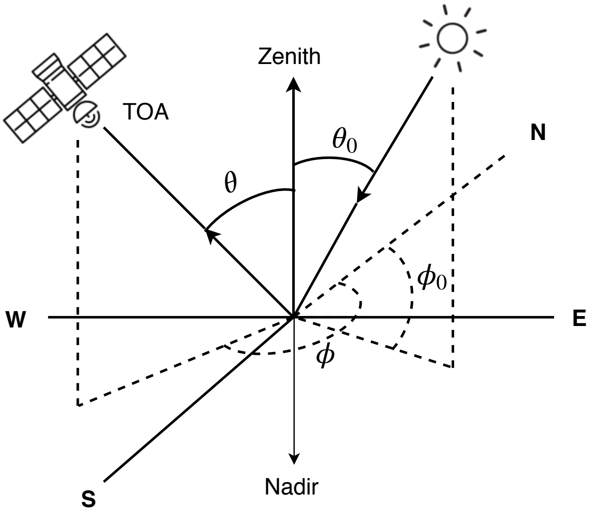
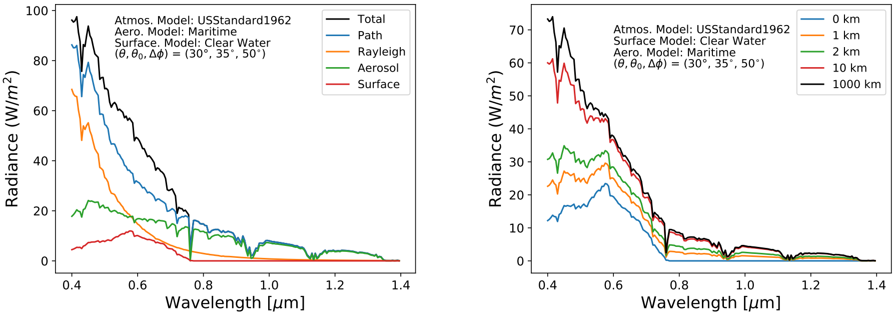


Figure 2.5: Illustration of viewing zenith and azimuth angles (θ,ϕ) and solar zenith and azimuth angles (θ_0,ϕ_0). Zenith is the direction away from an observed point, and nadir is the opposite direction of the observer's ascending (descending) vertical.

A plot of TOA, path, Rayleigh, and surface radiance is shown in Fig. 2.6 (**a**) for a simulated atmosphere using the radiative transfer model 6SV [20]. The surface radiance is the sum of water-leaving, whitecap, sky, and sun glint radiance. The plot shows that the surface contributions are very small compared to the atmospheric effects which normally contribute to between $70\,\%$ to $90\,\%$ of $L_t(\lambda)$. Fig. 2.6 (**b**) shows how the sensor measured radiance increases when the tangent height of the satellite increases due to an increased amount of atmospheric path radiance.



(a) TOA measured radiance with atmospheric path, Rayleigh and surface contributions. The surface contribution contains water glint, foam and water-leaving radiance.

(b) Total radiance at different tangent heights by a limb viewing satellite. The total radiance measured increases with increasing tangent height due to atmospheric path contributions.

Figure 2.6: TOA measured radiances from different contributions (**a**) and different tangent heights (**b**). The spectral data is based on simulations using Py6S [26].

The radiances from land and water surfaces are mainly contributions of reflected sun irradiance. The sun irradiance depends on the distance between the earth and the sun, thus it would, therefore, variate with the season. It can therefore often be more convenient to convert the water-leaving radiance ($L_w(\lambda)$), or any other radiance L(λ), to the dimensionless reflectance, (ρ_w). ρ_w is normalized by the extraterrestrial solar irradiance and contains all the absorption that takes place within the water and is an important parameter in ocean color. The definition of the reflectance ρ is given by

$$\rho_w(\lambda) = \frac{\pi L_w(\lambda)}{F_0 \cos(\theta_0)} \tag{2.8}$$

where $L_w(\lambda)$ is the water-leaving radiance in the given viewing direction, F_0 is the extraterrestrial solar irradiance and θ_0 is the solar zenith angle. Reflectance would describe the fraction of incident electromagnetic power that is reflected at an interface. The reflectance can then easily be compared with other reflectances measured in different spectral bands and at different times of the year or of the day. The goal of using such properties is often to achieve standard measurements from the satellite that is easier to compare with standard measurements $in\ situ$, which refers to measurements with instruments located directly at the point of interest, like just above sea level for measurements of water-leaving radiance [13]. Therefore, properties that are independent of atmospheric conditions, solar and view zenith angles, and the sea state and still contains useful information about the constituents of the water, like chlorophyll, is desired.

This has resulted in the concept of normalized water-leaving radiance which can be interpreted as the radiance that could be measured by a nadir-viewing instrument if the Sun were at the zenith and in the absence of any atmospheric loss, and when the Earth is at its mean distance from the Sun [13]. Mathematically, the normalized water-leaving radiance, $[L_w(\lambda)]_N$, is given by:

$$[L_w(\lambda)]_N \equiv \left(\frac{R}{R_0}\right)^2 \frac{L_w(\lambda)}{\cos(\theta_0)\, t_0(\lambda, \theta_0)} \tag{2.9}$$

where $L_w(\lambda)$ is the water-leaving radiance just above the sea surface, R is the Earth-Sun distance at the time of measurement, and R_0 is the mean Earth-Sun distance. The $(\mathrm{R}/\mathrm{R}_0)^2$ contribution corrects for the Earth-sun distance, which can vary up to $8\,\%$ over a year due to the Earth's elliptical orbit [13]. $t_0(\lambda, \theta_0)$ is the diffuse transmittance in the Sun's direction accounting for attenuation effects. The corresponding normalized water reflectance, $[\rho_w(\lambda)]_N$, is given as:

$$[\rho_w(\lambda)]_N \equiv \frac{\pi}{F_0}[L_w(\lambda)]_N = \left(\frac{R}{R_0}\right)^2 \frac{\pi\, L_w(\lambda)}{F_0 \cos(\theta_0)\, t_0(\lambda, \theta_0)} \tag{2.10}$$

where the denominator more commonly is written as $E_d(\lambda, \theta_0)$, and given by:

$$E_d(0^+, \lambda, \theta_0) = F_0 \left(\frac{R_0}{R}\right)^2 \cos\,(\theta_0)\, t_0(\lambda, \theta_0) \tag{2.11}$$

and is known as the spectral downward plane irradiance (radiant flux received by a surface per unit area). F_0 is the extraterrestrial solar irradiance and is expressed in W/m^2. It can be seen as the value of solar power reaching the earth without and disturbance from the Earth's atmosphere.

The goal with AC is to derive sea-level properties like water-leaving radiance, $L_w(\lambda)$, from the measured TOA radiance, $L_t(\lambda)$. A common procedure is to simulate and estimate the various surface and atmospheric-radiances terms in Eq. 2.4 and subtract them from the measured TOA radiance. AC algorithms vary in how they calculate the different terms in Eq. 2.5.

All the equations described so far give a quantitatively way of describing light itself, but it should further be explained how to describe and measure the optical properties of the medium where the light is propagating through. This is commonly known as optical oceanography.

2.2.1 Optical Oceanography

When it comes to aquatic research, the state of the water can be retrieved by looking at its optical properties [13]. There is a connection between the biological, chemical and geological constituents of natural waters and the optical properties. Optical properties of water can be divided into two mutually exclusive classes: Inherent Optical Properties (IOP) and Apparent Optical Properties (AOP). IOPs are properties that only depend upon the medium, thus are independent of the surrounding light field within the medium. The two fundamental IOPs are the volume scattering function and the absorption coefficient [13]. The properties that share the same properties as IOP, as well as the geometric (directional) structure of the surrounding light filed, are categorized as AOP. Also, AOP must display enough stability and features to be useful descriptors of the water body [13]. A widely used AOP is the spectral irradiance reflectance, which is given as:

$$R(z, \lambda) \equiv \frac{E_u(z, \lambda)}{E_d(z, \lambda)} \tag{2.12}$$

where E_d and E_u denotes the downwelling and upwelling irradiance, respectively, illustrated in Fig. 2.7. Radiometric variables such as radiance and irradiance depend on IOPs, but fail to be stable enough to be categorized as AOP. Measurements at the ground of E_d and E_u would both change drastically if there would appear a cloud that would block the light, thus they do not separately categorize as AOPs. On the other hand, R(z, λ) for the same case would not change much, thus it is also regarded as an AOP. Other common AOPs are average cosines, diffuse attenuation coefficient, and different reflectances. In later years, the AOP of choice for remote sensing of ocean properties [27, 28] has been the spectral remote sensing reflectance given as:

$$R_{rs}(\theta, \phi, \lambda) \equiv \frac{L_w(0^+, \theta, \phi, \lambda)}{E_d(0^+, \lambda)} \tag{2.13}$$

where 0^+ means that L_w and E_d are evaluated just above the sea surface. Notice, that Eq. 2.10 and Eq. 2.13 would give the following relationship between the normalized water-leaving reflectance and the remote sensing reflectance: $[\rho_w(\lambda)]_N = \pi R_{rs}(\lambda)$. $R_{rs}(\lambda)$ is the desired property because it is less sensitive to environmental conditions, such as sun angle or sky conditions when compared to R(z,λ). This property is the ratio of how much of the downwelling irradiance that would penetrate the sea surface and be backscattered by oceanic constituents and returned through the surface onto a small solid angle $\Delta\Omega$ centered on a particular direction (θ, ϕ) [13]. This is further illustrated in Fig. 2.8.

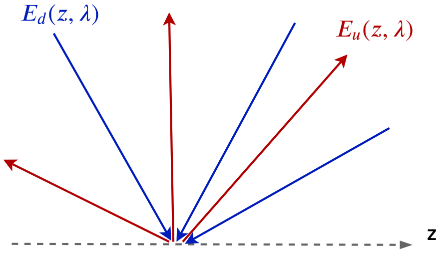


Figure 2.7: Illustration of the downwelling and upwelling irradiance.

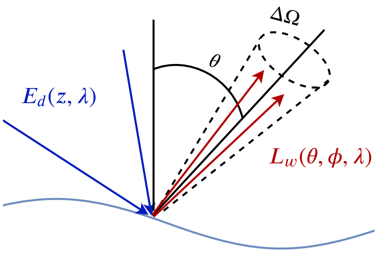


Figure 2.8: Illustration of the light rays contributing to $R_{rs}(\lambda)$.

Two things can happen when a photon interacts with matter, which are absorption and scattering. Absorption happens when the energy of the photon is being converted completely to another form like heat or energy contained in an energy bond. Scattering happens when the photon changes its direction and/or energy [1]. These two properties only depend on the water itself and the substances in it, and are regarded as the two fundamental IOPs [13]. More specifically, the two fundamental IOPs are called the absorption coefficient and the volume scattering function. If a small volume ΔV of water would be illuminated by a collimated beam of monochromatic light at some wavelength λ, then by conservation of energy, this could be written as:

$$\Phi_i(\lambda) = \Phi_a(\lambda) + \Phi_s(\lambda) + \Phi_t(\lambda) \tag{2.14}$$

where $\Phi_i(\lambda)$, $\Phi_a(\lambda)$, $\Phi_s(\lambda)$, and $\Phi_t(\lambda)$ are the incoming, absorbed, scattered and transmitted spectral radiant power (in W nm^{-1}), respectively [13]. These terms are illustrated in Fig. 2.9.

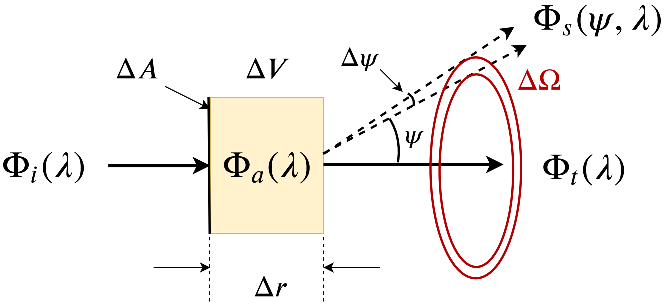


Figure 2.9: Illustration of absorption, scattering and transmission of an incoming beam with geometry used to define volume scattering and inherent optical properties.

Then the three properties absorptance, A(λ), scatterance, B(λ), and transmittance, T(λ), can be expressed as:

$$A(\lambda) \equiv \frac{\Phi_a(\lambda)}{\Phi_i(\lambda)}, \qquad B(\lambda) \equiv \frac{\Phi_s(\lambda)}{\Phi_i(\lambda)}, \qquad T(\lambda) \equiv \frac{\Phi_t(\lambda)}{\Phi_i(\lambda)} \tag{2.15}$$

where the A(λ) is the fraction of incident power that is absorbed within the volume, B(λ) is the fraction of incident power that is scattered out into all direction, T(λ) is fraction of non-interacting incident power and the sum of them should be 1 due to conservation of energy. The optical depth of a material is another used property in remote sensing and is defined as:

$$\tau(\lambda) = \ln\left(\frac{\Phi_i(\lambda)}{\Phi_t(\lambda)}\right) = -\ln(T(\lambda)). \tag{2.16}$$

where $\tau(\lambda)$ is the atmospheric optical path along a vertical path with nadir viewing direction for the sensor and includes all effects of atmospheric absorption and attenuation by all atmospheric constituents [13]. For a general path (off-nadir viewing direction θ), the direct transmittance (from Eq. 2.5) would be:

$$T(\theta, \lambda) = \exp\left(-\frac{\tau(\lambda)}{\cos{(\theta)}}\right). \tag{2.17}$$

Aerosol optical depth (AOD) is the optical depth coming from the aerosols in the atmosphere and is a well used property in remote sensing. More precisely, it is a dimensionless number indicating how much aerosol there is along a vertical column of the atmosphere and yields the amount of the direct sunlight that is prevented from going through the atmosphere.

In optical oceanography, the IOPs usually employed are the absorption and scattering coefficients and single-scattering albedo, a(λ), b(λ) and $\omega_0(\lambda)$ respectively. The absorption and scattering coefficient is defined as the absorbance and scattering per unit distance in the medium, respectively, given as:

$$a(\lambda) \equiv \lim_{\Delta r \to 0} \frac{\Delta A(\lambda)}{\Delta r} = \frac{dA(\lambda)}{dr}, \qquad b(\lambda) \equiv \lim_{\Delta r \to 0} \frac{\Delta B(\lambda)}{\Delta r} = \frac{dB(\lambda)}{dr} \qquad [\mathrm{m}^{-1}] \tag{2.18}$$

and the single scattering-albedo is given as:

$$\omega_0(\lambda) \equiv \frac{b(\lambda)}{c(\lambda)} \tag{2.19}$$

where $c(\lambda)$ is the beam attenuation coefficient defined as the sum of the absorption and scattering coefficient. ω_0 can be interpreted as the probability that a photon will be scattered rather than absorbed in any given interaction [13].

The second fundamental IOP is the volume scattering function (VSF), $\beta(\psi, \lambda)$. To define this property, two assumptions must hold. First, the medium must be isotropic, which would be reasonable for water where turbulence leads to randomly oriented particles. Second, the incoming light must be unpolarized. If these two assumptions hold, the scattering would only depend on the scattering angle ψ, and $B(\psi,\ \lambda)$ would be the fraction of the incident power scattered out of the beam through an angle ψ into a solid angle $\Delta\Omega$ centred on ψ. Lastly, the volume scattering function (VSF) is given as:

$$\beta(\psi, \lambda) \equiv \lim_{\Delta r \to 0} \lim_{\Delta\Omega \to 0} \frac{B(\psi, \lambda)}{\Delta r \Delta\Omega} = \lim_{\Delta r \to 0} \lim_{\Delta\Omega \to 0} \frac{\Phi_s(\psi, \lambda)}{\Phi_i(\lambda) \Delta r \Delta\Omega} \tag{2.20}$$

The VSF describes angular distribution of light that is scattered towards a direction, ψ, illustrated in 2.9, at a wavelength λ. The previously defined scattering coefficient, b(λ), is a measure of the overall magnitude of the scattered light, without regard to its angular distribution [13]. It therefore follow that the scattering coefficient would be the integral of $\beta(\psi, \lambda)$ over all solid angles (4π) [13]:

$$b(\lambda) = \int \beta(\psi, \lambda)\, d\Omega = 2\pi \int_0^\pi \beta(\psi, \lambda) \sin\psi\, d\psi \tag{2.21}$$

where ψ is the azimuth angle. In ocean optics, the scattering is often described by another property, known as the phase function [13]. The phase function, $\tilde{\beta}$, provides information about the shape of the VSF regardless of the scattered light intensity and is defined as the VSF normalised to the total scattering, b, given as:

$$\tilde{\beta} \equiv \frac{\beta}{b}. \tag{2.22}$$

This property is often desired of oceanic waters, and instruments have been build to measure them [13]. It can be interpreted as that if a scattering event has occurred, $\tilde{\beta}\ \mathrm{d}\psi/4\pi$ is the probability that a light beam traveling in the direction $\hat{\mathbf{\Omega}}$ is scattered into a cone of solid angle around $\mathrm{d}\psi$ the direction $\hat{\mathbf{\Omega}}'$. Here $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Omega}}'$ are unit vectors that are following the incoming direction of $\Phi_i(\lambda)$ and $\Phi_s(\lambda)$ (shown in Fig. 2.9), respectively. It follows from the unity vectors that their dot products yields, cos(ψ) = $\hat{\mathbf{\Omega}} \cdot \hat{\mathbf{\Omega}}'$. It is also usually desired to have an analytic formula that approximates the shape of an actual phase function. The Fournier-Forand phase function is commonly used in oceanography for this case to serve as such an approximation, as it is more realistic than other phase functions that have been used in the past [13].

IOPs together with initial conditions of the environment can be used to achieve radiometric variables and AOPs with radiative transfer theory.

2.2.2 Radiative Transfer Theory

Radiative Transfer Theory builds on the physical phenomena of electromagnetic radiation propagation thought different mediums, and aims to mathematically describe this propagation in terms of emission, absorption and scattering with RT equations. Quantities like energy and solid angle can be combined to describe light in terms of radiance. A combination of inherent optical properties and boundary conditions are used to formulate the RT equations that would describe this propagation. In other words, RT theory builds on boundary conditions like the physical environment of the water body, the depth and character of the water bottom, the surface waves and sky radiance to provide a connection between the IOPs and the AOPs. Consider a system where the IOPs only differ in the vertical direction denoted by z and where z increases upwards. The corresponding optical depth would then yield,

$$\tau(z) = \int_z^\infty [\alpha(z') + \beta(z')]dz' \tag{2.23}$$

where the absorption and scattering coefficient are described in 3.16 and 3.17. The diffuse radiance distribution L(z, θ, ϕ, λ) can further be described with the RT equations [13]:

$$\mu \frac{\mathrm{d}L(z,\theta,\phi,\lambda)}{\mathrm{d}\tau} = L(z,\theta,\phi,\lambda) - S(\tau,\theta',\phi',\lambda) \tag{2.24}$$

where

$$\begin{aligned} S(\tau,\theta',\phi',\lambda) = S^{*}(\tau,\theta',\phi',\lambda) - [1-\omega_0(\tau)]B(\tau) \\ -\omega_0 \int_{4\pi} \tilde{\beta}(\acute{\theta},\acute{\phi} \longrightarrow \theta,\phi) L(z,\acute{\theta},\acute{\phi},\lambda)\mathrm{d}\Omega(\acute{\theta},\acute{\phi}) \end{aligned} \tag{2.25}$$

where μ is the cosine of θ, θ is the polar angle, z is the height, L(z, θ, ϕ, λ) is the incoming radiance, ω_0 is the single scattering albedo as given in Eq. 2.19, B(τ) is the thermal radiation field given by the Planck function, ω_0 is single scattering albedo, ($\tilde{\theta}$, $\tilde{\phi}$) and (θ, ϕ) emphasizes the initial and final direction of the scattered light, respectively, with the angle between them is the same as the scattering angle ψ in Fig. 2.9 [29]. θ = 0° and θ = 180° is defined as light rays propagating towards zenith (straight up) and nadir (straight down), respectively. $\tilde{\beta}$ is the normalized VSF defined in equation 2.22. The differential vertical optical depth can be derived from Eq. 2.23:

$$\mathrm{d}\tau(z) = -[\alpha(z) + \beta(z)]dz \tag{2.26}$$

where the minus sign increases in the downward direction, whereas z increases in the upward direction. The scattering angle previously defined as ψ and the polar and azimuth angles have the following relationship:

$$\hat{\mathbf{\Omega}} \cdot \hat{\mathbf{\Omega}}' = \cos(\psi) = \cos(\theta)\cos(\theta') + \sin(\theta)\sin(\theta')\cos(\phi' - \phi) \tag{2.27}$$

S^{*}(z, θ, ϕ, λ) is an internal source term that can describe in inelastically scattered light from other wavelengths or an internal light source such as bioluminescence [4]. After solving the RT equations, other radiometric properties can be calculated from the radiance. E_d could, for instance, be calculated by integrating the radiance over all directions.

The equation would require different input parameters like absorption coefficient (a(λ)), volume scattering function (VSF), sun angle and such to produce the radiance. In other words, the RT equations would, in this case, account for the changes in radiance along the path from the sun through the atmosphere before it hits the remote sensor. Currently, there is no exact analytical solution to RT equations for a realistic medium with complex multiple scattering effects like the atmosphere, but there exists some for simplified situations like non-scattering environments. Thus, to solve these equations for photons in the atmosphere, numerical methods like Monte Carlo are required.

The MODerate-resolution atmospheric TRANsmission and radiance code (MODTRAN) is regarded as one of the most appropriated atmospheric modeling programs for hyperspectral imaging. It has a spectral resolution from 1 to $15\,\mathrm{cm}^{-1}$ and uses a high-resolution transmission molecular absorption database (HITRAN) of measured atmospheric gas properties for the calculation [13]. In addition, quantify atmospheric scattering processes by utilizing a multiple-scattering discrete-ordinate radiative transfer program for a multilayered plane-parallel medium model (DISORT) [15]. Further information as temperature, water vapor, ozone content, aerosol optical depth, and pressure are necessary for MODTRAN to simulate transmission conditions of solar radiation. A common way to solve this is to use pre-define atmospheric conditions and save them as standard atmospheric models and then use them for the simulations. This atmospheric information is something which would be very difficult to obtain for a satellite synchronized with when it would take the image [30]. In addition, to collect all this atmospheric information matching each picture would yield a lot of data to be saved.

MODTRAN solves the RT equations for light in the atmosphere, but does not solve for the light propagation underwater. Hydrolight is widely used for solving oceanographic radiative transfer problems. Typical input parameters for Hydrolight are absorption and scattering coefficients of all the constituents of the water body such as dissolved substances, mineral particles, and microbial particles as a function of depth and wavelength, the sea state, sky radiance going into the sea surface and scattering phase function. Hydrolight then solves Eq. 2.25 to obtain the radiance distribution L(z, θ, ϕ, λ) [4].

MODTRAN does, in contrary to Hydrolight, solve the radiative transfer in the atmosphere, which is perturbed by gaseous absorptive effects and scattering and aerosol scattering.

2.3 Atmospheric Effects

For remote sensing over oceans, the major contribution of the radiance measured by the satellite is atmospheric effects. The atmospheric effects are therefore crucial to handle correct with AC. Molecules in the atmosphere scatter and absorb most of the incoming sunlight, especially photons with small wavelengths. Most of the scattering are due to Rayleigh scattering, which is caused by scattering of small molecules. In the lower layer of the atmosphere, the air could also be filled with larger molecules, known as aerosols. These are more complex to predict, and can often have large variations locally in urban areas as a result of human created aerosols like dust, soot and smoke.

2.3.1 Gaseous Absorptive Effects

Photons may be absorbed both downwards towards the surface and upwards towards the remote sensor and most of the particles in the atmosphere and the ocean are also absorbers of radiation [1]. The absorption happens when the photons at certain wavelengths interact with atmospheric gases (like O_3, H_2O, O_2, CO_2, CH_4 and N_2O) and change the vibrational, rotational or electronic states due to the interaction with the photon. The vibrational changes often find a place in the NIR, whereas the rotational transition typically corresponds to weaker energies and lower frequencies located in the area microwave and SWIR. Electronic transitions occur with discrete values in the VIR and UV range and give rise to absorption and emission. This would be observable as spikes in a transmission plot as a function of wavelength and would reduce the measured TOA radiance as light would be lost to absorption. Water vapor (H_2O) (strong absorber in 5.5-7.0 μm range and

> 27 μm) and Ozone (O_3) (mostly UV range) are the main gases contributing to absorption and the concentration often depends on the time and location [31], whereas O_2, CO_2, N_2O and CH_4 can be assumed constant an uniformly mixed in the atmosphere [20]. The latter molecules have negligible absorption at the visible and NIR wavelengths relevant to ocean color remote sensing [13], which also can be seen in Fig. 2.11. There exist parts of the electromagnetic spectrum that can be transmitted through the atmosphere without absorption, known as atmospheric windows. The transmittance for a predefined atmosphere for these molecules as a function of wavelength is shown in Fig. 2.11. The figure also shows the total transmittance contribution from all the molecules, denoted as the Global transmittance and shows that a large amount of the light would be absorbed in the atmosphere, especially for some wavelengths.

The absorption bands are well known, therefore some sensors choose their bands with atmospheric windows. The MODerate Resolution Imaging Spectroradiometer (MODIS) [32] is an example of this, and is illustrated in Fig. 2.10. Here, some MODIS sensor bands are shaded in gray plotted with the transmittance of ozone and the combination of water and oxygen. One can observe that the sensor avoids the water and oxygen absorption by choosing bands with atmospheric windows. However, absorption of ozone is difficult to avoid and it illustrates that it can be difficult to avoid all the absorption bands.

Radiation may also be absorbed by aerosols, but this is not very important in the visible region. The primary attenuation to optical signals there are the scattering by aerosols.

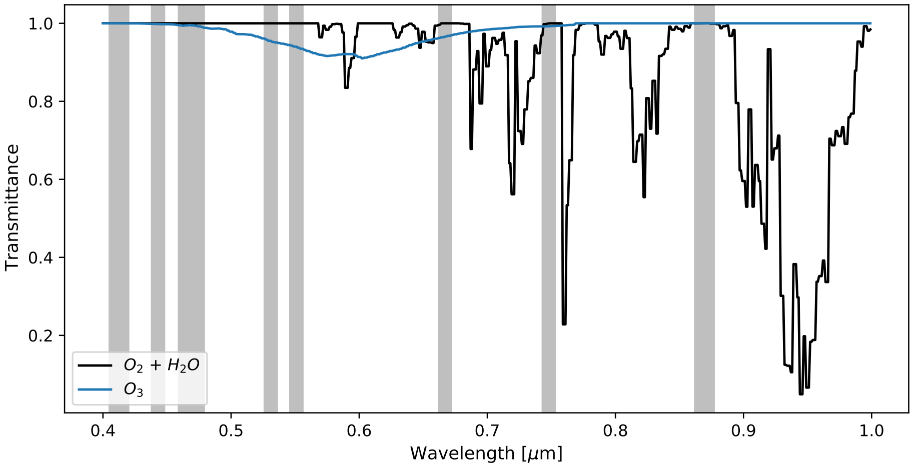


Figure 2.10: Transmittance for ozone (O_3) and combination of oxygen (O_2) and water (H_2O). The gray shaded areas are bands used by the MODIS sensor. The plot shows how MODIS avoid oxygen and water absorption by choosing atmospheric windows.

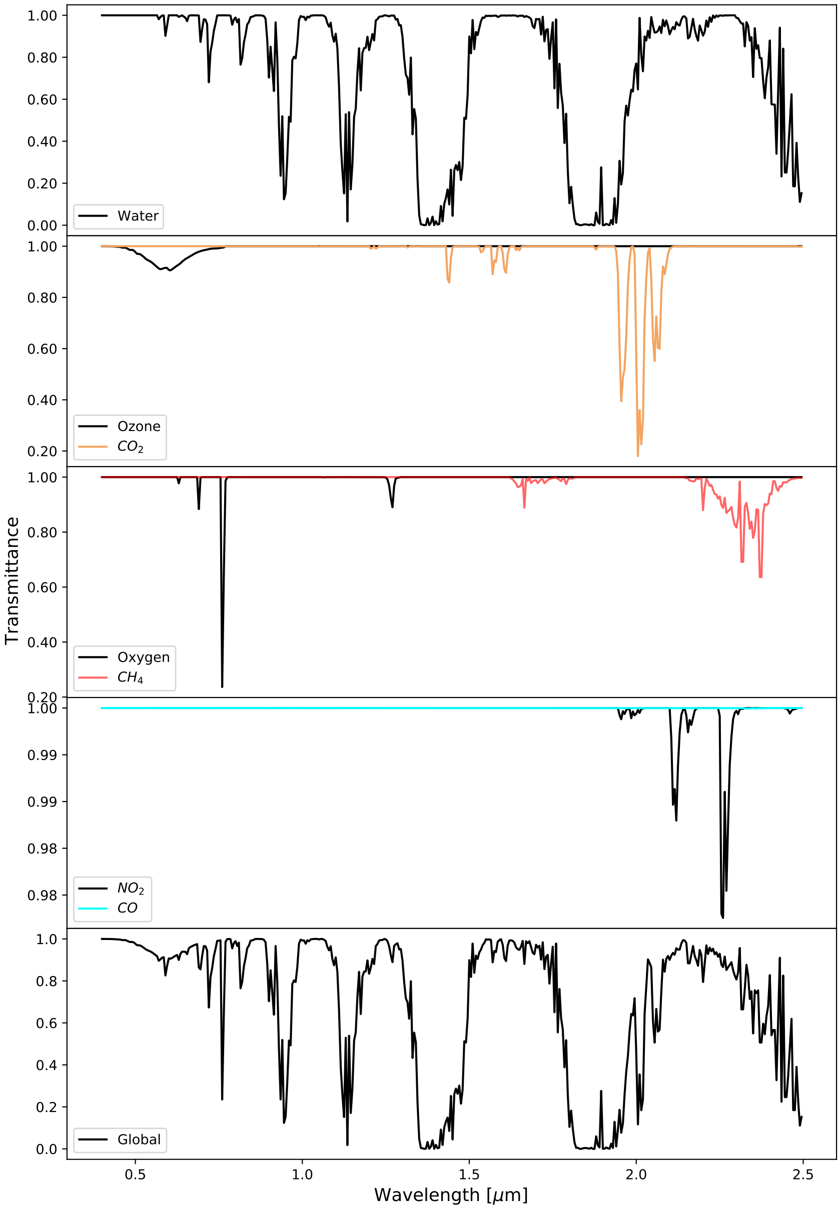
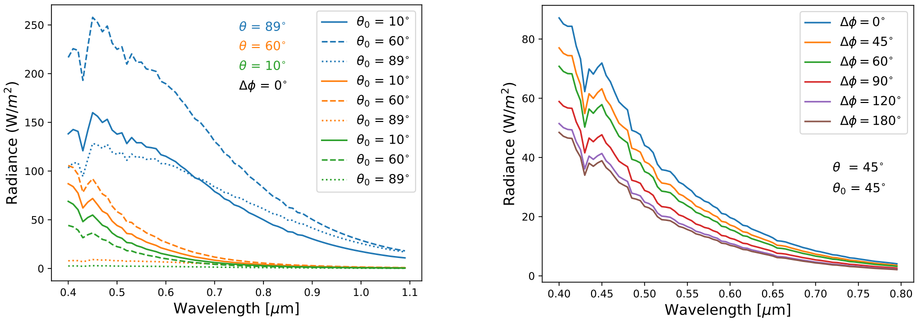


Figure 2.11: The transmittance for described atmospheric gases as a function of wavelengths from 0.4 μm to 2.5 μm. The plot is created using Py6S (Python interface of RT model 6SV) with U.S standard 1962 atmospheric model, Maritime aerosol model, Clear Water as the surface with solar zenith and solar azimuth angles to be 35 and 0°, respectively.



(a) TOA radiance for different solar and view zenith angles. $\Delta\phi$ was equal to 0° for all simulations.

(b) TOA radiance for different values of $\Delta\phi$ angles. Sun and view zenith angles were both $45°$.

Figure 2.12: Py6S simulations of TOA measured Rayleigh radiance for various combinations of sun-target-sensor geometries. Variation in sun-target-sensor angles, especially sun and view zenith angles can affect the TOA measured radiance significantly.

2.3.2 Rayleigh Scattering

Photons may also interact elastically with non-absorbing aerosols and molecules (mainly N_2 and O_2) in the atmosphere which could lead to re-direction of the original path, known as Rayleigh scattering.

The principles of Rayleigh scattering is used to calculate for scattering by particles small with respect to their wavelengths. This contributions is highly important to calculate accurately, as it constitutes a major proportion of the atmospheric radiance, especially around the VIS range [33]. This is often done by accurately model the contribution with a radiative transfer code, like MODTRAN. A set of spectral Rayleigh scattering as a function of radiant path geometry would normally be calculated and stored in Look-up-tables (LUTs). Atmospheric correction algorithms could retrieve values from the LUTs matching the geometry and parameters from the scene, using interpolation to retrieve values from the LUTs [11]. The Rayleigh contribution is heavily dependent on the sun-target-sensor angles and the altitude of the satellite. Fig. 2.12 (**a**) shows TOA radiance for different solar and view zenith angles and Fig. 2.12 (**b**) shows TOA radiance for different $\Delta\phi$ values. If the sunlight would have to propagate through much of the atmosphere before reaching the sensor, much would be scattered and result in a higher amount of path radiance. The blue lines in Fig. 2.12 (**a**) show a situation with large values for the solar and view zenith angles, and it therefore follows that the path radiance contributions are large.

2.3.3 Aerosols

Absorption by aerosols is often small [20] and remote sensors often tend to avoid the molecular absorption bands to reduce the perturbation in the TOA measured radiance. Nevertheless, the aerosols still scatter and absorb solar and terrestrial radiation. Aerosols are often more complex in coastal areas as they are highly variable both spatially and temporally where strongly absorbing aerosols like urban haze, smoke particles, desert dust, volcanic emissions, and sea salt comes with the wind from the continent [6]. Aerosols are then, unlike the Rayleigh scattering component, L_r, more difficult to predict and can not be predicted a priori [23].

Some well used AC algorithms are based on black water assumptions (totally absorbing waters) to retrieve the aerosol radiance. This assumption tends to fail in coastal areas with shallow water and highly absorbing aerosols, like dust and soot, which makes it important to consider them [33]. Coastal and inland waters may have more turbid waters and strongly-absorbing aerosols than models predict. This could break different assumptions like NIR black ocean assumption and violation of non- or weakly absorbing aerosols in coastal areas. Areas with a high amount of dust and smoke on the other hand, represent models with strongly absorbing aerosols. Fig 2.13 shows simulated aerosol, Rayleigh, and total scattering as a function of wavelength for a constructed atmospheric situation with Maritime aerosol model with clear water and the U.S. Standard 1962 atmospheric model. Both the Aerosol and Rayleigh curve are lowest for lower wavelengths. The Rayleigh transmittance is most significant for the lowest wavelengths, whereas aerosol transmittance is more important for higher wavelengths.

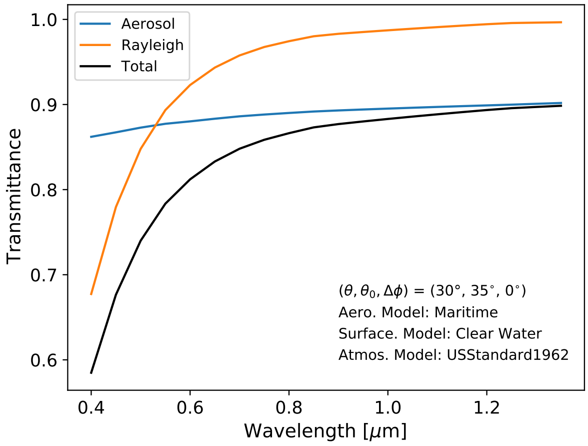


Figure 2.13: Aerosol, Rayleigh and total scattering transmittance as a function of wavelength. The simulations used U.S. standard 1962 atmosphere model, Clear Water surface model, Maritime aerosol model and solar zenith angle of 35° and $\Delta\phi = 0°$.

2.4 Ocean and Surface Effects

The atmospheric effects normally contribute to the major TOA radiance measured by the sensor. Even thought the contributions from the ocean are smaller, they must also be carefully considered when doing AC. This section will detail some of the most important ocean and surface effects and explain how they often are computed in AC algorithms.

2.4.1 Sun Glint

The specular reflection of direct sunlight by the surface is known as the sun glint and is not a part of the water-leaving radiance, thus must also be remove with AC. Some sensors have a built-in tilting function which can avoid the sun glint by tilting the sensor away from the direct reflection of the sunlight [34]. For sensors without tilting abilities, this contribution is typically calculated using the analytical Cox-Munk wind speed-wave slope distribution together with sun-target-sensor geometry [13]. Recall Eq. 2.5 where the total sun glint contribution was given as $T(\lambda)L_{sky}$. Here, L_{sky} can be rewritten as:

$$L_{sky} = F_0(\lambda)T_0(\lambda)L_{GN} \tag{2.28}$$

where $F_0(\lambda)$ and $T_0(\lambda)$ are extraterrestrial solar irradiance and atmospheric direct transmittance at the solar direction, respectively. L_{GN} is the normalized sun glint radiance and can be interpreted as the glitter radiance just above the sea surface, for an extraterrestrial solar irradiation $F_0 = 1\,\mathrm{W\,m^2\,nm^{-1}}$ with no atmosphere. The Cox-Munk model is based on the assumption that the sun glint radiance L_{sky} is a wavelength independent property computed as a function of wind speed W, given by:

$$L_{GN}(\theta_0, \theta, \phi, U) = \frac{R(\omega)}{4\,\cos^4(\beta)\,\cos(\theta)} P(Z(W)) \tag{2.29}$$

where P(Z(W)) is the Gaussian slope distribution as given by Cox and Munk [35] (Eq. 18), θ is the view zenith angle (shown in Fig. 2.5), β is the wave inclination angle. An illustration of the setup and the angles is shown in Fig. 2.14. $R(\omega)$ is the Fresnel reflectivity for unpolarized light, given by:

$$R(\omega) = \left(\frac{\sin^2(\omega - \omega^{'})}{\sin^2(\omega + \omega')} + \frac{\tan^2(\omega - \omega^{'})}{\tan^2(\omega + \omega')} \right). \tag{2.30}$$

For some sensors it is normal to mask out pixels with high values of $L_{GN}(\theta_0, \theta, \phi, U)$, as these pixels simply can be useful due to too much glint radiance compared to the water-leaving radiance. An empirical criterion is to remove pixels where $L_{GN} > 0.005\,\mathrm{sr}^{-1}$ and apply glint correction when $L_{GN} \leq 0.005\,\mathrm{sr}^{-1}$ [13]. The specular reflection of the direct sun glint will also experience atmospheric attenuation both downward towards the surface ($T_0(\theta_0, \lambda)$) and upward towards the sensor ($T(\theta, \lambda)$). Taking this into account, the glint corrected radiance, $L_t^{'}(\lambda)$, would in the end be:

$$L_t^{'}(\lambda) = L_t(\lambda) - F_0(\lambda)T_0(\lambda)T(\lambda)L_{GN} \tag{2.31}$$

where

$$\mathrm{T}_0(\theta_0, \lambda)\ \mathrm{T}(\theta, \lambda) = \exp\left\{-[\tau_R(\lambda) + \tau_a(\lambda)]\left(\frac{1}{\cos(\theta_0)} + \frac{1}{\cos(\theta)}\right)\right\} \tag{2.32}$$

where τ_R and τ_a are the Rayleigh and aerosol optical depth, respectively. Some satellite-borne sensors like SeaWiFS, OCTS, and CZCS can tilt the sensor 20° from nadir to minimize the sun glint. Another way to avoid it is to choose an appropriate time and place for when the data was obtained with a small amount of sun glint. In addition, there are different studies suggesting optimal viewing angles to avoid glint, where a viewing angle of 40° from nadir and 135° from the sun is optimal to minimize glint contribution [36].

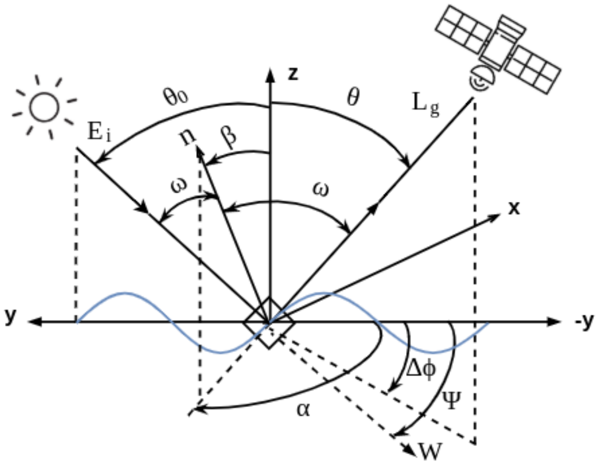


Figure 2.14: Illustration of the interaction of sunlight with ocean waves. W is pointing in the wind direction, ω is the angle between the incoming and outgoing direction of the sun rays from a normal vector (**n**) being perpendicular to the wave inclination.

2.4.2 Whitecap

The radiance contribution, $L_{wc}(\lambda)$, is the sunlight and skylight reflected from whitecaps on the surface. The two main factors to determine the whitecap radiance are the whitecap reflectance and the fraction of the sea surface covered with whitecaps (F_{wc}). The fraction can be calculated from photographs of the sea surface, but other methods are related to environmental parameters. A well-used method is based on the assumption that F_{wc} is related to the wind speed (W_{10}) [37]. Several statistical studies provide F_{wc} as a function of W_{10} measured 10 m above the sea surface. The normalized whitecap reflection can be written similarly to the normalized water-leaving radiance in Eq. 2.9, given as:

$$[\rho_{wc}]_N \equiv \frac{\pi}{F_0}[L_{wc}]_N = \left(\frac{R}{R_0}\right)^2 \frac{\pi\, L_{wc}(\theta_0)}{F_0\, cos(\theta_0)\, t_0(\theta_0)} \tag{2.33}$$

where $[L_{wc}]_N$ is the normalized whitecap radiance. $[\rho_{wc}]_N$ is independent of solar zenith angle and relative azimuth angles, and can then further be interpreted as the average reflectance of the sea surface that results from whitecaps in the absence of atmospheric attenuation [13]. There exist different functions for calculating the fraction of whitecap. One of them is given by [38]:

$$F_{wc} = 8.75 \times 10^{-5}(W_{10} - 6.33)^3 \tag{2.34}$$

where this holds for wind speeds in the range $6.33 \leq W_{10} \leq 12$ m s^{-1}. This equation builds on the assumption that the sea is not fully developed. Developed seas are most probably stormy and cloudy, thus also difficult to interpret for remote sensing [13]. The effective whitecap irradiance has been studied and measured $in\ situ$ and its mean value considered to be 0.22, albeit with ± 50 % error bars [39]. Recent studies have stated that the mean value changes with wavelength [40]. Thus, it can not be considered as spectrally flat in VIS and NIR region as previously assumed. The normalized whitecap reflectance is then given as the product of the mean reflectance value and the fraction of whitecap (F_{wc}). The final model for the normalized reflectance for whitecaps then results in:

$$[\rho_{wc}]_N(\lambda) = a_{wc}(\lambda) \times 0.22 \times F_{wc} \tag{2.35}$$

where $a_{wc}(\lambda)$ is the wavelength-dependent correction of the mean reflection value of 0.22. These values are taken from Frouin Figs. 3 and 4 [40] and are given as:

$\lambda =$	412	443	490	510	555	670	765	865
$a_{wc} =$	1.0	1.0	1.0	1.0	1.0	0.889	0.760	0.645

Fig. 2.15 shows $[L_{wc}]_N$ given by Eq. 2.35 for the different values of a_{wc} as a function of wind speed. F_{wc} is also plotted for comparison. When estimating $[L_{wc}]_N$, linear interpolation is done for wavelength values between the given values for $a_{wc}(\lambda)$.

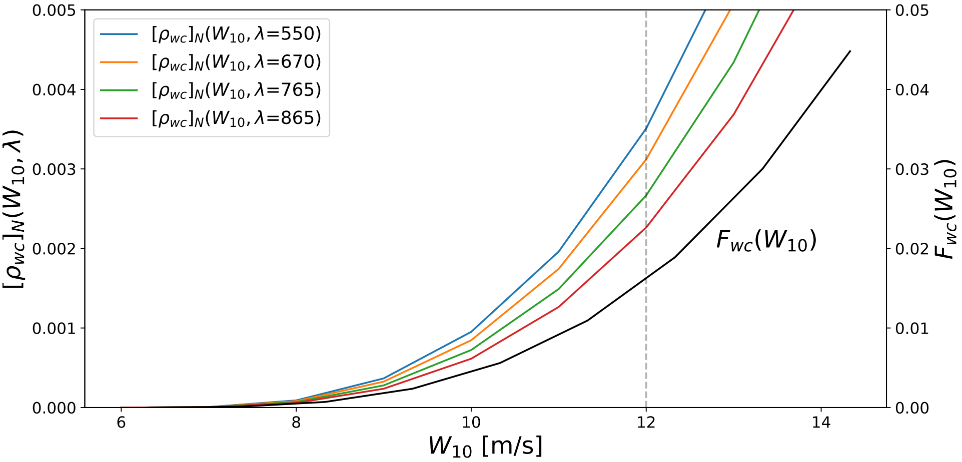


Figure 2.15: Normalized whitecap reflectance ($[L_{wc}]_N$) as a function of wind speed and wavelength given by Eq. 2.35 (colored lines), and fraction of whitecap on sea surface (F_{wc}) given by Eq. 2.34. The dashed line represents the end of the wind speed range.

2.4.3 Optical Constituents of the Ocean

The sun glint and whitecap radiances are surface dependent and do not hold any information about the oceanic constituents. The two most desired properties in ocean color remote sensing are the water-leaving radiance ($L_w(\lambda)$) and the remote sensing reflectance ($R_{rs}(\lambda)$), which both are water-leaving signals and properties of the ocean. When analyzing these properties in detail, it will be observed that several factors are influencing its shape. Direct and diffuse sunlight penetrating the water surface may not only be absorbed or scattered by the water molecules, but also by different suspended and dissolved materials present in the water [5]. The remote sensing reflectance is a property that can give valuable information about the constituents of the water and depends naturally on the constituents of the ocean which traditionally are divided into the following groups [13]: :

1. Sea water (water + inorganic dissolved materials)
2. Phytoplankton
3. Colored dissolved organic material (CDOM, also referred to as gelbstoff)
4. Organic (algae) particles and suspended inorganic (mineral) particles (also referred to as detritus)

It is important to mention that for shallow, clear waters, the remote sensing reflectance can also be affected by sunlight backscattered from the bottom of the ocean. Nevertheless, the sun rays rarely propagate the whole way down to the ocean bottom and back out again, but should be considered for very shallow clear waters. The shape of the remote sensing reflectance is therefore defined by these other constituents. The two different types of waters often addressed in ocean color, namely the Case 1 and Case 2 waters, are dependent on how the optical properties are determined by the different constituents. The definition of these cases commonly used today [41] are that Case 1 water are waters whose optical properties are determined primarily by the phytoplankton and related CDOM and detritus degradation products. In Case 2 waters the optical properties are heavily dependent on other constituents than chlorophyll such as mineral particles, CDOM and microbubbles, and where the concentrations do not covary with the phytoplankton concentration.

Fig. 2.16 shows a plot of the absorption coefficients for clear water (a_w), phytoplankton (a_{ph}), coloured dissolved organic matter (a_{cdom}), detritus/minerals (a_{dm}) and the sum of all of them (a_{tot}) for two different oceanic situations. The clear open ocean (upper plot) and coastal water (lower plot) are to illustrate the Case 1 and Case 2 waters, respectively. The plot of the Coastal water case shows how the particulate and dissolved organic matter dominates the blue and green parts of the spectrum in comparison to the clear open ocean. To date, remote sensing has focused a lot on the Case 1 waters and the major AC and IOP retrieval algorithms in use are built for Case 1 waters, and tend to fail on the complex Case 2 waters [5, 11, 33].

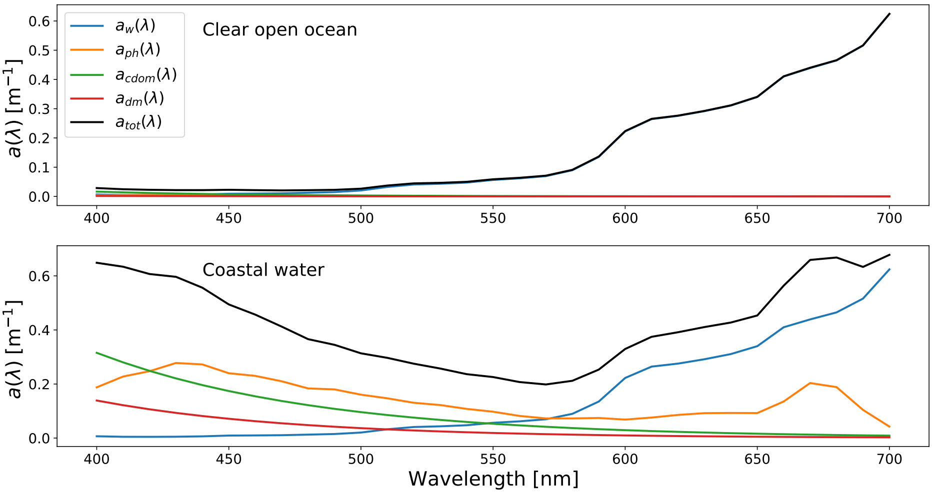


Figure 2.16: $a_w(\lambda)$, $a_{ph}(\lambda)$, $a_{cdom}(\lambda)$ $a_{dm}(\lambda)$ and $a_{tot}(\lambda)$ as a function of wavelength for clear open ocean (top) and coastal water (bottom). The absorption coefficients are extracted from the synthetic dataset from the IOCCG Report 5 [12].

Water

The water itself contributes to both absorption and scattering. When talking about water in this context, it is referred to as pure water indicating a hypothetical medium consisting of no other substances than water molecules and inorganic salts dissolved in the water [5]. In clear open oceans, the water effect on the ocean color in VIS must be taken into account, whereas for NIR it might not be present due to high absorption. The absorption of water is both temperature and salinity dependent, which must be considered when the optical properties of the water are calculated [13].

Phytoplankton

The phytoplankton category also includes other microscopic organisms, but the major influence on the remote sensing reflectance is influenced by phytoplankton. Therefore, it is more convenient to address the other microscopic organisms to this category since they would not affect the optical properties appreciable. Phytoplankton are single cell, free-floating organisms, and posses a pigment called chlorophyll that allows them to produce rich organic material by harvest sunlight through the process of photosynthesis [13]. Due to this, phytoplankton is an important component of global carbon cycle and forms the base of the aquatic food web and is often the primary reason to study ocean color. There exist thousands of phytoplankton species with different sizes, shapes, and physiologies and contains a large number of different pigments, not only chlorophyll. However, the main pigment, chlorophyll-a often serves as an index of the phytoplankton biomass and is often desired to measure. Nevertheless, phytoplankton may covary with the other substances leaving it to a complex problem to distinguish them from each other. The phytoplankton component in remote sensing context therefore also includes other microscopic organisms. This can be done as the highly pigmented phytoplankton normally dominate the signal from microscopic organisms [5]. The concentrations and species composition may change rapidly over time and space, leaving it to a difficult task to monitor the changes with small scale measurements. Satellites, however, capture pictures on a synoptic scale with comparable timescales, and therefore suits ocean color monitoring well. Chlorophyll absorbs light in the blue and red region and waters with high concentrations of phytoplankton would therefore shift towards green, accordingly, which can be observed by a satellite. However, a limitation with ocean color is that the satellites only can measure near-surface chlorophyll concentrations, which potentially could underestimate the total amount of phytoplankton present at all water depths [42].

Colored Dissolved Organic Matter

CDOM is a group of organic, dissolved substances consisting of humic and fulvic acids and is often referred to as yellow substances or gelbstoff in addition to CDOM. CDOM comes from various sources like degradation of phytoplankton cells and other organic particles [5]. In addition, they may be transported from distant regions. Rivers flowing through organic-rich soils or heavily wooded regions accumulate a load of CDOM transporting it into oceans. Human activities such as logging, agriculture, effluent discharge, and wetland drainage can also increase the amount of CDOM [43]. This property is there-

(a) Satellite picture of polluted rivers spilling into the Atlantic Ocean.

(b) Using picture from (**a**) to highlight CDOM with dark brown colors.

Figure 2.17: Landsat 8 CDOM imagery after Hurricane Florence's destruction. NASA scientists use this image to inform state and local agencies on water quality post-Hurricane Florence [44].

fore heavily dependent on the location, where for instance rural areas may lead to much higher concentrations of CDOM. In addition, locally sources of CDOM such as degradation of phytoplankton tend to accumulate more at depth than in the surface layers. CDOM transported from distant areas are more likely to be found in the upper layers, and leading to a higher impact on the ocean color measured by satellites. Due to this, it is more likely to find high, unpredictable concentrations of yellow substances in coastal areas. For open oceans, the yellow substances can also be predicted from the chlorophyll concentrations, as they tend to covary there because the major CDOM comes from degradation of phytoplankton cells. Pure water on the other hand absorbs longer wavelengths as red light, and will therefore in regions with low concentrations of CDOM appear blue. The absorption of CDOM is present in the short wavelength regions near the blue bands in VIS. Therefore areas with high amounts of CDOM will appear more brown and yellow. Fig. 2.17 shows a Landsat-8 imagery (**a**) capturing CDOM after Hurricane Florence's destruction. NASA scientists use this imagery to help inform state and local agencies on water quality post-Hurricane Florence [44]. The CDOM is highlighted as darker brown colors in Fig. 2.17 (**b**).

Inorganic Material and Minerals

Inorganic materials are also referred to as suspended materials, but refers to the inorganic suspended materials since phytoplankton and other microscopic organisms also are suspended materials. The category is a bit loose but can be summarized to include all inorganic particulate material that is not included in the phytoplankton category [5]. Examples of such are when waves and current brings up bottom sediments into suspension in shallow inland and coastal waters. These materials can then affect the ocean colors significantly. Regions with muddy rivers and estuaries, areas of large tidal excursions, and waters influenced by the outflow from rivers are regions where inorganic materials play an important role in ocean-colors. This category also includes minerals or detritus which are particles of rock as a result of erosion and weathering.

Water-leaving Radiance and Remote Sensing Reflectance

The spectral shape of both the water-leaving radiance and the remote sensing reflectance is defined by the composition of the different constituents of the water, like CDOM, minerals, chlorophyll, and water itself. These properties, therefore, consists of valuable information regarding the state of the water and are often the desired product of AC. Recall remote sensing reflectance given as $L_w(0^+, \lambda, \theta_0, \theta, \Delta\phi)/E_d(0^+, \lambda)$. The water-leaving radiance depends on the solar irradiance reaching the ocean, whereas $R_{rs}(\lambda)$ will be somewhat unaffected by this due to the division of the incoming irradiance. The amount of irradiance reaching the surface is highly affected by the solar zenith angle as this decides how long the solar rays are propagating through the atmosphere before reaching the water surface. The remote sensing reflectance will therefore to some extend remove this angular dependency. This is shown in Fig. 2.18 where the three different chlorophyll concentrations shown shapes with small variations. $R_{rs}(\lambda)$ is shown to be a good AOP due to that external environmental conditions affect the shape, while it still is very sensitive to the different IOPs [13]. $R_{rs}(\lambda)$ still refers to a particular viewing direction $(\theta, \Delta\phi)$ which could further be removed by converting the remote sensing reflectance into the normalized water-leaving reflectance ($[\rho_w]_N$). This is, however, a bit more complicated to calculate, but for RT models like Hydrolight, this can be calculated as [13]:

$$[\rho_w]_N = \pi R_{rs}(\text{HydroLight}; \theta_0 = 0, \theta = 0) \tag{2.36}$$

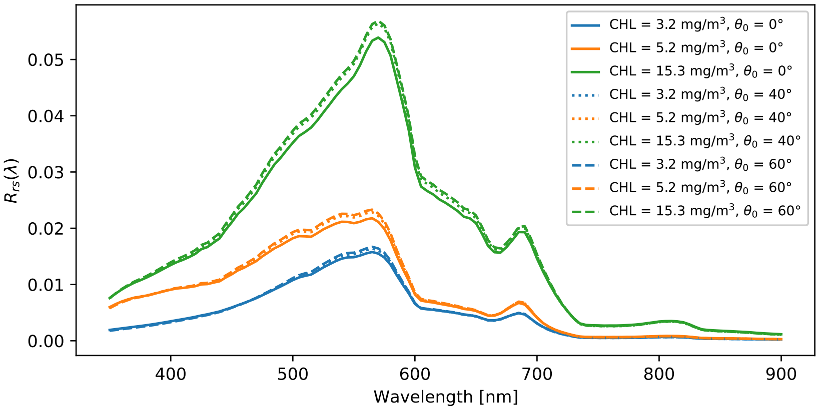


Figure 2.18: Three simulations of $R_{rs}(\lambda)$ as a function of wavelength for three different solar zenith angles, 0°, 40° and 60° and chlorophyll concentrations (CHL). The simulations are retrieved from the CCRR dataset [45].

2.5 Traditional AC algorithms and Limits

There are produced several different AC algorithms which often are categorized as either in-scene methods or model-based methods [15].

2.5.1 In-scene and Model-based Methods

The model-based methods use an RT model like MODTRAN to calculate the different radiance contributions based on some input parameters and use this to retrieve the surface or water-leaving radiance contribution. This kind of method requires several, precisely given parameters like pressure, temperature, and gaseous concentration for the calculation in order to produce accurate results. These parameters can be hard to obtain and could require someone in the field making meteorological measurements [13], even though some of these parameters often are constrained in a known range, making some of them feasible to estimate. Some sensors, like MODIS, also have specific bands that could capture information about these parameters. The other category is the in-scene method where the atmospheric properties are estimated without RT models by using a priori knowledge about the spectral nature of materials expected to be present in-scene (at ground) [15]. Some sensors are also constructed to reduce atmospheric contributions, like introducing tilting abilities to minimize sun glint and choosing a band with atmospheric windows to reduce absorption.

The computational time for RT models is high, and simulations would have to be done for all the pixels in a hypercube in order to do a full AC. The time complexity of the AC would highly depend on how complicate the atmospheric characterization is. The RT model Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV) has a rough estimation of around two seconds for one pixel at one wavelength [46]. Landsat is an American Earth observation satellite which normally creates a scene of size around 33 million pixels. This would then take more than two years to complete if each pixel would take 2 seconds [46].

The in-scene methods reduce computational time and abstract themselves from the need of $in\ situ$ measurements or atmospheric models. Several AC algorithms for ocean waters are based on the assumption that the water-leaving radiance for wavelengths larger than VIS is negligible, also known as black water assumption. Neural Networks (NN) and Machine Learning (ML) have been tested later years and can be highly accurate when provided enough training data [47, 33]. The computational time for AC when using ML approaches could be reduced significantly in comparison with RT calculations when they are trained. Machine learning models also tend to capture noise and outliers better than model-based methods [13].

2.5.2 Atmospheric Correction Algorithm Description

There are various methods and approaches for doing atmospheric corrections and further derive global ocean-color products.

AC with Black Ocean Assumption

The AC algorithms are often implemented towards a specific satellite ocean-color sensor, as the sensors are designed differently, like spectral range and spectral bandwidth. Several sensors have been produced and some are CZCS (1978), OCTS (1996), POLDER-1 (1996), SeaWiFS (1997), MODIS-Terra (1996), MODIS-aqua (2002), MERIS (2002), POLDER-2 (2002) and HICO (2009) [13], where the parentheses emphasize the launch year. These AC algorithms all use standard radiative transfer models to simulate $L_r(\lambda)$ and then remove it from the TOA radiance. SeaWiFS, MODIS-Aqua, and MERIS are still operational and have provided global ocean-color products continuously since their launch, with the use of different AC algorithms. These algorithms are designed specifically for Case-1 waters (often open ocean) and non- or weakly absorbing aerosols [12]. What differs in the algorithms is how the rest of $L_{path}(\lambda)$ is derived after removing $L_r(\lambda)$. Sunglint and whitecaps are often neglected in these methods [11]. The algorithms build on the assumption that open ocean can be considered as black, meaning that the measured TOA radiance for these wavelengths would purely consist of atmospheric contributions. This means that the contribution from $[L_w(\lambda)]_N$ would be neglected, and the rest of the path radiance can be extracted from measured TOA radiance in NIR. Further, an estimate of $L_a(\lambda)$ and $L_{ra}(\lambda)$ in VIS is made from the NIR contributions, and how this is done is the main difference of the algorithms.

Limits with Black Ocean Assumption and new Approach

The AC methods assume that the water is black in NIR, thus leaving no water-leaving radiance in the NIR [11]. This assumption holds for regions where the optical properties are purely dependent on chlorophyll, and that the other constituents covary with the chlorophyll concentration. However, this assumption breaks down for Case 2 waters where the optical properties have much more complex behavior and water-leaving radiance is found in NIR. The major percentage of the water-covered parts of the earth is classified as Case 1 water, but the fraction covering the Case 2 water is of high importance. The Case 2 waters can almost be seen as a synonym on coastal and inland areas, because the major occurrence of Case 2 waters are found here. New AC algorithm capable handling Case 2 waters are therefore needed [11].

Multilayer Neural Networks

The previously described algorithms are based on decoupling atmospheric and oceanic effects. A more recent AC approach consists of using a NN and has shown promising results [11, 33]. A typical way to use the NN is to achieve a direct mapping of the measured TOA radiance/reflectance to surface reflectance or geophysical parameters (inverse modeling). Another way to use the NN is by direct modeling, to model the radiative transfer in the ocean-atmosphere system. The pros with NN are the robustness to noise, good approximation properties and certainly the rapidity. In inverse modeling, the input parameters would be the measured TOA radiance together with solar and viewing geometries, surface pressure and wind speed in a fully, connected forward-feed NN. Different situations of Case 2 waters were tested with the NN and MERIS approach and showed substantial improvements in accuracy for the NN method compared to the MERIS algorithm [11]. On the other hand, NN requires a lot of data for tuning of the algorithm. It can be difficult to obtain large amounts of $in\ situ$ data, therefore, simulating needed data, like Rayleigh and water-leaving radiance could be a solution.

Chapter 3

Radiative Transfer Models & Atmospheric Correction Method

The main objective of this master project was to study AC for coastal areas and produce AC models capable of correcting strong aerosol containment and Case 2 waters using different ML models. This chapter will present my choice of how to acquire data for ML based on RT models, and how to generate data relevant for coastal areas with strong aerosol containment.

3.1 Atmospheric Correction Methods

There are different pros and cons when it comes to in-scene and model-based AC methods. Model-based methods can be computationally expensive and needed input can be hard to obtain. However, it can be fairly accurate when the correct inputs are given. In-scene methods try to avoid using RT models, but rest the accuracy on different assumptions about the environment that not always are correct. The gain of not using RT models and go towards statistical methods is the reduction in computational time, avoiding $in\ situ$ measurements or pre-defined models and robustness to noise [11]. As discussed in the section 2.5.2, NN has shown good results for AC, at least for multispectral data, both for Case 1 and Case 2 waters.

Applying ML to AC has become more used the recent years for several reasons. First of all, there has been a need for new AC algorithms [11]. The old models work well in open oceans, but not for coastal and inland waters, where black ocean assumption fails and the aerosol variations are high. Therefore different ML models would be interesting to test, to see how they perform based on accuracy, time complexity, and interpretability. To generalize well with ML, a lot of data from different cases would have to be obtained. However, it is difficult to achieve large amounts of ocean satellite data and corresponding metadata (like sun-target-sensor angles) together with correct and matching $in\ situ$

data for training, especially hyperspectral data. A solution to this is to use RT models to simulate a synthetic dataset for ML training and validation. This would produce as much data as desired with associated parameters like sun-target-sensor angles, wind speed, water IOPs, and aerosol properties. A similar approach with synthetic data simulation has been done by others, like [33, 48, 11]. If one would like to train ML models on $in\ situ$ satellite data, one would need both the input data and the corresponding data to predict, like water-leaving radiance or remote sensing reflectance. This data is often referred to as ground truth data, known as information provided by direct observations. When using RT models, one can use the generated data as ground truth, which can be used for both training and validation.

3.2 Radiative Transfer Discussion

An approach with data-generating based on RT models was also the desired path in this study. When choosing an RT model, some factors were important to consider as there exists plenty of RT models with a broad range of applications. First of all, the RT model needed to be relevant to the ocean-atmosphere case studied in this thesis. The RT code should also be a well-tested and validated code used in this scientific community. Advantages as user-friendliness and flexibility were also considered. Different RT models will be discussed in light of these factors.

MODTRAN is one of the most widely used RT models for different applications in Earth Observation by scientists and commercial organizations [49]. In 1987, the Air Force Research Laboratory (AFRL) started to collaborate with the research and development company Spectral Sciences Incorporated (SSI) with MODTRAN, and during the period of the last 30 years, this RT model has been extensively validated, both against $in\ situ$ measurements and other RT models [50]. Despite being regarded as one of the most appropriated atmospheric modeling programs for hyperspectral imaging, it does not solve RT for light propagation underwater, which is necessary for this study. Several other RT models like SBDART, Stream and LibRadTran are other examples of such typical RT models also only meant for atmospheric applications with no coupling to an underlying water body [51].

On the contrary, Hydrolight is an RT code widely used for solving exactly what these RT models do not solve, namely oceanographic radiative transfer problems [52]. This model, however, is a good tool for marine optics applications only, as the model serves no connection to an atmosphere. Nevertheless, it has been shown that combining two RT models like MODTRAN and Hydrolight can work for producing in-water and atmospheric radiances both just above the sea surface and at the top of the atmosphere [13]. The radiance propagating thought an atmosphere specified in MODTRAN could serve as an input as incoming radiance towards a water body specified in Hydrolight. This would, however, be a bit more complicated than for a code with a build-in coupled ocean-atmosphere system, as the connection and interface between the two models must be carefully considered.

6S is a radiative transfer model used in the MODIS atmospheric correction algorithm by calculating LUTs for the AC algorithm [53]. The model is based on the methods of successive orders of scatterings approximation (SOSA) for calculating Rayleigh and Aerosol scattering. 6SV is the vector version of 6S and can account for radiation polarization, which studies have shown can lead to an error up to $10\,\%$ between simulated and $in\ situ$ [54]. Besides, there exists a user-friendly Python interface of the 6SV code [26]. 6SV, in contrary to MODTRAN, can take into account ocean effects, as well as counting for the atmospheric effects, such as Rayleigh and aerosol contributions. MODTRAN is a purely atmospheric effect with a given surface with wavelength-dependent albedo. 6SV has a homogeneous ocean setup that can initialize a water surface parameterized by four parameters; wind speed, wind azimuth angle, salinity, and pigment concentration [26]. 6SV, therefore, seems to be constructed for the Case 1 waters.

A third RT model is AccuRT, which is a coupled ocean-atmospheric RT code which also can include effects from Case 2 waters. There are few reliable, well-tested, and user-friendly RT tools for coupled ocean-atmosphere systems available. AccuRT was made to provide a reliable, robust, versatile, and easy-to-use model for exactly this purpose [51]. AccuRT fits the problem of the thesis perfect, the user-friendliness is good and it is possible to parameterize aerosol, atmospheric, and ocean setups for a wide range of IOPs. AccuRT will, therefore, be the RT model used to generate hyperspectral data for this study.

3.3 AccuRT Method

AccuRT provides a coupled atmosphere-water system where the atmosphere and ocean further will be referred to as the upper and lower slab, respectively. Each of the two slabs is divided into a sufficiently large number of homogeneous horizontal layers to adequately resolve the vertical variation in its IOPs [29]. On the surface interface between the two slabs, Fresnel's equations and the law of reflection and Snell's Law are applied to determine the direction of reflected and refracted light. The interface between the upper slab and lower slab is assumed to be flat in the version of AccuRT used, but a wind-roughened water interface will be included in an upcoming version of it [29]. This means that whitecap (L_{wc}) and strong direct sun glint (L_{sun}) radiances are not included in the AccuRT simulations. The AccuRT setup is illustrated in Fig. 3.1. Discrete ordinate (DISORT) solutions to the RT equation are computed separately for each layer in the two slabs [29]. One can further include different input files describing atmosphere, aerosol and ocean properties often in terms of different IOPs. When all necessary input files are defined, RT simulations can be done to provide radiative quantities such as irradiances and radiances in the wavelength range between $280\,\mathrm{nm}$ and $4000\,\mathrm{nm}$. Irradiance and radiance (for given azimuth and polar angles) can be obtained at any location in the specified atmosphere or water. For a more detailed description of AccuRT, both when it comes to details around the physics and how to use it, see AccuRT review [29] and user manual "AccuRT-User-Manual_VM_v2.pdf" which can be found on the book webpage Stamnes et al. (2017) [1]. The physics and interpretations used in AccuRT most relevant to the data generations done in this thesis will be presented in the following subsections.

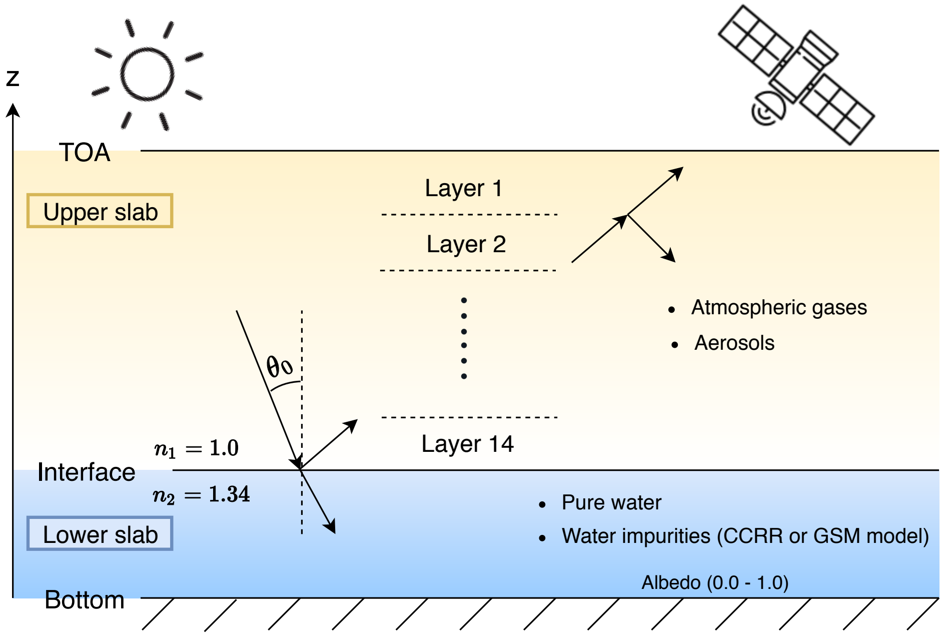


Figure 3.1: Illustration of the AccuRT setup with horizontal layers, upper and lower slap and surface interface. Various layer materials like atmospheric gases, aerosols, bottom albedos, pure water and water impurities can be applied to the model.

AccuRT will be used to generate simulation values of TOA radiance (L_t), Rayleigh and absorption corrected TOA radiance (L_{rac}), water-leaving radiance (L_w) and remote sensing reflectance (R_{rs}). The reason why L_{rac} is generated is that both the Rayleigh and absorption properties are quite standard and well known. A common procedure in standard AC algorithms is to retrieve the Rayleigh radiance from pre-calculated LUTs based on sun-target-sensor angels [11]. The absorption bands are also well known and would often also mainly depend on the sun-target-sensor angels.

To run the simulations, several input configurations must be specified before AccuRT can start to generate the data. The RT code and the physics behind it must, therefore, be closely investigated, to chose input values representative for actual atmospheric and oceanic environments. The different components of a coupled ocean-atmosphere system will be described further in the following subsections.

3.3.1 Atmosphere

AccuRT uses a stratified vertical structure defined by intensive properties of an atmosphere in hydrostatic balance [29]. These properties, also known as bulk properties, does not depend on the amount of material in the system or the system size. This balance is maintained as a result of countervailing pressure and gravitational forces. By equating these forces, one finds that:

$$dp(z) = -g\rho(z)dz \tag{3.1}$$

where g is the gravitational acceleration, assumed constant, p(z) is the air density, and dp is the differential change in pressure over small height interval dz. Further, we have the ideal gas law given by:

$$p(z) = \frac{NkT(z)}{V(z)} \tag{3.2}$$

where N is the number of molecules in the volume of gas at height z, V(z) at temperature T(z) and k is Boltzmann's constant. Combining (3.1) and (3.2), one finds $\frac{dp(z)}{p(z)} = -\frac{dz}{H(z)}$, where H(z) = kT(z) / $\overline{M}$g is the atmospheric scale height, $\overline{M}$ is $\rho(z)$ / n(z) and n(z) = N / V(z) [29]. A further integration of this equation would yield:

$$p(z) = p(z_0)\ \exp\left[-\int_{z_0}^{z} dz'/H(z')\right] \tag{3.3}$$

This equation can further be converted to a similar expression for the density $\rho(z)$ and concentration $n(z)$ by using the ideal gas law. The bulk properties for any height z can, therefore, be calculated with the knowledge of the surface pressure $\rho(z_0)$ and the variation of the scale height H(z). This applies to well-mixed gases in the atmosphere, but not to water which undergoes phase changes on short time scales and short-lived species such as ozone.

For the data generation, 14 different layers were used with the top of the upper slab equal to $100\,\mathrm{km}$ and the bottom (interface between ocean and atmosphere) equal to $0.0\,\mathrm{km}$. 14 layers in between the top and bottom were specified and found at 2, 4, 6, 8, 10, 12, 16, 20, 24, 30, 40, 50 and $70\,\mathrm{km}$, all with refractive index equal to 1.0. This atmosphere setup was also the default setup provided by AccuRT.

Gases in the Earth's Atmosphere

In AccuRT, the user is allowed to choose between six atmosphere models described by Anderson et al. (1986) [55]. These atmospheric constituent profiles are US Standard atmosphere 1976, tropical, midlatitude summer/winter, and subarctic summer/winter. The models contain profiles of temperature, pressure and concentrations of several gaseous constituents like H_2O, CO_2, O_3, CH_4 and NO_2 based on the best data available when they were published [29]. These gaseous constituents are the most important as the atmospheric gaseous absorption in the solar spectrum principally is due to these molecules. A spectral LOWTRAN/MODTRAN band model [56] is used to compute the IOPs of the atmospheric gases like the absorbing coefficients. The computation of the molecular, Rayleigh scattering coefficients is also included. The phase function describing the angular distribution of scattered unpolarized light is given by

$$\tilde{\beta}(\cos(\psi)) = \frac{3}{3+\mathrm{f}}(1 + \mathrm{f}\,\cos^2(\psi)) \tag{3.4}$$

with parameter f = $\frac{1-\rho}{1+\rho}$ where ρ is the depolarization ratio, included to account for the anisotropy of the molecules [29]. ρ = 0.0286 at $500\,\mathrm{nm}$ was used in line with [57]. The model used for the simulations is the US Standard atmosphere 1976 and its atmospheric constituent profile can be found in Appendix B.

3.3.2 Aerosol

When the size distribution and the refractive index of the aerosol particles are known, different aerosol models can be used to generate aerosol specific IOPs. It has been shown that atmospheric aerosols tend to have a bimodal distribution, with two types of particles, fine and coarse [29]. The smaller particles are referred to as the fine mode aerosols, with radii between 0.1 and $0.25\,\mu\mathrm{m}$. The larger, coarse mode particles are generally between 1.0 and $2.5\,\mu\mathrm{m}$. The volume size distribution used in AccuRT is an adopted bi-modal log-normal distribution given by [29]:

$$v(r) = \frac{dV(\ln r)}{dr} = \frac{1}{r}\frac{dV(\ln r)}{d\,(\ln r)} = \sum_{i=0}^{2} \frac{V_i}{\sqrt{2}\,\pi\sigma_i}\frac{1}{r}\mathrm{exp}\left[-\left(\frac{\ln r - \ln r_{vi}}{\sqrt{2}\sigma_i}\right)^2\right] \tag{3.5}$$

where subscript i represents the mode, either coarse or fine, V_i is the total volume of particles with mode i, r_{vi} is the volume geometric mean radius (or mode radius) and σ_i is the geometric standard deviation. It further follows:

$$\int_0^\infty \frac{dr}{\sqrt{2}\,\pi\sigma_i}\frac{1}{r}\mathrm{exp}\left[-\left(\frac{\ln r - \ln r_{vi}}{\sqrt{2}\sigma_i}\right)^2\right] = 1 \tag{3.6}$$

which yields after integrating over all sizes for both modes:

$$\int_0^\infty v(r)dr = V_f + V_c = V \tag{3.7}$$

where V is the total volume of fine (V_f) and coarse (V_c) particles. The volume fraction of fine mode particles, known as the aerosol fine mode fraction, would therefore be f_v = V_f / V. The equation can further be formulated in terms of number density:

$$n(r) = \frac{dN(r)}{dr} = \frac{1}{r}\frac{dN(r)}{d(\ln r)} = \sum_{i=0}^{2} \frac{N_i}{\sqrt{2}\,\pi\sigma_i}\frac{1}{r}\mathrm{exp}\left[-\left(\frac{\ln r - \ln r_{ni}}{\sqrt{2}\sigma_i}\right)^2\right] \tag{3.8}$$

where N_i is the number of particles and r_{ni} is the mean geometric radius. These two variables are related to V_i and R_{vi} as follows:

$$\ln r_{ni} = \ln r_{vi} - 3\sigma_i^2 \tag{3.9}$$

$$N_i = \frac{V_i}{\frac{4}{3}\pi r_{ni}^3} \exp(-4.5\sigma_i^2) \tag{3.10}$$

In the same manner as for Eq. 3.6, the total number of particles (N) can be found by integrating the number density function:

$$\int_0^\infty n(r)\, dr = N_f + N_c = N \tag{3.11}$$

where N_f and N_c is the number of fine and coarse particles, respectively. Further, the relative humidity (RH), which is the ratio of the maximum humidity that the air can handle versus the actual humidity or water vapor that is currently present in the air, will affect the size of the aerosol particles in addition to the refractive index. It is been shown that the radii of both the fine and coarse aerosol particles increase with RH, and also that the rate of increase is larger for RH > 70 %. For really large values of RH (> 95 %), the rate of increase becomes almost exponential [58]. The radius of the particle would be a function of RH from the wet-to-dry mass ratio given by:

$$r(a_w) = r_0 \left[1 + \rho \frac{m_w(a_w)}{m_0} \right]^{1/3} \tag{3.12}$$

where $m_w(a_w)$ is the mass of condensed water, m_0 is the dry particle mass where RH = 0, ρ is the particle density relative to that of water. a_w is the water activity of a soluble aerosol at radius r [µm] and can be expressed as:

$$a_w = \text{RH} \exp \left[\frac{-2\sigma\, V_m}{R_w\, T} \frac{1}{r(a_w)} \right] \tag{3.13}$$

where σ is the surface tension on the wet surface, V_m is the specific volume of water, R_w is the water vapor gas constant and T is the absolute temperature [K]. At the same time, the refractive index changes with RH and can be determined from

$$\tilde{m}_c = \tilde{m}_{c,w} + (\tilde{m}_{c,0} - \tilde{m}_{c,w}) \left[\frac{r_0}{r_{RH}} \right]^3 \tag{3.14}$$

where $\tilde{m}_{c,w}$ and $\tilde{m}_{c,0}$ are the complex refractive indices of water and dry aerosols respectively. r_0 and r_{RH} are the radii of the aerosols in the dry state and at a given RH, respectively. The magnitude of the particle growth and change of the refractive index depends on the water uptake of the aerosol (the ratio $m_w(a_w)/m_0$ in Eq. 3.12) and also on the size r_0 of the dry aerosol.

In AccuRT, the user must specify in which layers the aerosols are found together with the refractive index of the coarse and fine mode in addition to the effective radii and the variances. Lastly, the fine mode fraction (f_s), RH, and aerosol volume fraction (f_v) must be specified. The latter property is given by:

$$f_v \equiv \int_{r_{min}}^{r_{max}} \left(\frac{4\pi}{3}\right) r^3 n(r) dr = \mathrm{AMC}/\rho_a \tag{3.15}$$

where AMC is aerosol mass content (with unit $\mathrm{kg/m^3}$) and r_{min} and r_{max} are the minimum and maximum radius of the aerosols.

3.3.3 Bio-optical model for water

In the ocean-color community, the term "bio-optical model" may refer to at least two different meanings. In this study, a bio-optical model will refer to models that are based on various fundamental theories of optics together with a set of rigorous equations, which can be used to analyze, and then predict the optical properties of biological materials, such as phytoplankton [59]. For many oceanic environments, the optical properties of water bodies are essentially subordinated to the biological activity. For the Case 1 waters, the optical properties can be derived from the phytoplankton and their derivatives. However, bio-optical models purely derived from chlorophyll properties are insufficient for Case 2 waters. For these cases, the bio-optical models must be supplemented by accounting for the presence of mineral and organic sediments, and CDOM, not covarying with the algal biomass [59].

The two main IOPs employed in optical oceanography are often the total absorption coefficient $a_t(\lambda)$ and the total scattering coefficient $b_t(\lambda)$ [13]. Normally, the main contributions are from pure water, phytoplankton (embedded algal) and minerals (non-algal) particles and colored dissolved organic matter (CDOM) and can be considered as a sum of these partial contributions [29]:

$$a_t = a_w(\lambda) + a_{ph}(\lambda) + a_{dm}(\lambda) + a_{cdom}(\lambda) \tag{3.16}$$

$$b_t = b_w(\lambda) + b_p(\lambda) \tag{3.17}$$

where $a_w(\lambda)$ and $b_w(\lambda)$ are the parts of the total absorption and scattering coefficients from pure water, $a_{ph}(\lambda)$ is the absorption coefficient due to algal particles, $a_{cdom}(\lambda)$ is the absorption coefficient due to colored dissolved organic matters, and $a_{dm}(\lambda)$ is the absorption coefficient due to detritus/minerals (non-algal particles), and $b_p(\lambda)$ is the scattering coefficient due to embedded particles, including both algal and non-algal particles. The model used to construct IOPs for water in AccuRT is based on a bio-optical model used in the CoastColour Round Robin (CCRR) effort [60].

CoastColour Round Robin

The CCRR project was founded by the European Space Agency (ESA) to test various algorithms that retrieved water quality parameters with a variety of reference datasets. This to help end-user of remote sensing products to select the most accurate algorithm for their coastal region [45]. The CCRR bio-optical used in the CCRR effort served as a useful approximation for turbid areas observed in estuary areas [29]. The CCRR model consists of the three input parameters CHL, MIN and $a_{cdom}(443)$, in addition to predefined absorption and scattering coefficients from pure water, where $a_{cdom}(443)$ is the absorption coefficient for CDOM particles at $443\,\mathrm{nm}$ and CHL and MIN are chlorophyll and mineral concentration in $[\mathrm{g/m^3}]$ and $[\mathrm{mg/m^3}]$, respectively. The component including the mineral particles can also include non-algae particles not covaring with the algae particles. The optical model is described in the document "Coastcolor-RRP-V1.2.doc" [60], and the most relevant parts are further described in the following subsection.

Pure Water

For the setup of pure water in AccuRT, the real part of the refractive index of pure water from Segelstein (1981) [61] was used. In addition, the absorption coefficients $a_w(\lambda)$ used are the data published by Smith and Baker (1981) [62]. The scattering coefficients $b_w(\lambda)$ are based on the data published by Morel [63]. The absorption and backscattering coefficients for pure water as a function of wavelength are shown in Fig. 3.3 and Fig. 3.4, respectively.

Mineral Particles

The absorption coefficient for minerals with spectral variation is given by [64]:

$$a_{\text{dm}}(\lambda) = a_{\text{dm}}(443)\,[\exp(-0.0123(\lambda - 443))] \tag{3.18}$$

where λ is wavelength in nm and $a_{\text{dm}}(443)$ is the absorption coefficient for minerals at $443\,\mathrm{nm}$. This property is further given by:

$$a_{\text{dm}}(443) = 0.041 \times 0.75 \times \text{MIN}. \tag{3.19}$$

The spectral variation in the attenuation coefficient is given by:

$$c_{\text{dm}}(\lambda) = c_{\text{dm}}(555) \times (\lambda/\lambda_0)^{-c}; \qquad c = 0.3749, \qquad \lambda_0 = 555\,\mathrm{nm} \tag{3.20}$$

where $c_{\text{dm}}(555)$ is the sum of the absorption and scattering coefficients at the same wavelength, where $b_{\text{dm}}(555)$ is given in similar manner as for $a_{\text{dm}}(443)$ by:

$$b_{\text{dm}}(555) = 0.51 \times \text{MIN} \tag{3.21}$$

$c_{\text{dm}}(555)$ can further be calculated by summing the absorption and scattering coefficients:

$$\begin{aligned} c_{\text{dm}}(555) &= a_{\text{dm}}(555) + b_{\text{dm}}(555) \\ &= [0.041 \times 0.75 \exp(-0.0123(555-443)) + 0.51] \times \text{MIN} = 0.52 \times \text{MIN} \end{aligned}$$

Lastly, the spectral variation of the scattering coefficient follows from:

$$b_{\text{dm}}(\lambda) = c_{\text{dm}}(\lambda) - a_{\text{dm}}(\lambda) \tag{3.22}$$

The absorption and backscattering coefficients as a function of wavelength are shown in Fig. 3.3 and Fig. 3.4, respectively.

Colored Dissolved Organic Matter (CDOM)

The absorption by CDOM is given in the same manner as for the minerals:

$$a_{\text{cdom}}(\lambda) = a_{\text{cdom}}(443) \times \exp[-0.0176(\lambda - 443)] \tag{3.23}$$

where $a_{\text{cdom}}(443)$ is one of the inputs in the CCRR bio-optical model. The absorption coefficient as a function of wavelength is shown in Fig. 3.3.

Algae Particles

The correlation between the absorption coefficient for phytoplankton and its concentration is studied by Bricaud et al. (1998) and is formulated as:

$$a_{ph}(\lambda) = A_\phi(\lambda) \times [\text{CHL}]^{E_\phi(\lambda)} \tag{3.24}$$

where $A_\phi(\lambda)$ and $E_\phi(\lambda)$ are numerical coefficients, and CHL is the chlorophyll concentration. The spectral variation in the numerical coefficients are shown in Fig. 3.2.

As for the extinction coefficients for CDOM and minerals, the absorption coefficient for pigmented particles follows a similar manner, where in this case, the attenuation coefficient for pigmented particles at $660\,\text{nm}$ is used to calculate the spectral variation in the attenuation coefficient which is given by:

$$c_{ph}(\lambda) = c_{ph}(660) \times (\lambda/660)^\nu \tag{3.25}$$

where

$$\nu = \begin{cases} 0.5 \times [\log_{10} \text{CHL} - 0.3] & 0.02 < \text{CHL} < 2.0 \\ 0 & \text{CHL} > 2.0. \end{cases}$$

and $c_{ph}(660)$ is given by Loisel and Morel (1998) [66]:

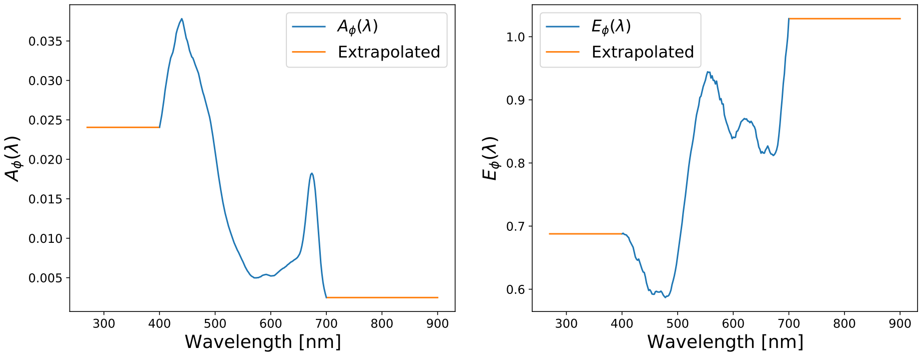


Figure 3.2: The numerical coefficients $A_\phi(\lambda)$ and $E_\phi(\lambda)$ as a function of wavelength, where blue and orange color refers to the original data provided by Bricuad et al. (1998) [65] and extrapolated values, respectively.

$$c_{ph}(660) = \gamma_0 \times [\text{CHL}]^{\eta}; \qquad \gamma_0 = 0.407; \qquad \eta = 0.795 \tag{3.26}$$

Lastly, the spectral variation of the scattering coefficient for pigmented particles follows from the difference in:

$$b_{ph}(\lambda) = c_{ph}(\lambda) - a_{ph}(\lambda) \tag{3.27}$$

It therefore follows that the total absorption and scattering coefficient for impure water are:

$$a_{\text{tot}}(\lambda) = a_{\text{dm}}(\lambda) + a_{\text{ph}}(\lambda) + a_{\text{cdom}}(\lambda) \tag{3.28}$$

$$b_{\text{tot}}(\lambda) \equiv b_{\text{p}}(\lambda) = b_{\text{dm}}(\lambda) + b_{\text{ph}}(\lambda) \tag{3.29}$$

The main ideas and models used in AccuRT are now presented. The input parameters are still to be specified by the user.

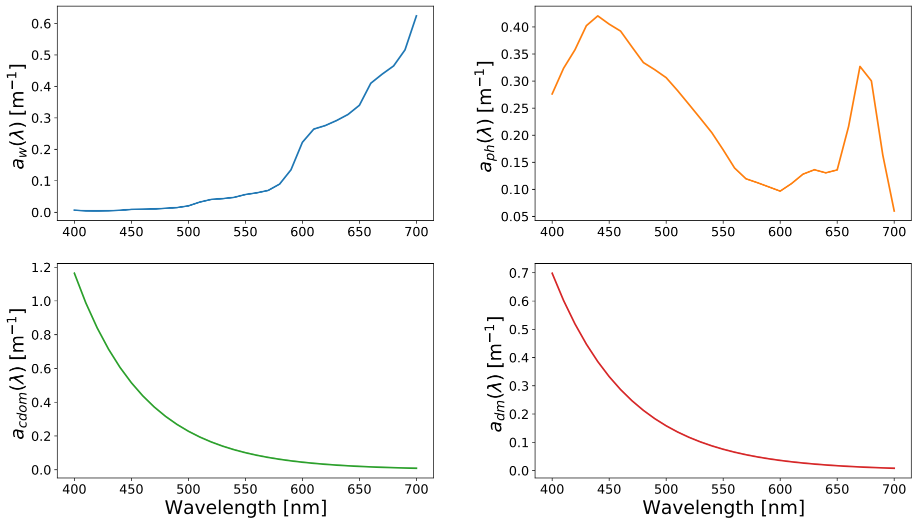


Figure 3.3: The absorption coefficients for pure water ($a_w(\lambda)$), phytoplankton pigments ($a_{ph}(\lambda)$), colored dissolved organic matter ($a_{cdom}(\lambda)$) and detritus/minerals ($a_{dm}(\lambda)$) as a function of wavelength.

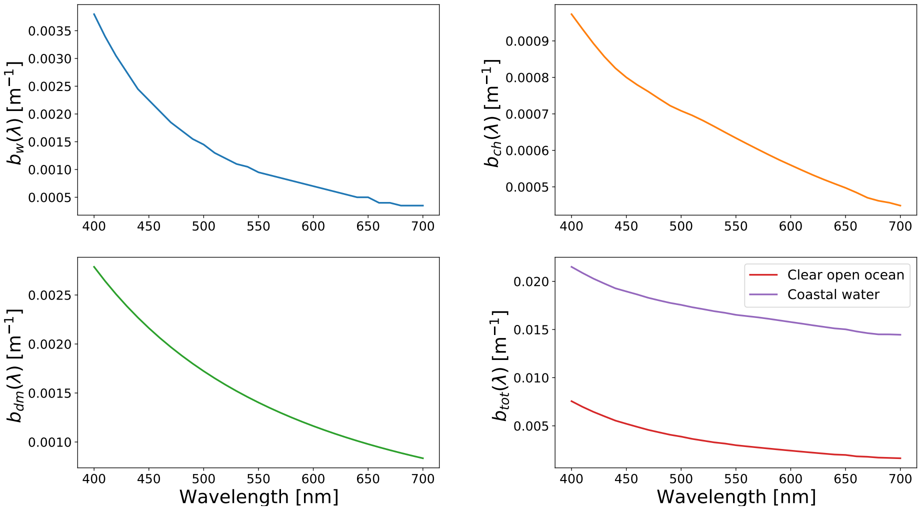


Figure 3.4: The backscattering coefficients for pure water ($b_w(\lambda)$), phytoplankton pigments ($b_{ph}(\lambda)$), detrius/minerals ($b_{dm}(\lambda)$) and total ($b_{tot}(\lambda)$) as a function of wavelength. The plot of the total backscattering shows two cases, one clear open ocean and coastal water .

3.4 Input Parameters for Data Generating with AccuRT

To run simulations with AccuRT, various input parameters used to describe the ocean, atmosphere and sun-sensor-geometries must be specified. To set up a coupled ocean-atmosphere environment as similar to actual ocean and atmospheric conditions as possible, each parameter must be studied and the input range must be chosen carefully. The useful spectral data for ocean-color lies in the wavelength region around 400-$800\,\mathrm{nm}$. This will also be the chosen wavelength region for the generated data in this study. Also, it is desired that the data generated would be hyperspectral. Therefore spectral data in the given wavelength region with a $5\,\mathrm{nm}$ spectral resolution will be used and will result in 81 bands.

3.4.1 Spectral Input and Resolution

To simulate the radiance/irradiance values for various depths, an input beam irradiance must be specified. One can define an own wavelength dependent input, or use a default solar spectrum defined in AccuRT. The default solar spectrum was used for the simulations in this study as this would be the closest to what would actually reach the atmosphere. The default spectra used in AccuRT are different input solar spectra for different wavelength regions from different sources, some used in other well known RT models like MODTRAN. TOA spectrum from ATLAS3, ATLAS2, MODTRAN (v3.5) and ASTM G173-03 are used for respectively $200 < \lambda < 407\,\mathrm{nm}$, $407.8 < \lambda < 419.9\,\mathrm{nm}$, $419.9 < \lambda < 800\,\mathrm{nm}$ and $\lambda > 800\,\mathrm{nm}$. The spectral resolution between 200 and $800\,\mathrm{nm}$, which is the interesting range for this project, is $0.05\,\mathrm{nm}$ [29]. The total input solar spectrum for wavelength between 300 to $900\,\mathrm{nm}$ is shown in Fig 3.5. The different colors represent the different sources

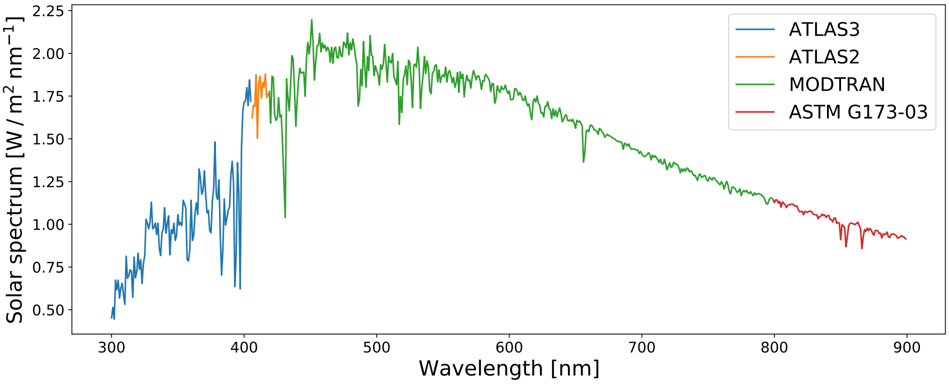


Figure 3.5: Input solar spectrum at the TOA used in AccuRT. The different colors represent from where the TOA spectrum is adopted from, which are ATLAS3, ATLAS2, MODTRAN (v3.5) and ASTM G173-03.

3.4.2 Sun-target-sensor Geometries

The solar and sensor zenith angles (θ_0, θ) can both maximally achieve 90° in the simulations. For airborne instruments, the flight path can be chosen to minimize glint where paths towards or away from the sun with solar zenith angles of 30°-60° are recommended. In addition, numerical simulations of radiative transfer suggest that a viewing angle of 40° from nadir and 135° from the sun is optimal for avoiding glint [36]. The range for solar and view zenith angles were chosen to be 0°-65° and 0°-70°, respectively, to include these desired angle ranges.

The last angle is the detector azimuth angle which can range from -180 ° to 180°. The radiance however, is symmetric around 0 so it is only the difference in azimuth angles that counts. Therefore this property can be named the relative azimuth angle ($\Delta\phi$). AccuRT will calculate the same result for -10 ° and 10 °, so the detector azimuth angle was in the range 0 to 180 °.

3.4.3 Aerosol Input

The AOD at $500\,\mathrm{nm}$ and $869\,\mathrm{nm}$ are often used in order to describe the aerosol contribution in remote sensing. To cover a large range of aerosol situations, it would be desired to use AOD(869) to be from 0.0001 to 0.4 [33]. The aerosol model used in AccuRT, however, do not let the user specify AOD(869) or AOD at any other wavelength as input. However, AOD at a desired wavelength can be calculated after the simulations are done with AccuRT, based on the output. Therefore, the predefined aerosol model described in 3.3.2, was used in various simulations with different values for volume fraction of aerosols (f_v), fraction of fine and coarse aerosols (f_s) and relative humidity (RH) as input. The AOD(869) was calculated afterwards and the desired AOD(869) range could be chosen based on the different ranges of volume fraction, fine fraction and RH. Nevertheless, some of this ranges could already be determined based on expected values for $in\ situ$ measurements of f_v, f_s and RH.

RH over coastal areas tends to be higher than over non-coastal areas. During daytime, a lot of evaporation occurs on the surface of the water body. The land is also getting warmer than the ocean, and as a result, heavy humid ocean air moves from the ocean towards coastal areas which is close to land areas where the air density is low. Observation of RH over the West Florida Continental Shelf was done over a 5 year period from 1998-2003 [67]. The monthly mean values of RH was around 75 % throughout the year, but with synoptic and daily variations between seasons, especially winter and summer seasons. June measurements of RH were normally between 70-80 %, whereas January measurements could vary between 40-100 %.

For the aerosol model described in section 3.3.2, the increase rate becomes exponential when RH goes towards $100\,\%$. Therefore the largest values of RH were omitted and RH range chosen for the simulations was between 30-95 $\%$. The fraction of fine and coarse aerosols can vary between 0 and 1, i.e. from a system with only fine to a system with only coarse aerosols. The input to the simulations was chosen to be from 0 to 1 with step size equal to 0.1.

The last parameter to determine was the volume fraction of aerosols to include. As discussed, this range would be chosen based on the desired range of AOD(869) chosen. To chose the range for the volume fraction of aerosols, various simulations with AccuRT were done with the extreme values for RH (30 and 95 $\%$) and f_v (0 and 1) and different values of aerosol volume fraction. AccuRT would calculate the scattering and absorption coefficient for each layer and each material in it. The aerosol fine and coarse mode absorption and scattering coefficient were therefore extracted and used to calculate the AOD for a given reference wavelength as the AOD was given by [68]:

$$\tau_m = c_m h = [f_s c_s + (1 - f_s) c_l]\,\mathrm{h} \tag{3.30}$$

where τ_m is the AOD, c is the extinction coefficient given as the sum of the absorption and scattering coefficients, h is the geometrical thickness of the layer, and the subscripts m, s, and l stands for mixture, small and large aerosol particles. 12 simulations with AccuRT for different combinations of volume fraction, fine mode fraction, and RH are shown in Fig. 3.6. The figure shows AOD as a function of wavelength between 400 and $800\,\mathrm{nm}$. The upper and lower limit of f_v which satisfied the desired range of AOD(869) was found to be around 10^{-10} and 10^{-12}, respectively.

One can also observe how the three different input parameters volume fraction, fine mode fraction, and RH are affecting the AOD(λ). Yellow, black and blue colors represent f_v equal to 1e-10, 5e-11 and 1e-12, respectively. Dotted and solid line style represent f_s equal to 1.0 and 0.0, respectively. Round and cross markers represent RH equal to 1.0 and 0.0, respectively. One can observe that small values of RH would yield increasing AOD and the other way around. More small aerosols compared to larger would yield larger values of AOD and the opposite. The largest values of f_v would yield a higher concentration of aerosols in the atmosphere and would also increase the AOD accordingly.

The aerosol input to AccuRT was applied to the bottom layer, namely between 0 to $2\,\mathrm{km}$. When generating the data with AccuRT, the values of f_s, f_v, and RH were chosen by randomly pick a value from their given ranges. After all AccuRT simulations were finished, the AOD(869) was calculated for each simulation, and the distribution achieved is shown in Fig. 3.8.

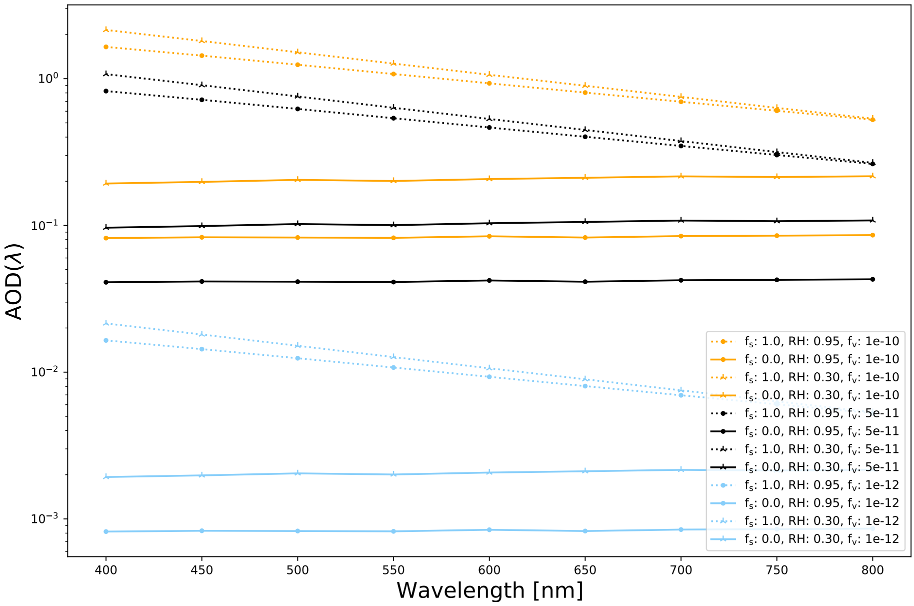


Figure 3.6: AOD(λ) (logarithmic scale) for different values of f_s, f_v and RH as a function of wavelength. Yellow, black and blue colors represent f_v equal to 1e-10, 5e-11 and 1e-12, respectively. Dotted and solid line style represent f_s equal to 1.0 and 0.0, respectively. Round and cross markers represent RH equal to 1.0 and 0.0, respectively.

3.4.4 Water IOPs Input

To simulate a representative synthetic dataset for the ML algorithms, water IOPs were extracted from $in\ situ$ field measurements. The water IOPs extracted were $a_{cdom}(443)$, chlorophyll (CHL) and mineral (MIN) concentrations. Data were extracted from the NASA bio-Optical Marine Algorithm Dataset (NOMAD) dataset and CCRR datasets. The goal was to extract water IOPs representative for both Case 1 and Case 2 waters. When generating data with AccuRT, the bio-optical model CCRR described in section 3.3.3 would be used with these water IOPs as input.

NOMAD Dataset

The NOMAD dataset is a global, high quality $in\ situ$ bio-optical dataset provided for satellite data product validation and ocean-color algorithm. The data is collected from over 3400 globally located stations, containing metadata such as date, time, location of data collection, ancillary data, sea surface temperature, and water depths in addition to several AOPs and IOPs [69]. The global distribution of where the data was collected is shown in Fig. 3.7. In addition to the original and latest NOMAD dataset, two additional datasets were also provided for this study. These are limited to NOMAD stations where High Performance Liquid Chromatography (HPLC) data are available and stations with coincident and valid SeaWiFS coverage, respectively. The distribution of CHL, MIN, and $a_{cdom}(443)$ are shown in Fig. 3.8 for the original and additional datasets, labeled as NOMAD (original) and NOMAD (addition), respectively.

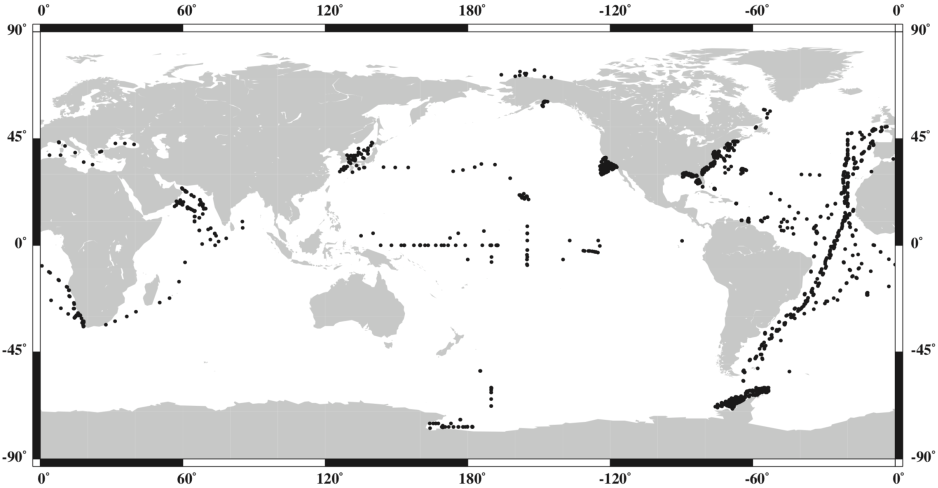


Figure 3.7: The global distribution of the NOMAD dataset (originally from [69]). Data is measured all around the globe for both open oceans and inland/coastal areas.

CCRR Field Measurements and Synthetic dataset

The CCRR project was designed to put together three reference datasets that could be used for research as validation and evaluation of different algorithms that provide coastal water quality data products from ocean-color satellite remote sensing [70]. The result was the dataset including the Medium Resolution Imaging Spectrometer (MERIS) level 2 match-ups, *in situ* reflectance measurements, and synthetic data generated by a radiative transfer model (HydroLight), were the data mainly are representing coastal waters [45]. The dataset consists of over 6000 marine reflectances both multispectral and hyperspectral for different sun-sensor geometries and sky and water conditions. Globally measured and synthetic simulated water IOPs such as Total Suspended Matter (TSM), CHL, and CDOM are also provided, and the distributions are shown in Fig. 3.8 as CCRR Synthetic and CCRR Filed, respectively.

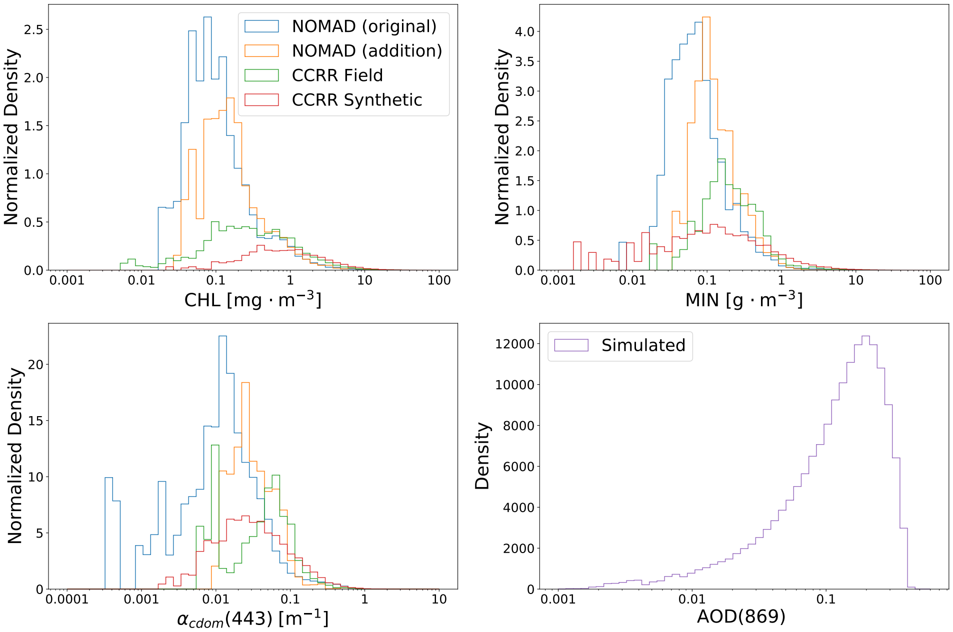


Figure 3.8: Distribution of CCRR inputs CHL, MIN and $a_{cdom}(443)$ extracted from the different field datasets used for the AccuRT simulations. The lower right plot shows the distribution of AOD(869) generated with AccuRT.

Chapter 4

Data Preparation & Machine Learning

The main objective of this study was to produce new AC methods for waters in coastal regions capable of correcting for the complex aerosol and ocean effects based on ML. Large amounts of $in\ situ$ data with corresponding metadata covering both Case 1 and Case 2 waters were difficult to obtain. Therefore the training data for the ML models were chosen to be generated with AccuRT as described in chapter 3. This chapter will describe the data generation, data pre-processing, the different ML models desired to use, and ML hyperparameter optimization. For simplicity for the rest of the thesis, the λ-notation will be removed from the radiance and reflectance terms.

4.1 Data Generation for the Machine Learning

It has been shown that there is a spectral similarity between aerosol and Rayleigh corrected radiance (L_{rac}) and water-leaving radiance (L_w) which has been studied for multispectral data [33]. Therefore, AccuRT will first be used to generate hyperspectral TOA L_{rac}, water-leaving radiance (L_w), and remote sensing reflectance (R_{rs}). L_{rac} is then a combination of aerosol, water-leaving, and reflected sky glint radiance, as AccuRT was not able to simulate whitecap and strong sun glint radiances.

A flowchart illustrating which input parameters used to generate data with AccuRT, which data that were generated and where ML was applied is shown in Fig. 4.1. The yellow boxes named ML1, ML2, ML3, and ML4 refer to the different ML approaches produced in this study. Different values of sensor geometries, aerosol properties, and the water IOPs chlorophyll and mineral concentrations and CDOM absorption coefficient at $443\,\mathrm{nm}$ ($a_{cdom}(443)$) were used as input to AccuRT when generating the data. First, it was desired to predict R_{rs} from L_{rac} (ML1). In the second phase, the same input parameters used to generate L_{rac} were used to additionally generate TOA radiance with contribution from aerosols, Rayleigh, the surface reflected sky radiance, absorption, and water-leaving radiance (L_t). The goal of the last simulation was to generate data as similar to what satellites would measure. Here, ML models were trained to correct for absorption and Rayleigh to get L_{rac} from L_t (ML2). One last AC method was to predict R_{rs} directly from L_t (ML3).

The AC methods will predict the remote sensing reflectance, which again can be used to determine oceanic constituents, like chlorophyll concentration (CHL). Several empirical and semi-empirical algorithms have been used to calculate these parameters, often by looking for ratios or combinations of radiance signals at different wavelengths in R_{rs}. One such IOP retrieval algorithm is the current operational chlorophyll algorithm for SeaWiFS (OC4v4) given by [12, 71]:

$$log_{10}(\mathrm{CHL}) = a_0 + \sum_{i=1}^{4} a_i \left(log_{10} \left(\frac{R_{rs}(\lambda_{blue})}{R_{rs}(\lambda_{green})} \right) \right)^i \tag{4.1}$$

where $R_{rs}(\lambda_{blue})$ is max($R_{rs}(\lambda_{443,490,510})$) and $R_{rs}(\lambda_{blue})$ is $R_{rs}(\lambda_{555})$, and values of a_{0-4} are 0.366, -3.067, 1.93, 0.649, and -1.532, respectively. The band wavelengths and a_0-a_4 are sensor-specific and could have other values for other remote sensing sensors [71]. However, these coefficient are derived from datasets that do not necessarily represent all natural variations, and the performance of such algorithms is always subject to compatibility between the waters under study and the waters from which data were obtained for algorithm development [12]. The datasets used in this study includes water IOPs from both Case 1 and Case 2 waters and water IOP retrieval models based on different ML models will also be tested. They will predict CHL, MIN and $a_{cdom}(443)$ from the remote sensing reflectance spectra (ML4 in Fig. 4.1).

As a conclusion, the generated data were used as training and validation data for different ML approaches, where the goal was to do different types of AC and water IOP retrieval. Different machine learning models as the neural network (NN), partial least square regression (PLSR), support vector regression (SVR) and stochastic gradient descent regression (SGDR) were compared to determine the approach best suited for AC, based on accuracy, time complexity, and interpretation capability.

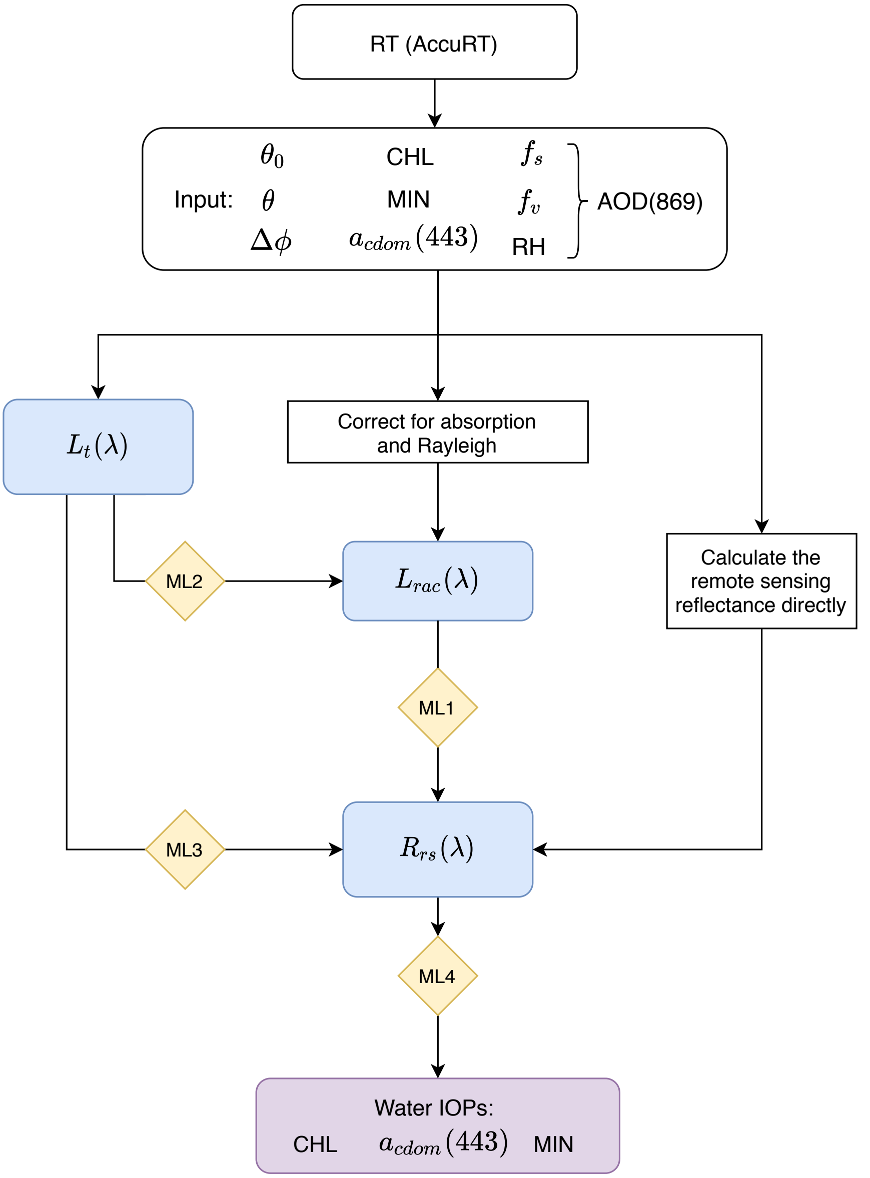


Figure 4.1: Flowchart illustrating the AC and water IOP retrieval approach in this study. Blue boxes represent data generated with AccuRT, yellow boxes represent where ML is applied and the purple box represents the water IOPs: CHL, MIN and $a_{cdom}(443)$. The input box shows the input parameters that varied when the different spectral data was generated. ML1, ML2 and ML3 refer to the three AC approaches, $L_{rac}(\lambda)$ to $R_{rs}(\lambda)$, $L_t(\lambda)$ to $R_{rac}(\lambda)$ and $L_t(\lambda)$ to $R_{rs}(\lambda)$, respectively. $L_{rac}(\lambda)$ refers to TOA radiance corrected for Rayleigh scattering and absorption. ML4 refers to the water IOP retrieval model where CHL, MIN and $a_{cdom}(443)$ were calculated from $R_{rs}(\lambda)$. Notice that AOD(869) not was a direct input to AccuRT, but a result of the combination of f_s, f_v, and RH.

4.2 Radiance Simulations

The setup described in section 3.3 with the described input parameters and ranges were used in order to generate L_{rac}, L_t and R_{rs}. Recall Eq. 2.13, where R_{rs} was given as the water-leaving radiance just above sea surface divided by the downwelling irradiance just above sea surface (E_d). Unfortunately, AccuRT did not give the water-leaving radiance (L_w) directly as an output after the simulation. Therefore, to compute the water-leaving radiance just above sea level, $L_w(\theta_0, \theta, \Delta\phi, \lambda)$, a small pre-calculation was done. AccuRT provided the water-leaving radiance just below the surface ($L_u(0^-, \theta_0, \theta, \Delta\phi, \lambda)$), whereas the radiance just above the surface, was the sum of both the water-leaving and the Fresnel reflected direct attenuated sunlight and skylight, $L_g(\theta_0, \theta, \Delta\phi, \lambda)$. The sea surface did also either transmit or reflect the radiance just below the surface, which would have to been taken into consideration as well. This is illustrated in Fig. 4.2.

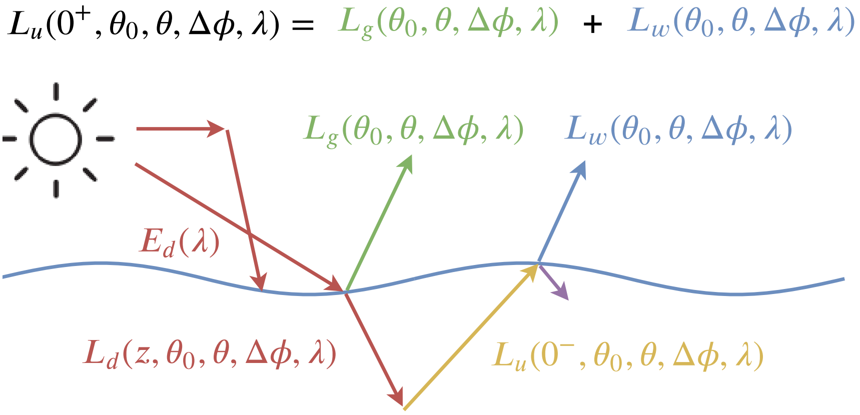


Figure 4.2: Illustration of the radiance just above the sea surface ($L_u(0^+, \theta_0, \theta, \Delta\phi, \lambda)$) as a sum of the water-leaving ($L_w(\theta_0, \theta, \Delta\phi, \lambda)$) and glint radiance ($L_g(\theta_0, \theta, \Delta\phi, \lambda)$). The red arrows represent direct and diffuse irradiance from the sun reaching the surface and the purple arrow represents water-leaving radiance reflected back into the water again.

Consider a case with black water (no scattering) where all radiance in the water is absorbed and nothing is backscattered. In this case, the radiance calculated just above the surface will be all contributions except the water-leaving radiance since no radiance would be scattered out of the water. $L_w(0^+, \theta_0, \theta, \Delta\phi, \lambda)$ would therefore be given by:

$$L_w(0^+, \theta_0, \theta, \Delta\phi, \lambda) = L_u(0^+, \theta_0, \theta, \Delta\phi, \lambda) - L_{u,black}(0^+, \theta_0, \theta, \Delta\phi, \lambda) \qquad (4.2)$$

where $L_{u,black}(0^+, \theta_0, \theta, \Delta\phi, \lambda)$ would be the upward radiance just above the ocean surface which included the Fresnel reflected direct attenuated sunlight and skylight, but no radiance from the water. The black ocean was constructed by setting a very shallow ocean with no bottom albedo (totally absorbing) and only clear water (no dissolved material).

Therefore, each simulation was done twice, one for a normal setup and one with black ocean setup. The water-leaving radiance was then calculated as described in Eg. (4.2) and the remote sensing reflectance was further calculated using Eq. (2.13).

Therefore, in total three separate data generations were done for each parameter setup. One parameter setup refers to on specific combination of the input values θ, θ_0, $\Delta\phi$, f_s, f_v, RH, CHL, MIN and a_{cdom} which would produce unique spectra of L_{rac}, L_t and R_{rs}. The different input parameters, their range, units, and how the parameters were selected are shown in Tab. 4.1. For the uniform selection, the parameters were drawn randomly from a uniform distribution of the given value range. For the distribution selection, the parameters were drawn from a predefined distribution based on the data stored from the field measurements as shown in Fig. 3.8.

First, L_t and L_{rac} were simulated by including and removing atmospheric gases, respectively. Then water-leaving radiance, and further R_{rs} were calculated by simulating radiance when applying black water.

Table 4.1: Different input parameters used for the AccuRT simulations, their ranges and how the parameters were selected.

Parameter	Value range	Unit	Selection
θ_0	0-65	[°]	Uniform
θ	0-70	[°]	Uniform
$\Delta\phi$	0-180	[°]	Uniform
RH	30-95	[%]	Uniform
f_s	0-1	unitless	Uniform
f_v	1e-12 - 1e-10	unitless	Uniform
CHL	0.006 - 98	[mg/m^3]	Distribution
MIN	0.002 - 99	[g/m^3]	Distribution
$a_{cdom}(443)$	0.0004 - 5	[m^{-1}]	Distribution

Data generated with AccuRT from four simulations are shown in Fig. 4.3 represented in four different colors as a function of wavelengths. Different input IOP values are highlighted in the labels and solid and dashed lines represent normal and black ocean, respectively. In this case, one simulation refers to generating L_t, L_{rac} and R_{rs} for one input parameter setup. In Fig. 4.3, one can see that $L_w(0^-, \theta, \Delta\phi, \lambda)$ is zero for the black ocean (dashed lines), as expected due to the black water setup. The dashed lines in the $L_u(0^+, \theta, \Delta\phi, \lambda)$ plot represent the radiance just above sea surface with a black ocean, and is therefore the sky glint contribution. One can observe that $L_w(0^-, \theta, \Delta\phi, \lambda)$ and $L_w(0^+, \theta, \Delta\phi, \lambda)$ have the same spectral shape, but that the radiance values are smaller just above water. This is because some of the radiance is reflected back into the water, as illustrated with the purple arrow in Fig. 4.2. Comparing the plots of $L_t(\theta, \Delta\phi, \lambda)$ and $L_{rac}(\theta, \Delta\phi, \lambda)$ show that the total radiance with atmospheric gases is increasing due to Rayleigh scattering, especially in the small wavelength region where the Rayleigh scattering is more significant, and decreasing at some wavelengths (760 nm) due to absorption.

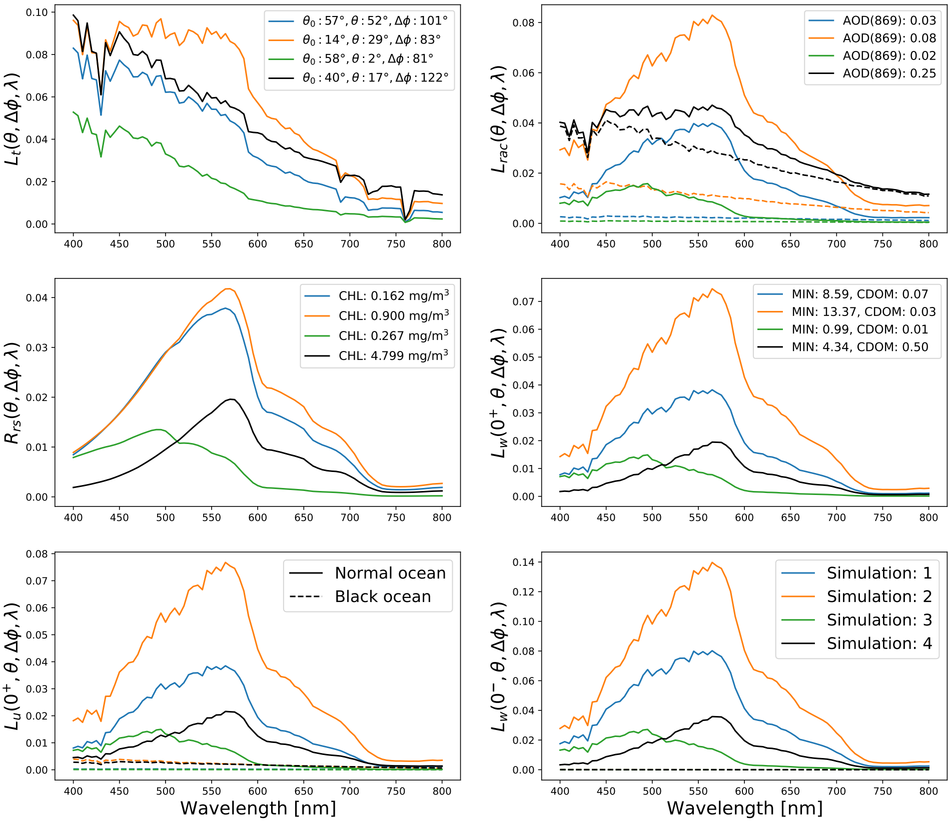


Figure 4.3: Four simulations, represented by four different colors, of various spectral data using AccuRT for different input parameters specified in the labels. Solid and dashed lines represent normal and black ocean, respectively. L_t and L_{rac} are evaluated at TOA and 0^+ and 0^- are evaluated just above and below water, respectively.

4.2.1 Assumptions and Limits of the Method

The two most difficult parts of AC over coastal waters are the aerosol and water body complexity. The latter problem is tried to be solved by generating data with large variations in water IOPs, such that it would cover both Case 1 and Case 2 waters. The aerosol problem is considered in the same way by generating AOP(869) for a large range. Therefore it would be most interesting to first study these contributions in (L_{rac}).

Rayleigh contributions are as discussed previously, often calculated from LUTs as a function of sun-target-sensor geometries. Since these parameters are given and will be used as input data for the ML models, this would most certainly be easier for the ML models to predict. Also, the absorption bands in the atmosphere are well known and the amount of absorption for a given atmospheric setup of absorptive molecules will highly depend on how long the path for the incoming solar flux. This is again dependent on the sun-target-sensor, if clouds, obstacles, and such are neglected. It must be therefore be specified that all data in this study is generated on the assumption of cloud-free environments.

The atmosphere setup used when generating L_t was the U.S. standard atmosphere with predefined height dependent temperatures, pressures and molecular concentrations. The same setup was used for all simulations of L_t, thus atmospheric variations found in the real atmosphere could be difficult to predict when training on this data. In AccuRT it is possible to adjust the amount of the different absorbing and scattering molecules in the atmosphere, so this problem could be taken into consideration in a further approach. However, this was not done in this study, and it would therefore not be expected that the complexity of the TOA radiance would increase significantly when including absorption and Rayleigh in this case with U.S. Standard setup without molecular variations. This is also a reason why the main focus in this thesis is on AC of L_{rac}.

The input irradiance spectra used was the default solar spectrum defined in AccuRT, which was more detailed defined in section 3.4.1. Recall Eq. 2.9, where the sun-Earth distance R is included for normalization. In the AccuRT configuration file, one can specify a source scaling factor, which is to take care of this problem. This scaling factor can be wavelength dependent or just a scalar that is to be multiplied to the input spectrum, this could account for the variation in the sun-Earth distance throughout a year. A limit with this method was that this scaling factor was set to 1.0 for all simulations, thus ignoring the effects from the sun-Earth distance variations. In addition, sea surface roughness and polarization are considered by AccuRT, which should be included for an even more realistic ocean-atmosphere set up. This is something which is under development for a newer version of AccuRT [33].

The last thing to discuss is the contributions of radiance from sunlight penetrating the water and that are reflected from the bottom. For deeper waters, most of the radiance going deeper into the water is absorbed, but for very shallow water, the water-leaving radiance can be heavily affected by radiance from the bottom. In these simulations, the water depth used was more than 50 meters for all the simulations. For a future approach, it would be better to also vary the depth and include very shallow waters.

4.3 Data Pre-processing

Investigating the data used to train the ML models could be important as pre-processing of the data may lead to improvements of the models [72]. The following subsections will detail the pre-processing steps used in this study. In addition to these pre-processing steps, two other steps were included. First, all the input and response data were checked for negative values, strange values, and spikes in the wavelength spectra. These data would further be removed. Secondly, all the data were randomly shuffled before used to train the models.

4.3.1 Standardization

The idea behind standardization is to transform the data such that its distribution will have a mean value equal to 0 and a standard deviation equal to 1 along each dimension of data. Given the distribution of the data, each value in the dataset will have the mean value subtracted, and then divided by the standard deviation of that feature. To standardize the datasets before applying them to many ML algorithms is often a common requirement as ML estimators may behave badly if the individual features do not more or less look like normally distributed data [72]. For multivariate data like this case, standardization makes the values of each feature in the data have zero-mean and unit variance. Mathematically this can be written as:

$$\mathrm{X}_{std} = \frac{\mathrm{X} - \overline{X}}{\sigma_X} \tag{4.3}$$

where X is data from one dimension, $\overline{X}$ is the mean value of that dimension and σ_X is the corresponding standard deviation of the same dimension. Standardization helps the model to not neglect variables with small variations and not to weight variables with large variations too heavy. Both the input and the output data were standardized before proceeding with the ML models.

4.3.2 Angles pre-processing

In addition to the TOA spectra, the solar and view zenith angles (θ_0, θ), as well as the relative azimuth angle ($\Delta\phi$), are also provided as input data for the ML models. For the ML1 and ML3 approach (see Fig. 4.1), the desired output is the remote sensing reflectance. This property is given by Eq. (2.13), where the denominator, E_d, is the irradiance measured just above the sea surface. This property is the product of the extraterrestrial irradiance, atmospheric diffuse transmittance for irradiance in the Sun's direction and the cosine of the solar zenith angle. Therefore, the input data consisting of the L_t and L_{rac} spectra were divided by the cosine of the solar zenith angle. This is something the ML models would have to learn, and since this data was already provided, this pre-processing could improve the result. One radiance spectrum of L_{rac} with (orange) and without (blue) dividing by the cosine of the solar zenith angle is shown in Fig. 4.4.

Also, the direct transmittances are functions of the cosine of the solar and viewing angles. Only the cosine of the zenith angles are used, thus as a pre-processing step, the cosine is applied to all the solar and viewing zenith angles used as input.

4.3.3 Savitzky Golay Filter

The Savitzky Golay filter (Savgol) is a filter that can be applied to a set of discrete, discontinuous data points with the purpose of smoothing the data to increase the precision of the data without distorting the signal tendency [73]. The smoothing is done by partly fit successive subsets of adjacent data points by the method of linear least squares with low-degree polynomials. The number of data points in each sub-set and the highest degree of the polynomial is further referred to as the window length (W_l) and the polynomial degree. This method has long been used in the absorption spectroscopy community to smooth and differentiate absorption spectra [74, 75]. Since the data used in spectroscopy bears some similarities with the TOA spectra used in this problem, it would be interesting to investigate its effect on the result as a pre-processing step.

For this study, the python library scipy.signal.savgol_filter [76] was used to apply a Savitzky Golay filter to different TOA spectra. To use the filter, three input parameters must be specified. First parameter is the W_l, which must be a positive odd integer. The second parameter is the order of the polynomial used to fit the samples, which must be less than window length. The last parameter is the order of the derivative to compute. The lower the polynomial degrees are, the more of the spectral shape is smoothed out. When the polynomial degree gets higher, the Savgol filter will be more and more equal to the original spectra. Three Savitzky Golay filters with window length and polynomial order equal to 31 and 10, respectively, for a simulated L_{rac} spectra with zeroth, first and second derivative are shown in Fig 4.4 as the green, red and purple line, respectively. The blue and orange lines represent the original spectra and spectra after divided by the cosine of the solar zenith angle, respectively. One can observe how the Savitzky Golay filter removes the spiky shapes of the radiance spectra and turns it into a smoother spectrum. The choice of W_l and polynomial degree must, therefore, be carefully chosen, as it might remove almost all spectral information from the input spectra.

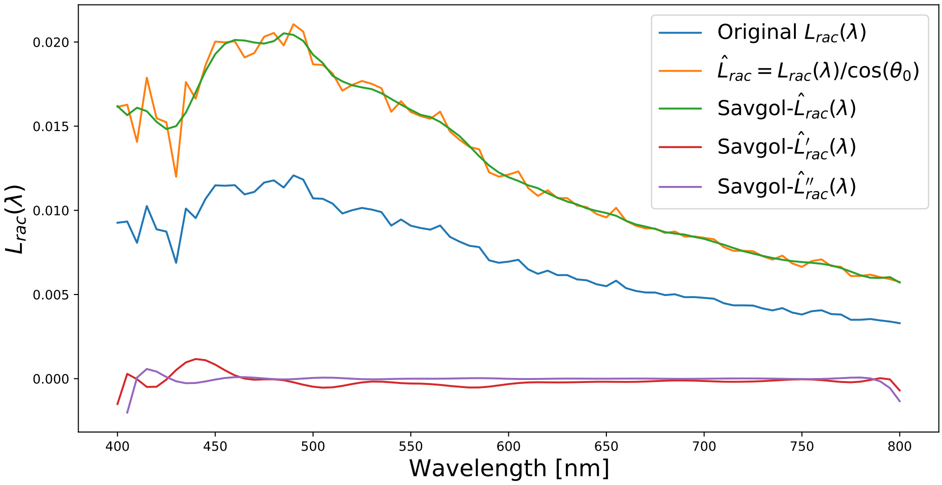


Figure 4.4: One simulation of $L_{rac}(\lambda)$ spectra with the different pre-processing steps. The original $L_{rac}(\lambda)$, $L_{rac}(\lambda)$ divided by the cosine of the solar zenith angle ($\hat{L}_{rac}(\lambda)$), and zeroth, first and second derivative of $\hat{L}_{rac}(\lambda)$ with a Savitzky Golay filter.

4.4 Machine Learning

NN, different regularization methods for linear models, SVR and SGDR were tested to decide which would perform AC best in terms of accuracy, interpretability, and time complexity. When testing ML models, one important factor is also to determine which hyperparameters to use, which are parameters whose value is used to control the learning process. This section will mainly present each of the four ML models and discuss hyperparameter optimization for each ML models separately. The final results of AC, discussion, and performance comparison of the different ML models will be presented in chapter 5.

Before proceeding, some basic terms and representations will be presented. The input and output data for the ML models with one dimension are denoted as $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively. The data points have following shape: $\langle(\mathbf{x}_i, \mathbf{y}_i)\rangle_{i=1}^{N}$, where N is the number of datapoints with $\mathbf{x}_i \in \mathbb{R}^{D_1}$ and $\mathbf{y}_i \in \mathbb{R}^{D_2}$. D_1 and D_2 are the dimensions of the input and output data. Further this will be noted as:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \dots \\ \mathbf{x}_N^T \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1^T \\ \dots \\ \mathbf{y}_N^T \end{pmatrix}$$

where $\mathbf{X}$ is a N $\times$ D_1 matrix and $\mathbf{y}$ is a N $\times$ D_2 matrix, and are the full input and output datasets, respectively. In the following subsections, different ML hyperparameters and pre-processing steps will be tested to find which are giving the best results. The results were based on validation metrics. In this chapter, AC of Rayleigh and absorption corrected TOA radiance (L_{rac}) will be tested because this was the main focus of this study. The input data will be L_{rac} with 81 different bands together with solar, view, and relative azimuth angles, giving D_1 = 84. The output dimension will be the full remote sensing reflectance spectra (R_{rs}) for the same bands, giving D_2 = 81.

The different ML models were validated with three statistical metrics, the squared of the Pearson correlation coefficient (R), the Root Mean Squared Difference (RMSD), and Normalized RMSD (NRMSD) formulated in Eq. 4.4, 4.5 and 4.6. These metrics and other metrics will be further described in Chapter 5 in Tab. 5.1 when the final optimized AC models will be validated against each other. In this chapter, these three metrics will be used to validate the different combinations of hyperparameters against each other for each ML model. Briefly, R^2 measures the linear correlation between two variables X and Y and the RMSD is an accuracy measure given as the square root of the mean of the squares of the deviations. NRMSD is RMSD divided by the difference of the maximum and minimum value of the validation data and is not dependent on the scale of the values.

$$\mathrm{R} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \bar{X_i}}{\sigma_X}\right)\left(\frac{Y_i - \bar{Y_i}}{\sigma_Y}\right) \tag{4.4}$$

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \tag{4.5}$$

$$\mathrm{NRMSD} = \frac{\mathrm{RMSD}}{Y_{max} - Y_{min}} \tag{4.6}$$

As mentioned for AC, the ML models would give 81 predicted outputs from the wavelength bands between 400 and $800\,\mathrm{nm}$. Metric values for each predicted wavelength band were calculated and would therefore yield 81 values for each metric ($\mathrm{R}^2_{400}, \mathrm{R}^2_{405}, \ldots, \mathrm{R}^2_{800}$). It can be difficult to compare all of them at once, so the optimal results of the different hyperparameter testings were therefore based on the mean of the 81 metric values. This will be represented as $\overline{\mathrm{R}^2}$, $\overline{\mathrm{RMSD}}$ and $\overline{\mathrm{NRMSD}}$.

4.4.1 Neural Network

It has been demonstrated that NN with one or more hidden layers can do predictions of non-linear functions that could be suitable for this problem when deriving remote sensing reflectance from various TOA radiances [33, 11].

Building the NN model

The NN used in this thesis is a feed-forward artificial NN, also known as the multilayer perceptron. The Python Deep Learning library Keras using TensorFlow as backend was used to build the NN models in this thesis, which were simple sequential models. The network was organized in layers with one input layer starting with the input data and one output layer giving the predicted result. The layers were connected via different numbers of nodes, also known as neurons. The NN of consisted several hidden layers between the input and output layer, where each layer served as an input for the upcoming layer. The structure of the input, hidden and output layers are illustrated on the left-hand side of Fig. 4.5.

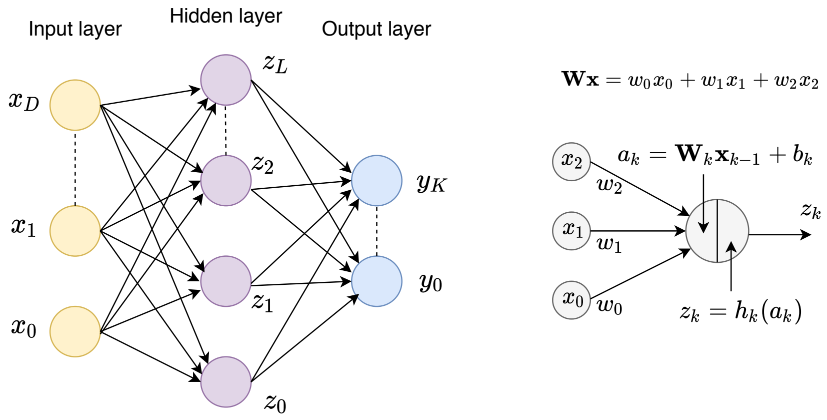


Figure 4.5: Illustration of the NN structure with input, output and hidden layers. The calculation of the input for the next layer is also illustrated.

For this case with D_1 and D_2 equal to 84 and 81, respectively, the number of neurons for the input and output layers would be the same (D=84, and K=81 in Fig. 4.5). The NN works in the way that the k-th layer serves as inputs of k + 1th layer, where each layer k only does simple computations. This is done as a linear function of previous layer's outputs $\mathbf{z}_{k-1}$ given by:

$$\mathbf{a}_k = \mathbf{W}_k\mathbf{z}_{k-1} + \mathbf{b}_k \tag{4.7}$$

where the next layers output is calculated with a nonlinear transformation through an activation function h_k given by:

$$\mathbf{z}_k = h_k(\mathbf{a}_k) = h_{relu}(\mathbf{a}_k) = \max(0, \mathbf{a}_k) \tag{4.8}$$

where the choice of activation function for this problem was chosen to be the Relu activation function indicating that it would be zero when $\mathbf{z}$ is less than zero and equal to $\mathbf{z}$ when $\mathbf{z} \geq 0$. This was specified as a good choice of activation function for regression [77]. The calculation of the input for the next layer is shown in the right part of Fig. 4.5. Also, a loss function, an optimizer, a validation split, and a metric to monitor during training and testing must be chosen. The loss function will measure the performance of the network and guide it in the right direction and the optimizer will help the network update itself based on the data it sees and its loss function. The loss function was set to the mean square error. Both the "adam" and "rmsprop" optimizers were highlighted as good options for NN regression [77]. The validation split would be a percentage number indicating how much data to be set aside for validation while the remaining data were used for training of the NN model. Two last parameters to decide are the epoch and batch size which are important factors when it comes to training time and accuracy. The batch size decides the number of training examples to be run through the network [78] (one pass) and one epoch is when all the training data has been run through the network. This means that a small batch size would yield more passes to achieve one finished epoch. The choice of hidden, layers, neurons, and optimizer would have to be further tested to optimize the result.

Choice of Hidden Layers and Neurons

One of the main issues when it comes to constructing NNs is to find the optimal number of hidden layers and neurons, something that is still an open research question [33]. There are many parameters the optimal numbers depend on, such as the size of input and output layers, the complexity of the function to be approximated, choice of activation function, and loss function, training algorithm and training sample, which leaves it to a very complex problem to decide [77]. There are some rules of thumb when it comes to the choice of hidden layers. For instance, should the number of hidden neurons be between the size of the input layer and the size of the output layer. Also, the number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer [78]. It can often be enough with only a couple of hidden layers to approximate a function. None hidden layers could solve linear separable functions, so it would be expected to have at least one hidden layer [78].

With these rules of thumbs in mind, different combinations of hidden layers and neurons were tested to see which would perform best. The number of hidden layers was varied between 0 and 30. For each specific number of hidden layers, 16 different neuron combinations were applied. The first hidden layer were tested with following initial values of neurons in first layer: 700, 400, 350, 300, 250, 200, 185, 150, 140, 125, 110, 100, 95, 90, 85 and 84. If the number of hidden layers was more than 1, a linear spaced list of neurons starting from one of the initial values and ending with the output shape, for this case D_2 = 81, was used. Consider a case with 3 hidden layers with the starting number of neurons equal to 700. This would yield 700, 493, and 287 neurons for the first, second and last hidden layer, respectively. All these combinations would again be done for the "adam" and "rmsprop" optimizers. Different validation metrics were calculated as the mean of three equal simulations for each combination and compared to see which one did the best. The setup of the network that yielded the best results due to accuracy was chosen.

The results, based on predicting R_{rs} from L_{rac}, are shown in Fig. 4.6 and shows $\overline{\mathrm{R}^2}$ (top plots) and $\overline{\mathrm{NRMSD}}$ (bottom plots) as a function of starting number of neurons in hidden layer 1 (y-axis) and the total number of hidden layers (x-axis). The left and right plots show the results with "adam" and "rmsprop" optimizer, respectively. Bright boxes represent good performance, in other words high and low values for $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$, respectively. The top 10 best values are marked with black boxes where the rank number is highlighted within the boxes. The mean value of the top 10 best values with respect to $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ for "adam" was found to be 0.9966 and 0.074. For "rmsprop" this was calculated to be 0.9964 and 0.081, thus the "adam" optimizer provided slightly better results and was the desired optimizer to use. One can observe from Fig. 4.6 that choosing the right number of hidden layers was more important than choosing the right number of neurons. Almost all the best results were found between 1 and 7 hidden layers, and especially between 1 and 3 hidden layers. However, the overall best result was found for the "adam" optimizer with 2 hidden layers and starting with 700 neurons in the first hidden layer. This was the desired structure to use further with the NN.

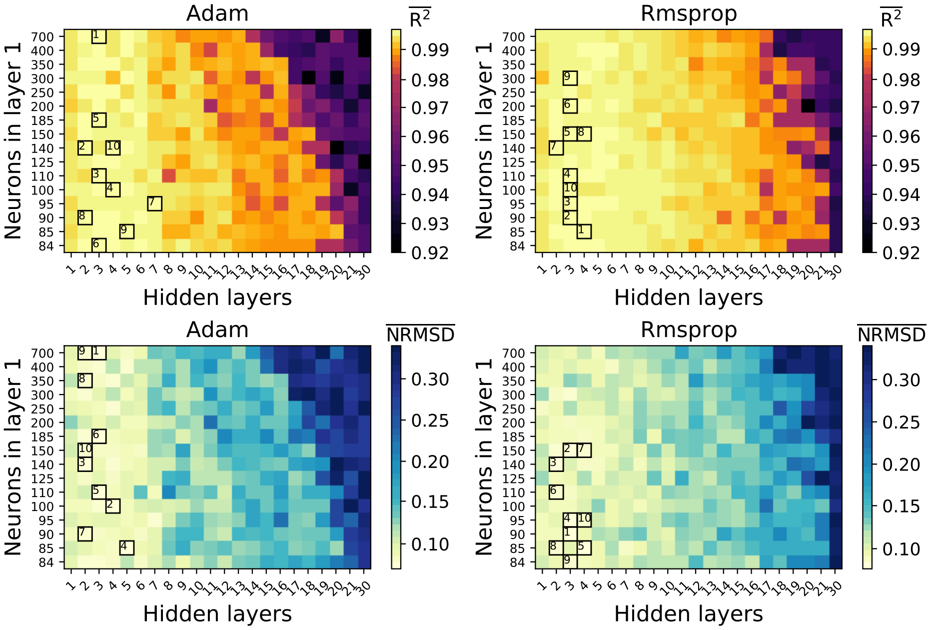


Figure 4.6: $\overline{\mathrm{R}^2}$ (top plots) and $\overline{\mathrm{NRMSD}}$ (bottom plots) as a function of starting number of neurons in hidden layer 1 (y-axis) and the number of hidden layers (x-axis) based on AC of L_{rac}. The left and right plots represent NN with the "adam" and "rmsprop" optimizer, respectively. The top 10 best values are marked with black boxes and the rank number is highlighted inside of the box.

NN with Savitzky-Golay filter

The further choices of loss function, optimizer, number of hidden layers and neurons used to construct the NN were inspired by the results from 4.4.1. Different Savgol filters were then tested on this NN model, were the only parameters to be varied were the three different variables in the Savgol filter, namely the window length, polynomial degree, and the derivative. The Savgol filter was applied to each band in the TOA spectra, in other words on 81 successive radiance values as shown in Fig. 4.4. The window length (W_l) was tested for all odd integers from 1 to 41, approximately half of the data points. The polynomial degree could never exceed the window length and was therefore for each W_l, varied from 0 up to W_l - 1. This meant that for W_l = 3, polynomial degrees of 0, 1, 2 were tested.

Additionally, this was also tested for the first and second derivative of the Savgol filtered spectra. $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ were calculated for for each each training. 120000 data points with L_{rac} and R_{rs} values were used and the validation data was equal to $5\,\%$ of the data. The NN would not train on the validation data. For each Savgol parameter combination, the NN was run three times, and the metrics shown are the mean of the metrics calculated for each of the three fittings. Epochs, batch size, validation split, and optimizer for the Savgol testing was set to 10, 15, 0.15, and "rmsprop", respectively.

Three plots of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ for each combination of window length (y-axis), polynomial degree (x-axis) and derivative are shown in Fig. 4.7. The top, middle and bottom plot represents Savgol filters with zeroth, first and second derivative, respectively. Each color bar represents one of the metrics used, specified with the color bar title. The $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ values are found in the lower left and upper right part of the plots, respectively. The two metrics are divided by a black diagonal line located in the middle of each plot. Light colors represents good results, in other words high values for $\overline{\mathrm{R}^2}$and small values for $\overline{\mathrm{RMSD}}$. The top 30 best values for both $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ are marked with black boxes and numbers ranking the values. In other words, if the number is 1, it had the best metric value, if 2 then the second-best, and so on. If the box is marked with a red square, it means that this value was among the 30 best values of both $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$. Red boxes, therefore, indicate a good combination of window length and polynomial degree.

The top and bottom values from the color bars in Fig. 4.7 shows that the best results are found for the first (middle plot) and second (bottom plot) derivative of the Savgol filters. For the case with no derivative (top plot), the best $\overline{\mathrm{R}^2}$ values were approximately 0.98, whereas, for the first and second derivatives, the best results were both higher than 0.99. Also, one can observe a pattern in the results for the first and second derivative, as the best values are sorted around some specific areas. One can also observe that there exist values among the top 30 for almost all window lengths (y-axis), whereas the top values for the different polynomial degrees (x-axis) are found only for a few of them. This means that one can find good results for almost all window lengths, as long as it is combined with the right polynomial degree. For the second derivative, it can be observed that all 30 maximum values are found between polynomial degrees between 2 and 12, whereas for the case with no derivative this range is between 0 and 24. Specifically for the second derivative (bottom plot), one can find good values in the whole range where window length and polynomial degree is between 23 to 41 and 4 to 11, respectively. The maximum value of $\overline{\mathrm{R}^2}$ and the minimum value of NRMSD were both found for the second derivative and were 0.9948 and 0.000359 respectively.

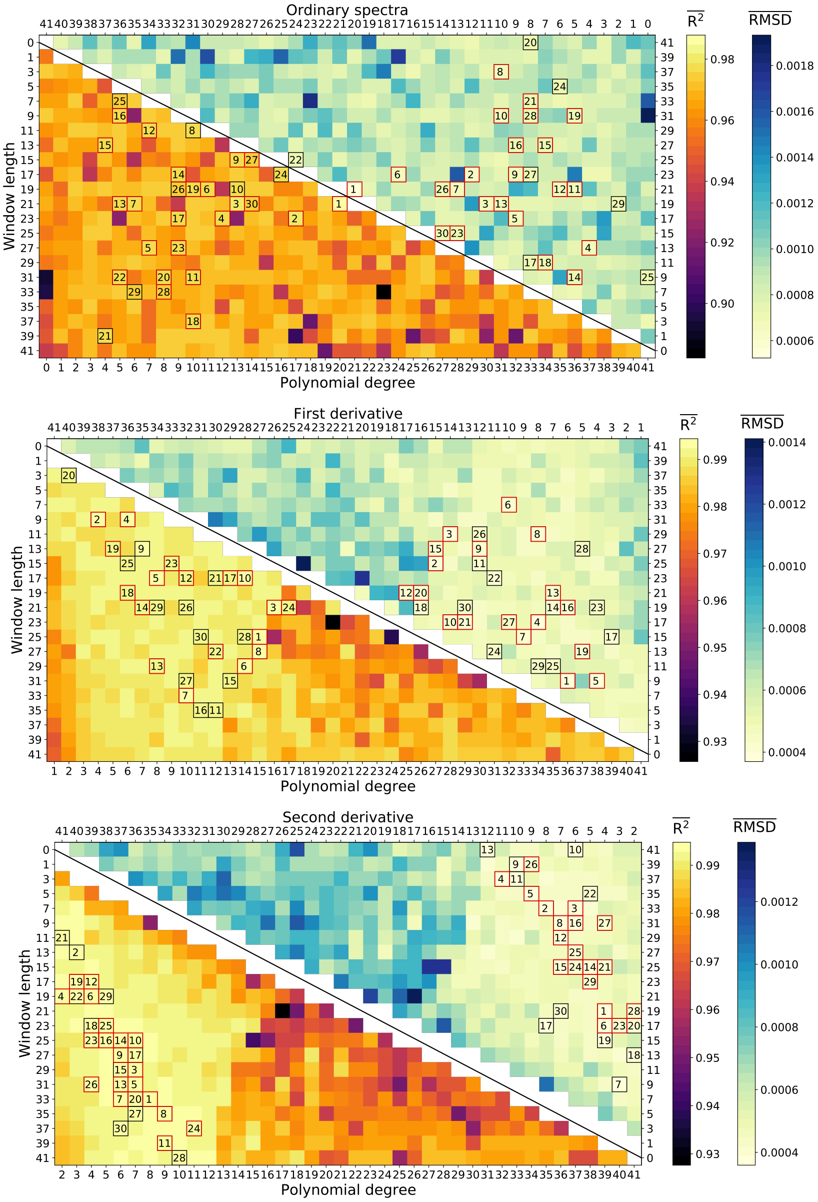


Figure 4.7: Plot of calculated $\overline{\text{R}^2}$ and $\overline{\text{RMSD}}$ values based on a NN approach when predicting R_{rs} from L_{rac} where Savgol filters with different combinations of window length (y-axis), polynomial degree (x-axis) and derivative were applied to the input data. Left and right color bars represent $\overline{\text{R}^2}$ and $\overline{\text{RMSD}}$, respectively. $\overline{\text{R}^2}$ and $\overline{\text{RMSD}}$ values are found in the lower left and upper right part of the plots, respectively. The top 30 best values for both $\overline{\text{R}^2}$ and $\overline{\text{RMSD}}$ are marked with a black box and the number within ranks the values. The red boxes represent values that are among top 30 in both $\overline{\text{R}^2}$ and $\overline{\text{RMSD}}$.

A corresponding plot to Fig. 4.7 is shown in Fig. 4.8. The plot shows the same calculated $\overline{\mathrm{R}^2}$ as a function of the ratio defined as the polynomial degree plus 1 divided by window length. Plus one is added because the polynomial degree could be 0, which would give the same ratio for all window lengths when the polynomial degree was 0. The derivatives are represented in different colors. The two arrows in the bottom of the plot indicate what would happen to the input spectrum for the different ratios. A radiance spectrum with a Savgol filter with ratio close to 1, would not change the radiance spectrum. When the ratio goes towards 0, the input radiance spectrum would be more and more smoothed out, and the spectral information would be removed accordingly. All $\overline{\mathrm{R}^2}$ values found when the polynomial degree was smaller than the derivative was removed.

The blue data points representing the zeroth derivative show that $\overline{\mathrm{R}^2}$, as a function of the ratio, is very unpredictable. This is because the data points change rapidly from smaller to larger values for almost all ratios, which is also possible to observe for the zeroth derivative in Fig. 4.7 where dark and bright values are spread all over the plot. In general, the highest values of $\overline{\mathrm{R}^2}$ were found for the first and second derivative since the orange and green data points are almost always higher than the blue data points. The maximum for the zeroth, first, and second derivatives were 0.9878, 0.9944, and 0.9948, respectively. However, when comparing the first and second derivative against each other, the major part of the best values of $\overline{\mathrm{R}^2}$ were found for the second derivative.

Interestingly, the best results for the first and second derivative were found for almost the same ratios, which was between 0.15 and 0.4, seen in Fig. 4.8. $\overline{\mathrm{R}^2}$ values found for ratios between 0.15 and 0.4 were constantly among the highest values, indicating that good performance always would occur if the ratio was between 0.15 and 0.4. On the other hand, when the ratio got higher than 0.5, the first and second derivative also showed unpredictable results as the data points were changing rapidly. These ratios are seen in Fig. 4.7 where the optimal results (black boxes) seem to follow a linear pattern for the first and second derivative.

In conclusion, the Savgol filter affects the performance, and the second derivative of the spectra with Savgol filter for window lengths and polynomial degrees with ratios between 0.15 and 0.4, showed the best performance. It must be mentioned that in general, the NN showed good performance for both R^2 and RMSD when predicting R_{rs} from L_{rac}.

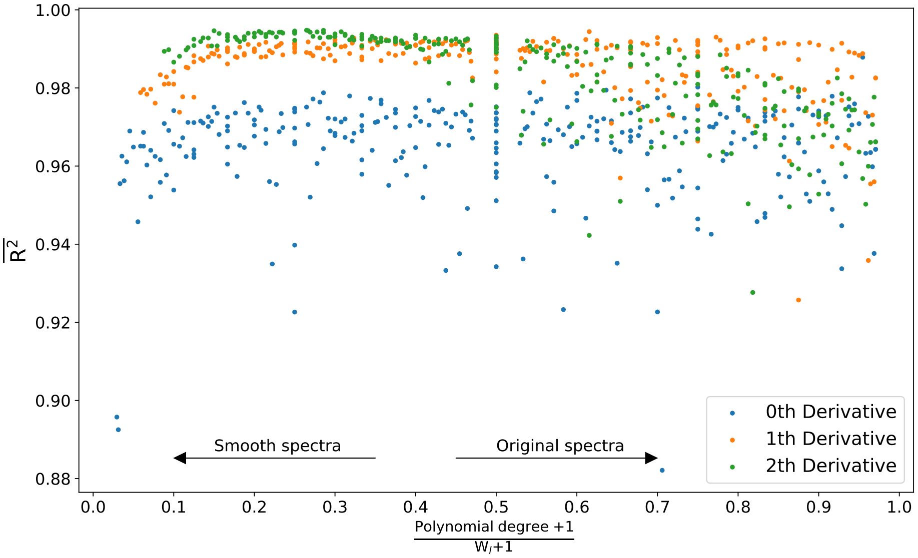


Figure 4.8: $\overline{\mathrm{R}^2}$ as a function of the ratio of polynomial degree plus 1 divided by the window length plus 1 for NN when predicting R_{rs} from L_{rac}. The window length was tested from 0 to 41, polynomial degree up to W_l - 1 for zeroth, first and second derivative. Ratios close to 1 yield almost no changes in the spectra, whereas ratios close to 0 mean that the spectra are very smooth and the spectral information could have been removed. The best results are found for the first (orange) and second (green) derivative for ratios between 0.15 and 0.4 because this $\overline{\mathrm{R}^2}$ values are almost constantly the highest.

4.4.2 Regularization for Linear Models

For this study, Principal Component Regression (PCR), Ridge Regression, and PLSR with and without variable selection were tested and investigated. Each method was tested for predicting R_{rs} from L_{rac}, where three different input datasets of L_{rac} were tested. The first input data followed the pre-processing steps described in section 4.3, without any Savgol filters, whereas the other two datasets had two different Savgol filters applied. These datasets were the first and second derivative of the TOA spectra, with window length and polynomial degree equal to 31 and 10, respectively. This approach was done to investigate how the different Savgol filters affected the regularization models. R^2 and NRMSD were used for validation for the different linear models to compare them against each other. If nothing else is specified, then for all the regularization methods presented in this section, 50.000 data points were used where $95\,\%$ and $5\,\%$ of the data was used for training and validation, respectively.

PCR

Principle Component Regression (PCR) is a well-used extension to Principle Component Analysis (PCA), when it comes to regression. PCA aims to reduce the input space, by removing dimensions in multidimensional data that display a high level of correlation, while still retaining the variation found in the data [74]. This means that if two or more input variables are dependent upon each other in such a way that one can linearly predict from the other with a high degree of accuracy, then some of these variables can be removed [79]. PCR is also a well-studied area when it comes to chemometrics for NIR spectroscopy. Because a reflectance/radiance value at one wavelength illustrates a high correlation with their neighboring wavelengths, selecting a few PC is often unavoidable, and results can be achieved with just a small part of the data [74]. Data from NIR spectroscopy share similarities with the spectral data in this thesis, which makes it interesting to test PCR. PCR can be interpreted as a two-step procedure [74]:

1. Find the principal components of the dataset by decomposing the independent variables with PCA
2. Run the regression on a proper subset of the principal components

An important factor is that the principal components (PC) found using PCA might not be optimal for the regression. The components are chosen without taking the ground truth **y** into account. This is done using PLSR which will be discussed later. The difficulty with PCR relay on deciding the number of principal components to use. To investigate that further, a calibration model was implemented. This created a linear regression between the input data, $\mathbf{X}_{red}$, and the output, $\mathbf{y}$. $\mathbf{X}_{red}$ denotes that the input space has been reduced to an x number of principle components using PCA. PCR fits a model by performing conventional least squares regression on the PCs. The model for linear regression can be written in augmented vector notation as:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \epsilon \tag{4.9}$$

where **w** is a vector with weighted components $(\omega_0, \omega_1, ..., \omega_{D_1})$, where D_1 is the index of the maximum dimension used in **X**, **ŷ=Xw** is the fitted model that will attempt to predict the outcome **y**, also denoted as f(**X**, **w**). $\boldsymbol{\epsilon}$ is the error residuals for which the model differ from the true values. Least squares aims to minimize the error squared, which mathematically can be described by:

$$\arg\min_{\mathbf{w}} ||\boldsymbol{\epsilon}||^2 = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||^2 \tag{4.10}$$

The squared Pearson coefficient (R^2) and the mean squared error (MSE) were used to evaluate the regression model. MSE is squared of the RMSD defined in Eq. 4.5. To avoid that the model was built to perfectly describe the input data, but nothing else (overfitting), 10 % of the data was used for validation, using cross-validation (CV). The calibration and CV models were then tested and validated for PC values from one up to the maximal number of dimensions (D_1 = 84), and both $\overline{\text{MSE}}$ and $\overline{\text{R}^2}$ for the calibration and CV were calculated. One example is shown in Fig. 4.9, and shows $\overline{\text{MSE}}$ and $\overline{\text{R}^2}$ as a function of numbers of PC included. Fig. 4.9 shows that the calibration metrics improve for an increasing number of PCs, especially for the first 10 PCs. For larger numbers of PCs, the metrics continue to improve, but not as much as for the first 10 PCs. The interesting part is the metrics based on CV, which give an indication of how well the model performed on untrained data. The optimal number of PC was therefore chosen to be where the minimum value of MSE with CV was found, which is illustrated in Fig. 4.9 with a black in the bottom plot.

The optimal numbers of PC were calculated to be 31, 40, and 77 for no Savgol filter, first derivative, and second derivative, respectively. The optimal solution was found for the original data without Savgol filters. Because the model predicts 81 values, also 81 metric values will be calculated. So, in a similar manner, as done with NN, the mean values of R^2 and MSE were used to compare the performance. The results from PCR are shown in Tab. 4.2. The prediction based on the original, first derivative, and second derivative datasets was almost the same for all three. In general, this linear method also showed good results with respect to $\overline{\text{R}^2}$ and $\overline{\text{MSE}}$, even though the $\overline{\text{R}^2}$ values were higher for the NN approach.

Table 4.2: Optimal results of $\overline{\text{R}^2}$ and $\overline{\text{MSE}}$ based on calibration and CV data when using PCR to predict $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$. Original refers to data pre-processed without Savgol filters, and 1st and 2nd derivative refer to the first and second derivative of the data with Savgol filters. The optimal numbers of PC are also specified. N$_{\text{train}}$ = 50.000.

Parameter	**Original**	**1st derivative**	**2nd derivative**
PC (optimal)	31	40	77
$\overline{\text{R}^2}$ calibration	0.971	0.970	0.971
$\overline{\text{R}^2}$ CV	0.968	0.965	0.967
$\overline{\text{MSE}}$ calibration	0.0287	0.0303	0.0290
$\overline{\text{MSE}}$ CV	0.0315	0.0346	0.0330

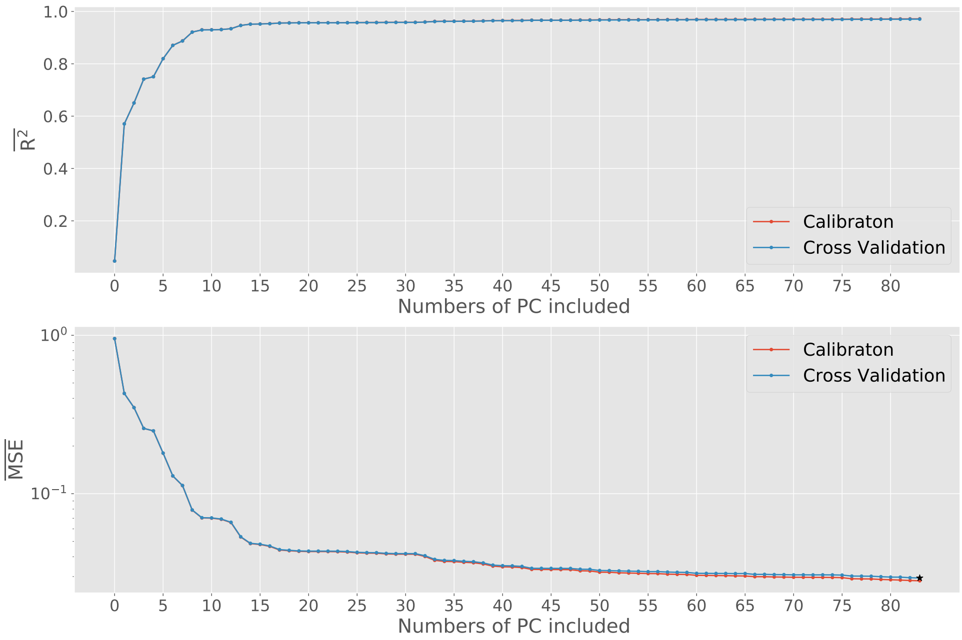


Figure 4.9: Calculated $\overline{\mathrm{R}^2}$ (top plot) and $\overline{\mathrm{MSE}}$ (bottom plot) with calibration and CV data for PCR with various numbers of PC included in the model when predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$. The optimal number is marked with a black start in the lower plot. These results are based on AC of L_{rac} with no Savgol filter applied with number of training data points = 20.000.

Ridge Regression

Ridge Regression differs from PCR as it uses data from all the input dimensions for the regression, whereas PCR only selects the most important dimensions. However, Ridge tries to penalize the least important dimensions, instead of removing them, by minimizing this function:

$$\arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||^2 + \alpha||\mathbf{w}||^2 \tag{4.11}$$

where the squared norm of **w** penalties solution with large values of the components of **w**. However, the value of the hyperparameter α is left to investigate, as this could both improve and aggravate the performance. A large value of α will result in extremely small coefficients and the prediction will end up with $\hat{\mathbf{y}} = 0$. Small α would make the Ridge regression behave as ordinary least squares. For very smooth data, this could cause a problem as ordinary least squares generally would not work on data which are strongly correlated [74]. The GridSearchCV function in sklearn (machine learning toolkit in Python) was used to optimize the value of α. The optimal value of α and different metrics validated on calibration and CV data are shown in Tab. 4.3. The results are very similar, but the best values for all the metrics are found for the original dataset without any Savgol filters. Ridge Regression and PCR are providing very similar results when comparing Tab. 4.3 and Tab. 4.2.

Table 4.3: Optimal results of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ based on calibration and CV data when using Ridge Regression to predict $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$. Original refers to data pre-processed without Savgol filters, and 1st and 2nd derivative refer to the first and second derivative of the data with Savgol filters. The optimal value of α and time for fitting the model are also specified. $\mathrm{N}_{\mathrm{train}}$ = 50.000.

Parameter	Original	1st derivative	2nd derivative
α (optimal)	0.0003162	0.0001	0.0001
$\overline{\mathrm{R}^2}$ calibration	0.972	0.971	0.970
$\overline{\mathrm{R}^2}$ CV	0.967	0.965	0.967
$\overline{\mathrm{MSE}}$ CV	0.0327	0.0355	0.0334

PLSR

PLSR is a widespread regression technique that shares some similarities with PCR. Both methods aim to build a linear model described by Eq. 4.9 and to get rid of correlated data. Normally, PCR reduces the input set **X** into a new reduced set, $\mathbf{X}_{red}$, with orthogonal components based on principle component analysis. In this case, only the input data **X** is considered. The main difference is that PLSR also considers the response data **y** and builds a model based on maximizing the covariance between **y** and $\mathbf{X}_{red}$.

To find an optimal solution with PLSR, one must find the optimal number of components to use when fitting the data to the model. A code similar to the one for optimizing components for PCR was implemented using the same metrics for the validation. The number of components varied from 1 to $D_1 = 84$, which was the number of dimensions of the training data. $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ for all the different numbers of components were calculated and saved. The optimal number of components to be used was determined by choosing the number of components resulting in the smallest $\overline{\mathrm{MSE}}$ with CV, as done with PCR. A plot of $\overline{\mathrm{MSE}}$ as a function of components is shown in Fig. 4.10 and the optimal number of components and different calibration and CV metrics are shown in Tab. 4.4. The results are also very similar to both PCR and Ridge Regression with the best value of $\overline{\mathrm{R}^2}$ equal to 0.967 and $\overline{\mathrm{MSE}}$ CV equal to 0.0330.

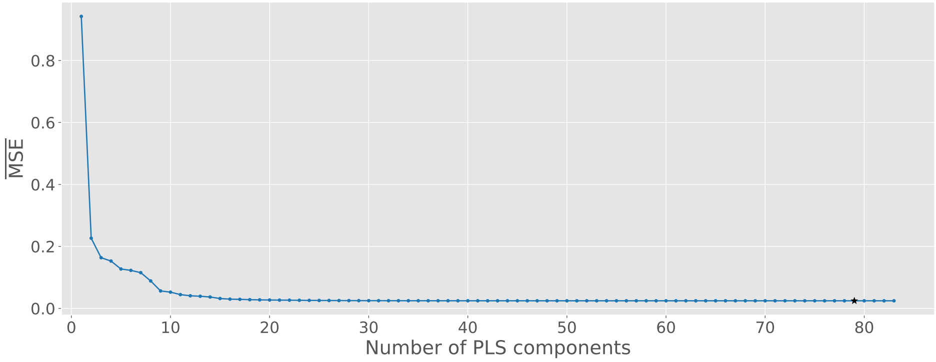


Figure 4.10: $\overline{\mathrm{MSE}}$ as a function of number of components for PLSR. The optimal numbers of PC is marked with a black star which represents the minimum value of $\overline{\mathrm{MSE}}$. These results are based on AC of L_{rac} with no Savgol filter applied with number of training data points = 50.000.

Table 4.4: Optimal results of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ based on calibration and CV data when using PLSR with optimal number of components to predict $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$. Original refers to data pre-processed without Savgol filters, and 1st and 2th derivative refer to the first and second derivative of the data with Savgol filters. $\mathrm{N_{train}} = 50.000$.

Parameter	**Original**	**1st derivative**	**2nd derivative**
Components (optimal)	79	40	77
$\overline{\mathrm{R}^2}$ calibration	0.972	0.970	0.971
$\overline{\mathrm{R}^2}$ CV	0.967	0.965	0.967
$\overline{\mathrm{MSE}}$ calibration	0.0280	0.0303	0.0290
$\overline{\mathrm{MSE}}$ CV	0.0331	0.0346	0.0330

Variable Selection with PLSR

One last method to investigate with PLSR is to test if the result shown in Tab. 4.4 for PLSR could be further improved by also considering to not train the model on all the input features. This was done by extending the code for the optimal number of components of PLSR, to also find the optimal number of input features to use. Possibly, not all the 81 wavelength bands used as input are necessary to fit the model.

The feature selection was done by first considering the full spectrum with all features to optimize the PLS regression, and then use the PLS regression coefficients ($\mathbf{w}$) to see which features were important. When predicting an output with dimension D_2 = 81, 81 sets of regression coefficients will also be calculated. Each set will have an equal number of coefficients as there were input dimensions, for this case D_1 = 84. In other words, each of the 84 input features would produce 81 corresponding coefficients that would indicate the importance of that particular feature to predict the 81 outputs. The interpretation of the coefficients is that coefficients with a large value (both negative and positive) would indicate that this feature is important for predicting the output. In order to decide which were the most important features, all the absolute values of the 81 regression coefficients for each input feature were added together. These sums would, therefore, give an indication of how important each input feature was in predicting the output spectra. Small absolute values of the regression coefficients would yield a low correlation with the quantity of the interest. Therefore, the wavelength band represented by the lowest absolute values of the PLS coefficients would iteratively be discarded at a time and the calibration model was rebuilt for each time and $\overline{\text{MSE}}$ was calculated. The number of removed wavelengths resulting in the $\overline{\text{MSE}}$ to be minimized would, therefore, be the optimal solution.

The optimal number of components and wavelengths and different calibration and CV metrics are shown in Tab. 4.5. As before, the optimal results were found when training on the dataset without Savgol filters. However, the results when using Savgol filters were very similar to the one without. The results when using variable selection were better than the results found when using PLSR without variable selection and Ridge Regression showed in Tab. 4.4 and Tab. 4.3. PLSR will, therefore, be the model used for AC when choosing from the linear models PCR, Ridge Regression, and PLSR.

Table 4.5: Optimal results of $\overline{\text{R}^2}$ and $\overline{\text{MSE}}$ when using PLSR to predict $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ with variable selection with optimal numbers of components and discarded wavelengths. CV and calibration represent training and validation data, respectively, and N_{train} = 50.000.

Parameter	Original	1. derivative	2. derivative
Components (optimal)	52	68	60
Discarded Wavelengths	9	15	19
$\overline{\text{R}^2}$ calibration	0.971	0.970	0.970
$\overline{\text{R}^2}$ CV	0.970	0.966	0.967
$\overline{\text{MSE}}$ calibration	0.0292	0.0301	0.0302
$\overline{\text{MSE}}$ CV	0.0298	0.0340	0.0327

A plot of the input spectra, **X**, as a function of wavelength (top plot), the discarded wavelength bands, and the absolute value of the PLS coefficients (bottom plot) are shown in Fig. 4.11. This result is based on the optimal result for the dataset without Savgol filter (Original) shown in Tab. 4.5. The red vertical lines in the top plot shows which bands could be discarded before training the PLSR model, and still get an optimal result. The corresponding discarded wavelength bands and sun-target-sensor angles are colored red in the bottom plot and represent features with a low correlation to the desired output. One can observe for this case that angles show low importance for the prediction and that some wavelength bands are more important than others.

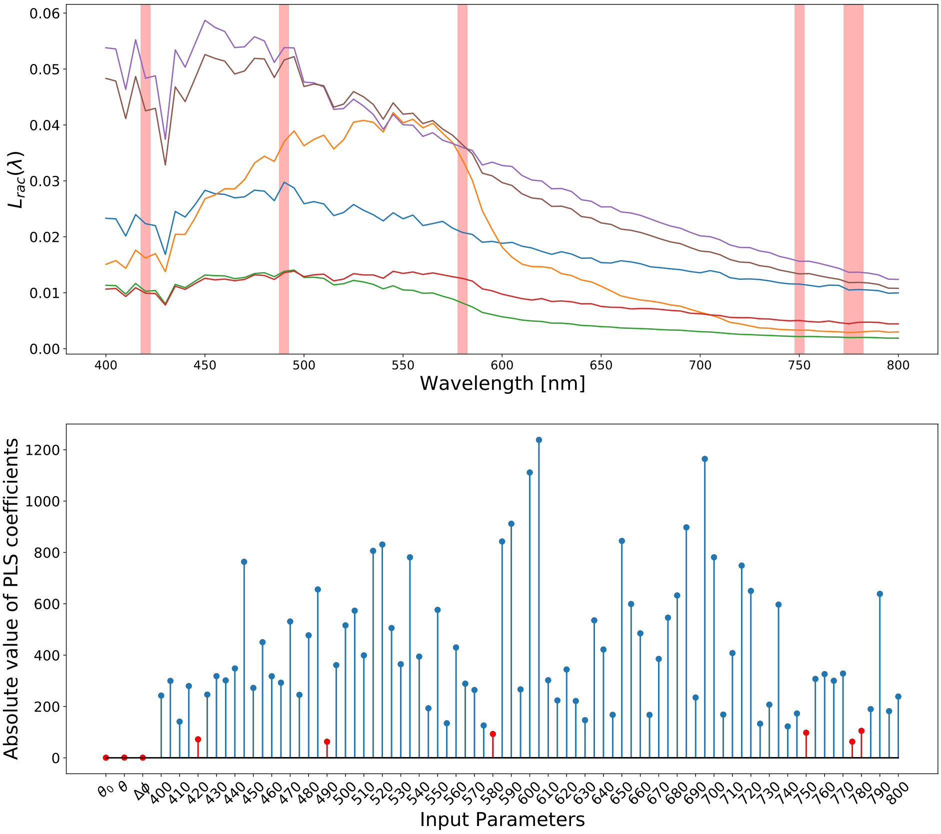


Figure 4.11: Upper plot shows the TOA spectra as a function of wavelength where discarded bands are shown with red vertical lines. The lower plot shows the absolute value of the PLS coefficients for all the input parameters: θ_0, θ, $\Delta\phi$ and 81 wavelength bands noted with the wavelength number. The red and blue dots represent the discarded and used parameters in the PLSR, respectively.

Results from the variable selection with $\mathrm{N}_{\mathrm{train}}$ = 120.000 are shown in Fig. 4.12. The plots show $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ as a function of input features used ($\mathrm{N}_{\mathrm{feat}}$) and number of components ($\mathrm{N}_{\mathrm{comp}}$) when predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ with PLSR. The first, second, third, 100th, 500th and 1000th best value of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ are highlighted in the plots for comparison. Savgol filters were not applied to the data used for training these PLSR models.

One can observe that the best performance (bright boxes in Fig. 4.12) was found for $\mathrm{N}_{\mathrm{feat}}$ between 82 and 84, and for $\mathrm{N}_{\mathrm{comp}}$ between 41 and 84. Nevertheless, from the results it is possible to observe that one can use less input features and number of components and still get performance close to the optimal results. The best and 1000th best performance gave $\overline{\mathrm{R}^2}$ values equal to 0.9759 and 0.9751, respectively. This is was not a very significant difference. However, the best result had $\mathrm{N}_{\mathrm{feat}}$ and $\mathrm{N}_{\mathrm{comp}}$ equal to 69 and 84, whereas the 1000th best result only had $\mathrm{N}_{\mathrm{feat}}$ and $\mathrm{N}_{\mathrm{comp}}$ equal to 55 and 58. In general, the plots show that one can reduce the amount of input features and number for components and still get results close to the optimal results. This can possibly reduce training time and data storage and shows that the optimal number of $\mathrm{N}_{\mathrm{feat}}$ and $\mathrm{N}_{\mathrm{comp}}$ not always is equal to D_1.

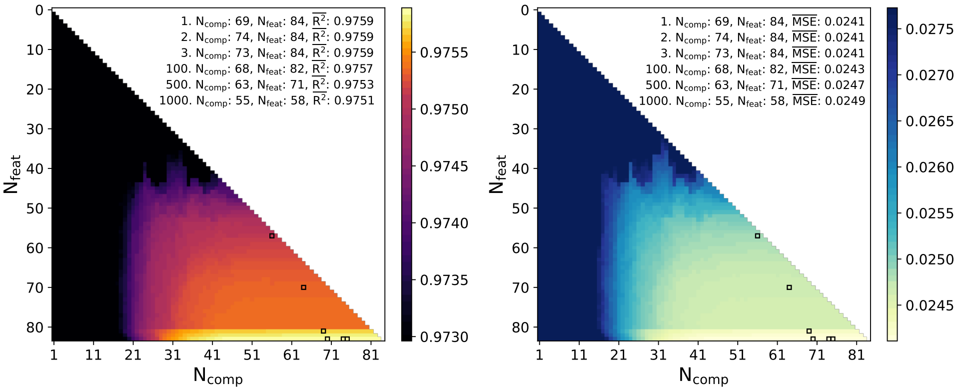


Figure 4.12: $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ as a function of input features used ($\mathrm{N}_{\mathrm{feat}}$) and number of components ($\mathrm{N}_{\mathrm{comp}}$) when predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ with PLSR and no Savgol filters. The first, second, third, 100th, 500th and 1000th best value of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{MSE}}$ are highlighted in the plots for comparison. $\mathrm{N}_{\mathrm{train}}$ = 120.000.

PLSR with Savgol Filter

The results found when using PLSR with component and input feature selection were very similar when training with and without Savgol filters. This was not the case for NN, where the second derivative of the spectra with some specific Savgol filters gave much better results than without. Therefore, the same variation of Savgol filters as used on the NN was tested out for PLSR to investigate how Savgol filters affected the performance. The number of components to use was set to the number of dimensions in the input data (84), all the input features were used to train the model, and the same pre-processing steps as before were applied. The results are shown in Fig. 4.13.

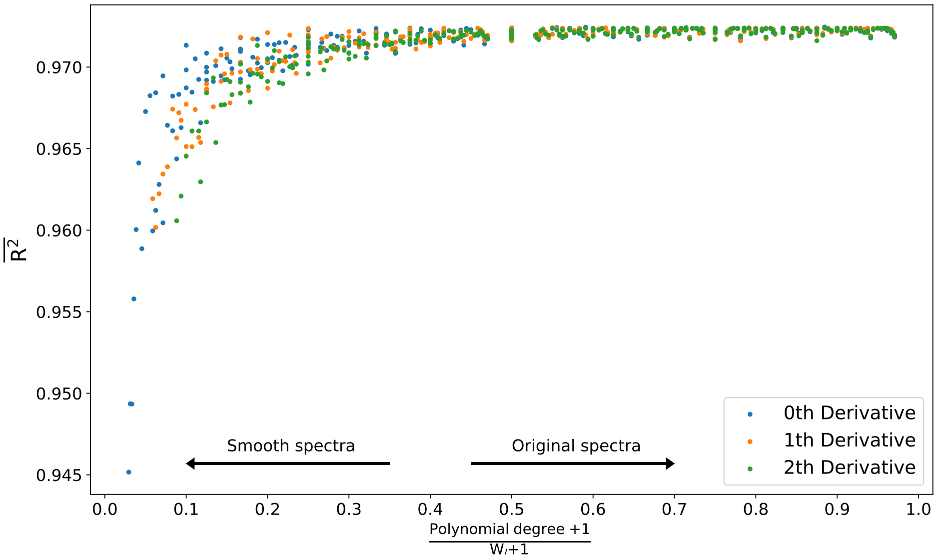


Figure 4.13: $\overline{R^2}$ as a function of the ratio of polynomial degree plus 1 divided by the window length plus 1 for PLSR when predicting R_{rs} from L_{rac}. The window length was tested from 0 to 41, polynomial degree up to W_l - 1 for zeroth, first and second derivative. Ratios close to 1 yield almost no changes in the spectra, whereas ratios close to 0 mean that the spectra are very smooth and the spectral information could have been removed. The best results are found for the first (orange) and second (green) derivative for ratios between 0.15 and 0.4 because this $\overline{R^2}$ values are almost constantly the highest.

The minimum values of $\overline{R^2}$ were found where the polynomial degree was equal to zero or close to it and were found to the left in Fig. 4.13. These regions where the polynomial degrees were small, especially if they were small compared to the window length, were regions where much of the spectral information is smoothed out by the Savgol filter. In the plot, the $\overline{R^2}$ values corresponding to a polynomial degree equal to 0 for the first and second derivative are removed as they gave no reasonable results. Two important things can observed from Fig. 4.13. First, on the contrary to Savgol filters used for NN, the results did not depend appreciable on the derivatives. The figure shows that the $\overline{R^2}$ calculated values stay almost the same for the zeroth, first, and second derivative. The second thing was that the highest values of $\overline{\mathrm{R}^2}$ were almost the same for each Savgol parameter setup, except regions where the polynomial degree was low. These regions are also expected to have worse results as the spectral information is removed.

The same results as shown in Fig. 4.13 are shown in Fig. 4.14, with the same plot layout as for the NN shown in Fig. 4.7. From Fig. 4.14 one can see that the best values of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ are found for a larger range of window lengths and polynomial degrees than for NN because one can find the bright boxes almost all over the plots. In addition, the top values are grouped more and can be found where the polynomial degree is close to the window length and the worst values were the polynomial degrees are small.

In conclusion, the Savgol filter was not an important pre-processing step when using PLSR as the best performance in Fig. 4.13 was found for ratios close to 1, where the Savgol filter would give back the original spectrum. This was also shown in Tab. 4.4 and 4.5 where the results for zeroth, first and second derivative were very similar. Also, the optimal results with NN were better than the optimal results for PLSR.

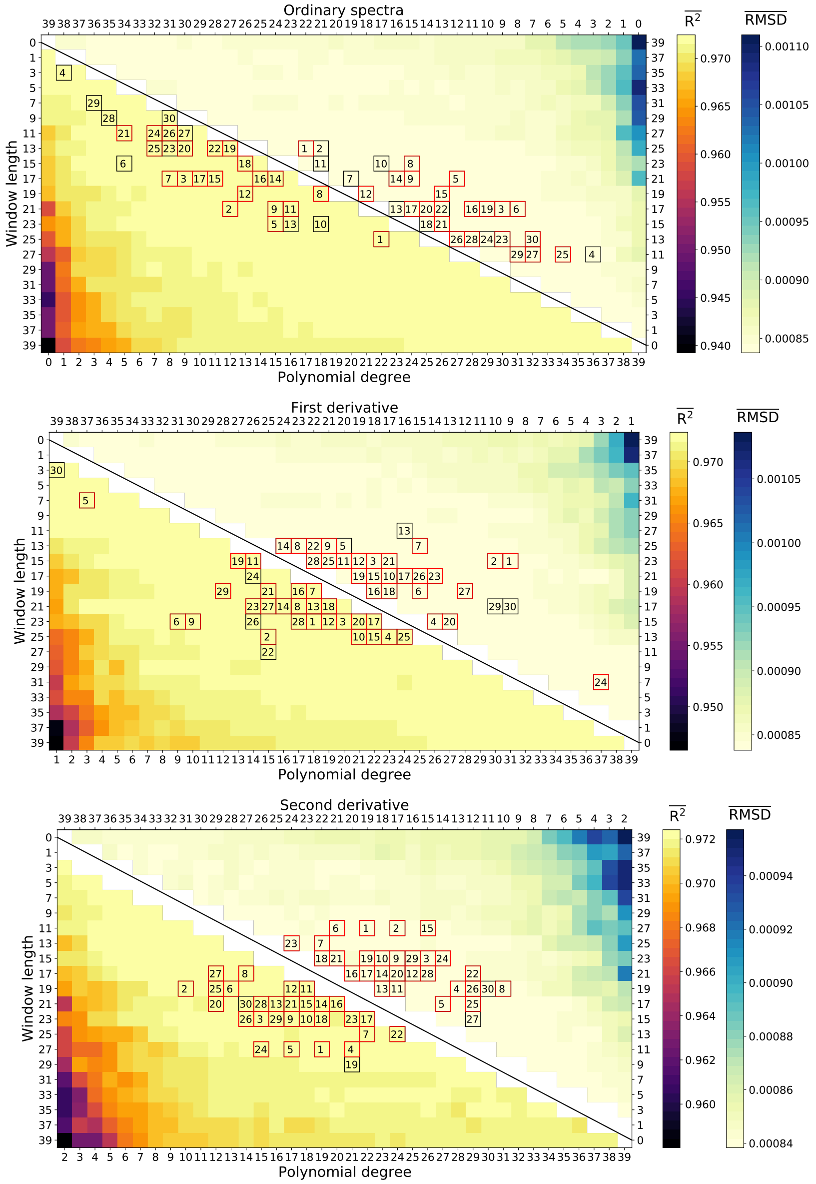


Figure 4.14: Plot of calculated $\overline{R^2}$ and $\overline{RMSD}$ values based on a PLSR approach when predicting R_{rs} from L_{rac} where Savgol filters with different combinations of window length (y-axis), polynomial degree (x-axis) and derivative were applied to the input data. Left and right color bars represent $\overline{R^2}$ and $\overline{RMSD}$, respectively. $\overline{R^2}$ and $\overline{RMSD}$ values are found in the lower left and upper right part of the plots, respectively. The top 30 best values for both $\overline{R^2}$ and $\overline{RMSD}$ are marked with a black box and the number within ranks the values. The red boxes represent values that are among top 30 in both $\overline{R^2}$ and $\overline{RMSD}$.

4.4.3 Support Vector Regression

Support Vector Machine is an ML model which can be used both for linear and non-linear classification and regression. SVR has also shown promising results for AC [80]. The NN which could account for non-linearity, did perform better than the linear model PLSR. Therefore, it would be interesting to study the performance of SVR.

SVR model

The general principle of Support Vector Machine is to construct a hyperplane or a set of hyperplanes in a high or even infinite-dimensional space to separate or fit the input data points. For the Support Vector Classification (SVC), it is desired to adjust the hyperplane in such a way that the data points in each class are as far away from the hyperplane as possible. The Support Vector Regression model, discussed and used in this thesis, is the ϵ-SVR model which has many similarities with SVC. The goal of the ϵ-SVR is to find a function f($\mathbf{X}$, $\mathbf{w}$) with at most ϵ deviation from actual targets $\mathbf{y}_i$ for all the training data and also keep the $\mathbf{w}$ values small (flatness). In some cases this would not be feasible, as there will be some errors larger than ϵ. Analogously to the soft margin loss function in SVM, one can introduce so-called slack variables (ξ_i, ξ_i^*) to cope with the infeasible constraints that could occur. All together can be formulated as a convex optimization problem aiming to minimize a given objective function ($\tau(\mathbf{w})$) with the slack conditions subjected to some constraints:

$$\begin{gathered}\underset{\mathbf{w}\in\mathcal{H},b\in\mathbb{R}}{\arg\min}\ \tau(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}(\xi_i+\xi_i^*)\\ \text{subjected to} = \begin{cases} y_i - \langle w, x_i\rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi, \xi_i^* \geq 0 \end{cases}\end{gathered} \tag{4.12}$$

where the constant C > 0 determines the trade-off between the amount up to which deviations larger than ϵ are tolerated and obtaining flatness in f($\mathbf{X}$, $\mathbf{w}$). This corresponds to dealing with a so-called ϵ-insensitive loss function $|\xi|_\epsilon$ described by

$$|\xi|_\epsilon := \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{otherwise} \end{cases} \tag{4.13}$$

This loss function and the ϵ-SVR model for a linear SVM are shown graphically in Fig. 4.15. The region between the two dashed lines are contributing zero to the loss function because the data points (blue) are never more than ϵ away from the hyperplane (solid black line). The cost function only adds up points outside the dotted lines, marked as red dots in the figure, which is contributing linearly to the loss function, shown in the right part of the figure.

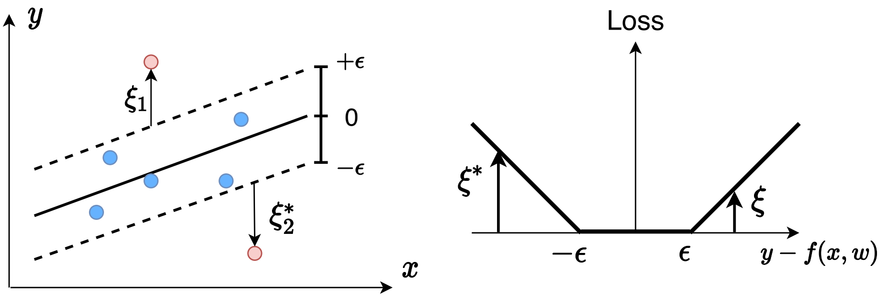


Figure 4.15: Illustration of a linear SVM Regression hyperparameter with corresponding ϵ and ξ constraints (left plot) and the ϵ-insensitive loss function (right plot).

Eq. 4.12 can be somewhat difficult to solve, but it turns out that this problem more easily can be solved with its dual formulation. It will also turn out that dual formulation is the key to solving nonlinear functions. The dual formulation is formulated by constructing a Lagrange function from the constraints and the objective function from Eq. 4.12. When solving the Lagrange problem, the original problem formulation changes to the dual formulation given by

$$\begin{aligned} \text{maximize} &= \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ -\epsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \end{cases} \\ \text{subject to} &= \begin{cases} \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) & = 0 \\ \alpha_i, \alpha_i^* & \in [0, C] \end{cases} \end{aligned} \tag{4.14}$$

and from the Lagrangian conditions a new formulation of f(x) can be written as

$$w = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\Phi(x_i) \text{ and further } f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{4.15}$$

where $\Phi(x_i)$ is a mapping function that maps the input space $\mathcal{X}$ into a higher dimensional feature space $\mathcal{H}$. The idea behind the mapping is to map a sample of data points that can not be separated linearly to a higher dimensional space where a linear separation can be done. The mapping can be computationally expensive, therefore, a different approach known as the kernel trick is normally applied in SVM. k(x_i, x) is known as the kernel function and replaces the linear dot product previously given as $\langle x_i, x\rangle$. The kernel function defines an inner product in the feature space and replaces the mapping $\Phi(x_i)$ and the following dot product operations by a calculation in the input space. The kernel function is defined as k(x, x') := $\langle\Phi(x), \Phi(x')\rangle$. Four different kernels were investigated in this thesis:

- linear := $\langle x, x' \rangle$,
- poly := $(\gamma\langle x, x' \rangle + r)^d$,
- rbf := exp($\gamma||x - x'||^2$),
- sigmoid := tanh($\gamma\langle x, x' \rangle + r$)

where γ > 0 is the kernel coefficient, d is the polynomial degree and r is an independent hyperparameter term. In addition to the ϵ-SVR a linear SVR version will also be tested. These two are found as the sklearn.svm.SVR and sklearn.svm.LinearSVR, respectively. The time complexity for the ϵ-SVR is expected to be between $\mathcal{O}(N_{features} \times N^2_{samples})$ and $\mathcal{O}(N_{features} \times N^3_{samples})$.

SVR Hyperparameter Optimization

SVR with linear, poly, sigmoid, and rbf kernels were all tested with different associated hyperparameters. In general, for all the kernels, no Savgol filter was applied to the input data, but all the other pre-processing steps discussed were applied. Due to the suffering of high time complexity for SVM for a high value of N_{train}, the number of training data points was reduced significantly compared to what was used for both NN and PLSR. The time complexity for the different kernels varied, so the number of training data points used for each kernel was be different and will be specified when they are discussed.

The sigmoid kernel showed significantly poorer performance than the other kernels, both concerning time and accuracy, so that kernel was left out for comparison. The poly kernel showed decent results but the time complexity was extremely high compare to the other kernels. Therefore, the linear and the rbf kernel were tested with the broadest ranges of hyperparameters, as they would take less time. The linear kernel provided coefficients like PLSR did, thus making it a bit more interpretable. The non-linear rbf and poly kernel increased the dimensionality for the input space making it less feasible to understand what is happening. However, this will make it easier to find hyperplanes that can separate the input feature in a better way than for the linear kernel. Some of the main hyperparameters used and their ranges are shown in Tab. 4.6.

Table 4.6: SVR hyperparameters and associated ranges.

Parameter	Range
γ	10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, "scale"
ϵ	10^{-1}, 10^{-2} , 10^{-3}, 10^{-4}
C	10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5

Linear kernel

The SVR model with linear kernel was trained and validated with $0.95\,\%$ and $0.05\,\%$ of 15000 data points, respectively. The two loss functions epsilon insensitive and squared epsilon insensitive were tested together with different C values shown in Tab. 4.6. The ϵ values were 0.5, 0.1, 0.01, 0.001 and 0.0. The results with respect to $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ from the linear kernel hyperparameter optimization are shown in Fig. 4.16. The left and right plots are the results using epsilon insensitive and squared epsilon insensitive, respectively. As before, the best values are ranked and highlighted with black boxes.

The results show that both loss functions gave almost the same results for the best values of $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$, which were found mainly for C = 10 and 100, for almost all ϵ values. The best result was found for the squared epsilon insensitive loss function, with C and ϵ equal to 100 and 0.0, respectively, which gave $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ equal to 0.973 and 0.187, respectively.

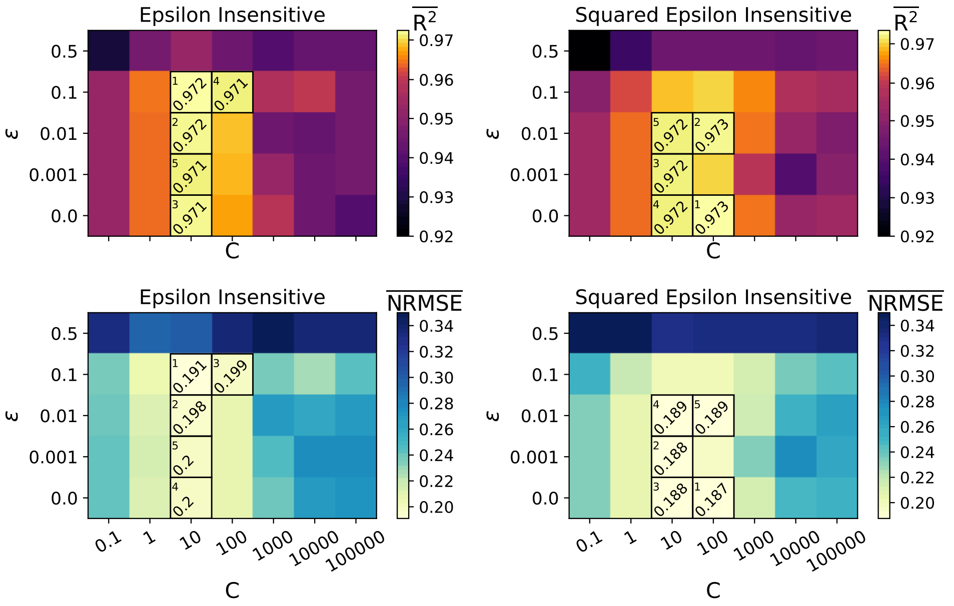


Figure 4.16: Calculated $\overline{\mathrm{R}^2}$ (top plots) and $\overline{\mathrm{NRMSD}}$ (bottom plots) using SVR with the linear kernel for various combinations of ϵ (y-axis) and C (x-axis). The best five values for each hyperparameter combination are highlighted with a black box with the metric value written in it. $\mathrm{N}_{\mathrm{train}}$ was set to 15000.

Poly kernel

The SVR model with poly kernel was trained and validated with $0.95\,\%$ and $0.05\,\%$ of 8000 data points, respectively, with the values of C and γ as shown in Tab. 4.6. ϵ was set to 0.1, r = 0.0 and 1.0 d = 2.0, 3.0 and 5.0. The results for $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ from the poly kernel hyperparameter optimization are shown in Fig. 4.17. When the results were calculated it was observed that similar results were found for the different values of d. Therefore, only d = 2 was plotted, as d = 3 and d = 5 would yield almost the same results. d = 2 was chosen as this gave the smallest training time. The two top values ranked as number 1 and 2 in the two rightmost plots used approximately 755 seconds each when d = 2. For d = 3 and d = 5, they used around 1000 seconds. On the contrary, the results are significantly improved when changing r = 0.0 to r = 1.0. This can be seen in the left and middle plots in Fig. 4.17. The two best results were found for C = 100 and γ = 0.001 and for C = 1000 and γ = 0.0001, where $\overline{\mathrm{R}^2}$ = 0.985 and 0.977, respectively. These two combinations also gave the best values for $\overline{\mathrm{NRMSD}}$ shown in the right plot in Fig. 4.17.

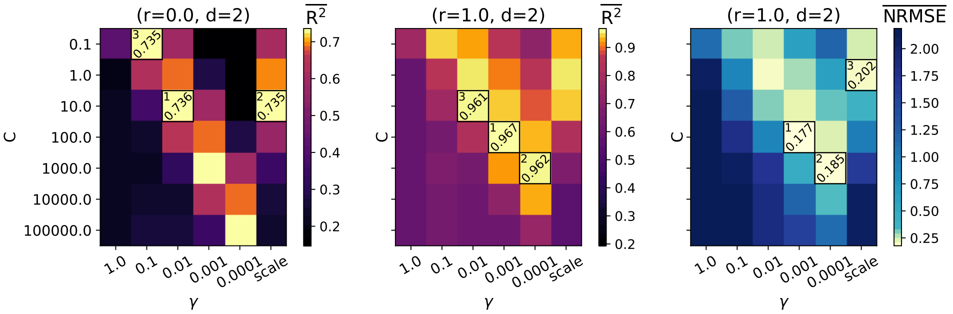


Figure 4.17: Calculated $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ using SVR with the polynomial kernel for various combinations of C (y-axis) and γ (x-axis) with ϵ = 0.1. Different values of r and d are highlighted it the title. The best three values for each hyperparameter combination are highlighted with a black box with the metric value written in it. $\mathrm{N}_{\mathrm{train}}$ was set to 8000.

rbf kernel

The SVR model with rbf kernel was trained and validated with $0.95\,\%$ and $0.05\,\%$ of 15000 datapoints, respectively, with different combinations of C, γ and ϵ values as shown in Tab. 4.6. The calculated $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ are shown in Fig. 4.18. The title refers to the ϵ value used, and the x- and y-axis represent γ and C, respectively. Light boxes represent good results with respect to the metrics, small for $\overline{\mathrm{NRMSD}}$ and high for $\overline{\mathrm{R}^2}$. The best value with respect to $\overline{\mathrm{R}^2}$ was found to be 0.985 for ϵ = 0.1, γ = 0.001 and C = 1000. The best value with respect to $\overline{\mathrm{NRMSD}}$ was found to be 0.126 for the same hyperparameters. The best performance was found for C between 100 and 1000, γ between 0.01 and 0.0001 for all the tested values of epsilon. The overall best results were found for ϵ = 0.1.

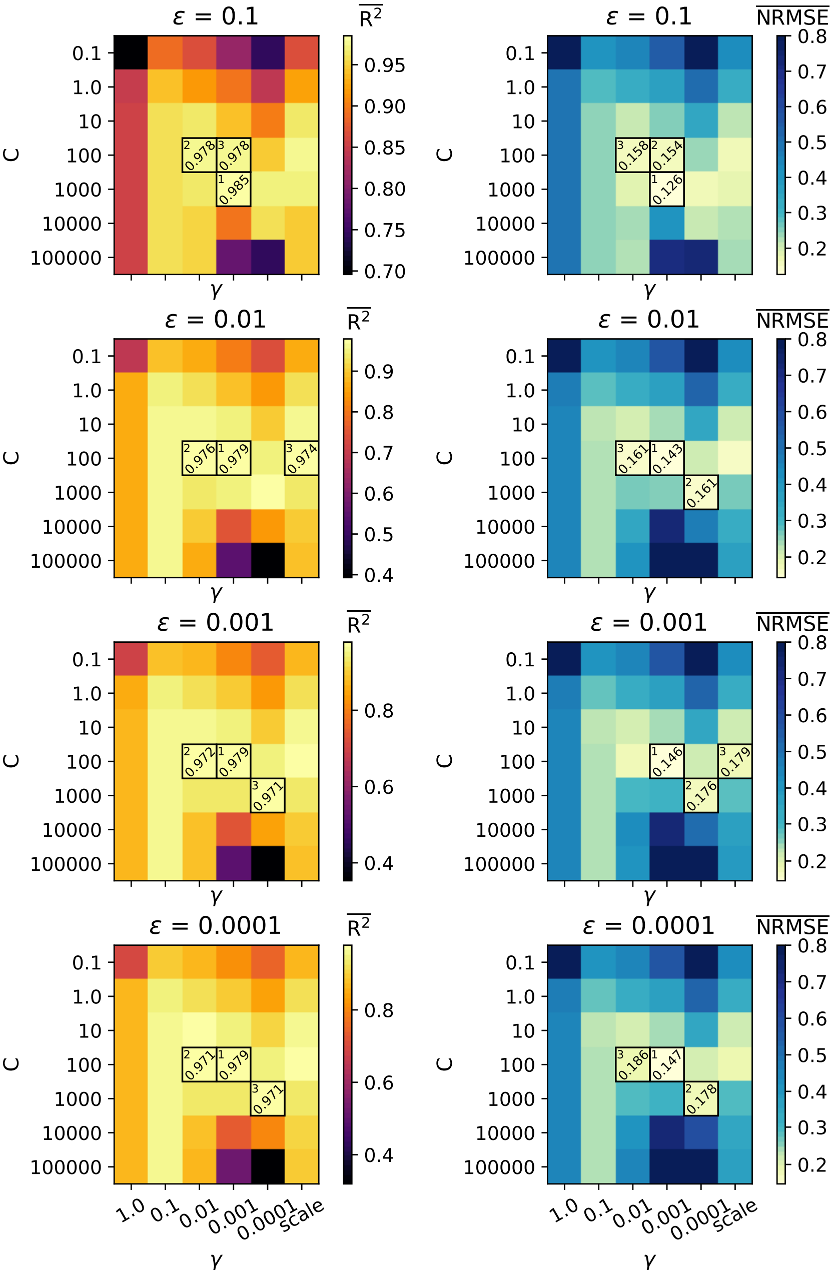


Figure 4.18: $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ calculated from SVR with rbf kernel for various combinations of C (y-axis), γ (x-axis) with $\epsilon = 0.1, 0.01, 0.001$ and 0.0001 (titles). The best three values are highlighted with a black box with the metric value written in it. $\mathrm{N}_{\mathrm{train}}$ was set to 12000.

Kernel Comparison

One last regression for each kernel was done using the optimal hyperparameters found, but now for the same number of N_{train} to get a more representative time comparison. The optimal results for each kernel with N_{train} = 14250 is shown in Tab. 4.7. The results show that the rbf kernel was performing best for all the metrics. This will also be the kernel used to represent SVR when the different ML models are compared in chapter 5. The effects of Savgol filters were not tested for SVR, simply because of the time complexity of the model.

Table 4.7: Optimal results using SVR with rbf, linear and poly kernel with N_{train} = 14250 based on $\overline{\text{R}^2}$, $\overline{\text{NRMSE}}$ and T_{fit}, when predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$.

Metric	**rbf**	**poly**	**linear**
$\overline{\text{R}^2}$	0.978	0.970	0.972
$\overline{\text{NRMSE}}$	0.1667	0.228	0.191
T_{fit} [min]	61.3	92.9	190.0

4.4.4 Stochastic Gradient Descent Regression

Stochastic gradient descent regression is an iterative method for optimizing an objective function with suitable smoothness properties [81]. As for the other ML methods, considering the problem of minimizing an objective function where this objective function is given by [82]:

$$E(\mathbf{w}) = \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} L(\mathbf{y}_i, f(\mathbf{x}_i)) + \alpha R(\mathbf{w}) \tag{4.16}$$

where L is a loss function, like (4.10), where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the training input and output, respectively, and f($\mathbf{x}_i$) is a linear prediction function, given by (4.9). α is a hyperparameter as used in Ridge Regression and R(w) is a regularization term that penalizes model complexity. For a given training dataset D = $\langle(\mathbf{X}_i, \mathbf{y}_i)\rangle_{i=1}^{N}$, the ML model aims to adapt the weights **w** to minimize a given loss function (L(**w**; D)). SGD tries to find the minimum of the loss function by looking at the gradients of the loss function for small batches of the data. The gradient of the loss function will give a clue on which direction to search for new **w** to update the model. The model will start with an initial guess for the weights, then compute the gradient of the loss function and whether the gradient is ascending or descending, the new weights will decrease or increase corresponding to the sign of the gradient. This is also known as gradient descent, and the updated values of **w** will follow the general principle mathematically given by:

$$\mathbf{w} = \mathbf{w} - \eta \nabla_{\mathbf{w}} (\frac{1}{m} \sum_{i=1}^{m} L(\mathbf{y}_i, f(\mathbf{x}_i)) + \alpha R(\mathbf{w})) \tag{4.17}$$

where $\nabla_{\mathrm{w}} E(w)$:= $[\frac{\partial(L+\alpha R)}{\partial w_1}, ..., \frac{\partial(L+\alpha R)}{\partial w_K}]$ for K weights. η is a non-negative hyperparameter known as the learning rate and m is the number of data points chosen in a small batch [82]. The choice of η affects how many iterations to do before finding the minimum of the loss function. However, a large value of η may lead to divergence, and a too small value can lead to too slow convergence. The choice of the learning rate for the regression model, was the inverse scaling given by [82]

$$\eta^{(t)} = \frac{eta_0}{t^{power\_t}} \tag{4.18}$$

where t is the time step (given as the product of the number of data points and iteration taken to reach the stopping criterion) and eta_0 and power_t are hyperparameters. These two values were chosen to be 0.01 and 0.25, respectively, which were also the default values [82]. How different η values are affecting the process of finding the minimum value of the loss function, is shown in the right plot in Fig. 4.19. The plot shows that a large (blue) and small (red) value of η could result in diverging or very slowly converging, whereas a good value, marked in orange, quickly lead to optimal results.

The different loss functions investigated were squared loss, Huber, epsilon insensitive, and squared epsilon insensitive. The squared loss refers to the ordinary least squares fit, and Huber is a modification of squared loss that will focus less on getting outliers. This is done by switching from squared to linear loss for values higher than epsilon. Epsilon insensitive, the loss function used in SVR, ignores errors to a lesser extent than epsilon, and is linear after that. Squared epsilon is insensitive, epsilon insensitive, but becomes squared loss past tolerance of epsilon [82]. The regularization terms included were:

- L2 norm: R(w) := $\frac{1}{2}\sum_{i=1}^{N} w_i^2$
- L1 norm: R(w) := $\sum_{i=1}^{N} |w_i|$,
- Elastic Net norm: R(w) := $\frac{\rho}{2}\sum_{i=1}^{N} w_i^2 + (1-\rho)\sum_{i=1}^{N} |w_i|$

where the L1 norm would lead to sparse solutions and the Elastic Net is a convex combination of L2 and L1. Here, ρ is given by 1 - L1_ratio and L1_ratio=0 would correspond to L2 penalty and L1_ratio=1 would yield L1. L1 and L2 are often referred to as the Lasso Regularization and Ridge Regression. An illustration of the different regularization terms is shown in Fig. 4.19.

Some of the advantages of SGD are the efficiency and ease of implementation. SGD is often used for spare and large-scale machine learning problems. In this case, the amount of data was around 100.000, which could fit SGD better than the other ML approaches when it comes to computational time. However, SGD requires a large number of hyperparameters.

SGDR Hyperparameter optimization

A set of different hyperparameters specific for SGDR was investigated and tested. The loss functions and penalties already mentioned were tested in addition to different parameter ranges for α, ϵ and η given in Tab. 4.8.

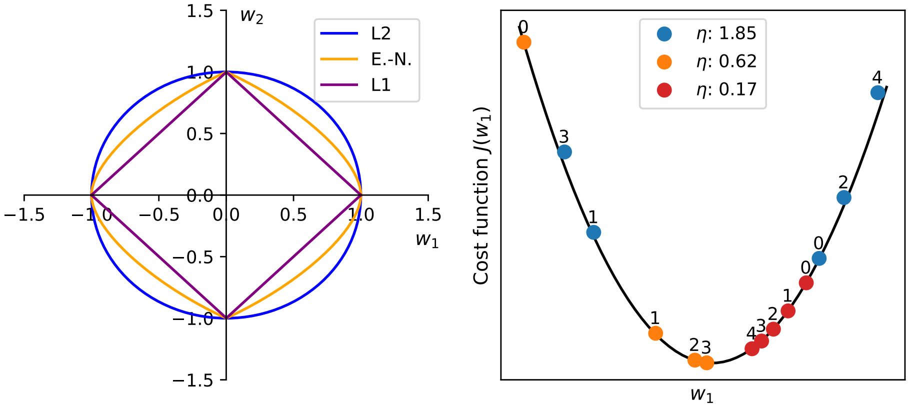


Figure 4.19: Plot of the L2, Elastic-Net, and L1 regularization terms where R(w) = 1 (left plot) and the cost function with gradient descent for three different η values (right plot). η equal to 1.85, 0.62, and 0.17 illustrates a diverging, slowly converging, and fast converging gradient descent, respectively. The numbers indicate which time step (t) of the iteration the gradient descent is on.

Table 4.8: SGDR hyperparameters (loss function, penalty, α, η and ϵ) and the ranges tested for hyperparameter optimization.

Parameter	Range
Loss	Squared Loss, Huber, Epsilon Insensitive, Squared Epsilon Insensitive
Penalty	L1, Elastic-Net, L2
α	10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}
η	10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}
ϵ	10^{0}, 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4}, 10^{-5}

The SGDRegressor from sklearn.linear_model was used to fit the models to the input data. For the hyperparameter testing, 120.000 data points were used of L_{rac} as input and R_{rs} as output with the pre-described pre-processing steps. Savgol filters were left out for this investigation, as there were enough other hyperparameters to check. Different Savgol filters were tested on the optimal hyperparameters found here and discussed in the next subsection. A total of 2520 regressions were done based on the different combinations of hyperparameters and $\overline{\mathrm{R}^2}$ was calculated for each one of them. After the calculations, it was interesting to investigate which hyperparameters that would influence the result most. Checking the influence of the choice of loss functions was then tested. There are four different loss functions to investigate. Consider now the four $\overline{\mathrm{R}^2}$ values calculated for one specific combination of all the other hyperparameters, for instance penalty=L1, α=0.1, η=0.1 and ϵ=1. The standard deviation for these four values would then be calculated. If all these $\overline{\mathrm{R}^2}$ values would be almost the same, the standard deviation would be small meaning that the choice of the loss function would not affect the outcome that much. The standard deviation for all other combinations of hyperparameters would then be calculated and divided on the number of combinations, referred to as the mean standard deviation given by:

$$\overline{\sigma}_i = \frac{\sum_{j\neq i\in\mathcal{H}}^{\mathrm{N_{comb}}}\left(\sqrt{\frac{1}{\mathrm{N}_i}\sum_{k=1}^{\mathrm{N}_i}(x_{j,k}-\overline{x}_j)}\right)}{\mathrm{N_{comb}}} \tag{4.19}$$

where subscript i indicated the i-th hyperparameter, $\mathcal{H}$ is the set of hyperparameters, $\mathrm{N_{comb}}$ is the number of combinations of hyperparameters without hyperparameter i, N_i is the number of different values of hyperparameter i and $x_{j,k}$ is the metric number k in combination j.

The mean standard deviations were calculated for the loss functions, penalties, α, η and ϵ for comparison. This is shown in Tab. 4.9. The results show that the choice of loss function affects the $\overline{\mathrm{R}^2}$ the most. Then α is also an important factor to consider, whereas ϵ stands out as the least import factor. Therefore, the chosen range of ϵ investigated seems to find a minimum for the loss function for all the values.

Table 4.9: The importance of the different SGDR hyperparameters based on the tested ranges shown in Tab. 4.8. The mean standard deviation ($\overline{\sigma}_i$) of one hyperparameter is based on how R^2 changes when this hyperparameter is changed and the other are held fixed. High and low values of $\overline{\sigma}_i$ indicate that this hyperparameter is important and less important for the result.

Hyperparameter	**Mean Standard Deviation**
Loss Function	0.151
Penalty	0.0153
α	0.0553
ϵ	0.00268
η	0.00166

Out of the 2520 regressions covering all the different combinations of hyperparameters, the 30 best values based on $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ were found. All these values had used epsilon insensitive as the loss function. 11, 13, and 6 of them used L1, Elastic-Net, and L2 as a penalty, respectively. The distribution of α, ϵ and η is shown in histograms, Fig. 4.20 and shows the range and frequency of these hyperparameters as a function of the three penalty functions. The best value of $\overline{\mathrm{R}^2}$ was found for the regression with squared epsilon insensitive loss function, L1 as a penalty, and α, ϵ and η to be 1e-05, 1e-05 and 0.001, respectively. However, the best and 30th best value of $\overline{\mathrm{R}^2}$ was calculated to be 0.951 and 0.948, respectively, so all combinations shown in Fig. 4.20 gave comparable results. The best and 30th best values of $\overline{\mathrm{RMSD}}$ were 0.00117 and 0.00120, respectively. From the figure, it can be seen that only three of the α values are present, which limits the choice of α to use. On the contrary, for the η, five of the η values are present, meaning it is not too important which η to choose. Both these observations also fit with the results from Tab. 4.9.

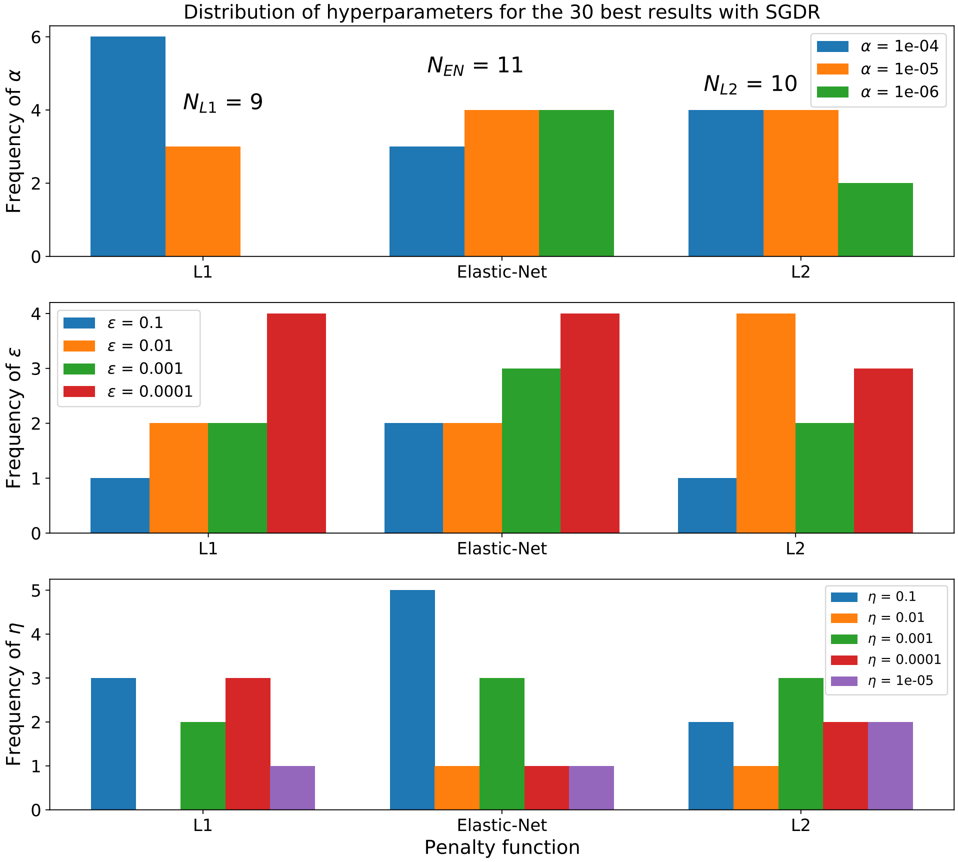


Figure 4.20: Histogram with the distribution of α, ϵ and η that shows the range and frequency of these hyperparameters as a function of the three penalty functions for the top 30 best regressions with respect to $\overline{\mathrm{R}^2}$. All the regressions used the squared epsilon insensitive loss function.

SGDR with Savgol filter

In a similar manner as for both NN and PLSR, the same variation of Savgol filters was tested out for SGDR. The choice of hyperparameters was the same that gave the optimal result in the previous subsection (4.4.4). The Savgol results are shown in Fig. 4.21 and Fig. 4.22. The minimums of $\overline{\mathrm{R}^2}$ were found where the polynomial degree was equal to zero or close to it, as before. In the plots, the $\overline{\mathrm{R}^2}$ values corresponding to a polynomial degree equal to 0 for the 1. and 2. derivative are removed as they gave no reasonable results. The results show that SGDR is affected by the Savgol filter in a similar manner as for NN. The regions in Fig. 4.22 is a bit comparable to the results found for NN in Fig. 4.7. In these two plots, for the zeroth derivative for both NN and SGDR, the results are very similar for all combinations of polynomial degrees and window lengths. The main difference is that the lowest value of $\overline{\mathrm{R}^2}$ for NN is found a bit randomly, whereas these values for SGDR are clearly found for small polynomial degrees. In general, the regions with the lowest and highest $\overline{\mathrm{R}^2}$ values are more prominent for SGDR than for NN, which can be seen in Fig. 4.22 for the bright and dark areas. However, the behavior of $\overline{\mathrm{R}^2}$ in the region where the ratio was between 0 and 0.2 was more comparable to the results found with the other linear model PLSR, shown in Fig. 4.13. The two linear models seem to struggle more when the spectra are smoothed out.

Now, consider the results based on the first and second derivatives, where the performance is better for both SGDR and NN compared to the zeroth derivative. The areas where the optimal results are found for both NN and SGDR, are comparable, and are found for almost the same ratios. For the NN model, the optimal ratio region is between 0.15 and 0.4 for the first and second derivative, whereas for the SGDR it is between 0.25 and 0.45. For these two regions, all the values of $\overline{\mathrm{R}^2}$ are among the highest values. In addition, the $\overline{\mathrm{R}^2}$ values in this ratio regions are almost constant. Both the first and second derivatives gave $\overline{\mathrm{R}^2}$ values equal to 0.967 ± 0.0003 for all ratios between 0.25 and 0.45. For comparison, $\overline{\mathrm{R}^2}$ values for the second derivative were found between 0.967 and 0.905 for ratios between 0.5 and 1.0. This random behavior of $\overline{\mathrm{R}^2}$ values for ratios higher than 0.5 was also found for NN.

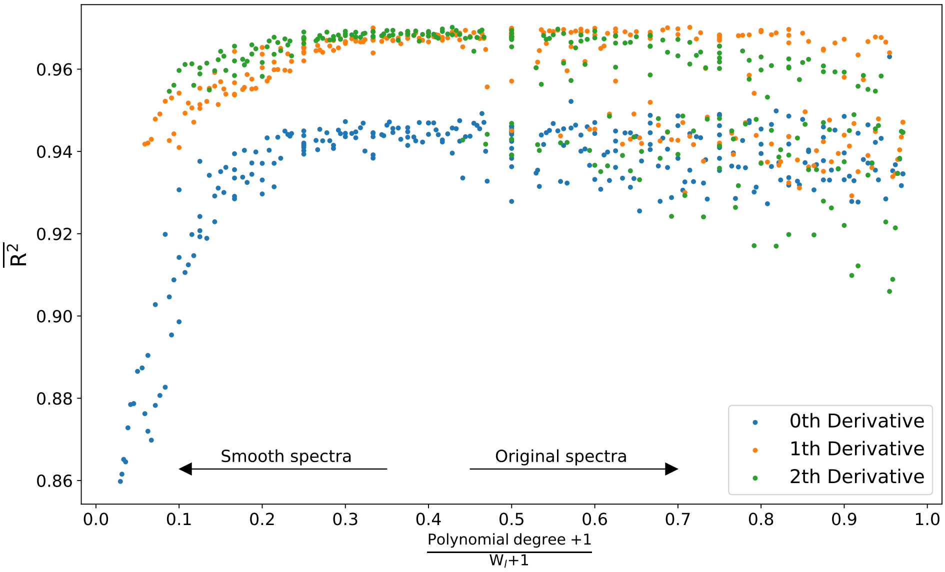


Figure 4.21: $\overline{\mathrm{R}^2}$ as a function of the ratio of polynomial degree plus 1 divided by the window length plus 1 for SGDR when predicting R_{rs} from L_{rac}. The window length was tested from 0 to 41, polynomial degree up to W_l - 1 for zeroth, first and second derivative. Ratios close to 1 yield almost no changes in the spectra, whereas ratios close to 0 mean that the spectra are very smooth and the spectral information could have been removed. The best results were found for the first (orange) and second (green) derivative for ratios between 0.25 and 0.45 because this $\overline{\mathrm{R}^2}$ values are almost constantly the highest.

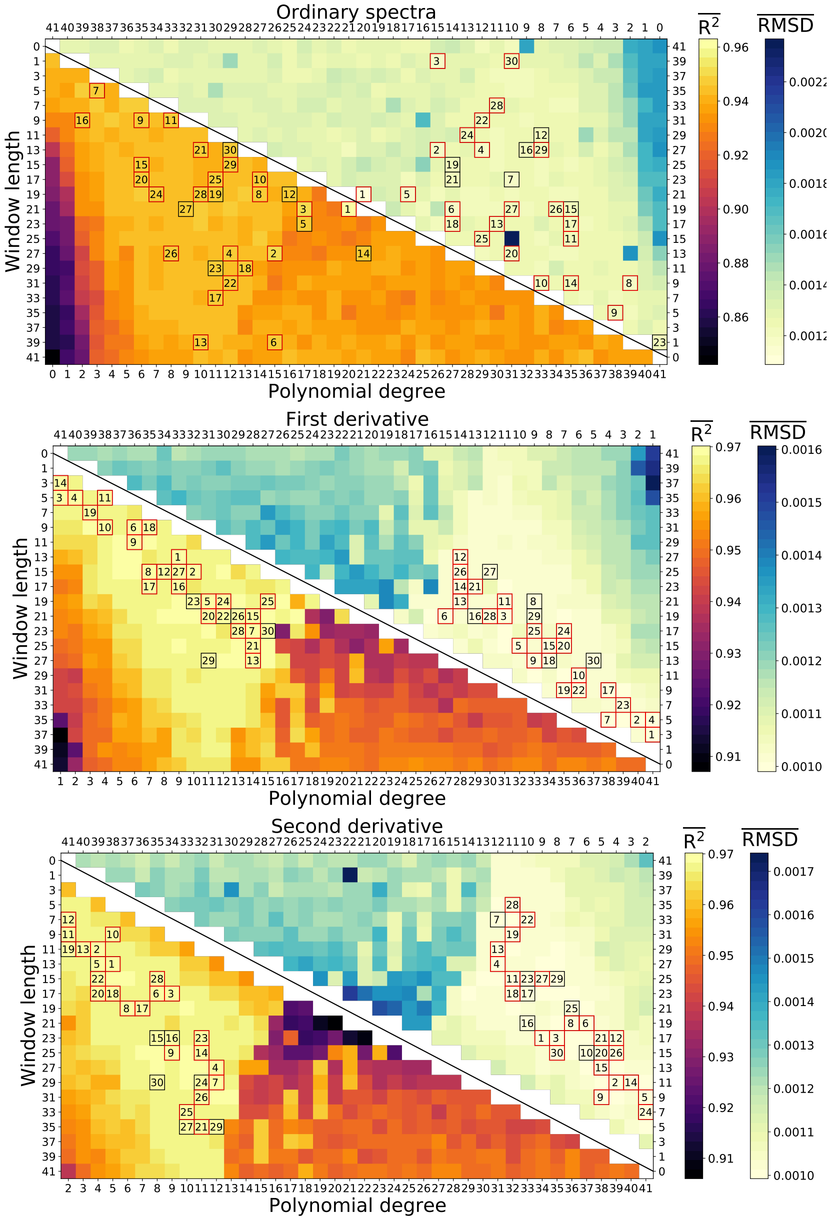


Figure 4.22: Plot of calculated $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ values based on a SGDR approach when predicting R_{rs} from L_{rac} where Savgol filters with different combinations of window length (y-axis), polynomial degree (x-axis) and derivative were applied to the input data. Left and right color bars represent $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$, respectively. $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ values are found in the lower left and upper right part of the plots, respectively. The top 30 best values for both $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$ are marked with a black box and the number within ranks the values. The red boxes represent values that are among top 30 in both $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{RMSD}}$.

Chapter 5

Validation and Discussion

This chapter will present the optimal results of AC and IOP retrieval using the different ML models NN, PLSR, SGDR and SVR. Five statistical metrics were used to evaluate the performance of the different ML models. These metrics were the squared Pearson correlation coefficient (R^2), the average percentage difference (APD), the mean percentage bias, the root mean square difference (RMSD), and the normalized root mean square difference (NRMSD). Metrics, abbreviations, formulas and interpretations are described further in Tab. 5.1.

5.1 Atmospheric Correction Validation and Discussion

Three different AC approaches were tested in this study, each correcting for different atmospheric effects. The correction of L_{rac} to R_{rs} (ML1 in Fig. 4.1) will first be discussed. Then correction of absorption and Rayleigh, where L_{rac} was predicted from L_t (ML2), and full AC of L_t to R_{rs} (ML3) will be discussed. As the main focus in this study was on ML1, the optimization and discussion of the results from approach ML2 and ML3 will not be described with the same detailed level as for ML1.

5.1.1 Atmospheric Correction of L_{rac} to R_{rs}

In this section, the results and discussion of AC of Rayleigh and absorption corrected radiance with all the four ML models will be presented, where the goal was to predict the R_{rs} from L_{rac} (ML1). The choice of the ML specific hyperparameters and Savitzky and Golay filters for this approach was based on the discoveries presented in section 4.4, where hyperparameter optimization based on AC of L_{rac} was done. The angle pre-processing step was also applied. The models are compared against each other concerning the metrics described in Tab. 5.1, time complexity and interpretation capability. One sample of the input $L_{rac}(\lambda)$ and output $R_{rs}(\lambda)$ is shown in Fig. 5.1 to illustrate the spectral differences.

Table 5.1: Metrics used for validation with abbreviation, formula and interpretation. X_i and Y_i are the predicted (forecasted) and simulated (actual) data, respectively. N is the number of data points that is validated.

Metrics	Formula	Interpretation
R	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\bar{X}_i}{\sigma_X}\right)\left(\frac{Y_i-\bar{Y}_i}{\sigma_Y}\right)$	Pearson correlation coefficient measures the linear correlation between two variables X and Y. The value ranges from -1 to 1, where -1 is total negative linear correlation, 0 is no linear correlation, and 1 is total positive linear correlation.
APD	$\frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{X_i-Y_i}{Y_i}\right\rvert \times 100\%$	Averaged percentage difference is the average of the absolute value of the relative change.
Bias	$\frac{1}{N}\sum_{i=1}^{N}\frac{X_i-Y_i}{Y_i} \times 100\%$	Bias is the average of percentage errors, by which forecasts of a model differ from actual values of the quantity being forecast.
RMSD	$\sqrt{\frac{\sum_{i=1}^{N}(X_i-Y_i)^2}{N}}$	Root mean squared difference is an accuracy measure given as the square root of the mean of the squares of the deviations. It is sensitive to outliers and depends on the scale of the numbers used.
NRMSD	$\frac{\mathrm{RMSD}}{Y_{max}-Y_{min}}$	Normalized RMSD is the RMSD divided by the difference of the max and minimum value. This property is not scale dependent and would be useful when comparing different bands.

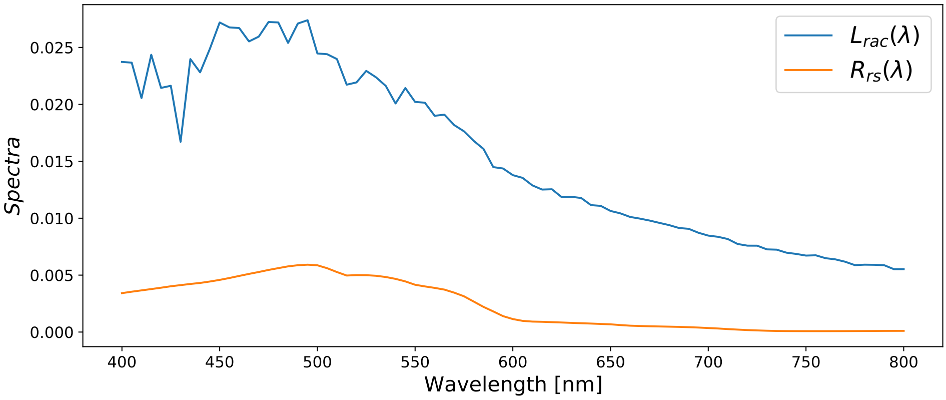


Figure 5.1: One sample of input $L_{rac}(\lambda)$ and output $R_{rs}(\lambda)$ to illustrate the spectral differences.

Validation based on metrics and time complexity

The training data for the ML models included 81 wavelength bands of L_{rac} ranging from 400 to $800\,\mathrm{nm}$ in addition to the three sun-target-sensor angles ($\theta_0, \theta, \Delta\phi$). The predicted output was the remote sensing reflectance for the corresponding 81 wavelength bands. Mean metric values for each ML model were calculated and the optimal results are shown in Tab. 5.2. Tab. 5.2 also shows the number of data points ($\mathrm{N_{train}}$) used for training and time for fitting the models ($\mathrm{T_{fit}}$) and predicting the output ($\mathrm{T_{pred}}$). The prediction time was the time it would take to predict all the validation data divided by the number of validation samples ($\mathrm{N_{val}}$). The results for the corresponding metrics for each wavelength band are also shown in Appendices A.1, A.2, A.3 and A.4.

Table 5.2: Optimal results for AC of Rayleigh and absorption corrected TOA radiance (R_{rac}) with NN, PLSR, SGDR and SVR based on the mean of different metrics. In addition, time to fit the model ($\mathrm{T_{fit}}$), time to predict the output ($\mathrm{T_{pred}}$) and the number of training data ($\mathrm{N_{train}}$) are given.

Metrics	**NN**	**PLSR**	**SGDR**	**SVR (rbf)**	**SVR (lin)**
$\overline{\mathrm{R}^2}$	0.999	0.974	0.968	0.995	0.968
$\overline{\mathrm{APD}}$ [%]	4.42	34.1	22.03	21.4	32.1
$\overline{\mathrm{Bias}}$ [%]	-0.40	9.20	5.70	3.62	11.8
$\overline{\mathrm{NRMSE}}$	0.045	0.197	0.223	0.080	0.229
$\mathrm{T_{fit}}$ [s]	675	166	424	1158	19743
$\mathrm{T_{pred}}$ [s/N]	1.3×10^{-3}	1.03×10^{-4}	1.18×10^{-2}	2.75×10^{-2}	9.66×10^{-3}
$\mathrm{N_{train}}$	91702	91702	91702	14250	14250

The NN model showed best results for all the metrics (Tab. 5.2). The linear models PLSR, SGDR and SVR(lin) performed very similarly for the metrics, even though PLSR performed a bit better for training and fitting time. SVR(rbf) gave better results on $\overline{\mathrm{R}^2}$, $\overline{\mathrm{Bias}}$ and $\overline{\mathrm{NRMSE}}$ than all the linear models, with $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSE}}$ values close to what was achieved with NN. In general, the two non-linear models NN and SVR (rbf) performed better for all the metrics than the linear models.

The training time for SVR was much higher than for all the other models when considering that it only trained on $15.5\,\%$ of the data compared to the other ML models, especially with the linear kernel. However, with much less data, the SVR(rbf) model still got metric values better than the linear models. PLSR was the best with respect to training time ($\mathrm{T}_{\mathrm{fit}}$) and prediction time ($\mathrm{T}_{\mathrm{pred}}$). Nevertheless, the training time of NN could be reduced to comparable times as the linear models, and still give better results concerning the metric values. The time of fitting the NN depended on how many epochs used in the training and the batch size. The NN would have a trade off on accuracy and training time, where higher training time would increase the accuracy. However, once the NN model is fully trained, it can be run very fast, and AC could be done without training the model each time. This would also be the case for the other linear models. SVR on the other side, might spend even more time on the training, as the input data would have to be changed with the non-linear kernel functions. This can also be seen from Tab. 5.2 where SVR(rbf) has the highest prediction time ($\mathrm{T}_{\mathrm{pred}} = 2.75 \times 10^{-2}$ s/N).

Fig. 5.2 presents scatterplots of predicted and simulated R_{rs} for wavelength band 400, 500, 600 and $700\,\mathrm{nm}$. These four wavelength bands were chosen as representative bands to illustrate the responses for individual bands in different parts of the wavelength range. In the figure, each row represents results from one ML model and each column represents one of the four wavelength bands. The orange lines represent where the predicted and simulated data were the same and would indicate R^2 equal to 1.0 and a perfect prediction. The wavelength bands and the corresponding metrics R^2, APD, and NRMSD are highlighted in the bottom right and top left in each plot, respectively. The red and blue dots represent data points classified as Case 1 and Case 2 waters. There are no clear definitions of how to separate the data into Case 1 and Case 2 waters based on the remote sensing reflectance, but for this study, the criterion of dividing the $R_{rs}(\lambda)$ into the two different cases were based on setting spectra with $R_{rs}(665) < 0.0005$ to Case 1 and the rest to Case 2, as done in [33].

From the scatterplots one can observe that the predictions closest to the optimal orange line are found for NN and SVR, which fit with the results shown in Tab. 5.2 where NN and SVR have the highest values of $\overline{\mathrm{R}^2}$. Also, the pattern of the scatterplots for PLSR and SGDR looks very similar, fitting the observation from the similar metric values in Tab. 5.2. What is similar for all the ML models is that the R^2 is lowest for wavelength band $400\,\mathrm{nm}$. Besides, one can observe that the different metric values NRMSD, APD, and R^2 vary with the different wavelength bands. The best values for APD and NRMSD are, for more or less for all the ML models, found at wavelength band $500\,\mathrm{nm}$. The performance of predicting the different bands varies for each ML models and should be investigated further.

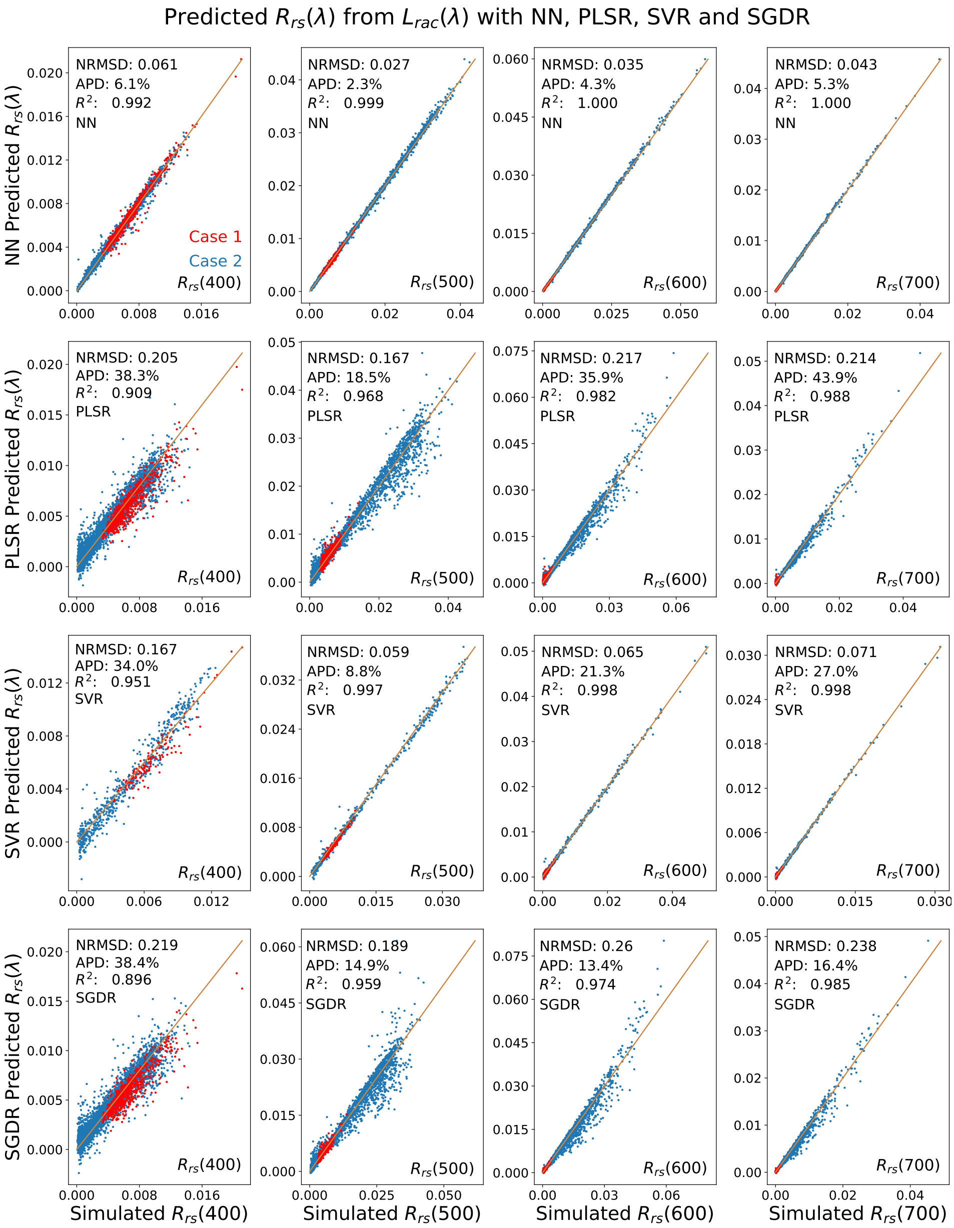


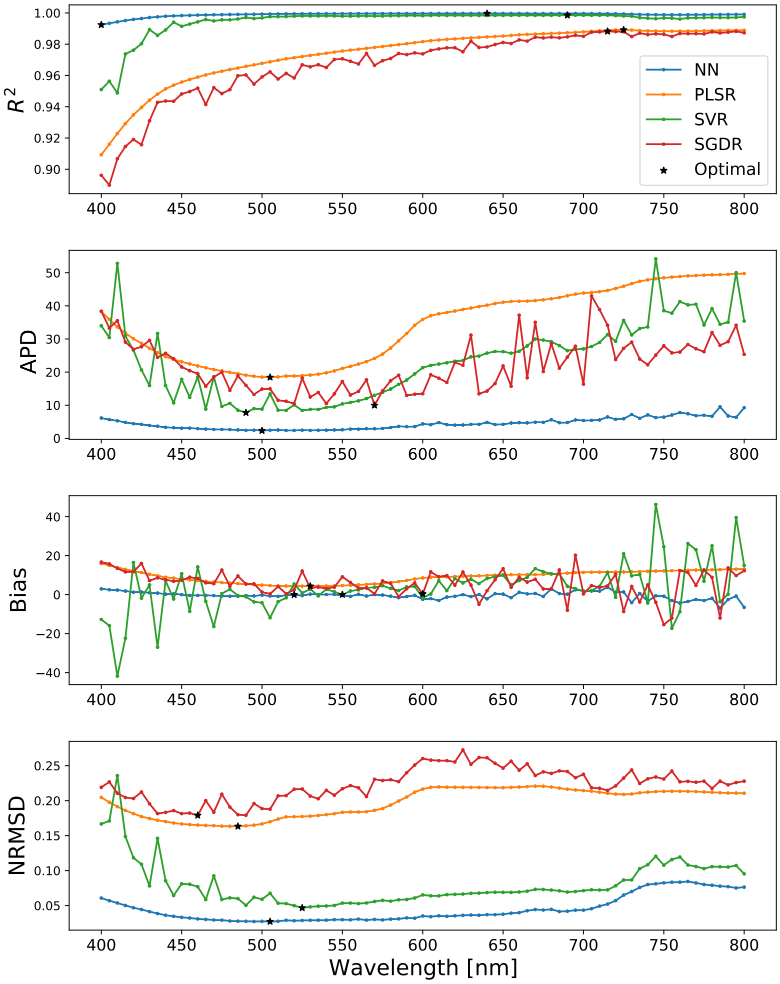
Figure 5.2: Scatterplots of predicted and simulated $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ for wavelength band 400, 500, 600 and 700 nm (indicated with text), with corresponding R^2, APD and NRMSD values. Red and blue dots represent Case 1 and Case 2 waters, respectively.

The different metrics calculated for each ML model as a function of the 81 bands from 400 to 800 nm are shown in Fig. 5.3. Each plot represents one of the metrics R^2, APD, Bias, and NRMSD, which is plotted as a function of the wavelength bands for the different ML models (indicated with different colors). The mean values of the metrics are shown in the tables on the right-hand side of the plot which are the same presented in Tab. 5.2. The best value of the metrics for each ML model in Fig. 5.3 is marked with a black star.

One can observe that the optimal results for all the ML models for APD and NRMSD are found in the wavelength region 470-560 nm. This region of wavelengths is also the region often used for water IOP retrieval algorithms like the OCx described in Eq. 4.1. The green wavelength region is often used because it typically contains information about the chlorophyll concentration. This region is found approximately between 520-560 nm, where the models work well.

The worst performance for APD and NRMSD was found for the longer wavelength bands (650-800 nm) and the very lowest (400-440 nm). For the R^2, the best values are found in the range 550-800 nm and worst in the range 400-440 nm. The reason why the models are struggling a bit more with the range 400-440 nm could be because this region also is most affected by the different ocean constituents. Both absorption and scattering from the different ocean constituents are much higher for short wavelengths compared to the longest wavelengths. This can be seen in Fig. 2.16 where the absorption coefficients for the shortest wavelengths for the coastal water are highly affected by all the ocean constituents, whereas the longest wavelengths are more or less only affected by phytoplankton and pure water.

From Fig. 5.3 it can be observed that the metric values as a function of wavelength bands are spikier for SVR and SGDR compared to the smooth graphs from NN and PLSR. Because of the spiky behavior in the metrics as a function of wavelength bands, it would be expected that the predicted remote sensing reflectance with SVR and SGDR also could share some of the spiky behavior.



	$\overline{R^2}$
NN	0.999
PLSR	0.974
SVR	0.995
SGDR	0.968

	$\overline{APD}$
NN	4.4
PLSR	34.1
SVR	23.4
SGDR	22.0

	$\overline{Bias}$
NN	-0.4
PLSR	9.2
SVR	3.6
SGDR	5.7

	$\overline{NRMSD}$
NN	0.045
PLSR	0.197
SVR	0.08
SGDR	0.223

Figure 5.3: R^2, APD, Bias and NRMSD as a function of wavelength bands with with NN, PLSR, SGDR, SVR for AC of $L_{rac}(\lambda)$. The mean values of the metrics are shown in the tables on the right hand side of the plots.

Visualization of the predicted remote sensing reflectance

A visualization of four randomly predicted and simulated R_{rs} spectra for the different ML models is shown in Fig. 5.4. The plots show that the NN and PLSR models (top plots) can predict the smooth shape of the R_{rs} spectra, whereas the SVR and SGDR models (bottom plots) predicted more spiky spectra, corresponding to the metric values in Fig. 5.3. The spiky spectra share a similar shape as the input radiance spectra, which could be caused by a lack of enough training data. The SVR was only trained on $15.5\,\%$ of the data compared to the other ML models, because of the high scaling of training time with an increasing number of training data. The same spiky behavior was also observed when training the NN with a smaller number of training data (15.000 data points or less).

The predicted spectra with SGDR were a bit smoother than for the SVR, but still struggled to achieve the smoothness found with PLSR and NN. SVR and SGDR got metric values comparable with the other ML models, but were still not able to obtain the smoothness which would be desired to get. Water IOP retrieval algorithms are often based on band ratios, like the OCx algorithm. Calculating band ratios with spiky spectral shape could yield strange ratios, which then could cause inaccurate water IOP predictions. Therefore, it is important to make sure that one have enough training data to avoid such spiky behavior.

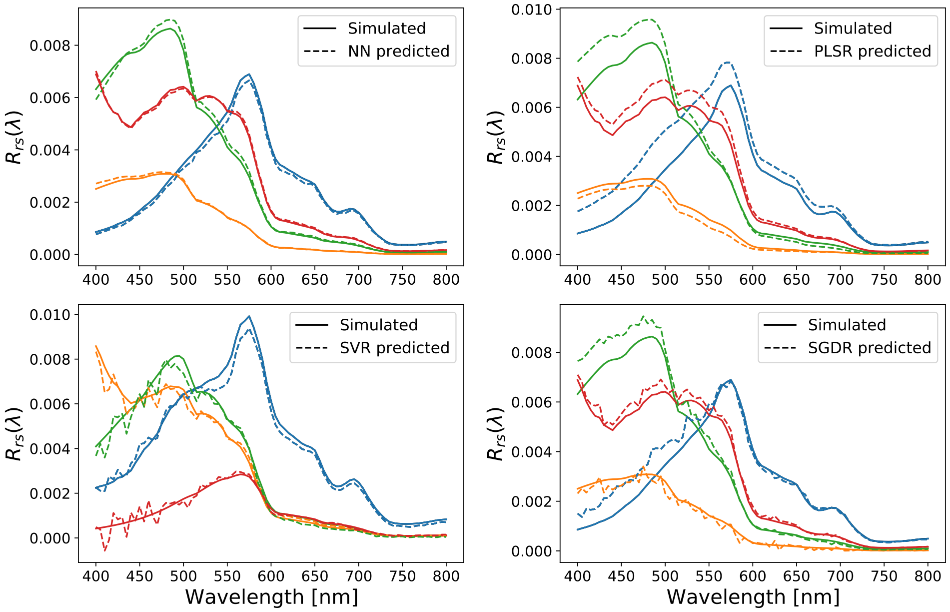


Figure 5.4: Four ML predicted and simulated $R_{rs}(\lambda)$ represented with different colors as a function of wavelength for the four different ML models (highlighted in the labels).

Pre-processing Validation

Pre-processing steps as dividing the input radiance spectra with the cosine of the solar zenith angle and applying Savgol filters have been done to get the optimal results shown in Tab. 5.2. The performance with and without these pre-processing steps for all the ML models with respect to $\overline{\mathrm{R}^2}$, $\overline{\mathrm{APD}}$ and $\overline{\mathrm{NRMSD}}$ is shown in Tab. 5.3. $\mathbf{X}_{\mathrm{prep}}$ and $\mathbf{X}$ represent the dataset with and without pre-processing, respectively. Savgol filters were not used for PLSR, so the only difference between $\mathbf{X}_{\mathrm{prep}}$ and $\mathbf{X}$ for PLSR is angle pre-processing. The "Ratio" column represents how many times better the ML model did it with pre-processed data than without.

The results in Tab. 5.3 show that the performance of all ML models improved for all the statistical metrics when pre-processing was applied. SVR had the second-highest improvement of all the metrics shown in Tab. 5.3 and was the ML model that improved the most in total with pre-processing compared to NN, PLSR, and SGDR. SGDR was the model which improved the least with the lowest ratio for both $\overline{\mathrm{APD}}$ and $\overline{\mathrm{NRMSD}}$. The highest ratio was found for NN with an improvement ratio of 3.55 for $\overline{\mathrm{APD}}$.

Table 5.3: Optimal results with the four different ML models with and without all pre-processing steps. This includes dividing by cosine of the solar zenith angle and Savgol filters. $\mathbf{X}_{\mathrm{prep}}$ and $\mathbf{X}$ represent the dataset with and without pre-processing. Ratio represents how many times better the ML model did it with pre-processed data than without.

	$\overline{\mathrm{R}^2}$			$\overline{\mathrm{APD}}$[%]			$\overline{\mathrm{NRMSD}}$		
ML model	$\mathbf{X}_{\mathrm{prep}}$	$\mathbf{X}$	Ratio	$\mathbf{X}_{\mathrm{prep}}$	$\mathbf{X}$	Ratio	$\mathbf{X}_{\mathrm{prep}}$	$\mathbf{X}$	Ratio
NN	0.999	0.992	1.007	4.42	15.7	3.55	0.0447	0.121	2.71
PLSR	0.975	0.923	1.056	34.1	73.2	2.14	0.197	0.38	1.93
SVR	0.995	0.969	1.027	23.4	62.9	2.69	0.080	0.210	2.63
SGDR	0.968	0.945	1.024	22.0	41.8	1.90	0.227	0.263	1.16

Interpretation of the Results

The best performance was obtained with SVR and NN, but it was difficult to say which features were the most important for the prediction. The SVR model transformed the input data with different kernel functions, increasing the feature dimensionality and also the possibility of understanding which features were most important for the prediction. NN also exhibited a similar problem, where the feature information was hidden in all the neurons and hidden layers. On the contrary, PLSR, SGDR and SVR(lin) provided weighted coefficients as output after the training. These coefficients gave information on which input parameters that contributed the most to the final results.

PLSR, SGDR and SVR(lin) coefficients as a function of input parameters (θ, θ_0 and $\Delta\phi$ and L_{rac} at 81 wavelength bands) are shown in Fig. 5.5. The coefficients were the absolute values of the sum of the coefficients for each band, as described in 4.4.2. The single-colored plots represented with blue, orange, and green colors show the PLSR, SGDR, and SVR(lin) coefficients, respectively. Savgol filter was applied to the input radiance spectra for SGDR, which is highlighted in the upper right corner of the plots, where W_l, Poly, and Der are window length, polynomial degree and derivative as described in 4.3.3. Therefore, the spectra coefficients for the SGDR results were not presenting the importance of the original input spectra, but the second derivative of the original spectra The absolute values of the coefficients for input spectra with another Savgol filter or without could, therefore, look completely different.

The bottom plot shows the sum of all the coefficients to clarify the importance of the different input parameters for each ML model. The plot also clarifies which input parameter was most important for each ML model. One can observe from Fig. 5.5 that the coefficients for θ, θ_0 and $\Delta\phi$ were almost zero, indicating low influence for the prediction. The explanation for this might be that the the input data was divided by the cosine of the solar zenith angle, removing some of the importance of the sun-target-sensor geometries. Also, the input data did not include Rayleigh and absorption contributions, which should be more dependent on the angles. Fig. 5.5 also shows that very few coefficients were almost zero, except for the angles, which meant that almost all the input parameters were necessary to predict the outcome. This could be because the models were predicting almost the same number of outputs (81) as input (84).

However, some of the features were more important than others. Wavelength bands 440, 445, 515, 585, 590, 600 605, 650, 695 and $700\,\mathrm{nm}$ had highest total coefficient values (bottom plot). The region 585-$610\,\mathrm{nm}$ was especially important. From Fig. 4.3, one can see that both L_{rac} and R_{rs} have sudden incremental changes in this wavelength region, indicating that this could be important input features. Also, one would seldom observe many small subsequent coefficient values, as a small coefficient often is found besides larger neighboring coefficient. This is most probably because neighbouring wavelength bands are highly correlated with each other. A lot of the spectral information can therefore be extracted from one band, as the neighboring bands most probably would posses much of the same information. This can be seen for PLSR, SGDR and SVR(linear) at wavelength band 490, 460 and $490\,\mathrm{nm}$, respectively.

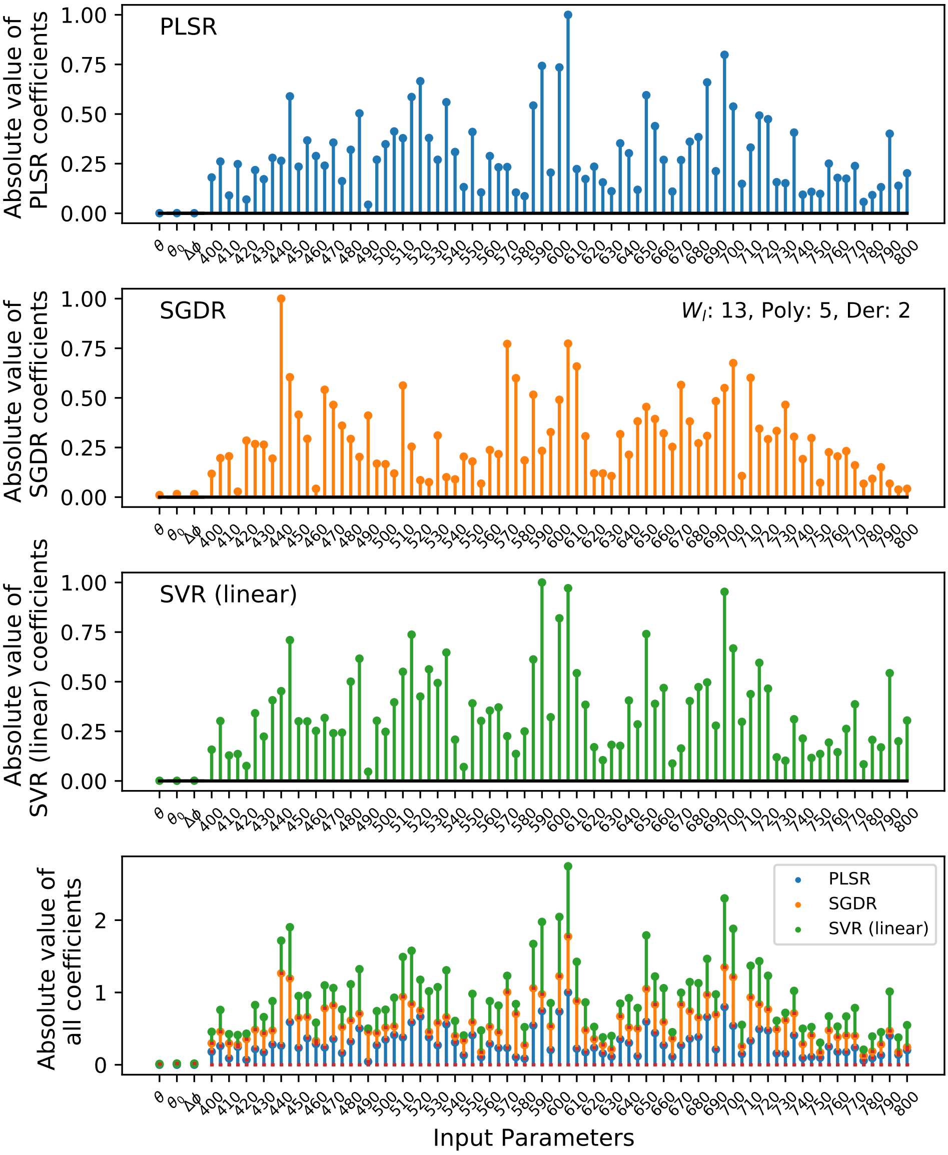


Figure 5.5: Absolute values of the PLSR, SGDR, and SVR (linear) coefficients as a function of the different input parameters: θ_0, θ, $\Delta\phi$ and 81 wavelength bands noted with the wavelength number. High values indicate that the input parameter is important for the prediction. The numbers in the x-axis represent the different λ values in (L_{rac}). Savgol filter was applied to the input radiance spectra for SGDR, which is highlighted in the upper right corner of the orange plot, where W_l, Poly, and Der are window length, polynomial degree and derivative as described in 4.3.3.

Case 1 and Case 2 Validation and Discussion

The AC approach was intended to work on both Case 1 and Case 2 waters. The results presented so far were validated on a combination of data classified as both Case 1 and Case 2 water. It would therefore be interesting to see how the AC algorithms would perform on the data classified as Case 1 and Case 2 waters, separately, by classifying spectra with $R_{rs}(665)$ < 0.0005 to Case 1 and the rest to Case 2, as done in [33]. The ML models trained on the same training data as discussed so far, but the validation data would be split into Case 1 and Case 2 water data and validated thereafter.

The results are shown in Tab. 5.4 and show that the ML models predicted significantly better for the data classified as Case 1 water compared to the Case 2 data. It turned out that about 65 $\%$ of the data was classified as Case 2 water data and the rest was Case 1 data. This could be the reason why the ML models performed better on the Case 2 validation data. To test if this was the reason, the data was separated into Case 1 and Case 2 data, and each ML model was then trained and validated separately on the two classes of data.

Table 5.4: Results when validating Case 1 and Case 2 data separately when predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ with NN, PLSR, SGDR, and SVR.

Metrics	NN		PLSR		SGDR		SVR	
	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
$\overline{\text{R}^2}$	0.940	0.997	0.672	0.974	0.843	0.987	0.792	0.967
$\overline{\text{APD}}$ [%]	15.1	10.1	45.4	32.0	44.1	32.0	31.0	24.1
$\overline{\text{Bias}}$ [%]	5.00	5.3	29.5	5.41	11.3	12.15	8.71	4.27
$\overline{\text{NRMSD}}$	0.148	0.0640	0.423	0.189	0.333	0.127	0.30	0.22

The results when of predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ when training the models on Case 1 and Case 2 data, separately is shown in Tab. 5.5. The same hyperparameters and Savgol filters were used as before. Improvement was observed for both cases compared to train them together, which makes sense as the ML models would train on datasets with smaller variations. Therefore, a solution for improved performance could be to train multiple ML models on several smaller datasets representing different types of atmospheric and marine conditions, like Case 1 and Case 2 waters. However, the challenge to find a way to select the useful ML model fitting the atmospheric and marine conditions. This could be done by choosing the useful ML models based on geographical locations. Consider Case 1 and Case 2 waters which more or less can be separated into inland/coastal and open ocean waters. Also, as discussed by Fan *et al* (2017) [33], a data analysis technique could be introduced to help select the ML model that would best match the atmospheric and marine condition based on the TOA satellite measurements. Realistically, it can be difficult to divide the ML models based on atmospheric and marine conditions. To avoid this, one should increase the number of training data points. However, increasing the sample number would also increase the training time.

Table 5.5: Results when of predicting $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ when training the models on Case 1 and Case 2 data, separately.

Metrics	NN		PLSR		SGDR		SVR	
	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
$\overline{\mathrm{R}^2}$	0.980	0.997	0.927	0.974	0.900	0.962	0.842	0.952
$\overline{\mathrm{APD}}$ [%]	6.56	7.05	13.5	17.9	14.8	15.34	24.2	41.0
$\overline{\mathrm{Bias}}$ [%]	-1.45	-2.84	3.71	2.60	3.71	2.45	-4.82	-20.1
$\overline{\mathrm{NRMSD}}$	0.070	0.068	0.123	0.150	0.148	0.180	0.225	0.237
$\mathrm{N}_{\mathrm{train}}$	37971	53731	37971	53731	37971	53731	14250	14250

Summary of Atmospheric Correction of L_{rac} to R_{rs}

The NN showed the best results of the ML models based on the metrics. SVR also provided good results, but both training time and predicting time was higher than for the other models. PLSR and SGDR performed almost equally, where PLSR performed best with respect to $\overline{\mathrm{R}^2}$ and $\overline{\mathrm{NRMSD}}$ and SGDR for $\overline{\mathrm{Bias}}$ and $\overline{\mathrm{APD}}$. However, the results show that PLSR was more than 100 faster than SVR(rbf) and SGDR with respect to $\mathrm{T}_{\mathrm{pred}}$, and 10 times faster than NN. Also, PLSR did not need a lot of hyperparameter tuning, especially compared to SVR and SGDR. PLSR was more or less independent of the Savgol filters used, which is positive as less pre-processing would have to be done. Without Savgol filters, the PLS coefficients would be based directly on the measured spectrum. Interpretation can be easier for coefficients based on the original spectrum, rather than some derivative of a filter. Both NN and SGDR should use Savgol filters as they significantly improved their results. The predicted spectra from SGDR and SVR became spiky, perhaps due to lack of enough training data. This could be problematic if one would use different band ratios to predict IOPs like done with the OCx algorithm. The PLSR model, therefore, stands out as the preferred linear model, when choosing between SGDR and SVR(lin).

In conclusion, the NN network was the overall best ML model as the metric values were significantly better than the other models. An equivalent plot as shown in Fig. 5.2, but for NN only, is shown in Fig. 5.6. It shows scatterplots of NN predicted and simulated $R_{rs}(\lambda)$ for each 25-th wavelength band from 400 to $775\,\mathrm{nm}$. The scatterplots show good correlations between the predicted and simulated remote sensing reflectance. A very similar study done by Fan $et\ al.$ (2017) [33] on AC of multispectral L_{rac} with MLNN produced $\mathrm{R}^2 > 0.993$ for 7 bands in VIS (412, 443, 488, 531, 547, 667 and $678\,\mathrm{nm}$) and $\overline{\mathrm{APD}} = 3.1\,\%$. The NN trained on hyperspectral data done in my study showed comparable results with $\mathrm{R}^2 > 0.992$ for all 81 bands and $\overline{\mathrm{APD}} = 4.4\,\%$. In fact, $\overline{\mathrm{R}^2}$ calculated with NN in my study was 0.999, which was higher than 0.996 reported by Fan $et\ al.$ (2017) [33]. These results imply that both models have been able to predict the spectral relationship between L_{rac} and R_{rs}.

The validation done in this thesis is based on data generated with AccuRT. The data generated do not include whitecap, strong sun glint, and polarization effects, which are effects present in $in\ situ$ measurements. However, these effects are often very small and sensors can also avoid these effects to some extend [13]. So far, it is not possible to know whether the models actually could work on real measured satellite data. Therefore, the models should be tested against $in\ situ$ measurements from turbid, coastal areas with strong aerosol properties. Besides, other standard AC algorithms should also be validated on the same data to compare them with the models produced in this study. However, it can be stated from the results that the ML models, especially NN, can learn the RTM very accurately, and find the spectral relationship between L_{rac} and R_{rs}. The spectral shape of $in\ situ$ measurements should be very similar to what has been generated in this study, and the promising results indicate that it could work on $in\ situ$ measurements. Fan $et\ al.$ (2017) [33] validated their MLNN against $in\ situ$ measurements and compared them with standard AC algorithms which were a part of the SeaDAS framework. They reduced the APD for blue bands in L_w by more than $60\,\%$ and $25\,\%$ for highly absorbing waters and highly scattering coastal waters, respectively, compared to the standard SeaDAS NIR algorithm [33]. This showed that a method build on data generated with RT models actually could perform better on $in\ situ$ measurements than standard AC algorithms, even though this was on multispectral data.

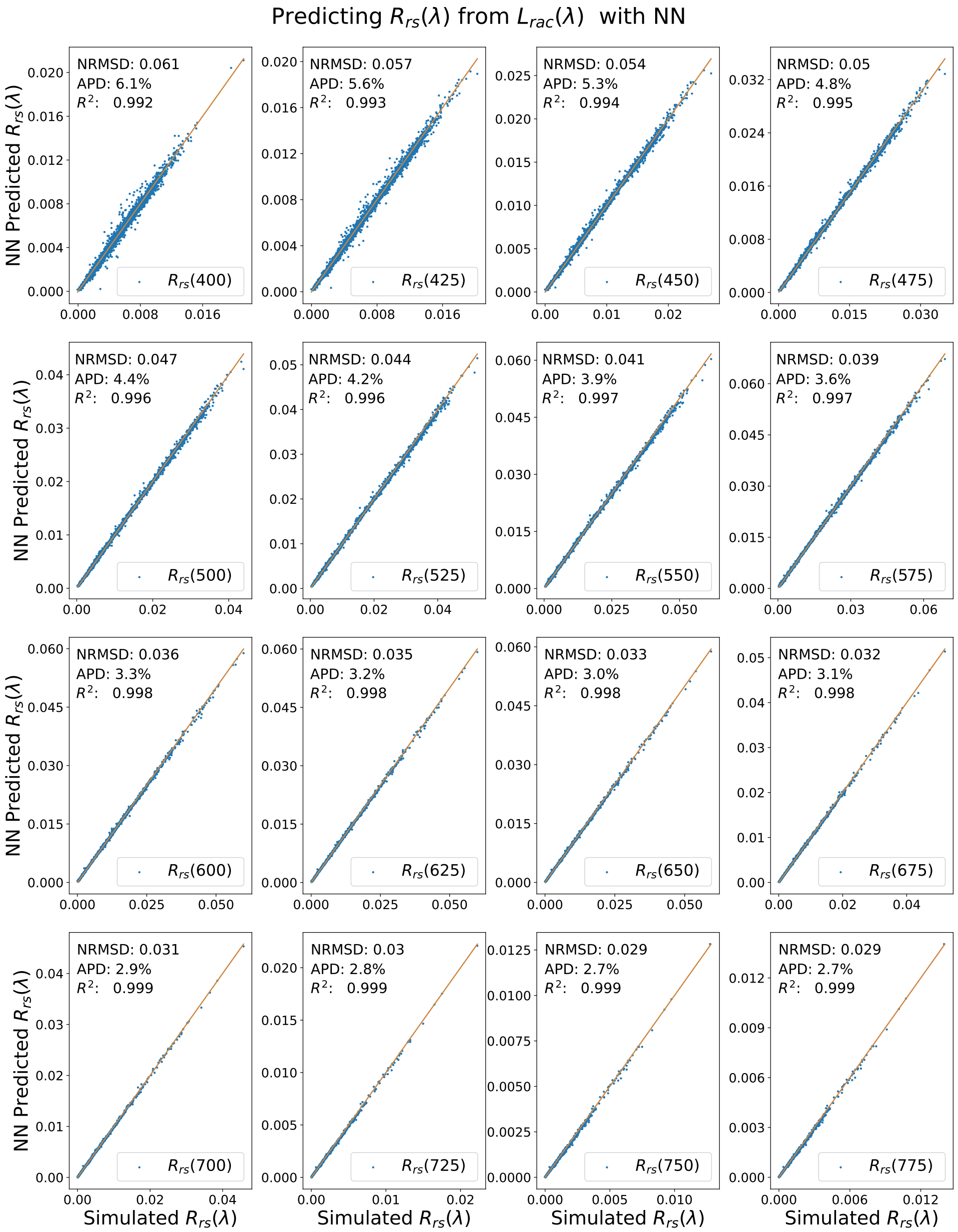


Figure 5.6: Scatterplots of NN predicted and simulated $R_{rs}(\lambda)$ from $L_{rac}(\lambda)$ for each 25-th wavelength band from 400 to $775\,\mathrm{nm}$. The orange line represents where the predicted and simulated data are exactly the same and would indicate a R^2 equal to 1. The wavelength is described in the labels and the metrics R^2, NRMSD and APD are given in the top left corners of the plots.

5.1.2 Correction of Rayleigh and Absorption Contributions

This section will present the results and discussion for the correction of Rayleigh radiance and absorption from L_t (ML2). For this part of the study, the data generated with AccuRT also included atmospheric molecules that both scattered and absorbed. The aerosol, sky glint, and water-leaving contributions were present in both the input radiance L_t and the desired output L_{rac}, so these contributions could be ignored by the ML models since only the difference between L_t and L_{rac} were to be calculated.

Input $L_t(\lambda)$ and output $L_{rac}(\lambda)$ from one simulation are shown in Fig. 5.7 and show that there are many similarities between the spectral shape of the input and output. The radiance values of L_t are higher because L_t also includes Rayleigh radiance, which are more present for shorter wavelengths. This can clearly be seen as the difference between L_t and L_{rac} is much more significant for the bands around 400-500 nm than for the bands around 700-800 nm. The absroption effects can also be seen in Fig. 5.7, for instance at The absorption bands found in the L_t spectra are shown as some sudden spikes at 760 nm where L_t has a sudden spike. The atmospheric gases used were extracted from the U.S. Standard setup, further described in Appendix B.

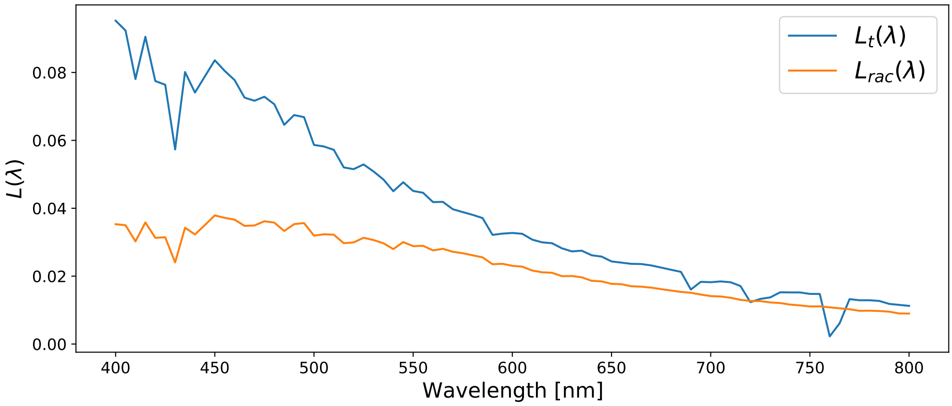


Figure 5.7: One sample of input $L_t(\lambda)$ and output $L_{rac}(\lambda)$ to illustrate the spectral differences.

The amount of the different absorbing gases did not vary in this study, simplifying the problem for the ML models. Therefore, the amount of absorption would depend mostly on how long the photons would travel in the atmosphere, again dependent on the various sun-target-sensor angles. The amount of Rayleigh scattering is also dependent on sun-target-sensor angles. Therefore, it was expected that the ML models performed more accurately on this problem than for the ML1 case, where L_{rac} to R_{rs} was discussed in section 5.1.1. Savgol filter and angle pre-processing were not used since the spectral shape of the input and output were similar. No hyperparameter optimization was done for this case and the hyperparameters were chosen based on the optimal hyperparameters found for the ML1 approach.

The results of the mean values of the metrics are shown in Tab. 5.6. The results for the corresponding metrics for each wavelength band are also shown in Appendices A.5, A.6, A.7 and A.8. The results show very good performance for all the ML models with the lowest $\overline{\mathrm{R}^2}$ found to be 0.9974 for SGDR and highest to be 0.9994 for NN. Again, the NN performed best considering the metrics. However, the linear model PLSR showed comparable results with the NN. PLSR was performing better than SVR for all the metrics. SGDR performed worst, but the results for SGDR were still good with $\overline{\mathrm{R}^2}$ = 0.9974 and $\overline{\mathrm{NRMSE}}$ = 0.043.

Table 5.6: Optimal results when predicting $L_{rac}(\lambda)$ from $L_t(\lambda)$ with NN, PLSR, SGDR and SVR based on mean metrics and time complexity.

Metrics	NN	PLSR	SGDR	SVR
$\overline{\mathrm{R}^2}$	0.9999	0.9990	0.9974	0.9988
$\overline{\mathrm{APD}}$ [%]	1.89	3.29	7.28	6.03
$\overline{\mathrm{Bias}}$ [%]	-1.39	0.35	0.24	1.91
$\overline{\mathrm{NRMSE}}$	0.012	0.0225	0.043	0.030
$\mathrm{T_{fit}}$ [s]	363	145	59	175
$\mathrm{N_{train}}$	91702	91702	91702	14250

Fig. 5.8 presents scatterplots of predicted and simulated L_{rac} for wavelength band 400, 500, 600 and 700 nm. These four wavelength bands were chosen as representative bands to illustrate the responses for individual bands in different parts of the wavelength range. Fig. 5.8 shows very good predictions for all the ML models with R^2 in the range 0.986-1.000 for all wavelength bands.

The different metrics calculated for each ML model as a function of the 81 bands from 400 to 800 nm, like shown in Fig. 5.3 for ML1, are shown in Fig. 5.9. The spiky behavior was still found for SVR and SGDR, which can be seen from the red and green lines. These results also showed that the worst performance was found for shorter wavelength bands, as found for AC of L_{rac}. A possible explanation can be that these regions are most affected by the Rayleigh scattering, and therefore would lead to larger variation more difficult to predict. Compared to AC of L_{rac}, this approach performed better for longer wavelengths bands. All of the ML models, except SVR, more or less improved the values of R^2, APD, and NRMSD for longer wavelength bands. The good performance of the PLSR model, in this case, should be noticed. R^2, Bias and NRMSD were all constantly improving for longer wavelengths, and the optimal values of R^2 APD and NRMSD were all found at 800 nm. The results with PLSR were slightly worse than for the NN for the shortest wavelengths, but one average were better than all the other ML models between 600-800 nm for all metrics.

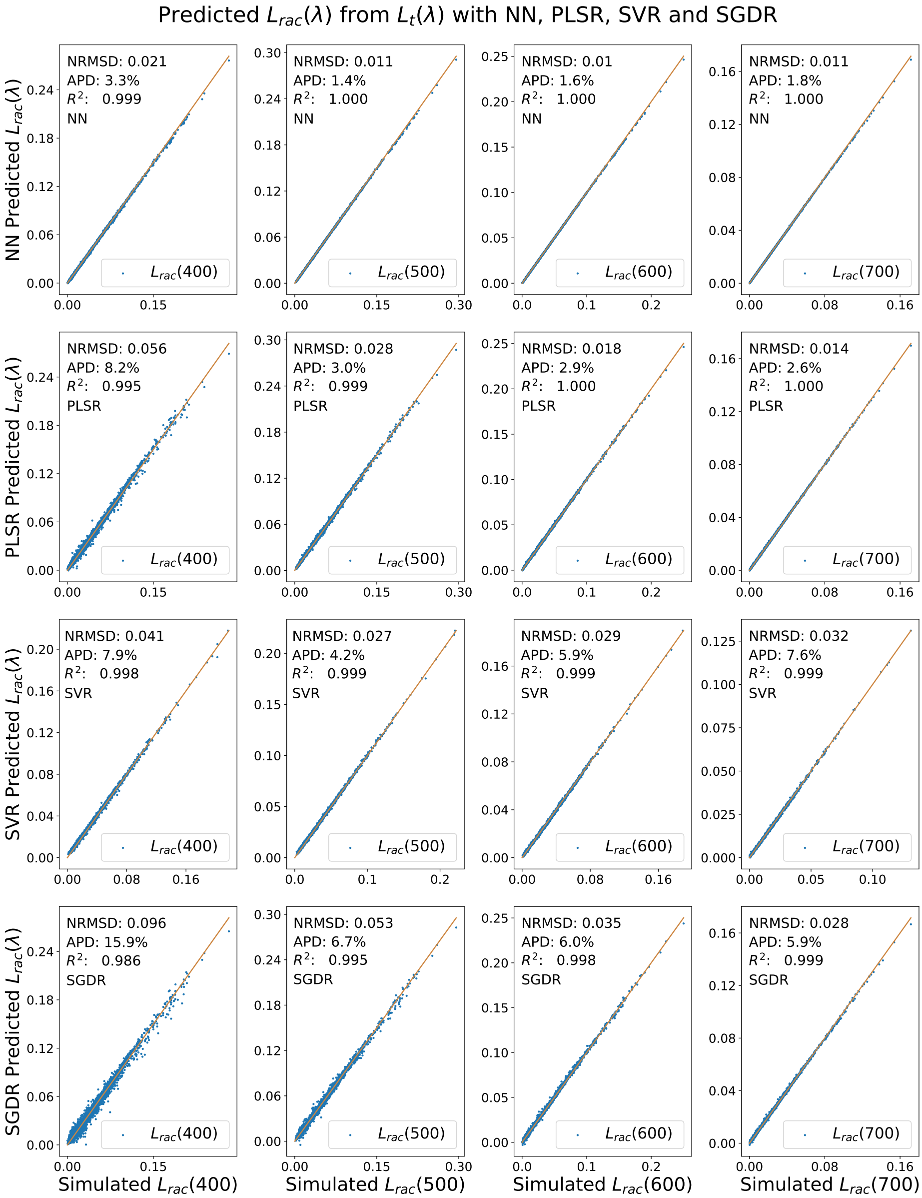
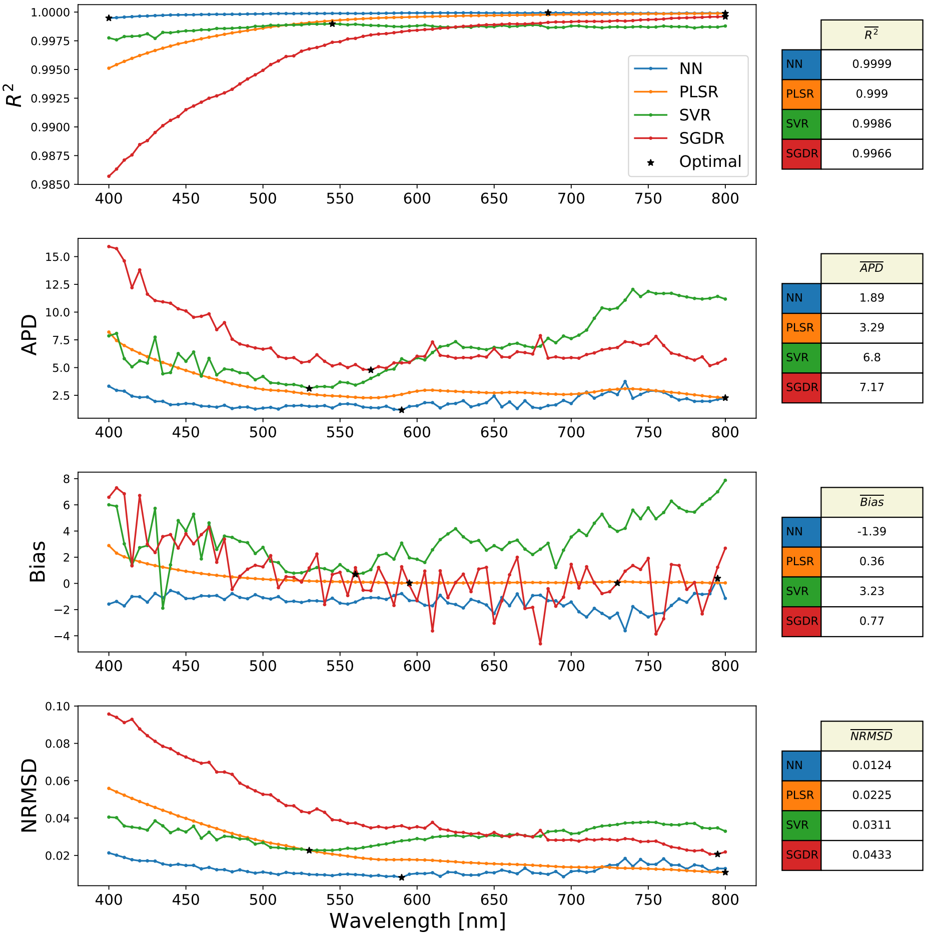


Figure 5.8: Scatterplots of predicted and simulated $L_{rac}(\lambda)$ for wavelength band 400, 500, 600 and 700 nm (indicated with text) with NN, PLSR, SVR and SGDR predicted from $L_t(\lambda)$, with corresponding R^2, APD and NRMSD values.



	$\overline{R^2}$
NN	0.9999
PLSR	0.999
SVR	0.9986
SGDR	0.9966

	$\overline{APD}$
NN	1.89
PLSR	3.29
SVR	6.8
SGDR	7.17

	$\overline{Bias}$
NN	-1.39
PLSR	0.36
SVR	3.23
SGDR	0.77

	$\overline{NRMSD}$
NN	0.0124
PLSR	0.0225
SVR	0.0311
SGDR	0.0433

Figure 5.9: R^2, APD, Bias and NRMSD as a function of wavelength for the four different ML models for predicting absorption and Rayleigh. The mean values of the metrics are shown in the tables on the right hand side of each plot.

5.1.3 Atmospheric Correction of L_t to R_{rs}

This section will discuss full AC where R_{rs} was to be predicted from L_t (ML3). The only difference between ML1 and ML3 was that the Rayleigh and absorption contributions also were included in ML3. Since the previous section showed that all the ML models performed very accurately when predicting Rayleigh and absorption effects, it would be expected that this approach would perform similarly as ML1. The reason for this assumption was that the most difficult parts to predict in both L_{rac} and L_t were the aerosols and the water-leaving radiance, which was already included in ML1.

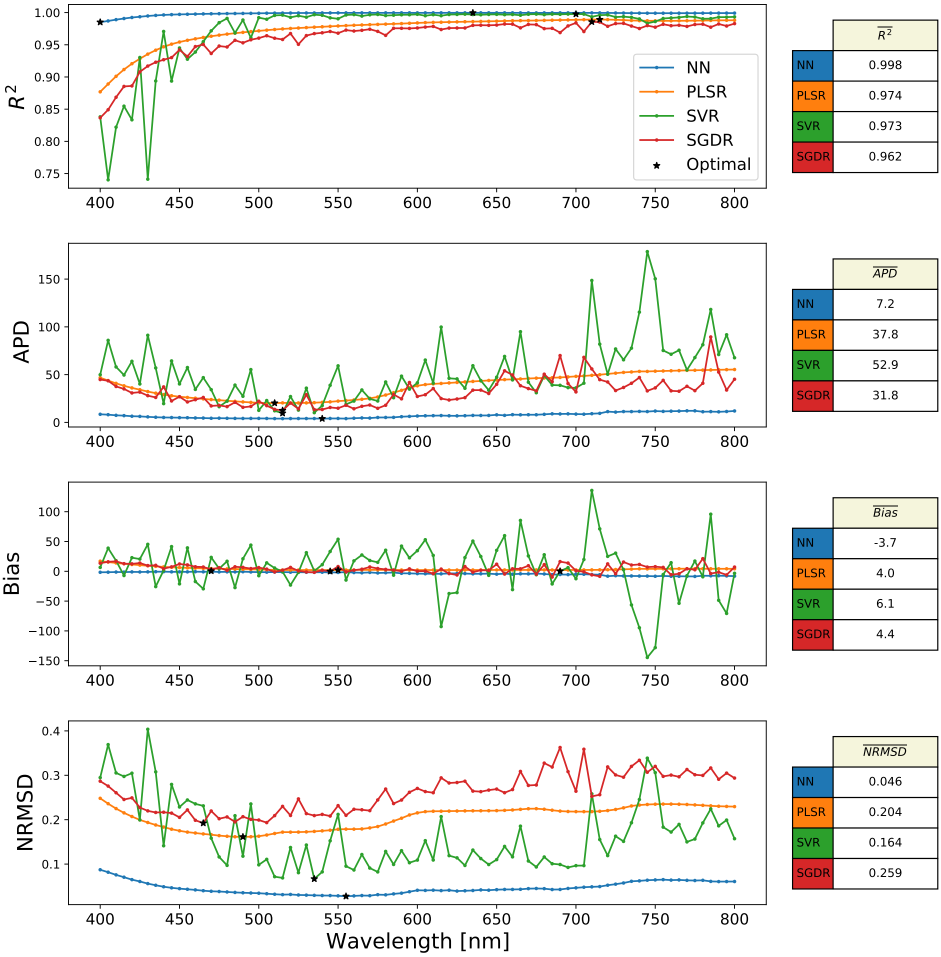
The results from the ML3 approach is shown in Tab. 5.6 and shows the mean metric values and the number of training data points for predicting R_{rs} from L_{rac}. As expected, the results were very similar to the results found for AC of L_{rac} shown in Tab. 5.2. NN gave the best results with respect to all metrics, whereas PLSR, SGDR, and SVR had decent and comparable results. SGDR and SVR still struggled with the spiky shapes as illustrated in Fig. 5.4. Fig. 5.11 presents scatterplots of predicted and simulated R_{rs} for wavelength band 400, 500, 600 and 700 nm. These four wavelength bands were chosen as representative bands to illustrate the responses for individual bands in different parts of the wavelength range. These scatterplots show very similar behavior as the scatterplots shown for ML1 in Fig. 5.2. The results for the corresponding metrics for all wavelength band are also shown in Appendices A.9, A.10, A.11 and A.12.

Table 5.7: Optimal results when predicting $R_{rs}(\lambda)$ from $L_t(\lambda)$ with NN, PLSR, SGDR and SVR based on mean metrics and time complexity.

Metrics	NN	PLSR	SGDR	SVR(rbf)
$\overline{\mathrm{R}^2}$	0.9984	0.9739	0.9559	0.9731
$\overline{\mathrm{APD}}$ [%]	7.21	37.8	27.28	52.9
$\overline{\mathrm{Bias}}$ [%]	-3.71	4.0	5.0	6.1
$\overline{\mathrm{NRMSE}}$	0.0461	0.204	0.263	0.164
$\mathrm{N}_{\mathrm{train}}$	112366	112366	112366	23750

The calculated values of R^2, APD, Bias, and NRMSD as a function of all 81 wavelength bands are shown in Fig. 5.10. The results were very similar to the ML1 results (Fig. 5.3). Actually, NN and PLSR were almost identical for ML1 and ML3, especially when considering R^2, APD, and NRMSD. All optimal results for APD, Bias and NRMSD for all ML models were also found at almost the same wavelength band as for ML1, namely between 470 and 560 nm. The best results of R^2 for both ML1 and ML3 were found between 550 and 800 nm. However, the calculated mean metrics for ML3 were all slightly worse than for ML1. The results show that adding Rayleigh and absorption only caused a very small increase in the prediction complexity. Furthermore, this mean that most of the complexity was caused by the aerosol and water properties.

As previously discussed in section 4.2.1, to make the ML3 more suitable for various atmospheric cases, different atmospheric setups would have to be used with AccuRT. This could be done by varying the amount of the different atmospheric gases or use another atmosphere setup than the U.S. Standard. Nevertheless, the ML3 approach was tested in this study to investigate how the ML models would perform on data similar to actual TOA measurements. The variations in atmospheric gases would most probably not be as complex as the aerosol and ocean properties, indicating that the ML models possibly could do well on even with more variation in atmospheric gases.



	$\overline{R^2}$
NN	0.998
PLSR	0.974
SVR	0.973
SGDR	0.962

	$\overline{APD}$
NN	7.2
PLSR	37.8
SVR	52.9
SGDR	31.8

	$\overline{Bias}$
NN	-3.7
PLSR	4.0
SVR	6.1
SGDR	4.4

	$\overline{NRMSD}$
NN	0.046
PLSR	0.204
SVR	0.164
SGDR	0.259

Figure 5.10: Calculated R^2, APD, Bias and NRMSD as a function of wavelength band for the four different ML models based on predicting $R_{rs}(\lambda)$ from $L_t(\lambda)$. The mean values of the metrics are shown in the tables on the right hand side of the plot.

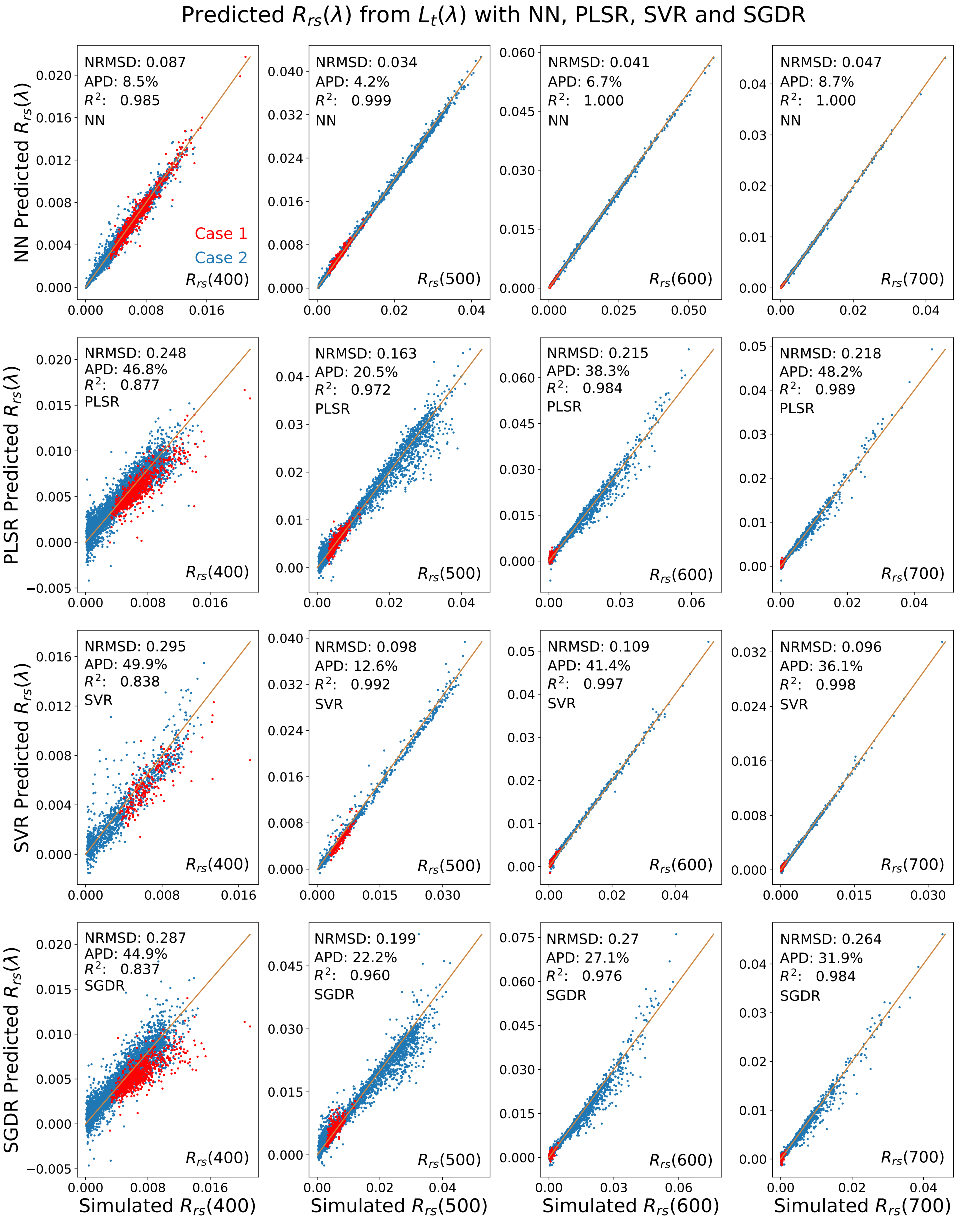


Figure 5.11: Scatterplots of predicted and simulated $R_{rs}(\lambda)$ for wavelength band 400, 500, 600 and 700 nm (indicated with text) from $L_t(\lambda)$, with corresponding R^2, APD and NRMSD values. Red and blue dots represent Case 1 and Case 2 waters, respectively.

5.2 Water IOPs retrieval Validation

Several ocean color algorithms are produced to predict IOPs from R_{rs} based on empirical relationships derived from $in\ situ$ measurements, like the non-linear OCx algorithm described in Eq. 4.1. It would be interesting to test ML models for water IOP retrieval. NN and PLSR were chosen to compare one linear and one non-linear model. This choice was also based on that NN and PLSR were the two methods providing the best overall results when investigating AC, when considering metrics values, time complexity, and interpretability. NN was superior when it came to accuracy, whereas one of the main advantages with PLSR is that it can be more intuitively interpretive and leave a better understanding of the relevance regarding the input features. In this study, R_{rs} data were generated, for a wide range of CHL, $a_{cdom}(443)$ and MIN, representative for both Case 1 and Case 2 waters. These data were used to train the two ML models to predicted CHL, $a_{cdom}(443)$ and MIN from R_{rs}, which is referred to as ML4 in Fig. 4.1.

5.2.1 PLSR with Variable Selection and Basis Functions

PLSR was first tested with variable selection. Here, the number of discarded features and optimal number of components were tested. Furthermore, different Savgol filters, as used for AC, were tested. As shown for AC with PLSR in Fig. 4.14, the Savgol filters did not affect the performance appreciable. The original input spectrum without Savgol filters provided the best results, therefore no Savgol filters were further used.

The initial results showed that PLSR showed decent linear predictions to CHL and MIN concentration, as R^2 for these two predictions both were higher than 0.94. However, when predicting $a_{cdom}(443)$, R^2 was never higher than 0.7, indicating that non-linear model would be needed. For this approach, I used a method capable of dealing with non-linearities when using PLSR. Recall that PLSR aims to produce a linear model f(**x**, **w**) like shown in Eq. 4.9. The limitation of linearities can be partially removed by considering the linear combinations of a fixed nonlinear function (basis function) of the input variables:

$$f(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^{D-1} w_i \phi_i(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) \tag{5.1}$$

where **x** in Eq. 4.9 is replaced with the nonlinear function $\phi(\mathbf{x})$ with $\mathbf{w} = (w_0, ..., w_{D-1})^T$ and $\phi = (\phi_0, ..., \phi_{D-1})^T$, where $\phi_0(\mathbf{x})$ = 1. f(**x**, **w**) is still a linear function of the weights, but is now a nonlinear function of the input dimensions (**x**).

The input and output data were compared and plotted against each other to look for non-linear behaviors, to choose which basis functions to use. To illustrate this, $a_{cdom}(443)$ as a function of the input data (R_{rs}) at two selected wavelength bands, 405 and 525 nm, are shown in the two left plots in Fig. 5.12. The top left plot shows that $a_{cdom}(443)$, as a function of input data $R_{rs}(405)$, has a nonlinear pattern similar to a 1/X function. The bottom left plot also follows a non-linear pattern with input data $R_{rs}(525)$. For the bottom left plot, it was more difficult to see which basis function to be used. Therefore, several basis functions were tested, and lg(X) gave the best results. It must be specified that for wavelength bands 525 to 800 nm, it was very difficult to find basis functions which resulted in good linear correlation. 1/X was chosen for wavelength bands 400 to 525 nm and lg(X) for wavelength 525 to 800 nm for the prediction of $a_{cdom}(443)$, shown in the right plots in Fig. 5.12. The orange lines in Fig. 5.12 are simple linear regression lines illustrating the linear correlation between the output and input data at wavelength 405 and 525 nm, where improvement was found when using basis functions, especially for $R_{rs}(405)$.

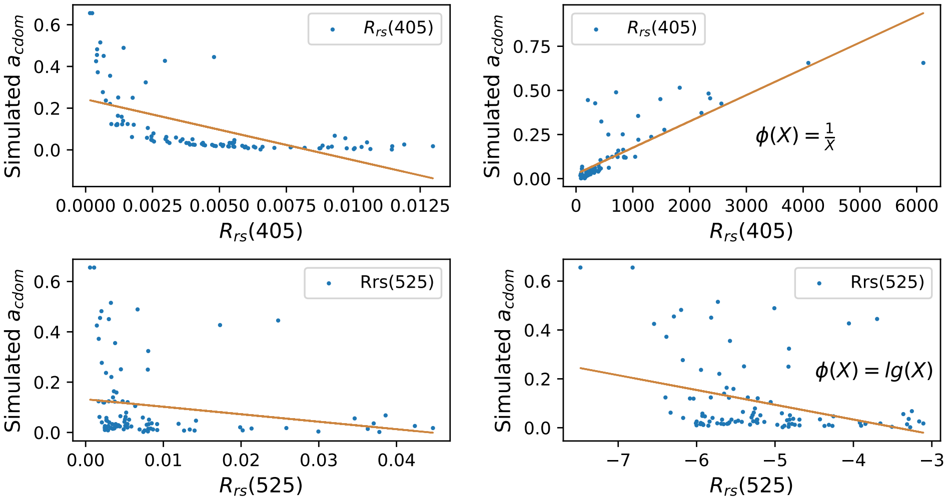


Figure 5.12: Scatterplot of simulated $a_{cdom}(443)$ as a function of the input data $R_{rs}(405)$ and $R_{rs}(525)$ without (left plots) and with basis functions (right). The basis functions are described in the right plots with text and the orange line shows a simple linear polynomial fit to the data to highlight the correlation.

Non-linear basis functions for CHL were also tested and the square root of the input data was chosen ($\phi(\mathbf{x}) = \sqrt{\mathbf{X}}$). No basis functions were used on the input data to predict MIN, as the results were good without. With the described basis functions, PLSR with variable selection was further tested to find the best combination of number of components and input features to be used. The number of components and features were tested from 0 to D_1=81. R^2 values as a function of used features (y-axis) and the number of components (x-axis) are shown in Fig. 5.13, where the bright colored boxes represent combinations of N components and features with high values of R^2. The dark colored boxes represent R^2 $\leq$ 0.98 and the top 25 values of R^2 are highlighted with black borders. The improvements with variable selection in this case for the metric values are not of major importance. The color bars show that the difference between the best and worst results for R^2 were only 0.002. Also important with variable selection was to reduce the number of features and components and still have optimal results. Reducing the number of components and features could reduce data and hopefully also reduce computational time.

Fig. 5.13 show that there were many combinations of components and chosen features that would lead to almost the same accuracy. For CHL, the optimal number of components was between 20-23. Also, changing the number of features from 81 to almost 65 was possible and still get the same level of accuracy for R^2. For $a_{cdom}(443)$ prediction the optimal number of features was 75 for several different numbers of components (31-74). Lastly, the optimal values of R^2 for MIN prediction were found for a wide range of both components (31-81) and features (55-81). Several good results were found for 76 features and more than 50 components.

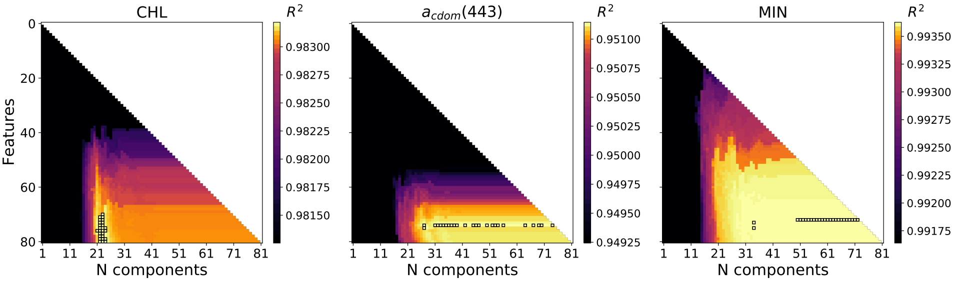


Figure 5.13: R^2 as a function of N components and features when using PLSR. The bright colored boxes show which combinations of N components and features giving best results for the prediction of CHL, $a_{cdom}(443)$ and MIN using PLSR. The dark colored boxes represent R^2 ≤ 0.98 and the top 25 values of R^2 are highlighted with black borders.

Optimal results when predicting IOPs from $R_{rs}(\lambda)$ with PLSR with (Original) and without (ϕ) basis functions with variable selection is shown in Tab. 5.8. Prediction of MIN was done without basis functions. Good improvements were obtained for the predictions of CHL and $a_{cdom}(443)$ where R^2 changed from 0.945 and 0.680 to 0.988 and 0.961, respectively.

Table 5.8: Optimal results when predicting IOPs from $R_{rs}(\lambda)$ with PLSR with (Original) and without (ϕ) basis functions. Prediction of MIN was done without basis functions.

	CHL		$a_{cdom}(443)$		MIN
Metrics	Original	ϕ	Original	ϕ	Original
R^2	0.945	0.987	0.680	0.961	0.994
APD [%]	259.9	127.1	265.4	84.5	65.0
Bias [%]	108.9	40.2	-86.2	-1.81	25.27
RMSD	1.26	0.649	0.100	0.0317	0.346
NRMSD	0.568	0.271	0.851	0.275	0.184

Fig. 5.14 illustrates the optimal results of water IOP retrieval using PLSR with basis functions and variable selection (ML4). The left plots show the absolute value of the PLS coefficients when predicting the corresponding IOP referred to in the right plots. The wavelength dependent vertical lines are colored in their corresponding RGB colors. Vertical black lines represent the features discarded before training the model. Number of features ($\mathrm{N_{feat}}$) and number of components ($\mathrm{N_{comp}}$) are highlighted with text. The right plots show scatterplots of predicted and simulated IOPs, whit calculated R^2 and NRMSD highlighted with text. Savgol filter was not used on the input spectra for any of the three IOP prediction models.

The results show that R^2 was highest when predicting CHL and MIN. The mineral predictions showed the highest value even without the nonlinear transformation, with R^2 equal to 0.994, whereas R^2 for chlorophyll concentration prediction was 0.988. The results of $a_{cdom}(443)$ on the other hand gave R^2 equal to 0.961, and without the non-linear transformation, only 0.680. However, with R^2 values between 0.961 and 0.994 for all the predictions, PLSR could predict IOPs with some accuracy. However, CHL and $a_{cdom}(443)$ would need basis functions. Also, the performance validated on APD, Bias, RMSD and NRMSD was not very good as shown in Tab. 5.8.

The absolute values of the PLS coefficients indicated which wavelength bands that were most important when predicting the various IOPs. When predicting CHL, the absolute values of the PLS coefficients were high for bands between 520 to $600\,\mathrm{nm}$, which cover the green wavelength region where the chlorophyll reflectance is most significant. The coefficients for CHL were very small for bands larger than $660\,\mathrm{nm}$ except for the red region in the range $685\text{-}705\,\mathrm{nm}$. This region is very prominent in the spectrum for the phytoplankton absorption coefficient as shown in Fig. 3.3. For MIN predictions, the wavelength band $655\,\mathrm{nm}$ showed a significantly high value of the absolute value of the PLS coefficient compared to the other bands. The other important bands for MIN were found in the red region in VIS. The red wavelength bands seemed to be more important for $a_{cdom}(443)$ and MIN than for CHL. Furthermore, $a_{cdom}(443)$ had prediction, was that coefficients with large values over the whole spectra were found. Also, the coefficients in the region $685\text{-}705\,\mathrm{nm}$ were more important, where VIS is ending and NIR is beginning. This yields that a high value of R_{rs} in this region would imply a higher value of $a_{cdom}(443)$. $In\ situ$ R_{rs} values in this region are often close to zero when the CDOM concentration is low. This could be the reason why they have higher expressive power in terms of determining $a_{cdom}(443)$.

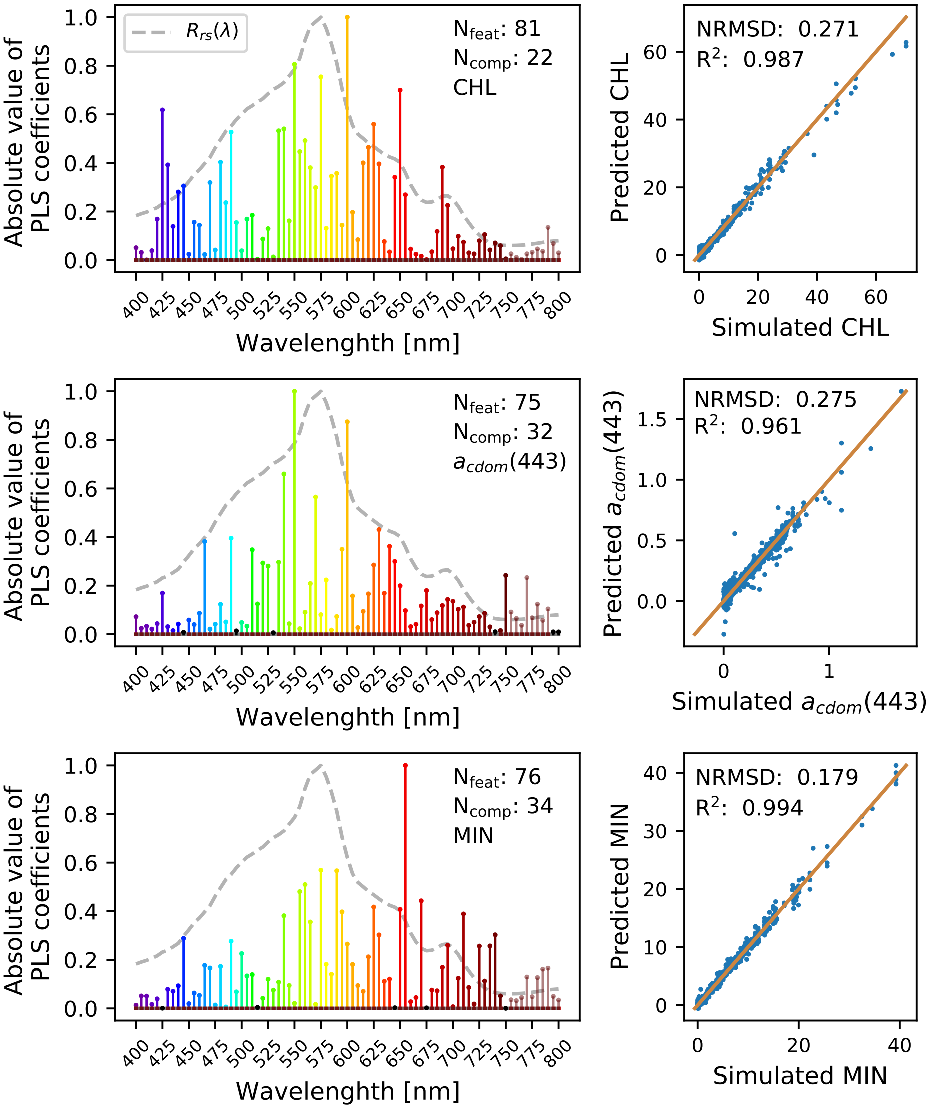


Figure 5.14: Normalized absolute values of PLS coefficients as a function of wavelength bands for predicting CHL (top plots), $a_{cdom}(443)$ (middle plots) and MIN (bottom plots) together with scatterplots of the predicted and simulated IOP. Number of features ($\mathrm{N_{feat}}$) and number of components ($\mathrm{N_{comp}}$) are highlighted in the left plots. The grey dashed curve represent one simulated $R_{rs}(\lambda)$. Vertical black lines represent the features discarded before training the model.

5.2.2 NN and PLSR Comparison

Three different NNs were trained to predict CHL, $a_{cdom}(443)$ and MIN from R_{rs}, respectively. Different Savgol filters were applied to the input data for each NN model. The optimal results using NN to predict IOPs are shown in Tab. 5.9. PLSR(ϕ) and PLSR refers to PLSR with and without data where basis functions have been used. One should notice that the results based on predicting MIN were comparable for NN and PLSR, with R^2 equal to 0.998 and 0.0994, respectively. However, the NN performed significantly better than PLSR when predicting CHL and $a_{cdom}(443)$, especially compared to PLSR without the basis functions, as can be seen from Tab. 5.9. All metric values for NN were many times better than for PLSR with and without basis function.

Table 5.9: Predicted chlorophyll concentration (CHL), $a_{cdom}(443)$, mineral concentration (MIN) from $R_{rs}(\lambda)$ with NN and PLSR validated with R^2, APD, Bias, RMSD and NRMSD. PLSR(ϕ) and PLSR refers to PLSR with and without data where basis functions have been used.

	CHL			$a_{cdom}(443)$			MIN	
Metrics	NN	PLSR(ϕ)	PLSR	NN	PLSR(ϕ)	PLSR	NN	PLSR
R^2	0.999	0.987	0.945	0.996	0.961	0.680	0.998	0.994
APD [%]	8.84	127.1	259.9	16.8	84.5	265.4	42.4	65.0
Bias [%]	2.53	40.2	108.9	-7.19	-1.81	-86.2	32.8	30.74
RMSD	0.0821	0.649	1.26	0.0383	0.0317	0.100	0.205	0.322
NRMSD	0.0353	0.271	0.568	0.0933	0.275	0.851	0.108	0.178

Scatterplots of NN predicted and simulated CHL, $a_{cdom}(443)$ and MIN from $R_{rs}(\lambda)$ are shown in Fig. 5.15. The orange lines illustrate perfect prediction and the calculated R^2 and NRMSD are highlighted in the top left corner of each plot. W_l, Poly and Der in Fig. 5.15 are the window length, polynomial degree and derivative, respectively, in a Savgol filter as described in section 4.3.3. From Fig. 5.15, one can see the good correlation between the predicted and simulated IOPs as the blue data points are very close to the orange line.

To conclude, an NN approach should be used as a water IOP retrieval algorithm, because the accuracy was significantly better for NN than PLSR. In addition, for the NN model, the best results were found when predicting CHL, which is important as CHL is the most desired property to predict in ocean color. The big advantage with PLSR was that it provides PLS coefficients that give a more interpretable derivation of the results.

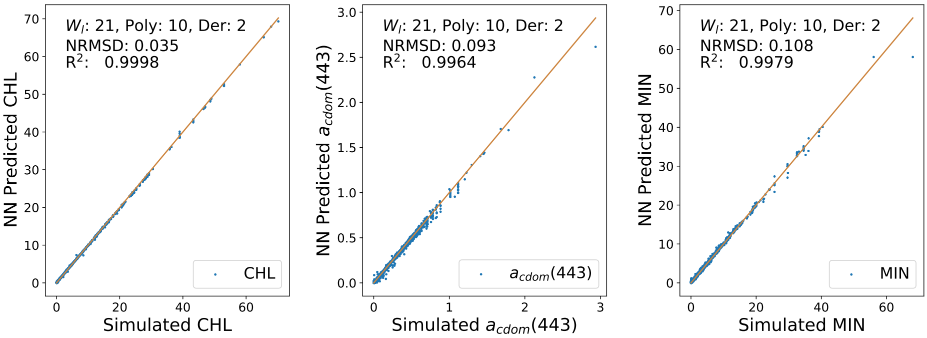


Figure 5.15: Scatterplots of NN predicted and simulated CHL, $a_{cdom}(443)$ and MIN from $R_{rs}(\lambda)$. The orange lines illustrates perfect prediction and the calculated R^2 and NRMSD are highlighted in the top left corner of each plot. W_l, Poly and Der are the window length, polynomial degree and derivative, respectively, in a Savgol filter as described in section 4.3.3.

Chapter 6

Conclusion and Further Work

AC in remote sensing for ocean color plays an important role as the view of the Earth's surface from satellites is degraded by the presence of the atmosphere. For coastal and inland waters, AC is more challenging than for open oceans due to the varying aerosol properties in the lower layers and complex water bodies. For these kinds of waters, several traditional AC algorithms fail and can predict negative water-leaving radiances as the black ocean assumption can be invalid. In this thesis, different AC and water IOP retrieval models, focusing on the challenging coastal waters, were tested based on four machine learning models NN, PLSR, SVR, and SGDR. The training and validation data were based on the simulations done with the coupled atmosphere-ocean radiative transfer model AccuRT.

AccuRT was shown to provide useful hyperspectral data representative for challenging coastal waters which further could be used to train ML models. When predicting the 81 $R_{rs}(\lambda)$ values from 400 to $800\,\mathrm{nm}$ from $L_{rac}(\lambda)$, all ML models resulted in $\overline{\mathrm{R}^2}$ > 0.968, indicating that they were able to predict the spectral relationship between $L_{rac}(\lambda)$ and $R_{rs}(\lambda)$. The best results were obtained with the NN algorithm ($\overline{\mathrm{R}^2}$ =0.999), especially compared to the linear models PLSR ($\overline{\mathrm{R}^2}$ =0.974) and SGDR ($\overline{\mathrm{R}^2}$ =0.968). On the other hand, the linear models provided interpretable derivation in form of coefficients, which provided a broader understanding of why the models gave the predictions, which input data that was important and which data that could be removed and still yield optimal results. Also, unlike many standard AC algorithms, these methods were capable of doing AC without the extra short-wave infrared (SWIR) bands, as they were trained on bands between 400 and $800\,\mathrm{nm}$. Furthermore, the NN approach provided R^2 > 0.9998 when predicting chlorophyll concentration from R_{rs}, indicating that the NN models also could be used for water IOP retrieval.

With the recent growth in cheaper, quick-to-launch small satellites, there has been a greater need for intelligent on-board data processing tools. The different AC algorithms became very fast once they had been trained and would therefore suit operational use and be part of the on-board data processing framework. A combination of AC and water IOP retrieval algorithms could reduce the data stored significantly, as one could store and downlink only relevant data, like chlorophyll concentration.

6.1 Further Work

There are many possibilities for applying ML to AC and water IOP retrieval in different ways. AccuRT or other relevant RT algorithms can be used to simulate different kinds of training data, and different corrections and IOP retrievals could be done. A large variation in absorbing and Rayleigh scattering molecules was not tested in this thesis and could be a possible further path to account for different $in\ situ$ atmospheric conditions. This study only considered the U.S. Standard atmospheric setup, but different other pre-defined atmospheric setups could also be tested. Also, a direct mapping of IOP retrieval from TOA radiance could also be tested, i.e. predict the chlorophyll concentration from the TOA radiance directly to reduce the data storage significantly.

A possible further path could be to use $in\ situ$ TOA radiance and water-leaving radiance or remote sensing reflectance as training data if large amounts of such data would be available. This could train models that are more representative of actual oceanic and atmospheric environments. However, this is only possible if enough data is available, which can be hard to find. This was the reason why AccuRT was used in this study. A positive aspect using RT models is the flexibility to simulate many different atmospheric and oceanic environments, and also the ability to generate almost unlimited amounts of data, making such an approach ideal for ML. The models in this study should also be validated against $in\ situ$ hyperspectral data and compared against standard AC algorithms.

In this study, AC of a single pixel was investigated. A further approach could be to study AC of more pixels at the same time, and include neighboring pixels as information when correcting several pixels simultaneously, or even the whole hypercube at the same time. A path dealing with what was studied in this thesis could be to implement one of the ML models as a part of an on-board processing framework. When correcting a whole scene, each pixel would have to be corrected iteratively, but they would all be independent of each other. Therefore, one should investigate parallel computing and maybe make use of GPU to decrease the computational time.

Bibliography

[1] K. Stamnes, G.E. Thomas, and J.J. Stamnes. *Radiative Transfer in the Atmosphere and Ocean*. Cambridge University Press, 2 edition, 2017. -Book webpage. See zip file for AccuRT info `http://www.rtatmocn.com/`.

[2] IOCCG. *Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology*, volume No. 7 of *Reports of the International Ocean Colour Coordinating Group*. IOCCG, Dartmouth, Canada, 2008.

[3] Ocean color. `https://science.nasa.gov/earth-science/oceanography/living-ocean/ocean-color`, 2020. Accessed: 2020-05-10.

[4] C.D. Mobley. *Radiative Transfer in the Ocean*, pages 2321–2330. In: J.H. Steele (ed), Academic Press Inc., 2001.

[5] IOCCG. *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters*, volume No. 3 of *Reports of the International Ocean Colour Coordinating Group*. IOCCG, Dartmouth, Canada, 2000.

[6] R. J. Frouin, B. A. Franz, A. Ibrahim, K. Knobelspiesse, Z. Ahmad, B. Cairns, J. Chowdhary, H. M. Dierssen, J. Tan, O. Dubovik, X. Huang, A. B. Davis, O. Kalashnikova, D.R. Thompson, L.A. Remer, E. Boss, O. Coddington, P. Y. Deschamps, B. C. Gao, L. Gross, O. Hasekamp, A. Omar, B. Pelletier, D. Ramon, F. Steinmetz, and P. W. Zhai. Atmospheric Correction of Satellite Ocean-Color Imagery During the PACE Era. *Frontiers of Earth Science*, (7):145, 2019.

[7] M. Soukup, J. Gailis, D. Fantin, A. Jochemsen, C. Aas, P.-J. Baeck, I. Benhadj, S. Livens, B. Delauré, M. Menenti, B. Gorte, S. Enayat Hosseini A., M. Esposito, and C. Dijk. Hyperscout: Onboard processing of hyperspectral imaging data on a nanosatellite. 2016.

[8] T. A. Johansen. Smallsat-related projects. `http://folk.ntnu.no/torarnj/Smallsat-prosjekt-2017-taj.pdf`, 2017. Accessed: 2020-05-05.

[9] A. Ibrahim, B. Franz, Z. Ahmad, R. Healy, K. Knobelspiesse, B.-C. Gao, C. Proctor, and P.-W. Zhai. Atmospheric correction for hyperspectral ocean color retrieval with application to the hyperspectral imager for the coastal ocean (hico). *Remote Sensing of Environment*, 204:60 – 75, 2018.

[10] M. Guelman and F. Ortenberg. Small satellite's role in future hyperspectral earth observation missions. *Acta Astronautica*, 64(11):1252 – 1263, 2009.

[11] IOCCG. *Atmospheric Correction for Remotely-Sensed Ocean-Colour Products*, volume No. 10 of *Reports of the International Ocean Colour Coordinating Group*. IOCCG, Dartmouth, Canada, 2010.

[12] IOCCG. *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*, volume No. 5 of *Reports of the International Ocean Colour Coordinating Group*. IOCCG, Dartmouth, Canada, 2006. -Synthesized dataset used in the report `http://www.ioccg.org/groups/OCAG_data.html`.

[13] C.D. Mobley. Ocean optics web book. URL `http://www.oceanopticsbook.info/`.

[14] M. Wang. A refinement for the rayleigh radiance computation with variation of the atmospheric pressure. *International Journal of Remote Sensing*, (26):5651–5663, 2005.

[15] M. T. Eismann. *Hyperspectral Remote Sensing*, volume PM210. SPIE Press, P.O. Box 10 Bellingham, Washington 98227-0010 USA, 2012.

[16] S. K. Meerdink, S. J. Hook, D. A. Roberts, and E. A. Abbott. The ecostress spectral library version 1.0. *Remote Sensing of Environment*, 230:1–8, 2019.

[17] A. M. Baldridge, S.J. Hook, C.I. Grove, and G. Rivera. The aster spectral library version 2.0. *Remote Sensing of Environment*, 113:711–715, 2009.

[18] Fundamemtals of hyperspectral remote sensing, 2013. URL `http://www.gisresources.com/fundamemtals-of-hyperspectral-remote-sensing_2/`.

[19] O. Hagolle. Radiometric quantities : irradiance, radiance, reflectance. `https://labo.obs-mip.fr/multitemp/radiometric-quantities-irradiance-radiance-reflectance/`, 2016. Accessed: 2020-02-13.

[20] E. Vermote, D. Tanre, J.L. Deuze, M. Herman, J.J. Morcrette, and S.Y.Kotchenova. Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV) .

[21] K. Moore, K. Voss, and H. Gordon. Spectral reflectance of whitecaps: Their contribution to water-leaving radiance. *Journal of Geophysical Research*, 105:6493–6499, 2000.

[22] H. R. Gordon. Atmospheric correction of ocean color imagery in the earth observing system era. *Journal of Geophysical Reseach: Atmospheres.*, (102):17081–17106, 1997.

[23] H. R. Gordon and M. Wang. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWIFS: a preliminary algorithm. *Applied Optics*, (33):443–452, 1994.

[24] C.D. Mobley, J. Werdell, B. Franz, Z. Ahmad, and S. Bailey. Atmospheric correction for satellite ocean color radiometry. 2016.

[25] J. Ma, S. Yang, X. Wang, and Y. Qiao. Atmospheric correction: Computing atmospheric diffuse transmittance. *Atmospheric Research*, 80(1):1 – 14, 2006.

[26] R.T. Wilson. Py6s: A python interface to the 6s radiative transfer model. *Computers and Geosciences*, 322(51):166–171, 2012.

[27] John O'Reilly, S. Maritorena, B.G. Mitchell, David Siegel, Kendall Carder, S.A. Garver, Mati Kahru, and Charles Mcclain. Ocean color chlorophyll algorithms for seawifs. *Journal of Geophysical Research*, 103:937–953, 1998.

[28] C.D. Mobley, L.K. Sundman, C.O. Davis, J.H. Bowles, T.V. Downes, R.A. Leathers, M.J. Montes, W.P. Bissett, D.D.R. Kohler, R.P Reid, E.M. Louchard, and A. Gleason. Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. *Applied Optics*, 44(17):3576–3592, 2005.

[29] K. Stamnes, B. Hamre, S. Stamnes, N. Chen, Y. Fan, W. Li, Z. Lin, and J. Stamnes. Progress in Forward-Inverse Modeling Based on Radiative Transfer Tools for Coupled Atmosphere-Snow/Ice-Ocean Systems: A Review and Description of the AccuRT Model. *Applied Sciences*, 8:2682, 2018.

[30] Y. Zhang. An improved 6s code for atmospheric correction based on water vapor content. *Advances in Remote Sensing*, 01:14–18, 2012.

[31] M. Montgomery, N. Malloy, C. Muhl, E. Degenstein, M. Kimble, D. Dean, D. Wright, J. Bauer, D. Wood, P. Lee, A. Flory, G. Newman, A. Daniels, T. Stohlgren, B. White, and L. Weisel. Atmospheric absorption & transmission. `http://gsp.humboldt.edu/olm_2016/courses/GSP_216_Online/lesson2-1/atmosphere.html`. Accessed: 2020-03-03.

[32] Terra & aqua moderate resolution imaging spectroradiometer (modis). `https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/`, 2019. Accessed: 2019-02-04.

[33] Y. Fan, W. Li, C. Gatebe, C. Jamet, G. Zibordi, T. Schroeder, and K. Stamnes. Atmospheric correction over coastal waters using multilayer neural networks. *Remote Sensing of Enviroment*, 199:218–240, 2017.

[34] R. Singh and P. Shanmugam. A robust method for removal of glint effects from satellite ocean colour imagery. *Ocean Science*, 11:2791–2829, 2014.

[35] C. Cox and W. Munk. Measurement of the roughness of the sea surface from photographs of the sun’s glitter. *Journal of the Optical Society of America*, 44(11):838–850, 1954.

[36] S. Kay, J.D. Hedley, and S. Lavender. Sun glint correction of high and low spatial resolution images of aquatic scenes: a review of methods for visible and near-infrared wavelengths. *Remote Sensing*, 1(4):697–730, 2009.

[37] H.R. Gordon and M. Wang. Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors. *Applied Optics*, 33(33):7754–7763, 1994.

[38] M. Stramska and T. Petelski. Observations of oceanic whitecaps in the north polar waters of the atlantic. *Journal of Geophysical Research*, 108, 2003.

[39] P. Koepke. Effective reflectance of oceanic whitecps. *Applied Optics - APPL OPT*, 23:1816–1824, 1984.

[40] R. Frouin, M. Schwindling, and P.Y. Deschamps. Spectral reflectance of sea foam in the visible and near-infrared: In situ measurements and remote sensing implications. *Journal of Geophysical Research*, 101(C6):14361–14371, 1996.

[41] A. Morel. Optical modeling of the upper ocean in relation to its biogeneous matter content (Case 1 waters). *Journal Geophysical Research*, 93, 1988.

[42] Changes in chlorophyll concentrations are affecting marine life. `https://www.globalchange.gov/browse/indicators/ocean-chlorophyll-concentrations`. Accessed: 2020-06-06.

[43] Colored dissolved organic matter. `https://en.wikipedia.org/wiki/Colored_dissolved_organic_matter`, 2020. Accessed: 2020-05-10.

[44] M. A. Jackson. Usgs’s landsat 8 captures debris imagery from hurricane florence, 2018. -Downloaded 06.06.2020 from `https://disasters.nasa.gov/hurricane-florence-2018/nasa-landsat-8-captures-debris-imagery-hurricane-florence`.

[45] B. Nechad, K. Ruddick, T. Schroeder, D. Blondeau-Patissier, N. Cherukuru, V. E. Brando, A. G. Dekker, L. Clementson, A. Banks, S. Maritorena, P. J. Werdell, C. Sá, V. Brotas, I. Caballero de Frutos, Y.-H. Ahn, S. Salama, G. Tilstone, V. Martinez-Vicente, D. Foley, M. McKibben, J. Nahorniak, T. D. Peterson, A. Siliò-Calzada, R. Röttgers, Z. Lee, M. Peters, and C. Brockmann. CoastColour Round Robin datasets, Version 1. PANGAEA, 2015.

[46] R.T. Wilson. Can I atmospherically-correct my images with Py6S? `http://blog.rtwilson.com/`. Accessed: 2020-01-03.

[47] D.R. Thompson, D.A. Roberts, B.C. Gao, R.O. Green, L. Guild, K. Hayashi, R. Kudela, and S. Palacios. Atmospheric correction with the bayesian empirical line. *Optical Express*, 24(3):2134–2144, 2016.

[48] A.D. Gerace, J.R. Schott, and R. Nevins. Increased potential to monitor water quality in the near-shore environment with Landsat's next-generation satellite. *Journal of Applied Remote Sensing*, 7(1):1–19, 2013.

[49] J. Vicent, J. Verrelst, N. Sabater, L. Alonso, J. P. Rivera-Caicedo, L. Martino, J. Muñoz Marí, and J. Moreno. Comparative analysis of atmospheric radiative transfer models using the Atmospheric Look-up table Generator (ALG) toolbox (version 2.0). *Geoscientific Model Development Discussions*, 2019:1–19, 2019.

[50] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, and J. Bosch. MODTRAN6: a major upgrade of the MODTRAN radiative transfer code. In M. Velez-Reyes and F.A. Kruse, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX*, volume 9088, pages 113 – 119. International Society for Optics and Photonics, SPIE, 2014.

[51] K. Stamnes, B. Hamre, S. Stamnes, and J. Stamnes. Accurt: A versatile tool for radiative transfer simulations in the coupled atmosphere-ocean system. *AIP Conference Proceedings*, 1810, 2017.

[52] J. Hedley. Introduction to hydrolight. `http://ioccg.org/wp-content/uploads/2018/07/hedley_hydrolight_intro.pdf`, 2018. Accessed: 2020-06-10.

[53] E. F. Vermote, D. Tanre, J. L. Deuze, M. Herman, and J.-J. Morcrette. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An Overview. *IEEE transactions on geoscience and remote sensing*, 35(3):675–686, 1997.

[54] M.I. Mishchenko, A.A. Lacis, and L.D. Travis. Errors induced by the neglect of polarization in radiance calculations for rayleigh-scattering atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 51:491–510, 1994.

[55] G. Anderson, Shepard Clough, F. Kneizys, J. Chetwynd, and Eric Shettle. AFGL Atmospheric Constituent Profiles (0.120km). *Environmental Research Papers*, 954:46, 1986.

[56] F.X. Kneizys, D.C. Robertson, L.W. Abreu, P. Acharya, G.P. Anderson, L.S. Rothman, J.H. Chetwynd, J.E.A. Selby, E.P. Shettle, W.O. Gallery, A. Berk, S.A. Clough, and L.S. Bernstein. The modtran 2/3 report and lowtran 7 model. 1996.

[57] B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser. On rayleigh optical depth calculations. *Journal of Atmospheric and Oceanic Technology*, 16(11):1854–1861, 1999.

[58] Z. Ahmad, B. A. Franz, C. R. McClain, E. J. Kwiatkowska, J. Werdell, E. P. Shettle, and B. N. Holben. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the seawifs and modis sensors over coastal regions and open oceans. *Applied Optics*, 49(29):5545–5560, 2010.

[59] A. Morel. *Bio-optical Models*, pages 317–326. In: S.A. Thorpe (eds). Academic press, San Diego, 2001.

[60] K. Ruddick. Coastcolour: Round robin protocol, version 1.2. *Brockmann Consult*, 2010. -found on `http://www.coastcolour.org/round_robin.html`.

[61] D.J. Segelstein. *The Complex Refractive Index of Water*. Department of Physics. University of Missouri-Kansas City, 1981.

[62] R. M. Pope and E. S. Fry. Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Applied Optics*, 36(33):8710–8723, 1997.

[63] A. Morel. *Optical properties of pure water and pure sea water*. Academic Press London, 1974.

[64] M. Babin, D. Stramski, G. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *Journal of Geophysical Research*, 108, 2003.

[65] A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research*, 103:31,033–31,044, 1998.

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

[67] J. Virmani and R. Weisberg. Relative humidity over the west florida continental shelf. *Monthly Weather Review*, 133:1671–1686, 2005.

[68] K. Zhang, W. Li, K. Stamnes, H. Eide, R. Spurr, and S.-C. Tsay. Assessment of the moderate-resolution imaging spectroradiometer algorithm for retrieval of aerosol parameters over the ocean. *Applied Optics*, 46(9):1525–1534, 2007.

[69] J. Werdell and S. Bailey. An improved bio-optical data set for ocean color algorithm development and satellite data product variation. *Remote Sensing of Environment*, 98:122–140, 2005.

[70] B. Nechad, K. Ruddick, T. Schroeder, K. Oubelkheir, D. Blondeau-Patissier, N. Cherukuru, V. Brando, A. Dekker, L. Clementson, A. Banks, S. Maritorena, J. Werdell, C. Sá, V. Brotas, I. Caballero, Y.-H. Ahn, M. Salama, G. Tilstone, V. Martinez-Vicente, and C. Brockmann. Coastcolour round robin datasets: A database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. *Earth System Science Data*, 7, 2015.

[71] Chlorophyll a (chlor_a). -found on `https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/`.

[72] Preprocessing data. `https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing`. Accessed: 2020-03-20.

[73] Savitzky-golay filter. `https://en.wikipedia.org/wiki/Savitzky-Golay_filter`, 2020. Accessed: 2020-03-30.

[74] The nirpy research blog. `https://nirpyresearch.com/`, 2020. Accessed: 2020-05-10.

[75] C. Ruffin and R. King. The analysis of hyperspectral data using Savitzky-Golay filtering-theoretical basis. 1. *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293), Hamburg, Germany*, 2:756–758, 1999.

[76] scipy.signal.savgol_filter. `https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol$_$filter.html`. Accessed: 2020-06-06.

[77] F. Chollet. *Deep Learning with Python*. Manning Publications Co., 20 Baldwin Road, PO Box 761, Shelter Island, NY 11964, 2018.

[78] J. Heaton. *Artificial Intelligence for Humans*. Createspace Independent Publishing Platform, 2015.

[79] T. Wu, H. Martens, P. Hunter, and K. Mithraratne. Emulating facial biomechanics using multivariate partial least squares surrogate models. *International Journal for Numerical Methods in Biomedical Engineering*, 30:1103–1120, 2014.

[80] J. C. Pyo, H. Duan, M. Ligaray, M. Kim, S. Baek, Y. Kwon, H. Lee, T. Kang, K. Kim, Y. Cha, and K. Cho. An integrative remote sensing application of stacked autoencoder for atmospheric correction and cyanobacteria estimation using hyperspectral imagery. *Remote Sensing*, 12:1073, 2020.

[81] Stochastic gradient descent. `https://en.wikipedia.org/wiki/Stochastic_gradient_descent`, 2020. Accessed: 2020-03-25.

[82] Stochastic gradient descent. `https://scikit-learn.org/stable/modules/sgd.html#sgd`, 2019. Accessed: 2020-03-25.

Appendices

Appendix A

Atmospheric Correction Results

Table A.1: Optimal results of AC of $L_{rac}(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $400\text{-}600\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	0.992	0.909	0.896	0.951	0.061	0.205	0.219	0.167
405 nm	0.993	0.916	0.89	0.956	0.057	0.198	0.227	0.171
410 nm	0.994	0.923	0.907	0.949	0.054	0.191	0.211	0.236
415 nm	0.995	0.929	0.915	0.974	0.05	0.186	0.204	0.149
420 nm	0.996	0.935	0.919	0.976	0.047	0.181	0.203	0.118
425 nm	0.996	0.94	0.916	0.98	0.044	0.177	0.212	0.109
430 nm	0.997	0.944	0.931	0.989	0.041	0.174	0.196	0.078
435 nm	0.997	0.948	0.943	0.986	0.039	0.172	0.182	0.146
440 nm	0.998	0.951	0.944	0.989	0.036	0.17	0.183	0.086
445 nm	0.998	0.954	0.943	0.994	0.035	0.168	0.186	0.064
450 nm	0.998	0.956	0.948	0.991	0.033	0.167	0.182	0.081
455 nm	0.998	0.957	0.95	0.993	0.032	0.166	0.182	0.08
460 nm	0.999	0.959	0.952	0.994	0.031	0.165	0.179	0.077
465 nm	0.999	0.96	0.941	0.996	0.03	0.165	0.2	0.059
470 nm	0.999	0.962	0.952	0.995	0.029	0.164	0.183	0.093
475 nm	0.999	0.963	0.948	0.996	0.029	0.164	0.209	0.059
480 nm	0.999	0.964	0.951	0.996	0.028	0.163	0.191	0.061
485 nm	0.999	0.965	0.96	0.996	0.028	0.163	0.18	0.06
490 nm	0.999	0.966	0.96	0.997	0.028	0.164	0.179	0.05
495 nm	0.999	0.967	0.954	0.996	0.027	0.165	0.196	0.062
500 nm	0.999	0.968	0.959	0.997	0.027	0.167	0.189	0.059
505 nm	0.999	0.969	0.962	0.998	0.027	0.17	0.188	0.068
510 nm	0.999	0.97	0.958	0.998	0.027	0.174	0.207	0.053
515 nm	0.999	0.971	0.961	0.998	0.029	0.177	0.207	0.053
520 nm	0.999	0.971	0.958	0.998	0.028	0.177	0.216	0.05
525 nm	0.999	0.972	0.967	0.998	0.029	0.177	0.217	0.047
530 nm	0.999	0.973	0.966	0.998	0.029	0.178	0.207	0.048
535 nm	0.999	0.974	0.967	0.998	0.029	0.179	0.203	0.049
540 nm	1.0	0.974	0.965	0.998	0.029	0.18	0.215	0.049
545 nm	1.0	0.975	0.97	0.998	0.03	0.181	0.208	0.05
550 nm	1.0	0.976	0.971	0.998	0.03	0.183	0.217	0.054
555 nm	1.0	0.976	0.969	0.998	0.03	0.184	0.222	0.054
560 nm	1.0	0.977	0.967	0.998	0.031	0.184	0.218	0.053
565 nm	1.0	0.977	0.974	0.998	0.029	0.184	0.206	0.054
570 nm	1.0	0.978	0.966	0.998	0.03	0.186	0.231	0.056
575 nm	1.0	0.979	0.969	0.998	0.029	0.189	0.229	0.057
580 nm	1.0	0.979	0.971	0.998	0.031	0.194	0.23	0.056
585 nm	1.0	0.98	0.974	0.998	0.031	0.2	0.227	0.058
590 nm	1.0	0.98	0.973	0.998	0.032	0.205	0.24	0.059
595 nm	1.0	0.981	0.974	0.998	0.032	0.212	0.251	0.061
600 nm	1.0	0.982	0.974	0.998	0.035	0.217	0.26	0.065

Table A.2: Optimal results of AC of $L_{rac}(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $605\text{-}800\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
605 nm	1.0	0.982	0.976	0.998	0.034	0.219	0.258	0.064
610 nm	1.0	0.983	0.977	0.998	0.035	0.22	0.257	0.064
615 nm	1.0	0.983	0.978	0.998	0.035	0.219	0.257	0.066
620 nm	1.0	0.983	0.978	0.998	0.035	0.219	0.255	0.066
625 nm	1.0	0.984	0.975	0.998	0.036	0.219	0.273	0.066
630 nm	1.0	0.984	0.982	0.998	0.036	0.219	0.252	0.067
635 nm	1.0	0.984	0.978	0.998	0.036	0.219	0.262	0.068
640 nm	1.0	0.985	0.978	0.998	0.037	0.219	0.261	0.069
645 nm	1.0	0.985	0.98	0.998	0.037	0.219	0.253	0.069
650 nm	1.0	0.985	0.981	0.998	0.038	0.219	0.247	0.069
655 nm	1.0	0.986	0.98	0.998	0.039	0.219	0.256	0.069
660 nm	1.0	0.986	0.983	0.998	0.04	0.219	0.244	0.069
665 nm	1.0	0.986	0.982	0.998	0.043	0.22	0.253	0.071
670 nm	1.0	0.986	0.985	0.998	0.044	0.221	0.236	0.073
675 nm	1.0	0.987	0.984	0.998	0.044	0.221	0.241	0.073
680 nm	1.0	0.987	0.985	0.998	0.045	0.22	0.239	0.072
685 nm	1.0	0.987	0.984	0.999	0.041	0.218	0.243	0.071
690 nm	1.0	0.987	0.985	0.999	0.042	0.216	0.242	0.069
695 nm	1.0	0.988	0.986	0.999	0.043	0.215	0.233	0.07
700 nm	1.0	0.988	0.985	0.998	0.043	0.214	0.238	0.071
705 nm	1.0	0.988	0.988	0.998	0.045	0.214	0.219	0.073
710 nm	1.0	0.989	0.988	0.999	0.049	0.212	0.218	0.072
715 nm	1.0	0.989	0.988	0.999	0.052	0.211	0.215	0.072
720 nm	0.999	0.989	0.988	0.998	0.057	0.209	0.221	0.078
725 nm	0.999	0.989	0.988	0.998	0.065	0.209	0.232	0.087
730 nm	0.999	0.989	0.985	0.998	0.07	0.21	0.244	0.087
735 nm	0.999	0.989	0.987	0.997	0.076	0.211	0.225	0.103
740 nm	0.999	0.989	0.986	0.997	0.08	0.212	0.231	0.108
745 nm	0.999	0.988	0.987	0.996	0.081	0.213	0.234	0.12
750 nm	0.999	0.988	0.986	0.997	0.082	0.213	0.231	0.108
755 nm	0.999	0.988	0.985	0.997	0.083	0.213	0.242	0.116
760 nm	0.999	0.988	0.987	0.996	0.083	0.213	0.227	0.119
765 nm	0.999	0.988	0.987	0.997	0.084	0.213	0.228	0.108
770 nm	0.999	0.988	0.987	0.997	0.082	0.213	0.226	0.106
775 nm	0.999	0.989	0.987	0.997	0.08	0.212	0.228	0.103
780 nm	0.999	0.989	0.988	0.997	0.079	0.212	0.217	0.106
785 nm	0.999	0.989	0.987	0.997	0.078	0.211	0.228	0.105
790 nm	0.999	0.989	0.988	0.997	0.077	0.211	0.223	0.105
795 nm	0.999	0.989	0.988	0.997	0.075	0.211	0.226	0.107
800 nm	0.999	0.989	0.987	0.997	0.076	0.211	0.228	0.095

Table A.3: Optimal results of AC of $L_{rac}(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands $400\text{-}600\,\mathrm{nm}$.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	6.1	38.29	38.38	33.98	2.99	15.97	16.73	−12.74
405 nm	5.62	35.91	33.29	30.42	2.55	15.02	15.81	−15.88
410 nm	5.26	33.64	35.52	52.85	2.38	13.99	13.49	−41.77
415 nm	4.78	31.73	29.05	30.87	1.87	13.03	11.72	−22.35
420 nm	4.4	30.07	26.89	26.63	1.3	12.15	11.8	16.57
425 nm	4.18	28.68	27.65	20.63	1.34	11.38	16.04	−1.76
430 nm	3.87	27.15	29.58	15.92	1.0	10.48	7.21	5.0
435 nm	3.63	25.87	24.43	31.67	0.83	9.72	8.65	−26.94
440 nm	3.29	24.65	25.67	15.9	0.29	8.98	7.69	7.12
445 nm	3.19	23.82	24.01	10.69	0.45	8.48	6.9	−2.2
450 nm	3.03	23.08	21.55	17.8	0.03	8.02	7.43	10.77
455 nm	3.05	22.47	20.4	12.35	−0.39	7.63	8.99	−8.54
460 nm	2.9	21.85	19.49	18.5	−0.37	7.22	8.37	14.23
465 nm	2.75	21.29	15.7	8.81	−0.37	6.85	6.1	−3.43
470 nm	2.65	20.74	18.47	18.56	−0.45	6.49	5.64	−16.36
475 nm	2.66	20.34	19.97	9.61	−0.63	6.24	12.63	0.6
480 nm	2.63	19.91	14.48	10.54	−0.79	5.96	4.36	2.8
485 nm	2.52	19.5	18.83	8.48	−0.81	5.71	9.57	−0.4
490 nm	2.39	19.06	15.97	7.74	−0.43	5.4	5.56	−1.08
495 nm	2.43	18.72	13.22	8.94	−0.63	5.16	5.6	−3.69
500 nm	2.3	18.48	14.86	8.81	−0.2	4.92	1.34	−4.15
505 nm	2.42	18.42	14.93	13.4	−0.65	4.7	0.41	−11.86
510 nm	2.48	18.53	11.48	8.46	−0.85	4.52	4.18	−3.57
515 nm	2.35	18.8	11.18	8.41	−0.22	4.43	0.54	−1.6
520 nm	2.36	18.81	10.42	10.09	0.0	4.4	2.52	5.46
525 nm	2.43	18.92	18.17	8.38	−0.47	4.38	12.14	0.86
530 nm	2.37	19.09	12.47	8.73	0.29	4.38	3.45	2.89
535 nm	2.4	19.37	13.84	8.71	0.06	4.39	3.87	−0.6
540 nm	2.43	19.76	10.49	9.33	0.09	4.43	3.24	2.82
545 nm	2.51	20.33	13.42	9.48	0.15	4.53	3.85	1.53
550 nm	2.56	21.1	17.15	10.38	−0.36	4.68	9.17	−0.08
555 nm	2.75	21.72	13.06	10.82	−0.85	4.84	6.28	1.9
560 nm	2.74	22.41	14.13	11.29	0.22	5.06	3.38	2.47
565 nm	2.88	23.15	17.63	11.98	−0.68	5.28	3.46	3.53
570 nm	2.87	24.15	9.95	12.98	0.03	5.51	0.49	3.78
575 nm	2.95	25.42	14.3	13.89	−0.18	5.83	7.05	4.58
580 nm	3.23	27.25	17.35	14.87	−0.75	6.23	5.84	3.24
585 nm	3.58	29.44	19.04	16.28	−1.54	6.76	−0.96	2.1
590 nm	3.5	31.64	12.93	17.58	−0.91	7.33	3.02	4.0
595 nm	3.52	34.14	13.31	19.45	−0.31	8.04	6.06	4.2
600 nm	4.32	35.91	13.44	21.33	−2.36	8.59	0.39	−1.58

Table A.4: Optimal results of AC of $L_{rac}(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands $605\text{-}800\,\mathrm{nm}$.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
605 nm	4.09	37.02	19.16	22.01	−2.01	8.94	11.69	0.37
610 nm	4.73	37.61	18.14	22.42	−2.98	9.11	9.21	7.15
615 nm	4.1	37.98	16.85	22.8	−1.2	9.19	9.86	2.09
620 nm	3.96	38.43	22.91	23.27	−0.82	9.31	4.89	8.24
625 nm	3.99	38.9	22.04	23.49	0.07	9.45	11.63	6.04
630 nm	4.18	39.35	31.15	24.59	−0.99	9.58	4.94	7.95
635 nm	4.19	39.77	13.43	24.81	0.08	9.71	−4.89	5.67
640 nm	4.79	40.21	14.25	25.73	−2.14	9.85	2.05	8.27
645 nm	4.13	40.7	16.54	26.21	0.53	10.02	7.65	9.28
650 nm	4.19	41.1	21.83	26.18	0.31	10.17	13.36	9.96
655 nm	4.59	41.35	15.7	25.7	−1.54	10.28	3.77	5.03
660 nm	4.7	41.38	37.18	26.27	1.25	10.31	8.98	7.25
665 nm	4.65	41.43	18.26	27.89	0.43	10.3	6.48	8.92
670 nm	4.83	41.58	35.05	29.96	0.64	10.31	7.93	13.4
675 nm	4.8	41.83	20.15	29.66	−0.88	10.4	2.97	11.76
680 nm	5.58	42.14	28.59	29.18	2.98	10.54	2.97	10.79
685 nm	4.7	42.53	21.17	28.0	0.6	10.77	12.7	10.73
690 nm	4.74	43.05	24.52	26.46	0.26	11.06	−7.99	4.34
695 nm	5.52	43.55	27.81	26.82	2.53	11.31	20.27	3.09
700 nm	5.34	43.87	16.35	26.99	1.95	11.48	0.41	1.72
705 nm	5.39	44.02	42.99	27.73	1.9	11.55	4.62	2.16
710 nm	5.5	44.29	38.89	28.93	1.85	11.58	3.76	4.26
715 nm	6.44	44.66	34.16	31.32	3.82	11.6	4.52	11.37
720 nm	5.7	45.24	23.77	29.28	1.26	11.61	10.5	−1.35
725 nm	5.89	45.94	27.16	35.63	1.4	11.67	−8.67	21.02
730 nm	7.17	46.72	29.03	31.23	−4.11	11.77	4.29	9.67
735 nm	6.02	47.43	23.94	33.13	0.64	11.89	−3.72	10.41
740 nm	7.05	47.86	22.21	33.64	−3.35	12.0	5.05	−4.25
745 nm	6.23	48.19	25.12	54.18	−0.56	12.13	−3.87	46.36
750 nm	6.38	48.48	27.92	38.54	−0.79	12.24	−15.39	24.6
755 nm	7.01	48.7	25.8	37.79	−2.96	12.35	−12.0	−17.21
760 nm	7.75	48.86	26.03	41.27	−4.32	12.46	12.38	−8.76
765 nm	7.39	49.05	28.33	40.3	−3.46	12.54	11.41	26.29
770 nm	6.81	49.18	27.04	40.48	−2.44	12.64	4.76	23.08
775 nm	6.94	49.27	26.14	34.22	−3.01	12.73	12.68	6.95
780 nm	6.63	49.37	31.96	39.17	−1.82	12.83	8.9	25.09
785 nm	9.49	49.43	28.08	34.42	−6.91	12.91	−11.97	−3.6
790 nm	6.73	49.52	29.18	35.08	−2.47	13.01	13.75	0.26
795 nm	6.31	49.67	34.15	50.03	−0.72	13.11	9.81	39.64
800 nm	9.23	49.78	25.33	35.43	−6.48	13.17	12.32	15.02

Table A.5: Optimal results of AC of $L_t(\lambda)$ to $L_{rac}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $400\text{-}600\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	0.9995	0.9951	0.9857	0.9978	0.021	0.056	0.096	0.041
405 nm	0.9995	0.9954	0.9863	0.9976	0.02	0.054	0.094	0.04
410 nm	0.9996	0.9957	0.9871	0.9979	0.019	0.052	0.091	0.036
415 nm	0.9996	0.996	0.9876	0.9979	0.018	0.051	0.093	0.035
420 nm	0.9996	0.9962	0.9885	0.9979	0.017	0.049	0.088	0.035
425 nm	0.9997	0.9964	0.9888	0.9981	0.017	0.047	0.084	0.034
430 nm	0.9997	0.9967	0.9895	0.9977	0.017	0.046	0.081	0.039
435 nm	0.9997	0.9969	0.9901	0.9982	0.015	0.044	0.078	0.036
440 nm	0.9997	0.997	0.9906	0.9982	0.015	0.043	0.077	0.032
445 nm	0.9998	0.9972	0.9909	0.9983	0.015	0.041	0.075	0.034
450 nm	0.9998	0.9974	0.9915	0.9984	0.015	0.04	0.073	0.033
455 nm	0.9998	0.9975	0.9918	0.9984	0.015	0.038	0.071	0.036
460 nm	0.9998	0.9977	0.9922	0.9985	0.013	0.037	0.069	0.029
465 nm	0.9998	0.9978	0.9925	0.9985	0.014	0.036	0.07	0.032
470 nm	0.9998	0.9979	0.9927	0.9986	0.012	0.034	0.065	0.028
475 nm	0.9998	0.9981	0.993	0.9986	0.012	0.033	0.065	0.03
480 nm	0.9998	0.9982	0.9933	0.9986	0.011	0.032	0.063	0.03
485 nm	0.9998	0.9983	0.9937	0.9986	0.012	0.031	0.059	0.029
490 nm	0.9998	0.9984	0.9942	0.9986	0.011	0.03	0.057	0.029
495 nm	0.9998	0.9985	0.9945	0.9988	0.011	0.029	0.055	0.026
500 nm	0.9998	0.9986	0.9949	0.9988	0.011	0.028	0.053	0.027
505 nm	0.9999	0.9987	0.9954	0.9989	0.01	0.027	0.052	0.024
510 nm	0.9999	0.9988	0.9957	0.9988	0.01	0.026	0.049	0.024
515 nm	0.9999	0.9989	0.9961	0.9989	0.011	0.025	0.047	0.024
520 nm	0.9999	0.9989	0.9962	0.9989	0.01	0.024	0.047	0.024
525 nm	0.9999	0.999	0.9966	0.9989	0.01	0.023	0.044	0.023
530 nm	0.9999	0.9991	0.9968	0.999	0.01	0.023	0.043	0.023
535 nm	0.9999	0.9991	0.9969	0.999	0.01	0.022	0.045	0.023
540 nm	0.9999	0.9992	0.9971	0.999	0.01	0.021	0.043	0.023
545 nm	0.9999	0.9993	0.9974	0.999	0.009	0.021	0.039	0.023
550 nm	0.9999	0.9993	0.9974	0.999	0.01	0.02	0.039	0.023
555 nm	0.9999	0.9993	0.9977	0.9989	0.01	0.02	0.037	0.024
560 nm	0.9999	0.9994	0.9977	0.9989	0.01	0.019	0.037	0.024
565 nm	0.9999	0.9994	0.9979	0.9989	0.01	0.019	0.036	0.024
570 nm	0.9999	0.9995	0.998	0.9988	0.009	0.018	0.035	0.025
575 nm	0.9999	0.9995	0.9981	0.9988	0.009	0.018	0.035	0.026
580 nm	0.9999	0.9995	0.9981	0.9988	0.009	0.018	0.035	0.026
585 nm	0.9999	0.9995	0.9982	0.9987	0.009	0.018	0.035	0.027
590 nm	0.9999	0.9995	0.9983	0.9987	0.008	0.018	0.036	0.028
595 nm	0.9999	0.9996	0.9984	0.9988	0.01	0.018	0.035	0.028
600 nm	0.9999	0.9996	0.9984	0.9988	0.01	0.018	0.035	0.029

Table A.6: Optimal results of AC of $L_t(\lambda)$ to $L_{rac}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $605\text{-}800\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
$605\,\mathrm{nm}$	0.9999	0.9996	0.9985	0.9989	0.01	0.018	0.035	0.028
$610\,\mathrm{nm}$	0.9999	0.9996	0.9985	0.9988	0.011	0.017	0.038	0.03
$615\,\mathrm{nm}$	0.9999	0.9996	0.9986	0.9987	0.009	0.017	0.034	0.03
$620\,\mathrm{nm}$	0.9999	0.9996	0.9986	0.9987	0.011	0.017	0.034	0.03
$625\,\mathrm{nm}$	0.9999	0.9997	0.9987	0.9987	0.011	0.017	0.032	0.031
$630\,\mathrm{nm}$	0.9999	0.9997	0.9988	0.9987	0.01	0.016	0.032	0.03
$635\,\mathrm{nm}$	0.9999	0.9997	0.9988	0.9987	0.009	0.016	0.032	0.031
$640\,\mathrm{nm}$	0.9999	0.9997	0.9989	0.9988	0.01	0.016	0.032	0.03
$645\,\mathrm{nm}$	0.9999	0.9997	0.9989	0.9987	0.011	0.016	0.031	0.031
$650\,\mathrm{nm}$	0.9999	0.9997	0.9989	0.9987	0.011	0.016	0.032	0.031
$655\,\mathrm{nm}$	0.9999	0.9997	0.9989	0.9988	0.012	0.015	0.03	0.03
$660\,\mathrm{nm}$	0.9999	0.9997	0.999	0.9987	0.011	0.015	0.03	0.031
$665\,\mathrm{nm}$	0.9999	0.9997	0.999	0.9988	0.01	0.015	0.031	0.031
$670\,\mathrm{nm}$	0.9999	0.9997	0.999	0.9988	0.013	0.015	0.031	0.031
$675\,\mathrm{nm}$	0.9999	0.9998	0.999	0.9989	0.011	0.015	0.03	0.03
$680\,\mathrm{nm}$	0.9999	0.9998	0.999	0.9988	0.01	0.015	0.033	0.03
$685\,\mathrm{nm}$	0.9999	0.9998	0.9991	0.9986	0.01	0.014	0.028	0.033
$690\,\mathrm{nm}$	0.9999	0.9998	0.9991	0.9987	0.011	0.014	0.028	0.033
$695\,\mathrm{nm}$	0.9999	0.9998	0.9991	0.9987	0.009	0.014	0.028	0.033
$700\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9988	0.011	0.014	0.028	0.032
$705\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9988	0.012	0.014	0.028	0.032
$710\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9987	0.011	0.014	0.029	0.034
$715\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9987	0.012	0.014	0.028	0.035
$720\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9986	0.014	0.014	0.029	0.036
$725\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9987	0.015	0.014	0.029	0.036
$730\,\mathrm{nm}$	0.9999	0.9998	0.9993	0.9987	0.015	0.013	0.028	0.037
$735\,\mathrm{nm}$	0.9999	0.9998	0.9992	0.9987	0.018	0.013	0.029	0.037
$740\,\mathrm{nm}$	0.9999	0.9998	0.9993	0.9987	0.014	0.013	0.029	0.038
$745\,\mathrm{nm}$	0.9999	0.9998	0.9993	0.9987	0.018	0.013	0.027	0.038
$750\,\mathrm{nm}$	0.9999	0.9998	0.9993	0.9987	0.015	0.013	0.028	0.038
$755\,\mathrm{nm}$	0.9999	0.9998	0.9994	0.9987	0.015	0.013	0.028	0.038
$760\,\mathrm{nm}$	0.9999	0.9999	0.9994	0.9988	0.018	0.013	0.026	0.037
$765\,\mathrm{nm}$	0.9999	0.9999	0.9995	0.9987	0.015	0.012	0.025	0.036
$770\,\mathrm{nm}$	0.9999	0.9999	0.9995	0.9987	0.015	0.012	0.024	0.036
$775\,\mathrm{nm}$	0.9999	0.9999	0.9995	0.9987	0.013	0.012	0.023	0.037
$780\,\mathrm{nm}$	0.9999	0.9999	0.9995	0.9986	0.015	0.012	0.022	0.037
$785\,\mathrm{nm}$	0.9999	0.9999	0.9995	0.9987	0.014	0.011	0.023	0.035
$790\,\mathrm{nm}$	0.9999	0.9999	0.9996	0.9987	0.012	0.011	0.021	0.034
$795\,\mathrm{nm}$	0.9999	0.9999	0.9996	0.9987	0.013	0.011	0.021	0.035
$800\,\mathrm{nm}$	0.9999	0.9999	0.9996	0.9988	0.013	0.011	0.022	0.033

Table A.7: Optimal results of AC of $L_t(\lambda)$ to $L_{rac}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands 400-$600\,\mathrm{nm}$.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	3.32	8.2	15.91	7.86	−1.58	2.89	6.58	6.0
405 nm	2.95	7.44	15.72	8.08	−1.38	2.31	7.3	5.89
410 nm	2.88	6.99	14.61	5.81	−1.72	2.03	6.85	3.02
415 nm	2.43	6.61	12.2	5.07	−1.0	1.83	1.34	1.39
420 nm	2.31	6.29	13.8	5.61	−1.02	1.66	6.71	2.73
425 nm	2.35	5.98	11.62	5.41	−1.43	1.5	2.99	2.89
430 nm	1.95	5.71	11.04	7.74	−0.77	1.36	2.36	5.73
435 nm	1.96	5.46	10.93	4.44	−1.11	1.24	3.58	−1.89
440 nm	1.66	5.23	10.8	4.55	−0.54	1.13	3.73	1.41
445 nm	1.68	4.97	10.3	6.26	−0.73	1.03	2.7	4.8
450 nm	1.76	4.75	10.11	5.57	−1.15	0.93	3.79	4.0
455 nm	1.73	4.52	9.53	6.4	−1.15	0.84	3.02	5.28
460 nm	1.53	4.3	9.62	4.23	−0.95	0.75	3.73	1.87
465 nm	1.51	4.1	9.84	5.83	−0.97	0.68	4.28	4.62
470 nm	1.43	3.92	8.42	4.34	−0.93	0.62	1.61	2.6
475 nm	1.6	3.72	9.05	4.87	−1.23	0.55	3.38	3.61
480 nm	1.31	3.55	7.56	4.8	−0.78	0.49	−0.46	3.52
485 nm	1.42	3.4	7.13	4.54	−1.07	0.45	0.54	3.21
490 nm	1.45	3.27	6.96	4.48	−1.17	0.41	1.09	3.11
495 nm	1.26	3.15	6.77	3.9	−0.86	0.36	1.38	2.28
500 nm	1.35	3.04	6.66	4.2	−1.1	0.33	1.27	2.75
505 nm	1.41	2.97	6.77	3.63	−1.19	0.29	2.12	1.69
510 nm	1.27	2.93	6.0	3.59	−1.01	0.27	−0.33	1.59
515 nm	1.56	2.89	5.83	3.46	−1.4	0.25	0.51	0.89
520 nm	1.56	2.78	5.9	3.48	−1.36	0.22	0.46	0.78
525 nm	1.6	2.69	5.46	3.34	−1.45	0.2	0.1	0.81
530 nm	1.51	2.61	5.55	3.11	−1.32	0.18	1.18	1.0
535 nm	1.51	2.55	6.15	3.28	−1.33	0.17	2.24	1.21
540 nm	1.57	2.49	5.57	3.29	−1.38	0.16	−1.62	1.12
545 nm	1.36	2.45	5.16	3.23	−1.13	0.14	0.68	0.93
550 nm	1.7	2.43	5.34	3.69	−1.51	0.13	0.85	1.43
555 nm	1.73	2.37	5.0	3.64	−1.57	0.12	−0.93	0.99
560 nm	1.66	2.32	5.27	3.43	−1.42	0.1	1.2	0.69
565 nm	1.44	2.28	4.84	3.67	−1.14	0.09	−0.52	0.78
570 nm	1.39	2.28	4.78	4.03	−1.09	0.08	−0.55	1.05
575 nm	1.37	2.29	5.08	4.36	−1.1	0.07	1.22	2.13
580 nm	1.52	2.36	4.95	4.77	−1.22	0.06	0.05	2.28
585 nm	1.24	2.47	5.42	4.87	−0.92	0.04	−1.68	1.85
590 nm	1.18	2.58	5.42	5.79	−0.78	0.03	1.27	3.08
595 nm	1.5	2.75	5.45	5.48	−1.31	0.03	−0.14	1.95
600 nm	1.55	2.88	6.03	5.89	−1.32	0.04	−1.31	1.85

Table A.8: Optimal results of AC of $L_t(\lambda)$ to $L_{rac}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands 605-800 nm.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
605 nm	1.84	2.96	6.01	5.69	−1.67	0.05	0.94	1.6
610 nm	1.84	2.97	7.3	6.36	−1.71	0.04	−3.63	2.56
615 nm	1.37	2.92	6.1	6.88	−0.91	0.04	0.96	3.34
620 nm	1.72	2.89	6.02	6.99	−1.5	0.05	−1.09	3.78
625 nm	1.76	2.86	5.85	7.34	−1.61	0.05	0.08	4.17
630 nm	2.01	2.81	5.9	6.81	−1.87	0.04	0.71	3.56
635 nm	1.47	2.8	5.88	6.83	−1.23	0.05	−0.63	3.15
640 nm	1.65	2.77	6.06	6.72	−1.4	0.05	1.1	3.28
645 nm	1.83	2.74	5.95	6.62	−1.64	0.05	1.22	2.54
650 nm	2.44	2.72	6.69	6.83	−2.3	0.04	−3.04	2.88
655 nm	1.46	2.74	5.96	6.76	−1.08	0.05	−1.34	2.59
660 nm	1.91	2.77	5.93	7.09	−1.71	0.06	0.66	3.12
665 nm	1.32	2.76	6.42	7.19	−0.8	0.07	2.0	3.3
670 nm	2.01	2.74	6.35	6.95	−1.83	0.07	−1.92	2.62
675 nm	1.42	2.7	6.22	6.82	−0.93	0.07	−1.83	2.21
680 nm	1.34	2.66	7.88	6.92	−0.9	0.06	−4.61	2.61
685 nm	1.57	2.62	5.85	7.64	−1.32	0.06	−0.41	3.07
690 nm	1.64	2.61	5.97	7.24	−1.33	0.06	−1.74	1.21
695 nm	2.04	2.57	5.85	7.84	−1.72	0.05	−1.05	2.54
700 nm	1.76	2.6	5.9	7.61	−1.42	0.06	1.46	3.54
705 nm	2.48	2.64	5.85	7.92	−2.17	0.06	−0.35	4.05
710 nm	2.79	2.73	6.18	8.38	−2.56	0.08	1.27	3.66
715 nm	2.25	2.8	6.32	9.44	−1.9	0.09	0.03	4.6
720 nm	2.57	2.94	6.61	10.38	−2.29	0.09	−0.77	5.28
725 nm	2.87	3.0	6.72	10.24	−2.64	0.13	−0.63	4.36
730 nm	2.56	3.05	6.81	10.37	−2.27	0.11	0.02	3.98
735 nm	3.75	3.1	7.33	11.08	−3.61	0.14	0.94	4.21
740 nm	2.24	3.08	7.27	12.05	−1.77	0.11	1.37	5.6
745 nm	2.58	3.05	7.02	11.4	−2.2	0.09	1.07	4.94
750 nm	2.88	2.99	7.18	11.86	−2.56	0.08	1.91	5.77
755 nm	2.91	2.92	7.81	11.67	−2.3	0.08	−3.86	4.94
760 nm	2.77	2.86	6.99	11.68	−2.26	0.08	−2.7	5.41
765 nm	2.41	2.77	6.3	11.7	−1.68	0.07	1.44	6.28
770 nm	2.08	2.71	6.14	11.5	−1.19	0.11	1.37	5.77
775 nm	2.21	2.62	5.9	11.37	−1.44	0.07	−0.46	5.5
780 nm	1.96	2.54	5.68	11.23	−0.77	0.06	0.07	5.44
785 nm	1.97	2.46	5.97	11.18	−0.84	0.06	−2.33	6.03
790 nm	1.96	2.38	5.18	11.24	−0.79	0.04	−0.57	6.46
795 nm	2.13	2.33	5.39	11.41	0.38	0.05	1.23	7.0
800 nm	2.22	2.27	5.75	11.18	−1.14	0.04	2.69	7.87

Table A.9: Optimal results of AC of $L_t(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $400\text{-}600\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	0.985	0.877	0.837	0.838	0.087	0.248	0.287	0.295
405 nm	0.987	0.889	0.849	0.74	0.082	0.236	0.276	0.369
410 nm	0.989	0.901	0.869	0.822	0.076	0.225	0.261	0.305
415 nm	0.991	0.912	0.885	0.855	0.07	0.215	0.245	0.297
420 nm	0.992	0.921	0.886	0.833	0.065	0.207	0.249	0.305
425 nm	0.994	0.928	0.908	0.93	0.061	0.2	0.227	0.199
430 nm	0.995	0.935	0.917	0.741	0.056	0.194	0.22	0.404
435 nm	0.996	0.942	0.923	0.894	0.052	0.188	0.216	0.308
440 nm	0.996	0.947	0.927	0.971	0.049	0.183	0.217	0.141
445 nm	0.997	0.951	0.93	0.894	0.047	0.179	0.215	0.279
450 nm	0.997	0.954	0.942	0.945	0.045	0.175	0.205	0.229
455 nm	0.997	0.957	0.932	0.928	0.043	0.172	0.222	0.244
460 nm	0.998	0.959	0.947	0.939	0.042	0.17	0.199	0.236
465 nm	0.998	0.961	0.95	0.955	0.04	0.168	0.192	0.231
470 nm	0.998	0.963	0.936	0.972	0.039	0.166	0.22	0.158
475 nm	0.998	0.965	0.948	0.984	0.038	0.164	0.202	0.116
480 nm	0.998	0.967	0.947	0.991	0.037	0.163	0.206	0.097
485 nm	0.999	0.968	0.957	0.968	0.036	0.162	0.195	0.209
490 nm	0.999	0.969	0.953	0.988	0.035	0.161	0.207	0.118
495 nm	0.999	0.97	0.958	0.961	0.035	0.162	0.201	0.235
500 nm	0.999	0.972	0.96	0.992	0.034	0.163	0.199	0.098
505 nm	0.999	0.973	0.964	0.99	0.033	0.165	0.193	0.111
510 nm	0.999	0.974	0.96	0.996	0.032	0.169	0.209	0.072
515 nm	0.999	0.975	0.958	0.996	0.031	0.172	0.23	0.069
520 nm	0.999	0.975	0.967	0.993	0.032	0.172	0.21	0.137
525 nm	0.999	0.976	0.951	0.995	0.03	0.172	0.247	0.081
530 nm	0.999	0.977	0.964	0.993	0.03	0.173	0.214	0.143
535 nm	0.999	0.977	0.967	0.997	0.03	0.174	0.209	0.067
540 nm	0.999	0.978	0.969	0.996	0.029	0.175	0.212	0.082
545 nm	0.999	0.979	0.971	0.992	0.029	0.176	0.208	0.152
550 nm	1.0	0.979	0.968	0.99	0.029	0.178	0.232	0.212
555 nm	1.0	0.98	0.973	0.996	0.028	0.179	0.21	0.095
560 nm	1.0	0.98	0.971	0.997	0.028	0.179	0.223	0.087
565 nm	1.0	0.981	0.972	0.995	0.029	0.179	0.222	0.122
570 nm	1.0	0.981	0.974	0.997	0.029	0.181	0.22	0.091
575 nm	1.0	0.982	0.971	0.997	0.031	0.185	0.244	0.082
580 nm	1.0	0.982	0.965	0.995	0.031	0.19	0.269	0.129
585 nm	1.0	0.982	0.976	0.996	0.033	0.197	0.236	0.099
590 nm	1.0	0.983	0.976	0.997	0.035	0.203	0.244	0.13
595 nm	1.0	0.983	0.975	0.997	0.038	0.211	0.262	0.103
600 nm	1.0	0.984	0.976	0.997	0.041	0.215	0.27	0.109

Table A.10: Optimal results of AC of $L_t(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with R^2 and NRMSD for wavelength bands $605\text{-}800\,\mathrm{nm}$.

	R^2				NRMSD			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
605 nm	1.0	0.984	0.977	0.995	0.041	0.218	0.263	0.153
610 nm	1.0	0.985	0.979	0.997	0.041	0.219	0.261	0.109
615 nm	1.0	0.985	0.974	0.995	0.04	0.219	0.294	0.207
620 nm	1.0	0.985	0.976	0.997	0.041	0.219	0.283	0.119
625 nm	1.0	0.986	0.976	0.997	0.039	0.22	0.284	0.114
630 nm	1.0	0.986	0.976	0.998	0.04	0.22	0.287	0.097
635 nm	1.0	0.986	0.98	0.997	0.041	0.22	0.264	0.132
640 nm	1.0	0.986	0.98	0.997	0.042	0.22	0.263	0.112
645 nm	1.0	0.986	0.98	0.997	0.041	0.22	0.267	0.099
650 nm	1.0	0.987	0.981	0.997	0.042	0.22	0.269	0.11
655 nm	1.0	0.987	0.982	0.997	0.043	0.221	0.261	0.14
660 nm	1.0	0.987	0.982	0.997	0.043	0.222	0.268	0.116
665 nm	1.0	0.987	0.976	0.996	0.043	0.223	0.309	0.185
670 nm	1.0	0.988	0.982	0.997	0.045	0.225	0.277	0.107
675 nm	1.0	0.988	0.981	0.998	0.045	0.225	0.278	0.094
680 nm	1.0	0.988	0.975	0.997	0.045	0.224	0.327	0.116
685 nm	1.0	0.988	0.976	0.998	0.043	0.221	0.319	0.101
690 nm	1.0	0.988	0.969	0.997	0.043	0.219	0.363	0.099
695 nm	1.0	0.989	0.979	0.998	0.045	0.218	0.308	0.092
700 nm	1.0	0.989	0.984	0.998	0.047	0.218	0.264	0.096
705 nm	1.0	0.989	0.971	0.998	0.048	0.218	0.359	0.097
710 nm	1.0	0.989	0.986	0.993	0.049	0.218	0.252	0.258
715 nm	1.0	0.989	0.985	0.996	0.049	0.219	0.256	0.155
720 nm	0.999	0.989	0.979	0.996	0.052	0.221	0.319	0.119
725 nm	0.999	0.989	0.983	0.993	0.055	0.223	0.301	0.163
730 nm	0.999	0.988	0.983	0.994	0.057	0.227	0.296	0.151
735 nm	0.999	0.988	0.979	0.993	0.061	0.231	0.32	0.193
740 nm	0.999	0.987	0.975	0.99	0.062	0.233	0.334	0.245
745 nm	0.999	0.987	0.98	0.983	0.064	0.234	0.307	0.339
750 nm	0.999	0.987	0.977	0.986	0.064	0.235	0.32	0.306
755 nm	0.999	0.987	0.982	0.991	0.065	0.235	0.298	0.183
760 nm	0.999	0.987	0.98	0.992	0.064	0.235	0.3	0.172
765 nm	0.999	0.987	0.98	0.993	0.064	0.235	0.297	0.189
770 nm	0.999	0.987	0.978	0.994	0.063	0.234	0.313	0.15
775 nm	0.999	0.987	0.981	0.993	0.063	0.233	0.301	0.156
780 nm	0.999	0.988	0.982	0.99	0.063	0.232	0.3	0.193
785 nm	0.999	0.988	0.977	0.991	0.061	0.231	0.316	0.224
790 nm	0.999	0.988	0.982	0.993	0.061	0.23	0.291	0.186
795 nm	0.999	0.988	0.979	0.993	0.061	0.23	0.305	0.199
800 nm	0.999	0.988	0.983	0.993	0.061	0.229	0.294	0.157

Table A.11: Optimal results of AC of $L_t(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands $400\text{-}600\,\mathrm{nm}$.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
400 nm	8.49	46.78	44.86	49.91	−1.64	16.98	13.62	6.53
405 nm	8.07	43.77	43.55	85.88	−1.5	15.52	15.94	38.84
410 nm	7.4	40.84	37.61	58.06	−1.25	14.01	16.7	17.81
415 nm	7.08	38.3	35.05	49.58	−1.37	12.65	12.83	−7.06
420 nm	6.46	36.1	30.82	63.95	−0.98	11.44	12.51	22.86
425 nm	6.27	34.22	31.67	40.21	−1.14	10.42	13.88	20.53
430 nm	5.79	32.26	27.99	91.18	−0.95	9.26	9.58	45.27
435 nm	5.4	30.6	26.12	57.01	−0.67	8.29	10.06	−25.67
440 nm	5.15	29.06	37.15	19.65	−0.62	7.37	4.47	1.59
445 nm	5.07	27.89	22.44	64.35	−0.93	6.77	7.49	41.45
450 nm	4.94	26.9	26.89	40.26	−0.92	6.22	12.27	−21.1
455 nm	4.88	26.08	21.4	57.41	−0.92	5.78	10.74	39.32
460 nm	4.65	25.32	23.76	34.53	−0.63	5.32	7.61	−17.16
465 nm	4.54	24.6	26.1	46.85	−0.86	4.93	7.18	−29.38
470 nm	4.31	23.89	17.24	34.26	−0.57	4.54	3.47	23.35
475 nm	4.41	23.27	17.76	16.43	−1.04	4.3	6.08	4.43
480 nm	4.29	22.64	16.49	22.18	−0.91	4.03	1.49	16.84
485 nm	4.19	22.05	21.02	39.04	−0.97	3.78	7.62	−27.33
490 nm	4.11	21.43	15.78	27.35	−0.67	3.48	6.62	21.02
495 nm	4.16	20.91	16.7	55.26	−1.11	3.23	4.54	43.93
500 nm	4.18	20.48	22.16	12.6	−1.41	2.98	6.25	−7.38
505 nm	4.1	20.26	17.53	22.87	−1.42	2.68	3.71	14.19
510 nm	3.96	20.26	13.64	12.15	−1.31	2.39	−1.1	5.48
515 nm	3.92	20.43	12.47	9.64	−1.3	2.18	1.88	−0.56
520 nm	4.02	20.34	20.27	27.04	−1.53	2.15	6.51	−23.07
525 nm	4.04	20.35	14.18	12.86	−1.52	2.14	0.77	−1.76
530 nm	4.02	20.44	28.99	35.85	−1.65	2.12	−1.17	31.26
535 nm	4.03	20.65	13.39	10.17	−1.65	2.09	−1.73	0.52
540 nm	3.92	20.99	13.49	17.59	−1.53	2.06	1.79	11.14
545 nm	4.11	21.51	15.63	38.87	−1.81	2.04	0.0	33.21
550 nm	4.13	22.26	14.86	59.19	−1.9	2.01	7.99	53.77
555 nm	3.97	22.84	17.92	18.89	−1.38	2.09	−0.96	−14.68
560 nm	4.36	23.47	14.26	22.32	−2.09	2.23	1.92	17.51
565 nm	4.67	24.18	16.86	33.95	−2.52	2.35	3.5	27.39
570 nm	4.53	25.23	18.18	24.66	−1.88	2.38	7.25	18.13
575 nm	5.13	26.6	14.73	22.38	−2.78	2.41	4.69	15.06
580 nm	5.17	28.64	17.93	42.31	−2.44	2.31	3.21	35.38
585 nm	5.24	31.1	29.18	25.84	−2.15	2.22	2.03	−6.52
590 nm	5.96	33.57	24.66	48.41	−3.37	2.17	0.06	42.31
595 nm	6.28	36.37	41.77	34.81	−3.45	2.13	4.11	22.61
600 nm	6.68	38.33	27.15	41.42	−3.89	2.13	0.21	34.6

Table A.12: Optimal results of AC of $L_t(\lambda)$ to $R_{rs}(\lambda)$ with NN, PLSR, SGDR and SVR validated with APD and Bias for wavelength bands 605-800 nm.

	APD [%]				Bias [%]			
Band	NN	PLSR	SGDR	SVR	NN	PLSR	SGDR	SVR
605 nm	6.92	39.6	28.96	65.26	−4.09	2.12	0.37	52.98
610 nm	7.05	40.3	35.44	41.07	−4.2	2.1	−3.93	26.64
615 nm	7.16	40.76	24.95	99.77	−4.24	2.09	3.72	−92.71
620 nm	6.93	41.29	23.44	46.16	−3.51	2.1	−3.84	−37.36
625 nm	6.84	41.84	24.48	45.54	−3.6	2.12	−6.16	−35.68
630 nm	7.0	42.37	25.89	35.7	−3.79	2.15	7.95	23.02
635 nm	7.29	42.87	33.74	59.33	−4.24	2.16	−1.48	50.71
640 nm	7.24	43.35	33.81	43.79	−4.14	2.19	1.81	24.84
645 nm	7.27	43.9	30.22	33.69	−3.96	2.26	1.54	−6.8
650 nm	7.93	44.36	39.87	47.3	−4.77	2.29	11.92	35.13
655 nm	7.34	44.78	54.0	69.05	−3.94	2.25	−4.62	59.81
660 nm	8.1	45.08	49.65	44.46	−4.83	2.2	4.22	−30.72
665 nm	7.93	45.45	38.39	94.93	−4.17	2.22	4.17	85.35
670 nm	8.01	45.93	35.39	42.06	−4.3	2.3	9.37	25.81
675 nm	8.0	46.31	33.07	31.05	−3.79	2.34	−4.76	−5.75
680 nm	8.44	46.53	50.46	48.97	−4.49	2.28	11.24	27.76
685 nm	8.98	46.7	42.57	39.09	−5.55	2.2	−9.63	−21.2
690 nm	8.83	47.05	69.95	38.81	−5.35	2.15	16.27	0.19
695 nm	8.96	47.59	40.68	36.41	−5.54	2.2	13.85	6.18
700 nm	8.69	48.18	31.89	36.11	−4.86	2.26	1.0	−12.44
705 nm	8.57	48.6	68.09	41.05	−4.92	2.3	−0.84	19.74
710 nm	9.1	49.25	55.95	148.56	−5.38	2.34	−6.05	135.66
715 nm	9.45	49.91	44.74	81.91	−5.72	2.46	−8.44	71.21
720 nm	11.28	50.74	42.3	50.38	−8.18	2.68	12.47	25.04
725 nm	10.77	51.52	33.34	76.87	−7.36	2.97	−4.14	30.49
730 nm	11.34	52.31	36.55	65.53	−7.84	3.4	15.53	2.52
735 nm	11.3	52.81	40.94	77.71	−7.83	3.69	10.51	−56.42
740 nm	11.47	53.36	46.96	115.39	−7.96	4.16	11.49	−94.74
745 nm	11.32	53.36	32.97	178.74	−8.0	4.03	7.08	−144.56
750 nm	11.86	53.6	36.12	150.31	−8.33	4.13	7.78	−128.04
755 nm	11.51	53.81	44.04	75.31	−7.72	4.23	6.59	−5.41
760 nm	11.75	54.02	33.02	71.33	−8.34	4.27	−5.7	14.44
765 nm	11.82	54.23	32.6	75.54	−8.33	4.31	−4.1	−53.7
770 nm	12.11	54.48	37.78	54.89	−8.47	4.38	4.28	−7.34
775 nm	12.11	54.57	33.23	67.87	−8.6	4.29	2.58	17.18
780 nm	11.16	54.74	40.95	81.3	−7.46	4.25	21.2	−9.18
785 nm	11.29	54.89	89.42	118.19	−7.55	4.24	−3.76	95.82
790 nm	11.06	55.05	52.66	71.06	−7.35	4.21	−1.47	−48.65
795 nm	11.41	55.21	33.86	91.72	−7.74	4.21	−6.0	−70.68
800 nm	11.96	55.34	45.15	67.66	−8.08	4.21	7.1	−3.33

Appendix B

U.S. standard atmosphere 1976

Table B.1: AFGL atmospheric constituent profiles, U.S. standard atmosphere 1976 (AFGL-TR-86-0110), part 1 [1].

z(km)	p (mb)	T (K)	air (cm^{-3})	O_3 (cm^{-3})	O_2 (cm^{-3})	$\mathrm{H}_2\mathrm{O}$ (cm^{-3})	CO_2 (cm^{-3})	NO_2 (cm^{-3})
120.0	3e-05	360.0	5.110260E+11	2.557000E+02	3.707650E+10	1.022800E+05	1.789900E+07	7.7221E+01
115.0	4e-05	300.0	9.681328E+11	4.844000E+03	9.106720E+10	2.325120E+05	3.875200E+07	1.5016E+02
110.0	7e-05	240.0	2.142688E+12	1.072000E+05	2.572800E+11	6.003200E+05	1.286400E+08	3.4090E+02
105.0	0.00014	208.8	5.029784E+12	1.006600E+06	7.046200E+11	1.711220E+06	5.536300E+08	8.2541E+02
100.0	0.00032	195.1	1.187967E+13	4.756001E+06	1.902400E+12	4.756001E+06	2.318550E+09	2.0213E+03
95.0	0.00076	188.4	2.921760E+13	2.046800E+07	5.263200E+12	1.578960E+07	7.894800E+09	5.1462E+03
90.0	0.00184	186.9	7.130506E+13	4.995200E+07	1.355840E+13	6.065600E+07	2.212160E+10	1.3059E+04
85.0	0.00446	188.9	1.710073E+14	8.555000E+07	3.422000E+13	2.275630E+08	5.475200E+10	3.2851E+04
80.0	0.0105	198.6	3.829322E+14	1.149600E+08	8.008879E+13	7.855600E+08	1.256896E+11	7.7406E+04
75.0	0.024	208.4	8.341139E+14	2.086750E+08	1.744523E+14	2.358028E+09	2.754510E+11	1.7946E+05
70.0	0.0522	219.6	1.721670E+15	5.169000E+08	3.601070E+14	6.030500E+09	5.685899E+11	3.9801E+05
65.0	0.109	233.3	3.383947E+15	2.370200E+09	7.076740E+14	1.422120E+10	1.117380E+12	8.5666E+05
60.0	0.219	247.0	6.421832E+15	7.068600E+09	1.343034E+15	3.052350E+10	2.120580E+12	1.8314E+06
55.0	0.425	260.8	1.180302E+16	2.125800E+10	2.468290E+15	6.023100E+10	3.897300E+12	4.0036E+06
50.0	0.7978	270.7	2.134605E+16	6.621600E+10	4.464240E+15	1.116060E+11	7.048799E+12	9.4625E+06
47.5	1.09	270.6	2.917498E+16	1.197200E+11	6.102800E+15	1.533000E+11	9.636000E+12	1.9447E+07
45.0	1.491	264.2	4.087489E+16	2.147250E+11	8.548099E+15	2.137025E+11	1.349700E+13	4.7035E+07
42.5	2.06	257.3	5.798815E+16	3.597860E+11	1.212827E+16	2.988545E+11	1.914990E+13	1.2593E+08
40.0	2.871	250.4	8.304447E+16	6.066300E+11	1.736790E+16	4.175775E+11	2.742300E+13	3.3489E+08
37.5	4.15	242.9	1.237464E+17	9.656401E+11	2.587420E+16	6.128100E+11	4.085400E+13	7.7499E+08
35.0	5.746	236.5	1.759731E+17	1.380096E+12	3.680490E+16	8.628900E+11	5.811300E+13	1.2820E+09
32.5	8.01	230.0	2.522415E+17	1.860945E+12	5.275160E+16	1.217830E+12	8.329200E+13	1.8198E+09
30.0	11.97	226.5	3.827699E+17	2.509799E+12	8.004700E+16	1.809675E+12	1.263900E+14	2.3593E+09
27.5	17.43	224.0	5.635873E+17	3.272892E+12	1.178760E+17	2.580300E+12	1.861200E+14	2.7128E+09
25.0	25.49	221.6	8.331283E+17	4.266877E+12	1.742433E+17	3.689123E+12	2.751210E+14	3.1180E+09

Table B.2: AFGL atmospheric constituent profiles, U.S. standard atmosphere 1976 (AFGL-TR-86-0110), part 2 [1].

z(km)	p (mb)	T (K)	air (cm^{-3})	O_3 (cm^{-3})	O_2 (cm^{-3})	H_2O (cm^{-3})	CO_2 (cm^{-3})	NO_2 (cm^{-3})
24.0	29.72	220.6	9.757872E+17	4.518265E+12	2.040885E+17	4.198950E+12	3.222450E+14	2.9881E+09
23.0	34.67	219.6	1.143492E+18	4.768192E+12	2.390960E+17	4.804800E+12	3.775200E+14	2.9515E+09
22.0	40.47	218.6	1.340895E+18	4.894274E+12	2.804780E+17	5.455230E+12	4.428600E+14	2.8987E+09
21.0	47.29	217.6	1.574064E+18	4.769100E+12	3.291750E+17	6.260625E+12	5.197500E+14	2.7720E+09
20.0	55.29	216.7	1.847990E+18	4.768571E+12	3.864410E+17	7.211100E+12	6.101700E+14	2.5701E+09
19.0	64.67	216.7	2.161503E+18	4.390890E+12	4.520670E+17	8.327550E+12	7.137900E+14	2.2928E+09
18.0	75.65	216.7	2.528494E+18	4.015110E+12	5.287700E+17	9.677249E+12	8.349000E+14	1.9506E+09
17.0	88.5	216.7	2.957987E+18	3.513520E+12	6.186399E+17	1.139600E+13	9.768000E+14	1.5362E+09
16.0	103.5	216.7	3.459340E+18	3.012286E+12	7.235580E+17	1.367490E+13	1.142460E+15	1.1044E+09
15.0	121.1	216.7	4.047595E+18	2.634525E+12	8.464500E+17	2.025000E+13	1.336500E+15	6.9255E+08
14.0	141.7	216.7	4.736121E+18	2.383717E+12	9.904510E+17	2.808805E+13	1.563870E+15	3.5448E+08
13.0	165.8	216.7	5.541629E+18	2.132992E+12	1.159114E+18	6.017410E+13	1.830180E+15	2.4680E+08
12.0	194.0	216.7	6.484174E+18	2.008345E+12	1.356201E+18	1.236803E+14	2.141370E+15	2.0440E+08
11.0	227.0	216.8	7.583652E+18	1.630876E+12	1.586101E+18	2.741906E+14	2.504370E+15	1.9883E+08
10.0	265.0	223.3	8.595457E+18	1.129443E+12	1.797818E+18	6.017959E+14	2.838660E+15	2.0473E+08
9.0	308.0	229.7	9.711841E+18	8.910379E+11	2.031271E+18	1.538518E+15	3.207270E+15	2.2548E+08
8.0	356.5	236.2	1.093179E+19	6.526804E+11	2.286460E+18	4.011698E+15	3.610200E+15	2.5162E+08
7.0	411.1	242.7	1.226845E+19	6.151052E+11	2.566520E+18	7.024160E+15	4.052400E+15	2.8244E+08
6.0	472.2	249.2	1.372429E+19	5.645776E+11	2.869570E+18	1.270574E+16	4.530900E+15	3.1579E+08
5.0	540.5	255.7	1.531006E+19	5.772576E+11	3.201880E+18	2.140204E+16	5.055600E+15	3.5236E+08
4.0	616.6	262.2	1.703267E+19	5.771448E+11	3.561360E+18	3.677232E+16	5.623200E+15	3.9192E+08
3.0	701.2	268.7	1.890105E+19	6.274337E+11	3.952190E+18	6.017162E+16	6.240300E+15	4.3493E+08
2.0	795.0	275.2	2.092331E+19	6.778279E+11	4.376460E+18	9.697315E+16	6.910200E+15	4.8162E+08
1.0	898.8	281.7	2.310936E+19	6.779402E+11	4.834170E+18	1.404222E+17	7.632900E+15	5.3199E+08
0.0	1013.0	288.2	2.545818E+19	6.777680E+11	5.325320E+18	1.973426E+17	8.408400E+15	5.8604E+08

Ole Martin Borge

NTNU
Norwegian University of
Science and Technology