HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation
Department of Radio Science and Engineering

Jaime Arroyo Pedrero

Evaluation of MERIS Case-II Water Processors in the Baltic Sea

Final Project

Supervisor: Dr. Sampsa KOPONEN
Espoo, February 25, 2009

Abstract

Author:	Jaime Arroyo Pedrero
Name of the Thesis:	Evaluation of MERIS Case-II Water Processors in the Baltic Sea
Date:	February 25, 2009
Number of Pages:	120
Faculty:	Faculty of Electronics, Communications and Automation
Department:	Department of Radio Science and Engineering
Professorship:	Space Technology
Supervisor:	Dr. Sampsa Koponen

Abstract: Four MERIS Case-II Water Processors are studied, compared and evaluated: Coastal Case 2 Regional Processor, Boreal Lakes Processor, Eutrophic Lakes Processor and FUB/Wew Water Processor. *In situ* data from the Baltic Sea have been used to evaluate the water constituent estimations. In addition, the effect of adjacency effect ICOL on the estimation has been analyzed.

For this purpose, a set of tools has been developed to automatise the evaluation process and make the comparisons with large amounts of MERIS products and *in situ* data.

Results show that the processors are far from being accurate. Both absolute errors and estimation concentrations are in the same order of magnitude. Coastal Case 2 Regional provides the best performance in the studied area of the Baltic Sea.

Keywords: MERIS, Case 2, Remote Sensing, Water Quality, Baltic Sea.

Preface

This Thesis has been carried out in the Department of Radio Science and Engineering at the Helsinki University of Technology (TKK) during 2008-2009 under the supervision of Dr. Sampsa Koponen.

First of all, I would like to thank Sampsa Koponen for his patient assistance and constant supervision during all my thesis. He worked hard to help me specially correcting this report in the last months. I owe gratitude to William Martin as well, for all the scientific writing support.

I acknoledge the Finnish Environmental Institute (SYKE) for providing the data that made this work possible and Seppo Kaitala for providing AERONET data from the Helsinki Lighthouse.

For financial support I would like to thank AGAUR and the *Univeristat Politècnica de Catalunya* (UPC), they gave me the grant and the opportunity to go to Finland under the Erasmus framework. I would like to mention also the administrative staff from TKK who helped me with all the paperwork.

I am also grateful to the colleagues who have shared time and the same working place, and helped me during the development of this thesis, Marc and Mònica.

Last but not least, I would like to give my warmest thanks to my parents, Casto and Conchita, and my sister Adelia, for supporting me every day, even from a great distance. The same for my friend Carolina who is always there to make me feel happy.

Huge thanks go to all of you and to the ones I have not mentioned by name in this preface.

February 25, 2009
Jaime Arroyo Pedrero

List of Acronyms

5S	Simulation of the Satellite Signal in the Solar Spectrum
aCDOM	Absorption of Colored Dissolved Organic Matter
AOP	Aparent Optical Properties
AOT	Aerosol Optical Thickness
API	Application Programming Interface
AERONET	Aerosol Robotic Network
ASAR	Advanced Synthetic Aperture Radar
BEAM	Basic ENVISAT Toolbox for (A)ATSR and MERIS
C2R	Case 2 Regional
CCD	Charge-Coupled Device
chl a	Chlorophyll a
DU	Dobson Units
EEC	European Economic Community
ENVISAT	Environmental Satellite
ERS	European Remote Sensing Satellites
ESA	European Space Agency
EU	European Union
FOV	Field of View
FR	Full Resolution
FRS	Full Resolution Swath
FUB/WEW	Freie Universitaet Berlin/Institut f ¨ur Weltraumwissenschaften
GKSS	Gesellschaft f ¨ur Kernenergieverwertung in Schiffbau und Schiffahrt mbH
GMT	Greenwich Mean Time
ICOL	Improved Contrast between Ocean and Land
IOP	Inherent Optical Properties
IPF	Instrument Processing Facility.
MAVT	MERIS-ATSR validation and calibration team
MEGS	Meris Ground Segment data processing prototype

MERIS	Medium-Spectral Resolution Imaging Spectrometer
METOP	Meteorological Mission
MOMO	Matrix Operator Method Code
MSL	Mean Sea Level
NIR	Near Infrared
NN	Neural Network
LMS	Least Minimum Square
PCA	Principal Component Analysis
pdf	Probability Density Function
POEM	Polar Orbiting Earth Observation Mission
psf	Point-Spread Function
REVAMP	Regional Validation of MERIS Chlorophyll products
RMSE	Root Mean Square Error
RR	Reduced Resolution
RTE	Radiative Transfer Equation
SeaWiFS	Sea-viewing Wide Field Sensor
SYKE	Finnish Environment Institute
TOA	Top Of Atmosphere
TOSA	Top Of Standard Atmosphere
TSM	Total Suspended Matter
TSS	Total Suspended Solids
YEL	Yellow Substance

List of Symbols

α, β, γ	Empirically determined regression parameters
a	Absorption coefficient
b	Scattering coefficient
b_b	Back-scattering coefficient
b_u	Upward-scattering coefficient
$\beta(\lambda)$	Volume scattering function
C	Concentration
d	Desired output in a NN
δ	Error signal in a NN
$E_0(\lambda)$	Extraterrestrial Solar Irradiance
$E_{0d}(\lambda, z)$	Scalar Irradiance
$E_d(\lambda, z)$	Downwelling Irradiance
$E_u(\lambda, z)$	Upwelling Irradiance
f	point-spread function
F	Primitive of the point-spread function
F_k	Activation Function
j_v	Emission coefficient
$L(\lambda, z, \theta, \phi)$	Radiance
K_d, K_u	Diffuse attenuation coefficient for downwelling/upwelling irradiance
N	Number of Samples
p	Probability density function
$P_a(\alpha, \lambda)$	Atmosphere-Scattering Phase Function
$r(\alpha)$	Fresnel Reflectance
R^2	Coefficient of Determination
$R(\lambda, z), \rho$	Reflectance
R_{RS}, RL	Radiance Reflectance (Remote Sensing Reflectance)
s	Shape Factor
s_k	Total Input to a NN
$\Im$	Source Function
T	Direct Transmittance
t	Diffuse Transmittance

τ	Optical Thickness
v	frequency
λ	Wavelength
μ	Cosine of Zenith Angle
ϕ	Azimuth angle
θ	Zenith angle
θ_k	Offset term in a NN
σ	Standard deviation

Contents

1 Introduction

In the recent decades, concerns regarding the state of the environment have increased in the society. Human made pollution not only affects the ecosystem, but the quality of living of people. Clean water is maybe the most important ingredient for human life. Indeed, rivers, seas and oceans are probably the largest ecosystems containing priceless diversity of flora and fauna.

Traditional monitoring programs provide essential and accurate results on water monitoring but they are expensive and inefficient (See Section 1.1). That's why satellite remote sensing of water quality (See Section 2.1.1) can provide an alternative solution.

The Medium Resolution Imaging Spectrometer (MERIS) is an instrument aboard the *Environmental Satellite* (Envisat), its objective is, among others, remote sensing of water quality.

Algorithms developed for that purpose process MERIS images and are reliable far from land and on open oceans (These type of water are called *Case 1 Waters*). However, in coastal and lake waters (*Case 2 Waters*) where high concentrations of reflective particles can be present in the water surface, traditional algorithms fail. Alternative and more sophisticated algorithms have been developed to deal with these complications [26].

In the present thesis, four of these processors will be described, analyzed and compared in a Case 2 Waters scenario:

1. Coastal Case 2 Regional Water Processor (C2R)[41, 42].
2. Boreal Lakes Water Processor [42, 43].
3. Eutrophic Lakes Water Processor[42, 43].
4. FUB/Wew Water Processor [51].

The scope of the present work is the Baltic Sea (See Section 1.2). The primary objective of this study is to choose the processor which is the most suitable for this area. In addition, several Matlab tools have been developed to automate the comparison process of several processors with a large amount of MERIS products and *in situ* measurements, in the framework of BEAM (*Basic ENVISAT Toolbox for (A)ATSR and MERIS*)[8].

1.1 Water quality monitoring

In traditional water quality monitoring, water samples are regularly taken from the same sampling sites and analyzed for various water quality pa-

rameters such as physical and chemical parameters or biological parameters. Standarised methods are used so that the results are comparable.

The first milestone of water monitoring in Finland was the Water Act (1962), implementing a local pollution control monitoring network based on the "polluter pays" principle. Other monitoring programs followed such as those for harmful substances, ground water, or wastes.

In this way, the role of water quality monitoring in addressing environmental questions is increasing. This is reflected, for example, in the policy of the European Union, where numerous directives relating to the environment have been put into practice. Early European water legislation began with standards for those of the rivers and lakes used for drinking water extraction in 1975, and culminated in 1980 in setting binding quality targets for drinking waters. It also included quality objective legislation on fish waters, shellfish waters, bathing waters and groundwaters. Its main emission control element was the 76/464/EEC (Dangerous Substances Directive).

A more recent example was the 2000/60/EC (EU Water Framework Directive) which required changes in the way rivers, lakes and coastal water were monitored. The emphasis was on assessing the biological status of water instead of using purely physical and chemical parameters.

Lately, the European Parliament approved on June 17th, 2008 the directive 2008/56/EC (Marine Strategy Framework Directive) establishing a framework for community action in the field of marine 1al policy. Thus, Member States shall take necessary measures to achieve or maintain good environmental status in the marine environment by the year 2020 at the latest. The following qualitative descriptors for determining a good environmental status are defined in its ANNEX I:

1. Biological diversity is maintained. The quality and occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climatic conditions.

2. Non-indigenous species introduced by human activities are at levels that do not adversely alter the ecosystems.

3. Populations of all commercially exploited fish and shellfish are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock.

4. All elements of the marine food webs, to the extent that they are known, occur at normal abundance and diversity and levels capable of ensuring the long-term abundance of the species and the retention of their full reproductive capacity.

5. Human-induced eutrophication is minimized, especially adverse effects thereof, such as losses in biodiversity, ecosystem degradation, harmful algae blooms and oxygen deficiency in bottom waters.

6. Sea-floor integrity is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected.

7. Permanent alteration of hydrographical conditions does not adversely affect marine ecosystems.

8. Concentrations of contaminants are at levels not giving rise to pollution effects.

9. Contaminants in fish and other seafood for human consumption do not exceed levels established by Community legislation or other relevant standards.

10. Properties and quantities of marine litter do not cause harm to the coastal and marine environment.

11. Introduction of energy, including underwater noise, is at levels that do not adversely affect the marine environment.

In addition, after the determination of the environmental status and the establishment of environmental targets, Member States shall implement coordinated monitoring programmes for the ongoing assessment of the environmental status of their marine waters. Biological characteristics such as phytoplankton, biomass or fish population among others, and their dynamics and seasonal variability have to be monitored.

The following marine regions or subregions, which are are covered by the sovereignty or jurisdiction of the EU, are in the focus of the mentioned directive:

1. The Baltic Sea.

2. The North-east Atlantic Ocean.

   (a) The Greater North Sea, included the Kattegat, and the English Channel.

   (b) The Celtic Seas.

   (c) The Bay of Biscay and the Iberian Coast.

(d) In the Atlantic Ocean, the Macaronesian biogeographic region, being the waters surrounding the Azores, Madeira and the Canary Islands.

3. The Mediterranean Sea.
   (a) the Western Mediterranean Sea.
   (b) the Adriatic Sea.
   (c) The Ionian Sea and the Central Mediterranean Sea.
   (d) the Aegean-Levantine Sea.
4. The Black Sea.

However, all these requirements makes the traditional method, which consist of collecting water samples and analyzing the samples in a laboratory and making on-site measurements, a very slow and expensive process. Hence, remote sensing appears as an obvious alternative solution since one of the advantages of remote sensing is that the measurements can be performed from a great distance, which means that large areas on ground can be covered easily. In addition, the frequency of the measurements can be much higher (even several times per day) allowing the analysis of constituents dynamics and variability. For further information on remote sensing of water quality see Section 2.1.1.

1.2 Description of study area (Baltic Sea)

The study area of this work is the Baltic Sea. This election is justified by the fact that it is in the scope of the European directives (see Section 1.1) mentioned above. Additionally, its special characteristics, that are explained in this section, makes possible to compare different processors such as those meant to be applied to fresh waters (i.e. Boreal and Eutrophic lakes) and those meant to be applied to salty waters (i.e. Case 2 Regional Coastal Waters and FUB/Wew water processor).

The Baltic Sea (See Figure 1) is a brackish inland sea, alleged to be the largest body of brackish water in the world [14], located from 53°N to 66°N latitude and from 20°E to 26°E longitude.

The low salinity of the Baltic Sea has led to the evolution of many slightly divergent species, such as the Baltic Sea Herring [56]. The benthic fauna consists mainly of *Monoperia affinis*, which is originally a freshwater species. The most common fish species are goldfish, herring, hake, plaice, flounder, sea trout, eel and turbot [56].

Approximately a quarter of its total area is a variable dead zone. The more saline water remains on the bottom, isolating it from surface waters and the atmosphere. This leads to decreased oxygen concentrations within the zone. It is mainly bacteria that grow in it, digesting organic material and releasing hydrogen sulfide. Because of this large anaerobic zone, the sea floor ecology differs from that of the neighbouring Atlantic.

Baltic Sea occupies a basin formed by glacial erosion. It is therefore quite shallow and due to its small volume susceptible to pollution. Cyanobacterial blooms have become common in some areas during late July and August.

All the *in situ* and MERIS data used in the present work are collected from two subbasins of this sea. Those are:

Gulf of Finland

It is the easternmost arm of the Baltic Sea that extends between Finland and Estonia all the way to Saint Petersburg in Russia, where the river Neva drains into it. As mentioned before, severe eutrophication is taking place in this part of the Baltic Sea. Algal blooms, which occur during summers, can cover large areas.

Figure 1: Map of the Baltic Sea [34].

Gulf of Bothnia

It is the northernmost arm of the Baltic Sea. It is situated between Finland's west coast and Sweden's east coast. A number of rivers flow into the gulf from both countries. Therefore, a salinity gradient exists from north to south. Actually, the north part is frozen over five months every year and the salinity is so low that salt cannot be tasted in the water.

The Baltic Sea is an endangered around it water body due to the intensive industrial activity around it. It is surrounded by former soviet countries such as Russia or Estonia, where collapsing infrastructures and institutions caused systematically neglection of any kind of control on pollution. In addition, Saint Petersburg port is one of the major trade gateway of Russia specializing in oil and gas trade. Indeed, Gazprom, the largest extractor of natural gas in the world, announced in 2005 that the Gazprom City business center is going to be built on the right bank of the Neva River in Saint Petersburg.

2 Theoretical Background

2.1 Remote Sensing

Remote sensing consists of a set of techniques to gather data and information about the physical world by detecting and measuring radiation associated to objects located beyond the immediate vicinity of the sensors.

Remote sensing measurements are usually based on interactions between electromagnetic radiation and matter. Three different basic strategies can be adopted:

1. The instrument records the reflection of solar radiation from target's surface. This kind of remote sensing is very near to everyday experience. For example when we look at an object of interest with our eyes, we see the radiation from the sun after it is reflected from it.

2. The instrument records radiation emitted by the target itself. Since emitted energy is strongest in the far infrared spectrum, this kind of remote sensing requires special instruments designed to record these wavelengths. Emitted energy is mainly derived from energy previously absorbed by the matter that is rerariated at longer wavelengths.

3. The instrument generates its own energy, recording the reflection of the radiated energy. This kind of remote sensing is called *active* remote sensing in the sense that it provides its own energy, so it is independent of solar or terrestrial radiation. Therefore, the other previous strategies are called *passive* remote sensing.

2.1.1 Remote Sensing of water quality

The kind of remote sensing to be discussed here is devoted to observation of the earth's water surfaces by means of reflected electromagnetic energy. Therefore, in this work, all the remote sensed data have been obtained following the first strategy of those mentioned before. *Passive* systems for remote sensing of ocean colour, make use of a sensor with a narrow field of view, capable of monitoring the radiometric flux reaching the sensor at several selected wavelengths in the visible and near-infrared bands of the electromagnetic spectrum. The sensor instruments, mounted on a satellite, are aimed at a point on the surface of the earth. Scanning devices on the sensor, and the movement of the platform itself combine to acquire data from different points on the earth. The sensors operate during the daylight

hours, and the ultimate source of the light reaching the sensor is the sun. However, the photons from the sun can follow five different pathways before they reach the remote detector (See Figure 2):

1. Radiation upwelling from the water surface after back-scattering in water. This radiation is attenuated on its journey from the sea surface to the sensor, due to absorption and scattering by the intervening atmosphere.

2. Radiation reflected from the bottom.

3. Radiation reaching the sensor after specular reflection of direct sunlight at the sea surface.

4. Radiation reaching the sensor after scattering of photons by the atmosphere.

5. Radiation reflected from a target that is close to the FOV of the sensor and scattered to the sensor by the atmosphere. This component is produced by the so called Adjacency Effect (See Section 2.1.3 for information about this effect).

Remote sensing involves analyses of the variations in magnitude and spectral shape of the water-leaving radiation to derive quantitative information about the type of substances present in the water and their concentrations. However, it is only the upwelling light from the sea surface path way (1) that carries any useful information about the water body. In water, the radiation is scattered and absorbed by water molecules and substances that are suspended or dissolved in water. The atmospheric contributions and specular reflection path ways (2 to 5) at the sea surface constitute noise in this context, and have to be corrected for.

The complete description of light is provided by *radiance* (L), which is a measure of flux per unit area and per unit solid angle. Therefore $L(\lambda, z, \theta, \phi)$ is the radiance in the direction defined by zenith angle θ and azimuth angle ϕ in some particular height defined by z and for wavelength λ. By removing the angle dependences it is possible to define *scalar irradiance*, $E_{0d}(\lambda, z)$, which measures flux received by a spherical collector from all directions [26]:

$$E_0(\lambda, z) = \int_0^{2\pi} \int_0^{\pi} L(\lambda, z, \theta, \phi) sin\theta d\theta d\phi \tag{1}$$

Changing the integration limits adequately, this magnitude can be split into downwelling and upwelling components, E_{0d} and E_{0u} respectively.

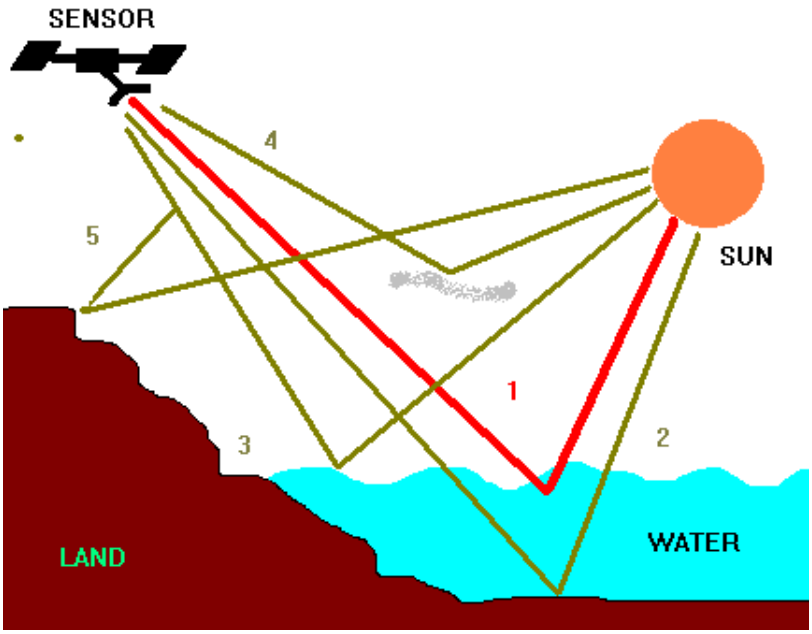


Figure 2: (1) Radiation upwelling from the water surface after backscattering in water. (2) Radiation reflected from the bottom. (3) Radiation reaching the sensor after specular reflection of direct sunlight at the sea surface. (4) Radiation reaching the sensor after scattering of photons by the atmosphere. (5) Radiation reflected by Adjacency Effect.

Similarly, fluxes received by a flat collector with a cosine response facing downwards or upwards, are also called *downwelling irradiance* and *upwelling irradiance*:

$$E_d(\lambda, z) = \int_0^{2\pi} \int_0^{\pi/2} L(\lambda, z, \theta, \phi) cos\theta sin\theta d\theta d\phi \tag{2}$$

$$E_u(\lambda, z) = \int_0^{2\pi} \int_{\pi/2}^{\pi} L(\lambda, z, \theta, \phi) cos\theta sin\theta d\theta d\phi \tag{3}$$

At this point, the intrinsic colour of the water is defined by the spectral variation in *reflectance* at the sea surface, that is the ratio between the upwelling irradiance and the downwelling irradiance. Therefore, at any depth z we have:

$$R(\lambda, z) = \frac{E_u(\lambda, z)}{E_d(\lambda, z)} \tag{4}$$

However, since remote sensing instruments measure *radiance* L, in the context of remote sensing, it is common to deal with *remote sensing reflectance* (R_{RS}) or radiance reflectance, that makes use of upwelling radiance rather than irradiance, and is defined as:

$$R_{RS}(\theta, \phi, \lambda, 0) = \frac{L(\theta, \phi, \lambda, 0)}{E_d(\lambda, 0)} \tag{5}$$

Note that remote sensing reflectance is not an intrinsic property of water since it depends on the angles of the incident light (See Section 2.2 to understand the importance of this fact).

In general, the are two different kind of approaches to the retrieval of water constituents information from sensed radiances at the top of the atmosphere: Empirical approaches and Model-based (Analytical) approaches.

Empirical Approach

This approach establishes a relationship between optical measurements and the concentration of constituents based on experimental datasets. The most common relationship is the so called *colour ratio*:

$$\hat{p} = \alpha(\frac{R_1}{R_2})^{\beta} + \gamma \tag{6}$$

where $\hat{p}$ is the physical quantity to be estimated such as chlorophyll, suspended matter, yellow substance, etc, and R_1 and R_2 are reflectances in different spectral bands. Thus, coefficients α, β and γ are derived from regressions between the radiance ratios and the desired property and are based on experimental data sets.

The advantages of empirically derived algorithms are:

1. They are simple.
2. Easy to derive even from a limited number of measurements.
3. Easy to implement and test.
4. They have a short computing time due to their mathematical simplicity.
5. They yield stable results, always inside the scope of the algorithm. Which means, in the regional location where they were derived.

They have, however, several limitations:

1. These algorithms are particularly sensitive to changes in the composition of water constituents (e.g. regional or seasonal effects) and the atmosphere.

2. The resulting errors may quickly exceed acceptable limits.

3. The lack of mathematical formulation, makes it difficult to analyze the source of the errors.

Model-based Approach

Model-based algorithms use bio-optical models to describe the relationship between water constituents and spectra of water-leaving radiance and reflectance, as well as radiative-transfer models (See Section 2.3.1) to simulate the light propagation through the water and the atmosphere. All the processors studied in this work use this approach. However they have some differences regarding to the forward model used: C2R, as well as, Boreal Lakes Processor and Eutrophic Lakes Processor are based on a Monte Carlo Radiative Transfer Code. On the other hand FUB/Wew Water Processor is based on the Matrix Operator Method (See Section 2.3.1). However, they both use Neural Networks to implement the inverse model. Here there is a list of other possible model-based techniques:

1. Algebraic methods. An ocean-colour model is first implemented using empirical data. In fact it is called *semianalytical*. The result is a set of algebraic equation that can be solved sequentially. As the number of unkown parameters increase, it becomes very difficult to implement.

2. Non-linear optimization techniques. In this method, a forward model is inverted directly by minimizing the differences between the calculated values and the measured radiances. It requires substantial computation time.

3. Principal component approach (PCA). In this method, water constituents are directly computed from TOA radiances. A segmentation of the range space of derived parameters should be needed to keep accuracy in this scheme.

The advantages of these approaches are:

1. They use bio-optical models that allow to better understand the underlying processes.

2. They can be implemented in a more global scale because they are based on a more general theoretical basis.

And their disadvantages:

1. The models are more complex mathematically and computationally than empirical models.

2. Increasing the number of unknown variables may make the mathematical solution of the inverse problem unstable, due to its nonlinear nature.

2.1.2 Atmospheric correction

The goal of the atmospheric correction algorithms applied to remote sensing of water is to retrieve the water-leaving radiance at the sea level from the total radiance recorded at the *top of the atmosphere* (TOA) by a satellite sensor. This radiance is made of photons that have crossed the atmosphere down to the ocean, then have twice crossed the air-sea interface before reaching the sensor after a second atmospheric travel. The spectrum of the water-leaving radiances carries information about the water constituents in which they were scattered. However, as mentioned at the beginning of this section, the total radiance received $L_t(\lambda_i)$ by a sensor in a spectral band centered at a wavelength λ_i will be the sum of several components:

$$L_t(\lambda_i) = L_{path}(\lambda_i) + T(\lambda_i)L_g(\lambda_i) + t(\lambda_i)L_{wc}(\lambda_i) + t(\lambda_i)L_w(\lambda_i) \qquad (7)$$

where $L_{path}(\lambda_i)$ is the radiance generated along the optical path by scattering in the atmosphere and by specular reflection of atmospherically scattered light from the sea surface, $L_g(\lambda_i)$ the contribution arising from specular reflection of direct sunlight from the sea surface, $L_{wc}(\lambda_i)$ the contribution arising from sunlight and skylight reflecting from the sea surface, and, $L_w(\lambda_i)$, the desired water-leaving radiance. In addition, T and t are the direct and diffuse transmittance of the atmosphere, respectively. Converting the Eq. 7 it becomes:

$$R_t(\lambda_i) = R_{path}(\lambda_i) + T(\lambda_i)R_g(\lambda_i) + t(\lambda_i)R_{wc}(\lambda_i) + t(\lambda_i)R_w(\lambda_i) \qquad (8)$$

Thus, measuring $R_t(\lambda)$, an atmospheric correction algorithm must provided accured estimations of $R_{path}(\lambda_i)$, $T(\lambda_i)R_g(\lambda_i)$, $t(\lambda_i)R_{wc}(\lambda_i)$ and $t(\lambda_i)$ in order to retrieve $R_w(\lambda_i)$. The diffuse transmittance of the atmosphere can be computed as [23]:

$$t(\lambda_i) = e^{-\frac{(\tau_r/2+\tau_{oz})}{\cos\theta}} e^{-\frac{(1-w_a(\lambda)F_a(\lambda))\tau_a(\lambda)}{\cos\theta}} \qquad (9)$$

where F_a is the probability that a photon scattered by the aerosol will be scattered through and angle less than 90^o, τ_r and τ_a are the optical thicknesses of atmosphere and aerosols, respectively, and τ_{oz} is the optical thickness of the ozone. In addition, pixels affected by sunlight specular reflection, this is so hight compared with the other components that they usually have to be dropped out (Sun glitter mask) [24]. Away from them, the largest remaining term is $R_{path}(\lambda_i)$, which can be decomposed into three components:

$$R_{path}(\lambda_i) = R_r(\lambda_i) + R_a(\lambda_i) + R_{ra}(\lambda_i) \tag{10}$$

where R_r is the reflectance resulting from multiple scattering by air molecules (Rayleigh scattering) in the absence of aerosols, R_a is the reflectance resulting from multiple scattering by aerosols in the absence of air molecules, and R_{ra} is the interaction term between molecular and aerosol scattering, that is, scattered photons by aerosols that then are scattered by air or the other way round. R_{path} used to be difficult to estimate due to the highly variable concentration and optical properties of aerosols. However, under a single scattering approximation (which is seen to work well for small optical depth and non absorbing aerosols, typically the case over the open ocean) $R_{ra} \approx 0$. According to [22] and [23], R_r can be accurately computed given the surface atmospheric pressure and the surface wind speed with the next expression:

$$R_r(\lambda_i) = \frac{\tau_r(\lambda_i) p_r(\theta_v, \phi_v; \theta_0, \phi_0; \lambda_i)}{4 \cos\theta_v \cos\theta_0} \tag{11}$$

where

$$p_a(\theta_v, \phi_v; \theta_0, \phi_0; \lambda_i) = P_a(\theta_-, \lambda_i) + (r(\theta_v) + r(\theta_0)) P_a(\theta_+, \lambda_i) \tag{12}$$

and

$$\cos\theta_\pm = \pm \cos\theta_0 \cos\theta_v - \sin\theta_0 \sin\theta_v \cos\phi_v - \phi_0 \tag{13}$$

where $P_a(\alpha, \lambda)$ is the atmosphere-scattering phase function for a scattering angle α; $r(\alpha)$ the Fresnel reflectance of the interface for an incident angle α; θ_0 and ϕ_0 are the zenith and azimuth angles of a vector from the point on the water surface to the sun, and likewise, θ_v and ϕ_v are the zenith and azimuth angles of a vector from the water surface to the satellite. Therefore under the *black pixel assumption*, R_a can be measured in two bands near infrared where water is completely absorbent and be extrapolated to the blue-green spectral range considering the spectral extinction of aerosol to be an exponential function characterized by the Angstrom coefficient.

Those hypothesis are no longer valid in Case 2 waters though, where the backscattering of water cannot be neglected in the NIR bands. The

problem can be solved by inverse modelling of the radiative transfer (See Section 2.3.1), where the concentrations of water constituents as well as of aerosols are modified and, thus, determined with the help of an optimization procedure, which is used to minimize the deviation between the measured and the modelled radiance spectra. However, even with a simple model, the inversion method requires an amount of computational time which is not acceptable for the mass processing of satellite scenes. One possibility, to combine a realistic description of the processes in the atmosphere using a detailed radiative transfer model with the required high computational efficiency, is to use neural networks (see Section 2.3.2).

Other problems that are still not solved so far for a fully operational atmospheric correction of ocean colour data are [42]:

1. Atmospheric correction over turbid water, where also the near infrared spectral bands are influenced by scattering of suspended particles.

2. The scattering by thin or subvisible cirrus clouds including aged jet trails.

3. Specularly reflected sun light, which is present even in the nadir radiances.

2.1.3 Adjacency effect

In remote sensed radiance images, an increase of the radiances measured over water and in the vicinity of vegetated coasts can often be observed, especially in the near infrared (NIR) bands. This effect occurs when photons are reflected and scattered towards the field of view of the sensor and where substantial contrast exists between the water surface and its surrounding. This happens because land areas have a higher reflectance than inland waters, especially in the NIR region of the spectrum. Even the darker land areas are usually much brighter than water. Due to this effect it is necessary to choose between two options:

1. Remove the data located near land.

2. Perform a correction to minimize its effect.

A detailed study [46] of adjacency effects was carried out for 5S code [53], in which the total surface contribution in nadir view of a uniform disc

of radius R (see Figure 3), reflectance ρ_t, and a uniform, infinite environment of reflectance ρ_e, is calculated as

$$\rho_t^* = \frac{T(\mu_s)}{1 - \langle\rho\rangle\, s}(\rho_t e^{\frac{-\delta}{\mu_v}} + \langle\rho\rangle\, t_d(\mu_v)) \tag{14}$$

with

$$\mu_s = \cos\theta_s \;\; \mu_v = \cos\theta_v \tag{15}$$

where T is the total transmittance, s the spherical albedo of the atmosphere, δ is the total optical thickness, t_d the diffuse transmittance (the diffuse irradiance at the surface level for a black surface, normalized by the direct solar irradiance at the top of the atmosphere (TOA)), $\langle\rho\rangle$ the average reflectance, θ_s the sun zenith angle and θ_v the viewing zenith angle. The averaged reflectance is generally expressed in Cartesian coordinates as:

$$\langle\rho\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\rho(x,y)dxdy \tag{16}$$

where $f(x,y)$ is usually called *atmospheric point-spread function*(psf). In order to evaluate the psf theoretically, it can be used the Monte Carlo method to solve the transfer equation. Unfortunately, using Monte Carlo codes is time-consuming. However, the simple uniform disc proposed in S5 allows to compute the averaged reflectance as follows:

$$\langle\rho\rangle = \rho_t F(R) + \rho_e(1 - F(R)) \tag{17}$$

where $F(R)$ is simply the primitive of $f(r)$. In this case, this function can be modelled so that the Rayleigh scattering and aerosol scattering are decoupled:

$$F(R) = ae^{-\alpha R} + be^{\beta R} \tag{18}$$

where a,b,α and β are experimentally computed parameters which values are different for molecules or continental aerosols.

However this simple model is not very realistic. The formulation of the problem with an off-nadir view is more complex because there is not symmetry in the azimuth. Additionally, for sea observation it is necessary to consider Fresnel reflection. The explanation of these cases is beyond the scope of the present work. It is recommended to look at [46] for a deeper understanding of the mathematical formulation of the effect.

Based on the theory exposed in this section, in Section 3.2.5 it will be explained how the *Improve Contrast between Ocean and Land* (ICOL) processor performs the correction of the effect.

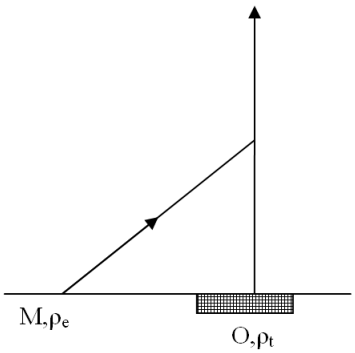


Figure 3: Satellite sensor observes at nadir a pixel O of reflectance ρ_t. By atmospheric scattering, a fraction of the incoming radiation can originate from the neighborhood of this pixel. At a given point M, reflectance ρ_e generally differs from the reflectance of the pixel that is directly observed by the sensor [46].

2.2 Optical Properties of Coastal Waters

Optical properties of water are determined by scattering and absorption of visible light by all the constituents of water:

1. Pure water itself.
2. Organic and inorganic substances.
3. Particulate and dissolved material.

In the case of Coastal waters (Case 2 waters), these substances must be considered separately since they can vary independently. In addition, coastal water colour maybe affected by bottom reflectance when water is shallow and clear.

Inherent and Apparent Optical Properties

Apparent optical properties (AOP) are those optical properties that are influenced by the angular distribution of the light field, as well as by the nature and quantity of substances present in the medium. On the other hand, inherent optical properties (IOP) are solely determined by the type and concentration of the substances. Understanding the relationship between them is therefore a key goal in order to successfully retrieve water constituent concentrations from remote sensed reflectances.

Although the most important apparent optical property of water is its reflectance, some others are often used in the models and therefore, they have to be mentioned.

One apparent property of interest is the *diffuse attenuation coefficient for downwelling irradiance*, K_d, which defines the rate of decrease of downwelling irradiance with depth. It is used in models of light penetration to compute source light as a function of light at some depth.

$$\frac{dE_d(\lambda, z)}{E_d(\lambda, z)} = -K_d(\lambda)dz \tag{19}$$

In this equation, E_d stands for the downwelling irradiance and z is the vertical depth. Solving it requires a known value of the irradiance:

$$E_d(\lambda) = E_{d0}(\lambda)e^{-K_d(\lambda)z} \tag{20}$$

where E_{d0} is the downwelling irradiance at the water surface. In the same way the *diffuse attenuation coefficient for upwelling irradiance* is also of interest in ocean-colour studies. However two types are defined:

1. Rate of exponential decrease in upwelling irradiance with increasing depth, K_u defined as:

$$\frac{dE_u(\lambda, z)}{E_u(\lambda, z)} = -K_u(\lambda)dz \tag{21}$$

2. Rate of decrease in an upwelling irradiance at a certain depth, with decreasing depth, κ, which has gained usage in the last decade. For more information on this parameter see [33].

These apparent optical properties are usually measured using flat-plate collectors facing directly upwards or downwards.

Inherent optical properties are those that determine the manner in which photons propagate through a natural water body. Two processes can be identified:

1. Absorption, which removes photons from the light field.

2. Scattering, which influences the light budget of the medium by changing the direction of photon propagation. Furthermore, this process can be divided in:

   (a) Elastic scattering, where the scattered photon has the same wavelength as the incident photon.

(b) Inelastic scattering, that implies a change in the wavelength. *Raman Scattering* or *Fluorescence emission* are examples of this kind of scattering.

Therefore, the relevant inherent optical properties in this context are:

1. Absorption coefficient, $a(\lambda)$, which defines the exponential rate of decay of flux per unit path length of light in the water, and per unit of incident flux, due to the process of absorption.

2. Scattering coefficient, $b(\lambda)$, which defines the exponential rate of decay of the flux due to scattering. However, since scattering implies a change in direction, a distribution of the scattered flux as a function of the scattering angle must be defined. This distribution is called *volume scattering function*, $\beta(\lambda, \chi)$ and:

$$b(\lambda) = \int \beta(\lambda, \chi) d\chi \tag{22}$$

In addition, some other parameters are defined in a similar way:

1. Back-scattering coefficient, $b_b(\lambda)$, that would be defined as the integral of $\beta(\lambda, \chi)$ over all backward directions ($\chi > 90^o$).

2. Upward-scattering coefficient, b_u which represents an integral over all upward directions. However it refers to scattering in natural conditions of illumantions, and hence is an apparent property.

Finally, it is usually necessary, when linking inherent and apparent optical properties, to have a description of the angular distribution of the light field given the dependency of the apparent properties of the light conditions. The most important are then:

1. Mean cosine for the downwelling irradiance:

$$\mu_d = \frac{E_d}{E_{0d}} \tag{23}$$

where E_d stands for the downwelling irradiance and E_{0d} is the downwelling scalar irradiance.

2. Mean cosine of the upwelling irradiance:

$$\mu_u = \frac{E_u}{E_{0u}} \tag{24}$$

where the subscript u refers to the upwelling direction.

Therefore, these mean cosines are the weighted averages of the cosines of the incident angles of the photon flux, the weighting function being the magnitude of the radiance from each incident direction.

Optical Models

The relationship between the apparent and inherent optical properties introduced before are determined by Optical Models. These models are obtained simulating radiative transfer processes in a realistic fashion. Several techniques have been applied for this purpose:

1. Monte Carlo Method [38].
2. Two-flow method [35].
3. Quasi-single scattering approximation [20].
4. Method of successive orders of scattering [30].

An interesting result is that all these studies led to a robust and consistent expression which estates that, in the absence of inelastic processes, the remote sensing reflectance at the surface can be expressed as a simple function of the back-scattering coefficient $b_b(\lambda)$ and the absorption coefficient $a(\lambda)$ [34]:

$$R_{RS}(\lambda, 0) = f\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \tag{25}$$

Furthermore, reflectance at the sea surface can also be expressed as function of attenuation coefficients for downwelling and upwelling fluxes according to [33]:

$$R_{RS}(\lambda, 0) = \frac{b_u}{\mu_d(K_d + \kappa)} \tag{26}$$

If now we note that $b_u = sb_b$ (being s a shape factor) and that

$$K_d \approx \frac{a + b_b}{\mu_d} \tag{27}$$

and

$$\kappa \approx \frac{a + b_b}{\mu_u} \tag{28}$$

it can be deduced the following value for the f parameter in eq. 25:

$$f = \frac{s\mu_u}{\mu_d + \mu_u} \tag{29}$$

Bottom Effect: In coastal waters however, the possibility that the water may be optically shallow must be considered. In this case, the eq. 26 has to be modified. Thus, water-leaving radiance will have an additional component: light reflected off the bottom. In [28], the following expression is suggested for a water column limited by a Lambertian (uniformly diffuse) reflector at the bottom H with albedo A (which is defined in the same terms as reflectance as the ratio of upwelling irradiance to downwelling irradiance at the bottom of the water column):

$$R(0,H) = \frac{b_u}{K_d + \kappa} + (A - \frac{b_u}{K_d + \kappa})e^{-(K_d+\kappa)H} \tag{30}$$

Optical Properties of Water Constituents

The IOP's used so far represent the combined effects of water and all its constituents. This general parameters, called *bulk optical properties*, can be expressed as the sum of contributions for each of the constituents. In the case of absorption we have the following:

$$a_{tot}(\lambda) = a_w(\lambda) + a_y(\lambda) + C_{ph}a_{ph}(\lambda) + C_s a_s(\lambda) \tag{31}$$

where C_{ph} and C_s are the concentrations of phytoplankton and suspended inorganic matter, respectively, and the subscripts w, y, ph and s stand for water, yellow substance, phytoplankton and suspended matter. Similarly, the back-scattering coefficient can be partitioned into three components:

$$b_{b,tot}(\lambda) = b_w(\lambda) + C_{ph}b_{ph}(\lambda) + C_s b_s(\lambda) \tag{32}$$

where it has been assumed that the backscattering coefficient of phytoplankton do not contribute significantly to the bulk back-scattering coefficient [6]. It is therefore clear, that in order to retrieve concentrations, it is necessary to measure the specific inherent optical properties of all the constituents of water that appear in these previous equations.

Phytoplankton: In a phytoplankton cell, the absorption of solar radiation takes place in photosynthetic pigments, of which the most common is chlorophyll a (chl a). It has two absorption peaks at 430nm and 665 nm (See Table 2) and a quite low scattering coefficients. It also has fluorescence emission with a peak at about 685 nm [26].

Inorganic Suspended Material: An impediment to the interpretation of ocean-colour data in Case 2 waters is the lack of sufficient information on the specific inherent optical properties of terrigenous suspended matter and therefore the Total Suspended Matter (TSM) concentration. Spectra specific absorption and scattering coefficient of inorganic matter are dependent upon particle shape, particle size distribution, and refractive index [26].

Yellow Substances: Colored dissolved organic matter (CDOM) has been found (with excellent agreement) to have an exponential function as spectral form of the absorption [26]:

$$a_y(\lambda) = a_y(\lambda_0)e^{-S(\lambda_0-\lambda)} \tag{33}$$

where $\lambda_0 = 400nm$ is a reference wavelength and $S = 0.016nm^{-1}$ is an empirically-determinated slope of the exponential. In modelling absorption by the yellow-substance component, variations in the magnitude of this component are parametrised through $a_y(\lambda_0)$, rather than through a more conventional measure of concentration, like chl a an TSM.

Pure Water: The scattering and absorption properties of pure water are well known and [40] represents the best measurements to date in the visible band. These absorption measurements in pure freshwater are taken to be valid for pure seawater, since the dissolved salts are not expected to have significant impact, given that their major absorption bands are in the ultraviolet and infrared domains.

2.3 Inversion Procedures

The inversion problem in remote sensing is to determine biogeochemical parameters from the upwelling radiance spectrum. In theory, it has been shown that the radiance distribution and its derivative can be inverted to obtain the inerhen optical properties of water. However in remote sensing only the radiance at the surface in a few directions is known. Thus, the version must depend on approximations. The processes can be conceptually divided in two steps:

1. The derivation of IOPs from the radiance.
2. The derivation of water constituents from the IOPs.

In this section, different techniques that allow to implement this procedures are presented. In the one hand, Radiative transfer calculations allow to use complex models of the atmosphere to build a forward models capable of emulating accurately the behaviour of the atmosphere. On the other hand, Neural Network techniques and its training, using previous radiative transfer simulations, allow to implement computationally efficient processor that can operate quickly.

2.3.1 Radiative Transfer Modelling

The theory of radiative transfer was mainly developed in the first half of the twentieth century with the aim to extract information on stellar and planetary atmospheres from measured light spectra. According to the history of its development, the theory of radiative transfer uses macroscopic quantities such as the beam attenuation coefficient or the volume scattering function to describe the interaction of light with matter.

The theory starts from the Radiative Transfer Equation (RTE) which describes the radiance change at some frequency v in a medium characterized by an absorption coefficient k_v and an emission coefficient j_v. Counting up the gains and losses when the radiation travels through an infinitesimal distance ds, we have [9]:

$$\frac{dL_v}{ds} = -k_v \rho L_v + j_v \rho \tag{34}$$

where ρ is the density of the material. If now we define the *source function* as the ratio of the emission to the absorption:

$$\Im_v = \frac{j_v}{k_v} \tag{35}$$

we can rewrite the equation in the form:

$$-\frac{dI_v}{k_v \rho ds} = I_v - \Im_v \tag{36}$$

The formal solution of the RTE is readily written down:

$$L(s) = L(0)e^{-\tau(s,0)} + \int_0^s \Im(s')e^{-\tau(s,s')} k\rho ds' \tag{37}$$

where $\tau(s, s')$ is the *optical thickness* of the material between the points s and s':

$$\tau(s, s') = \int_{s'}^{s} k\rho ds \tag{38}$$

The physical meaning of the solution (Eq.37) is clear: It expresses the fact that the intensity at any point and in a given direction results from the emission at all previous points, s', reduced by the factor $e^{-\tau(s,s')}$ to allow for the absorption by the intervening matter. One common particular case of RTE is the so called equation of transfer for plane-parallel problem, which can be written as:

$$\mu \frac{dL(\tau,\mu,\phi)}{d\tau} = L(\tau,\mu,\phi) - \Im(\tau,\mu,\phi) \tag{39}$$

where μ is the $cos(\theta)$ and τ is the *normal optical thickness*:

$$\tau = \int_z^\infty k\rho dz \tag{40}$$

where z is the distance normal to the plane of stratification. The solution is then $(1 > \mu > 0)$:

$$L(\tau,+\mu,\phi) = L(\tau_1,\mu,\phi)e^{-(\tau_1-\tau)/\mu} + \int_\tau^{\tau_1} \Im(t,\mu,\phi)e^{-(t-\tau)/\mu}\frac{dt}{\mu} \tag{41}$$

and $(-1 > \mu > 0)$

$$L(\tau,+\mu,\phi) = L(\tau_1,-\mu,\phi)e^{-\tau)/\mu} + \int_0^{\tau_1} \Im(t,-\mu,\phi)e^{-(\tau-t)/\mu}\frac{dt}{\mu} \tag{42}$$

for a finite atmosphere which is bounded on two sides at $\tau = 0$ and at $\tau = \tau_1$.

Monte Carlo Radiative Transfer Code

A Monte Carlo Radiative Transfer Code employs Monte Carlo methods for numerical solution of the *radiative transfer equation* (RTE) [29]. It is the most general technique since it is applicable to time dependent and three dimension radiative transfer equation with arbitrary boundary geometry. The idea underlying all Monte Carlo Methods is this: if we know the probability of occurrence of each single event in a sequence of events, then we can determine the probability that the entire sequence of events will occur.

Consider a photon that enters the atmosphere, with a direction $\hat{\xi}$. We must determine how far the photon travels in direction $\hat{\xi}$ before it encounters an air molecule or other particle. If we think of a photon not as one photon but as a beam of many photons, this beam has some radiance L($\hat{\xi}$) which decreases with distance according to:

$$\frac{dL(\hat{\xi})}{dz} = -k\rho L(\hat{\xi}) \tag{43}$$

or

$$L(\hat{\xi}) = L(0)e^{-l} \tag{44}$$

in terms of *optical path length* $l = k\rho z$. This decrease in radiance can be explained in terms of the probability of any particular photon being absorbed or scattered out of the beam between optical path lengths l and $l + dl$ which would have the next *probability density function* (pdf):

$$p_l(l) = e^{-l} \tag{45}$$

To generate such a distribution starting from a uniformly distributed number R, it is possible to use the *inverse function method*, that is:

$$l = P_l^{-1}(R) \tag{46}$$

where $P_l(l)$ is the *distribution function* of $p_l(l)$:

$$P_l(l) = \int_0^l p_l(l')dl' = 1 - e^{-l} \tag{47}$$

Hence,

$$l = -\ln(1 - R) \tag{48}$$

Suppose that a photon travels a distance l and then interacts with the medium. It has to be decided at random if the interaction is to be absorption or a scattering event. This is easily done by drawing another random number N uniformly distributed between [0,1] and comparing it with the albedo of single scattering w_0. Thus, if $N > w_0$ the interaction will be an absorption event, otherwise it will result in scattering. w_0 is therefore often called *probability of photon survival*. If the interaction results in absorption, the photon is terminated. On the other hand, if the interaction is a scattering event, new photon direction has to be determined by use of the phase function $\tilde{\beta}(\hat{\xi})$. Therefore, now the procedure will have to generate the pdf's of $\hat{\xi} = (\theta, \phi)$ from a uniformly distributed number Z. That is to solve the following:

$$Z = P_{\theta,\phi}(\theta, \phi) = \int_0^\theta \int_0^\phi \tilde{\beta}(\theta, \phi) \sin\theta d\theta d\phi \tag{49}$$

which in general must be solved numerically.

The same type of reasoning is easily extended. Thus, different kind of particle can be considered, choosing their pdf's with certain probability.

As many photons are traced, the solution of the RTE is being built one photon at a time. Counting the photons that reach the bottom allows to measure the final value of energy or radiance. Some strengths of this method are:

1. It is conceptually simple.
2. It is instructive, in the sense that the method highlights the fundamental radiative transfer process of absorption and scattering, making a connection between the photon level and the energy level formulations.
3. It is very general. It is applicable to any geometry, incident lighting, scattering phase function, distribution of IOPs,etc.
4. It is simple to program.

And some weaknesses:

1. It provides no insight into the underlying mathematical structure of radiative transfer theory.
2. It can be computationally extremely inefficient since almost all the photons generated are "wasted" because they never reach the bottom.
3. There is presence of statistical sampling noise in the estimated values whose standard deviation is proportional to $n^{-1/2}$. Therefore, if a 1% of accuracy is desired, we need

$$n = \frac{1}{0.01^2} = 10^4 \tag{50}$$

samples reaching the bottom. If for example we consider that only 1% of the samples at the start survive, we will need to trace a total of 10^6 samples.

Matrix Method Radiative Transfer

The matrix operator theory [39] is an entirely rigorous method for the solution of the equations radiative transfer in a plane parallel medium. The position within the medium is denoted by the *optical depth* τ. Let's consider a medium divided in three layers which boundaries are τ_0,τ_1 and τ_2 from the top to the bottom, respectively.

Now $L^+(\tau)$ is the *specific downward radiance* at the optical depth τ. In matrix operator theory $L^+(\tau)$ is a column matrix

$$L^+(\tau) = \begin{pmatrix} L^+(\tau, \mu_1) \\ L^+(\tau, \mu_2) \\ \vdots \\ L^+(\tau, \mu_m) \end{pmatrix} \tag{51}$$

where $L^+(\tau, \mu_i)$ is the downward radiance at the angle $\theta_i = \cos^{-1}\mu_i$. Similarly $L^-(\tau)$ is the column matrix that represents the upward radiance. Consider the radiance values of the radiation emerging from a layer whose boundaries are at τ_0 and τ_1. These radiances, $L^+(\tau_1)$ and $L^-(\tau_0)$, depend linearly on the incident radiances to the layer, $L^+(\tau_0)$ and $L^-(\tau_1)$ and the contribution from the sources within the layer, $J^+(\tau_0, \tau_1)$ and $J^-(\tau_1, \tau_0)$, the downward radiance at τ_1 and the upward radiance at τ_0 due to the sources within the layer, respectively. Thus we have (simplifying the notation) [24]:

$$L_1^+ = t_{01}L_0^+ + r_{10}L_1^- + J_{01}^+ \tag{52}$$

$$L_0^- = r_{01}L_0^+ + t_{10}L_1^- + J_{10}^- \tag{53}$$

where $t_{01} = t(\tau_0, \tau_1)$ is a diffuse transmission operator and $r_{01} = r(\tau_0, \tau_1)$ is a diffuse reflection operator. These operators are mxm matrices that multiply the column vectors L.Therefore, the objective is the derivation of the expression for the diffuse reflection and transmission operators for a combined layer (from τ_0 to τ_2) from the known operators for the separate layers (from τ_0 to τ_1 and from τ_1 to τ_2. This method can be extended to any number of layers.

Now, the emerging radiance of the layer from τ_1 to τ_2 can be written in the same way as before:

$$L_2^+ = t_{12}L_1^+ + r_{21}L_2^- + J_{12}^+ \tag{54}$$

$$L_1^- = r_{12}L_1^+ + t_{21}L_2^- + J_{21}^- \tag{55}$$

and also for the combined layer from τ_0 to τ_2:

$$L_2^+ = t_{02}L_0^+ + r_{20}L_2^- + J_{02}^+ \tag{56}$$

$$L_0^- = r_{02}L_0^+ + t_{20}L_2^- + J_{20}^- \tag{57}$$

Now, we operate with Eq. 52 and 55 as follows:

$$L_1^+ = t_{01}L_0^+ + r_{10}L_1^- + J_{01}^+$$
$$r_{12}L_1^+ = r_{12}(t_{01}L_0^+ + r_{10}L_1^- + J_{01}^+)$$

and we sum it to the Eq. 55 and multiply the resulting equation from the left by $(I - r_{12}r_{10})^{-1}$:

$$
\begin{aligned}
&r_{12}L_1^+ + L_1^- = r_{12}(t_{01}L_0^+ + r_{10}L_1^- + J_{01}^+) + r_{12}L_1^+ + t_{21}L_2^- + J_{21}^-,\\
&(I - r_{12}r_{10})^{-1}(r_{12}L_1^+ + L_1^-) = (I - r_{12}r_{10})^{-1}(r_{12}(t_{01}L_0^+ + r_{10}L_1^- + J_{01}^+)+\\
&+ r_{12}L_1^+ + t_{21}L_2^- + J_{21}^-),\\
&(I - r_{12}r_{10})^{-1}(r_{12}L_1^+ + L_1^-) = (I - r_{12}r_{10})^{-1}(r_{12}r_{10}L_1^- + r_{12}L_1^+)+\\
&(I - r_{12}r_{10})^{-1}(r_{12}t_{01}L_0^+ + r_{12}J_{01}^+ + t_{21}L_2^- + J_{21}^-),\\
&(I - r_{12}r_{10})^{-1}(r_{12}L_1^+ + L_1^- - r_{12}r_{10}L_1^- - r_{12}L_1^+) = (I - r_{12}r_{10})^{-1}(r_{12}t_{01}L_0^+ + r_{12}J_{01}^+ +\\
&+ t_{21}L_2^- + J_{21}^-),\\
&(I - r_{12}r_{10})^{-1}(I - r_{12}r_{10})L_1^- = (I - r_{12}r_{10})^{-1}(r_{12}t_{01}L_0^+ + r_{12}J_{01}^+ + t_{21}L_2^- + J_{21}^-),\\
&L_1^- = (I - r_{12}r_{10})^{-1}(r_{12}t_{01}L_0^+ + r_{12}J_{01}^+ + t_{21}L_2^- + J_{21}^-) \qquad (58)
\end{aligned}
$$

Similarly, if we multiply Eq. 55 by r_{10} from the left, add to Eq. 52, and multiply the resulting equation from the left by $(I - r_{10}r_{12})^{-1}$ we obtain:

$$L_1^+ = (I - r_{10}r_{12})^{-1}(r_{10}t_{10}L_2^- + r_{10}J_{21}^- + t_{01}L_0^+ + J_{01}^+) \qquad (59)$$

Finally, we can substitute the expression for L_1^+ from Eq.59 into Eq.54:

$$
\begin{aligned}
L_2^+ &= t_{12}(I - r_{10}r_{12})^{-1}(r_{10}t_{10}L_2^- + r_{10}J_{21}^- + t_{01}L_0^+ + J_{01}^+) + r_{21}L_2^- + J_{12}^+\\
&= t_{12}(I - r_{10}r_{12})^{-1}t_{01}L_0^+ +\\
&+ (t_{12}(I - r_{10}r_{12})^{-1}r_{10}t_{21} + r_{21})L_2^- +\\
&+ t_{12}(I - r_{10}r_{12})^{-1}(J_{01}^+ + r_{10}J_{21}^-) + J_{12}^+
\end{aligned}
$$

A comparison of the coefficient of these quantities with Eq.56:

$$L_2^+ = t_{02}L_0^+ + r_{20}L_2^- + J_{02}^+ \qquad (60)$$

yields the following expressions:

$$
\begin{aligned}
t_{02} &= t_{12}(I - r_{10}r_{12})^{-1}t_{01}\\
r_{20} &= t_{12}(I - r_{10}r_{12})^{-1}r_{10}t_{21} + r_{21}\\
J_{02}^+ &= t_{12}(I - r_{10}r_{12})^{-1}(J_{01}^+ + r_{10}J_{21}^-) + J_{12}^+
\end{aligned}
$$

And the substitution of the expression for L_1^- from Eq.58 into Eq. 53 and comparison of this result with Eq.57 yields:

$$
\begin{aligned}
t_{20} &= t_{10}(I - r_{12}r_{10})^{-1}t_{21}\\
r_{02} &= t_{10}(I - r_{12}r_{10})^{-1}r_{12}t_{01} + r_{02}\\
J_{20}^- &= t_{10}(I - r_{12}r_{10})^{-1}(J_{21}^- + r_{12}J_{01}^+) + J_{10}^-
\end{aligned}
$$

Hence, this procedure can be extended to any number of layers where the combined expressions of *diffuse transmittance* and *diffuse reflectance* can be analytically computed. In addition, these two parameters can be directly related to *IOPs* of the medium such as the *absorption coefficient* or the *scattering coefficient*. The matrix operator theory has the following advantages compared with other approaches:

1. All orders of the reflection and transmission matrices are calculated at once with a corresponding reduction in computer time over methods involving iterations.

2. Layers of any thickness may be combined, so that realistic model of the atmosphere may be developed from any arbitrary number of layers of any predetermined thicknesses, each with different properties.

3. Calculations can readily be made for large optical depths and with highly anisotropic phase functions.

4. All fundamental equations can be interpreted immediately in terms of the physical interactions appropriate to the problem.

And these disadvantages:

1. Its mathematical complexity makes difficult to implement accurate models.

2. The assumption of the single scattering approximation.

2.3.2 Neural Networks

An artificial neural network is a set of simple processing units which communicate by sending signals to each other over a larger number of weighted connections. The task performed by a single unit is very simple: it receives input from neighbours or external sources and use this to compute an output which is propagated to other units. According to the location of the unit inside the network, there a three types of processing units:

1. Input units, which receive data from outside the neural network.

2. Output units, which send data out of the neural network.

3. Hidden units, whose input and output signals remain within the neural network.

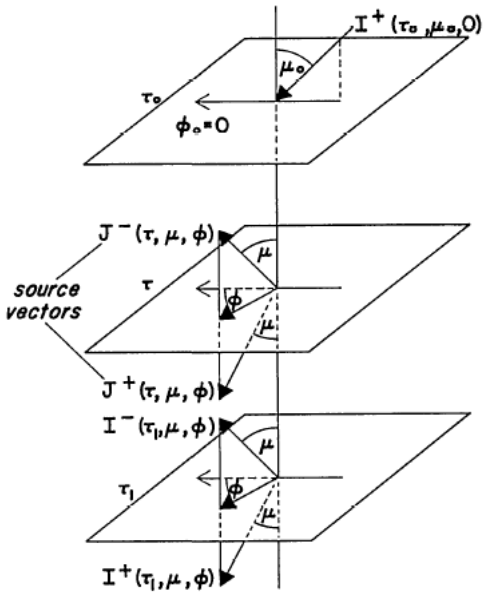


Figure 4: Graphic representation of two layers with three boundaries and its matrix operator vectors [39].

In most cases, the total input is defined as a linear combination of the inputs plus an *offset* term [36]:

$$s_k = \sum_j w_{jk} y_j + \theta_k \tag{61}$$

where s_k is the total input to the unit k, j is the index of the input, y_j its value, w_{jk} the weight of the input j to the unit k and θ_k the offset term. Therefore the output is obtained after applying a nondecreasing *activation function*. Thus:

$$y_k = F_k(s_k) = F_k(\sum_j w_{jk} y_j + \theta_k) \tag{62}$$

Depending on the application, the activation function can be a hard limiting threshold function (a Heaviside function) or some smoothly limiting threshold such as a sigmoid (See Figure 5):

$$y_k = F(s_k) = \frac{1}{1 + e^{-s_k}} \tag{63}$$

Neural networks can be classified according to their topology in two groups:

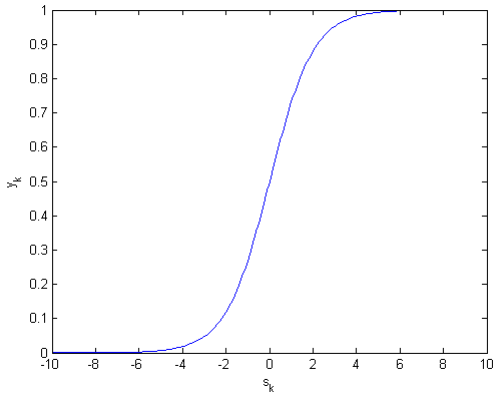


Figure 5: Sigmoid function (See Eq. 63).

1. Feed-forward networks, where the data flow from input to output units is strictly feed-forward. There can be multiple layers but no feedback connections are present, which means that there are not connections extending from outputs of units to inputs of units in the same layer or previous ones.

2. Recurrent networks that do contain feedback connections.

Before it can be used, a neural network has to be configured (i.e. set values to all the weights) such that the application of a set of inputs produces the desired set of outputs. The typical procedure, called *supervised learning*, is to "train" the neural network by providing it with input and matching output patterns and letting it change its weights according to some learning rule.

One of the most common training algorithm used is the *back-propagation learning rule* suitable for non-linear activation functions and multilayered networks. Let's consider a multi-layer feed-forward neural network (See Figure 6) with a differentiable activation function:

$$y_k^p = F(s_k^p) \tag{64}$$

where y_k^p denotes the output of the unit k, p stands for some specific input vector (or pattern) to the neural network and s_k^p is the total input of the considered unit k for the input pattern to the neural network p. In addition, s_k^p will be defined like in the Eq. 61, that is:

$$s_k^p = \sum_j w_{jk} y_j^p + \theta_k \tag{65}$$

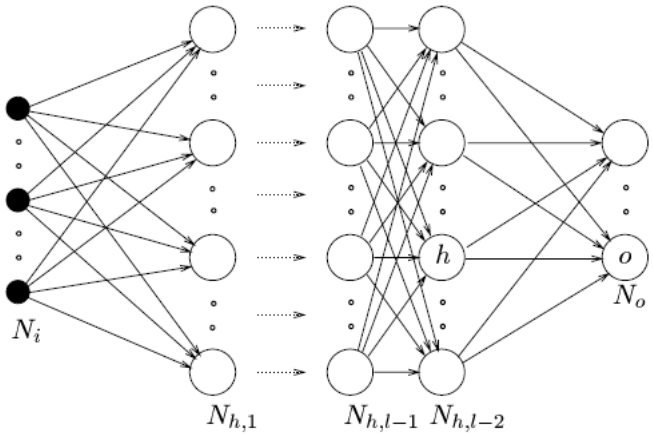


Figure 6: A multi-layer feed-forward neural network with l layers of units [36].

Suppose we want to train the neural network so that for a given input sample patterns x^p we obtain a desired output values d_0^p. For every given input sample, the output of the network differs from the target by $(d_0^p - y_0^p)$, where y_0^p is the actual output for this pattern p. The back-propagation learning rule define the cost function to minimise as the *summed squared error*. That is:

$$E = \sum_p E^p = \sum_p \frac{1}{2} \sum_{o=1}^{N_o} (d_o^p - y_o^p)^2 \tag{66}$$

where N_o are the total number of output units, the index p ranges over the set of input patterns and E^p represents the error on pattern p. The procedure to iteratively minimise the cost function is therefore the *least minimum square* procedure (LMS) by a method called *gradient descent*. The idea is to make a change in the weight proportional to the negative of the derivative of the error as measured on the current pattern with respect to each weight:

$$\Delta_p w_{jk} = -\gamma \frac{\partial E^p}{\partial w_{jk}} \tag{67}$$

Now, applying the chain rule:

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \tag{68}$$

and deriving in Eq. 65, the second factor is:

$$\frac{\partial s_k^p}{\partial w_{jk}} = y_j^p \tag{69}$$

and the first one is defined as δ_k^p:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} \tag{70}$$

Therefore, we can express Eq.67 in this way:

$$\Delta_p wjk = \gamma \delta_k^p y_j^p \tag{71}$$

There is a simple recursive computation of these δ_k^p for each k in the network which can be implemented by propagating error signals backward through the network. Thus, applying the chain rule to δ_k^p we have:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} = -\frac{\partial E^p}{\partial y_k^p}\frac{\partial y_k^p}{\partial s_k^p} \tag{72}$$

From Eq. 64 we can express the second factor as:

$$\frac{\partial y_k^p}{\partial s_k^p} = F'(s_k^p) \tag{73}$$

which is simply the derivative of the activation function F for the unit k, evaluated at the total input s_k^p to that unit.

On the other hand, to compute the first factor, there are two possible cases. In the first one, k is an output unit $k = o$ of the network. Therefore, deriving the Eq.66 we simply have:

$$\frac{\partial E^p}{\partial y_o^p} = -(d_o^p - y_o^p) \tag{74}$$

and we get the δ_o^p for the output units:

$$\delta_o^p = (d_o^p - y_o^p) F_o'(s_o^p) \tag{75}$$

In the second case, if k is not an output unit but a hidden unit $k = h$ of the previous layer, we do not readily know the contribution of the unit to the output error of the network. However, the error measure can be written as a multivariate function of the net inputs from hidden to output layer:

$$E^p = E^p(s_1^p, s_2^p, ..., s_j^p, ..., s_{N_o}^p) \tag{76}$$

Therefore, the chain rule can be applied:

$$\frac{\partial E^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial s_o^p}{\partial y_o^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial}{\partial y_o^p} \sum_{j=1}^{N_h} w_{ko} y_j^p = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} w_{ho} = -\sum_{o=1}^{N_o} \delta_o^p w_{ho} \tag{77}$$

and we get the δ_h^p for the last hidden units:

$$\delta_h^p = F_o'(s_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} \tag{78}$$

Applying this procedure iteratively, it is clear that we can obtain now all the weight changes of all the layers in the network, starting from the output to the input. Finally, let's consider a sigmoid activation function. Its derivate will be:

$$\begin{aligned} F'(s^p) &= \frac{\partial}{\partial s^p} \frac{1}{1+e^{-s^p}} \\ &= \frac{e^{-s^p}}{(1+e^{-s^p})^2} \\ &= y^p(1-y^p) \end{aligned}$$

Hence, the weight change for an output unit can be written as:

$$\Delta w_{jo} = \gamma(d_o^p - y_o^p) y_o^p (1-y_o^p) y_j^p \tag{79}$$

and the weight change for a hidden unit:

$$\Delta w_{jo} = \gamma y_h^p (1-y_h^p) y_j^p \sum_{o=1}^{N_o} \delta_o^p w_{ho} \tag{80}$$

To sum up, the application of the back-propagation learning rule involves two phases: During the first phase the input x is presented and propagated forward through the network to compute the output values y_o^p for each output unit. This output is compared with its desired value d_o, resulting in an error signal δ_o^p for each output unit. The second phase involves a backward pass through the network during which the error signal is passed to each unit in the network and appropriate weight changes are calculated.

3 Materials and methods

3.1 MERIS

This work deals with the comparison of different water processors (See Section 3.2) developed to retrieve water constituents from MERIS products in the context of remote sensing applied to water quality monitoring (See Section 2.1.1). An historical review of the different elements involved are presented in this section.

Envisat

The European Space Agency (ESA) started the development of the *Environmental Satellite* (Envisat) and its payload instruments (see Table 1) after December 1993 where the Polar Orbiting Earth Observation Mission (POEM-1) was split into the Environmental Mission (Envisat) and the Meteorological Mission (METOP-1). The Envisat satellite (see Figure 7) was launched on March 1st, 2002 aboard an Ariane 5 into the Sun synchronous polar orbit.

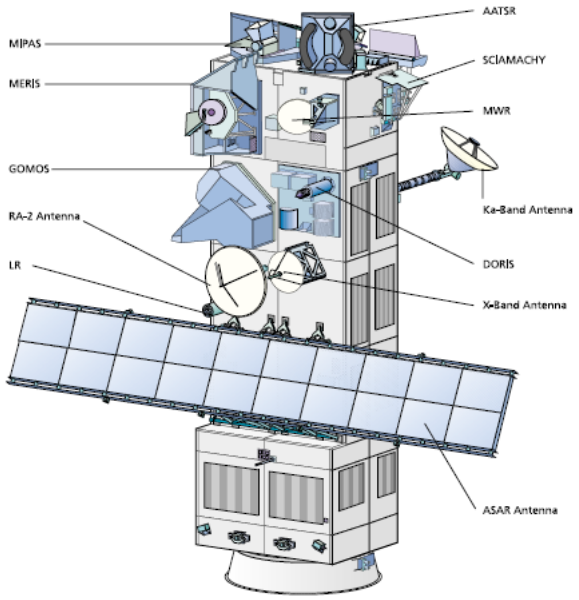


Figure 7: Envisat Satellite Configuration [16]. See Table 1 for a description of the instruments.

The primary objectives of this mission are [16]:

1. To continue and improve upon measurements initiated by the *European Remote Sensing Satellites* (ERS-1 and ERS-2).
2. To enhance the ERS mission, notably the ocean and ice mission.
3. To extend the range of parameters observed to meet the need of increasing knowledge of the factors determining the environment.
4. To make a significant contribution to environmental studies, notably in the area of atmospheric chemistry and ocean studies.

Furthermore, during six years, Envisat has delivered many remarkable results [55]:

1. In November 2002, Envisat revealed the extent of the oil slick after the "Prestige" tanker accident.
2. It was demonstrated during the Hurricane Katrina in summer of 2005 that the synergy of measurements is one of the advantages of a multi-instrument platform. Wind field measurements revealed the warm water body in the Gulf of Mexico feeding the hurricane with energy.
3. MERIS Full Resolution data are the basis for a new global land cover map to a resolution three times sharper than any previous satellite map (the Globecover Project).
4. The ASAR Background Mission revealed the blind fault of the Bam earthquake. It provides, besides others, subsidence measurements which form the basis of applications such as urban planning.
5. Envisat helps our understanding of anthropogenic greenhouse gases. Recently, discovery of high amounts of methane over tropical rain forests point the possibility of a not yet known source of methane (or to significantly underestimated known source). The measurements of NO_2 show an NO_2 concentration increase by 50% over China during last 8 years.
6. Envisat ASAR reveals ice loss such as exemplified with the Larsen B ice shelf over the past years. In 2005, the Arctic sea ice covered 5.3 million km^2, the lowest ever recorded extent by satellites.

Name	Description
ASAR	Advanced Synthetic Aperture Radar
MERIS	Medium-Spectral Resolution Imaging Spectrometer
AATSR	Advanced Along Track Scanning Radiometer
RA-2	Radar Altimeter
MWR	Microwave Radiometer
GOMOS	Global Ozone Monitoring by Occultation of Stars
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
DORIS	Doppler Orbitography and Radio-positioning Integrated by Satellite
LRR	Laser Retro-Reflector

Table 1: ENVISAT payload instruments.

7. The results of more than a decade of ERS and Envisat observations of the Earth show a global sea level rise of $3.2mm/year$ compared to $1.6mm/year$ in the past. Such results have fed the recently published UN climate change assessment report.

One of the instruments of the Envisat, the so called MERIS, is specially suitable to be used in water quality monitoring (See section 1.1). Here are explained some of its main features.

MERIS

The Medium Resolution Imaging Spectrometer (MERIS) is one of the payload instruments of Envisat (see Table 1) and provides data needed to estimate water quality parameters. It operates in the solar reflective spectral range of visible and near infrared light. Fifteen spectral bands can be selected by ground command, each of which has a programmable width and a programmable location between 390 nm and 1040 nm (See Table 2). MERIS measurement data are used to derive information including:

1. Characteristics of clouds, such as top pressure and optical thickness.

2. The presence of vegetation and derived indexes to characterize vegetation vigor.

3. Atmosphere parameters such as aerosol optical thickness, aerosol type, and water vapor column contents.

4. Ocean colour parameters in open waters such as chlorophyll and coastal waters parameters such as chlorophyll (chl a), dissolved organic matter (aCDOM), and suspended solid matter (TSM) concentrations.

The MERIS instrument has a cross-track Field of Vision (FOV) of 68º. Therefore it produces an image swath width of approximately 1150 km. The swath is comprised of 5 imaging spectrometer modules. However, it may operate at Full Resolution (FR) of approximately 0.3 km pixels at nadir, or at Reduced Resolution (RR) of approximately 1.2 km pixels at nadir:

Full Resolution (FR): In Full Resolution mode, it pixel has an Instantaneous Field of Vision (IFOV) of 0.019º, with a nadir spatial sampling of 0.26 km across track by 0.29 km along track. The data are processed *On Request* from the acquired Level 0 segments, on a floating scene basis. MERIS Full Resolution Swath (FRS) data are processed offline.

Reduced Resolution (RR): In Reduced Resolution mode, it has a maximum length of 43.5 minutes (all of full sunlight orbit), producing approximately 17400 km of coverage. Each pixel is approximately 1.04 km across track by 1.16 km along track at nadir. The processing is done systematically.

The instrument has 16 bits of digital dynamic range [15], with selectable gain settings. It also has on-board shutters and reflectance panels for on-orbit radiance and wavelength calibration, and collects a 16th "smear" band of data for compensation of radiance spectral smearing during CCD charge transfer during readout.

Each CCD module readout may "bin" the spectral detectors (each detector with 2.5 nm resolution and at 1.25 nm detector spacing) into selectable bandwidth bands at selectable wavelength positions. The "nominal" 15 spectral bands are selected later from the 16 bands from the Table 2.

The output, or products of MERIS are image packages of several parameters such as radiances at the top of the atmosphere (TOA) in different bands. Different products are created according to the parameters plotted and their resolution. In the Table 3 all the MERIS products are listed with

Centre (nm)	Bandwidth(nm)	Use
412.5	10	Yellow substance and detrital pigments.
442.5	10	Chlorophyll absorption maximum.
490	10	Chlorophyll and other pigments.
510	10	Suspended sediment, red tides.
560	10	Chlorophyll absorption minimum.
620	10	Suspended sediment.
665	10	Chlorophyll absorption and fluorescence reference.
681.25	7.5	Chlorophyll fluorescence peak.
705	10	Fluorescence reference, atmospheric corrections.
753.75	7.5	Vegetation, cloud.
760.625	3.75	O_2R-branch absorption band.
865	20	Vegetation, water vapour reference.
885	10	Atmosphere corrections.
900	10	Water vapour, land.

Table 2: MERIS bands wavelength, bandwidth and use[17].

their respective ID's and the working mode of the instrument: Reduced Resolution (RR), Full Resolution (FR) or Full Resolution full Swath (FRS).

It is possible to estimate different water constituents such as chlorophyll a, total suspended matter or aCDOM from the radiances at the top of the atmosphere provided by the MERIS instrument products (See Section 2.1.1). Several algorithms have been developed with this objective.

Therefore, in the present work several MERIS products of type MER_FR__1P and MER_FRS_1P are processed with different water processors (see Section 3.2) to retrieve water constituents as automatically as possible using the tools developed for this purpose (See Section 3.4.2). After this, the different processed estimations will be compared with matching *in situ* data (See Section 4) and within them (See Section 3.4) in order to compare the behaviour of the processors studied in terms of reliability and validity. The results of the comparison will be discussed in Section 5 and the conclusions stated in Section 6.

Mode	Product ID	Description
RR	MER_RV__0P	MERIS Level 0 Reduced Field of View
RR	MER_CA__0P	MERIS Level 0 Calibration (all calibration modes)
RR	MER_RR__0P	MERIS Level 0 Reduced Resolution
RR	MER_RR__1P	Reduced Resolution Geolocated and Calibrated TOA Radiance (stripline)
RR	MER_RR__2P	Reduced Resolution Geophysical Product for Ocean, Land and Atmosphere (stripline)
RR	MER_LRC__2P	Extracted Cloud Thickness and Water Vapour for Mete Users Level 2 Product generated from MER_RR_2P (Cloud thickness and water vapour content for the Meteo at reduced resolution 5 km) (stripline)
RR	MER_RRC__2P	Extracted Cloud Thickness and Water Vapour (non-Meteo Users) Level 2 product extracted from MER_RR__2P (Cloud thickness and water vapour content at nominal RR resolution) for NRT distribution (stripline)
RR	MER_RRV__2P	Extracted Vegetation Indexes Level 2 product extracted from MER_RR__2P (Vegetation indexes including atmospheric corrections for selected land regions) for NRT distribution (stripline)
RR	MER_RR__BP	Browse (covers FR and RR requirements) (stripline)
FR	MER_FR_0P	MERIS Level 0 Full Resolution
FR	MER__1P	Full Resolution Geolocated and Calibrated TOA Radiance
FR	MER_FR__2P	Full Resolution Geophysical Product for Ocean, Land and Atmosphere
FRS	MER_FRS_1P	Full Resolution Full Swath Geolocated and Calibrated TOA Radiance
FRS	MER_FRS_BP	Full Resolution Full Swath Browse product
FRS	MER_FRS_2P	Full Resolution Full Swath Geophysical Product for Ocean, Land and Atmosphere

Table 3: MERIS products description[17].

3.2 MERIS water quality processors

In this section, the studied processors are presented. The first three have been developed by the *Gesellschaft für Kernenergieverwertung in Schiffbau und Schiffahrt mbH* (GKSS) and the last one by the *Freie Universitaet Berlin/Institut für Weltraumwissenschaften* (FUB/Wew). All of them can be conceptually divided in two main parts: the atmospheric correction algorithm and the water algorithm. The first part, determines the water leaving radiance reflectance spectrum $RLw(\lambda)$ from the top of atmosphere radiance reflectance spectrum $RL_{toa}(\lambda)$ measured by MERIS; the second part, uses the water leaving radiances as input and computes different products that provide information about the quality of the water.

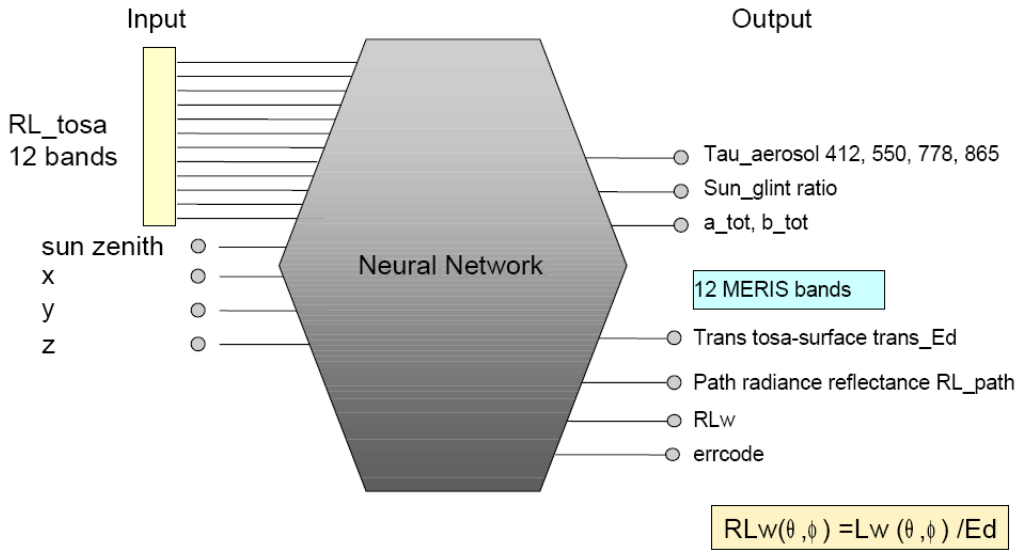


Figure 8: Outline of the Atmospheric Correction NN included in C2R [42].

3.2.1 Coastal Case 2 Regional Processor (C2R)

Atmospheric Correction Algorithm

The core of the atmospheric correction procedure is based on radiative transfer simulations according to the angular resolving ocean-atmosphere photon tracing Monte Carlo radiative transfer code developed by GKSS and partly based on [24, 29, 31]. However, to reduce computational time, its water part has been replaced by a forward neural network (fwNN) *55x20_2295.4.net* trained by the HYDROLIGHT radiative transfer code [29].

The simulated radiance reflectances in multiple bands, as shown in Table 4, as well as solar zenith angles and viewing zenith angles are then used to train a fully connected *25x30x35x40_4016.9.net* neural network (NN) with the GKSS Neural Network Simulator FFBP v1.0 [48] that parametrizes the relationship between RL_{tosa} and RLw (See Figure 8).

Finally, the angstrom coefficient is computed from the aerosol optical thicknesses that are direct output of the NN.

An atmosphere model (see Figure 9) that can be divided in two parts is used:

1. A layer between the top of atmosphere (TOA) and the top of standard atmosphere (TOSA) that directly computes RL_{tosa} from RL_{toa} which is previously calculated from MERIS L1b data (solar irradi-

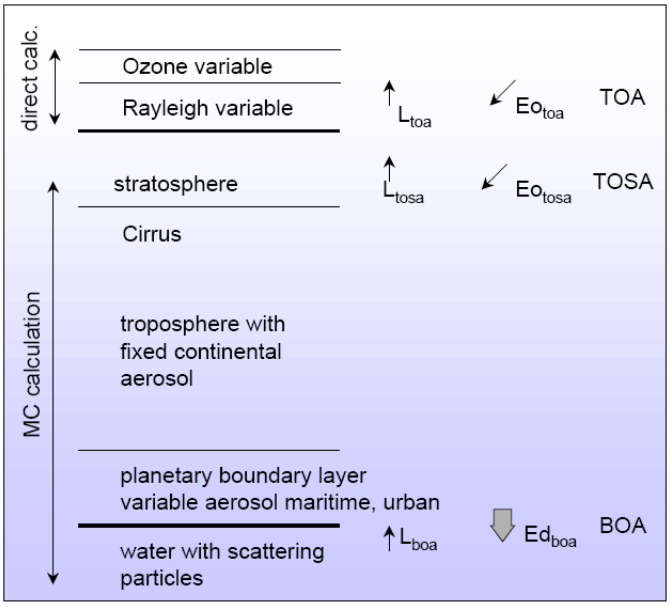


Figure 9: Model atmosphere, TOA is top of atmosphere, TOSA top of standard atmosphere and BOA bottom of atmosphere [42].

ance, cosine of the sun zenith and the downwelling irradiance at TOA). It contains therefore the differences between the standard atmosphere and real atmosphere concerning:

(a) Actual air pressure taken from MERIS L1b data used to correct the Rayleigh scattering.

(b) Actual ozone column content taken from MERIS L1b data used to correct the Ozone absorption.

(c) Water vapour influence correction on band 9 (708 nm) as performed in the MERIS Instrument Processing Facility. (IPF)

2. A standard atmosphere, used in the simulations to train the NN, which include 50 layers, 1km of thick each, with the next layer dependent parameters:

(a) Four different aerosols models as shown in Table 5, ranged and vertically distributed according to AERONET measurements [4].

(b) Thin cirrus cloud particles at top of the troposphere adapted from [27] simulation according to statistical parameters from [12], ranged and vertically distributed according to [4].

Band No.	Wavelength (nm)	GKSS Atm. Corr.	FUB/Wew Atm. Corr.
1	412	X	X
2	443	X	X
3	490	X	X
4	510	X	X
5	560	X	X
6	620	X	X
7	665	X	X
8	681	X	
9	708	X	X
10	753	X	X
11	760		
12	778	X	X
13	865	X	X
14	885		X
15	900		

Table 4: MERIS bands used for the Atmospheric Correction in the different processors (GKSS applies for Coastal, Boreal Lakes and Eutrophic Lakes processors; FUB/Wew applies for FUB/Wew Water processor.

(c) Rough and wind dependent sea surface using formulation of [10], with specular reflectance computed using Fresnel reflection formula with $n_{water} = 1.34$.

And the next constant parameters:

(a) Rayleigh scattering coefficient profile given by [13] for a standard air pressure at sea level of 1013.25 hPa.

(b) Ozone absorption profile taken from [13] for a surface pressure of 1013.25 hPa and a total ozone column content of 350 Dobson units.

Water Algorithm

The case 2 coastal water algorithm derives from directional water leaving radiance reflectances in eight spectral bands, four inherent optical properties (IOP):

1. Absorption coefficient of phytoplankton pigments.
2. Absorption coefficient of yellow substance

Aerosol Model	W-S(%)	D-L(%)	Soot(%)	H_2SO_4 (%)	Oceanic (%)
Continental	29	70	1	0	0
Maritime	5	0	0	0	95
Urban/Industrial	61	17	22	0	0
Stratospheric	0	0	0	100	0

Table 5: Composition of the different aerosols models used in the standard atmosphere model. Adopted from [1, 25, 52, 57]. Components: Water-Soluble (W-S), Dust-Like (D-L), Soot, H_2SO_4 and Oceanic. The properties of the aerosols components are defined in [25, 57].

3. Absorption of total suspended matter.

4. Scattering coefficient of total suspended matter.

that are directly converted to *chlorophyll a* concentrations (in μg/l) using empirical relations confirmed by Sorensen et al. (2008) and to *TSM dry weight* concentrations (in mg/l) in agreement with [3]. The mean conversion factors and equations are based on the results of the projects COASTLOOC, COLORS/MAP, REVAMP and MAVT which follow protocols set up by ESA MAVT, COLORS project and SeaWiFS project.

To derive the IOPs, it uses a Neural Network trained by a large table (550K entries) of simulated data generated by the forward model based on a HYDROLIGHT radiative transfer code[29].

The NN is composed of an *invNN* that deduce water concentrations from reflectance and geometry parameters, and a *forwNN* that deduce the reflectances from concentration and geometry parameters. Using them sequentially, for a given geometry, it is possible to calculate the error between the initial reflectances and the estimated ones [49]. This error can be therefore minimized iteratively adjusting the estimated concentration in a loop using the Levenberg-Marquardt algorithm which improves the quality of the retrieval [50].

To generate the training data, the forward model is fed with measured IOPs from cruises in the North Sea, partly in the Baltic Sea, Mediterranean Sea and North Atlantic. Its parameter ranges are listed in Table 6.

The following environmental conditions are defined: infinite deep water, vertical homogenous distribution of all water constituents, rough sea surface for a wind speed of $3m/s$ and an atmosphere with 50 layers for 17 solar zenith angles. Finally, no Raman scattering as well as no polarisation effects are considered.

Component/property	Coastal range	Boreal range	Eutrophic range
Yellow substance absorption wavelength exponent	$0.014 \pm 0.002nm^{-1}$	$0.016 \pm 0.0015nm^{-1}$	0.022 (fulvic), 0.008 (humic)
Bleached TSM absorption wavelength exponent.	$0.008 \pm 0.005nm^{-1}$	$0.010 \pm 0.001nm^{-1}$	
TSM scattering wavelength exponent.	0.4 ± 0.2	0.705	0.705
White particle scattering wavelength exponent.	0.0	0.0	0.0
Phytoplankton pigment absorption.	Random spectra from figure 10.	According to spectra from figure 11.	According to spectra from figure 12.
Yellow substance absorption at 442 nm.	$0.005 - 5m^{-1}$	$0.25 - 10.0m^{-1}$	$0.1 - 3.0m^{-1}$
Particle scattering bp at 442 nm.	$0.005 - 30.0m^{-1}$	$0.001 - 0.01m^{-1}$	$0.25 - 30m^{-1}$
White particle scattering bpw at 442 nm.	$0.005 - 30.0m^{-1}$	$0.001 - 0.01m^{-1}$	$0.25 - 30m^{-1}$
Phytoplankton pigment absorption at 442 nm.	$0.001 - 2.0m^{-1}$	$0.024 - 0.84m^{-1}$	$0.032 - 4.56m^{-1}$
Sun zenith angle	0-80º	0-80º	0-80º
Viewing zenith angle	0-50º	0-50º	0-50º
Difference between sun and viewing azimuth angle	0-180º	0-180º	0-180º

Table 6: Variability of the optical properties and range used for the simulation of $RL_w(\lambda)$

3.2.2 Boreal Lakes Processor

Atmospheric Correction Algorithm

The atmospheric correction method of the Boreal Lakes Processor is the same algorithm explained in the Coastal Case 2 Regional Processor section 3.2.1. The only difference refers to the data and its ranges [32] used in the HYDROLIGHT radiative transfer code to train the forward Neural Network (fwNN 55x20_2295.4.net) adopted instead of the water part of the Monte Carlo radiative transfer code as explained in section 3.2.1.

Water Algorithm

The Boreal Lake Water Algorithm and the Coastal Case 2 Regional Water Algorithm are the same except the bio-optical data, extracted from Finnish Lakes (see Figure 11) [32] and its ranges (See Table 6) used to train the Neural Network and the expressions that relate the output IOPs of the NN with the water concentrations (See Table 7).

Concentration	Coastal alg.	Boreal alg.	Eutrophic alg.
Chl a $[mg/m^3]$	$21 * a_{pig}(442)^{1.04}$	$62.6 * a_{pig}(442)^{1.29}$	$26.32 * a_{pig}(442)$
TSM $[g/m^3]$	$1.72 * b_{tsm}(442)$	$1.042 * b_{tsm}(442)$	$1.7 * b_{tsm}(442)$

Table 7: Equations that relate output IOPs of the processors with the water concentrations (Chl a: Chlorophyll a concentration in water; TSM: Total Suspended Matter concentration in water; $a_{pig}(442)$: Absorption of phytoplankton pigment at 442 nm MERIS band $[m^{-1}]$; $b_{tsm}(442)$: Total Suspended Matter scattering at 442nm MERIS band $[m^{-1}]$).

Concentration	Coastal	Boreal	Eutrophic	Water
Chl a $[mg/m^3]$	0.016 - 43.18	0.5 - 50	1 - 120	0.05 - 50.0
TSM $[g/m^3]$	0.0086 - 51.6	0.1 - 20	0.005 - 5	0.05 - 50.0
CDOM $[m^{-1}]$	0.005 - 5	0.25 - 10	0.1 - 3.0	0.005 - 1.0

Table 8: Concentration ranges used in the different processor's simulations (Chl a: Chlorophyll a concentration in water; TSM: Total Suspended Matter concentration in water; CDOM: Absorption of yellow substance).

3.2.3 Eutrophic Lakes Processor

Atmospheric Correction Algorithm

The atmospheric correction algorithm of the Eutrophic Lakes Processor is the same algorithm explained in the CR2 3.2.1.

Water Algorithm

The Eutrophic Lake Water Algorithm and the Coastal Case 2 Regional Water Algorithm are the same except the bio-optical data, extracted from Spanish Lakes (see Figure 12), [32] and its ranges (See Table 6) used to train the Neural Network and the expressions that relate the output IOPs of the NN with the water concentrations (See Table 7).

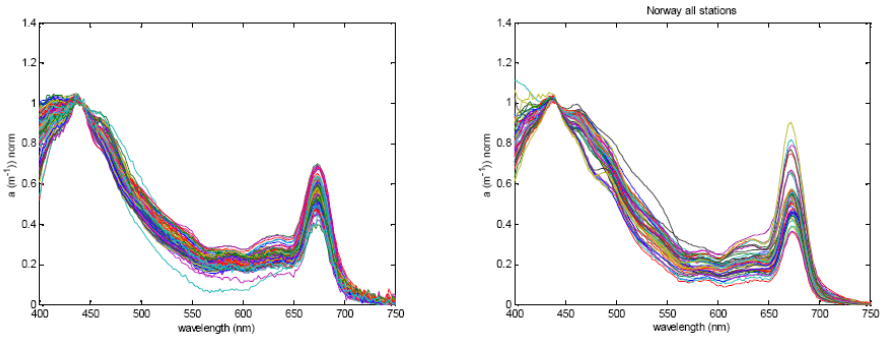


Figure 10: Absorption spectra of phytoplankton pigments used in C2R Processor. The left graph belongs to the Southern North Sea. The one at the right shows spectra of Norwegian coastal waters [42].

3.2.4 FUB/Wew Water Processor

Atmospheric Correction Algorithm

The core of the atmospheric correction algorithm is based on two large databases generated by a radiative transfer code based on the matrix-operator method according to [18, 19] using the MOMO code. The azimuthally resolved upward radiances just above the sea surface and at the top of the atmosphere for a variety of sun and observing geometries are then used to train a three layer fully connected feedforward networks trained by the backpropagation algorithm according to [45]. Multiple MERIS bands are used, as shown in Table 4.

The atmospheric correction scheme can be divided into two parts:

1. A Rayleigh-Ozone correction. This network is trained with 12000 randomly chosen vectors from simulations of a pure Rayleigh-Ozone atmosphere (AOT = 0) to calculate Rayleigh-Ozone path radiances (MERIS bands 4,9,10,12,13 and 14) at the top of the atmosphere for the actual pressure and ozone loading of 344DU for each pixel. Gaussian noise is added to all vectors according to the assumed uncertainty of the associated parameters. A principal component analysis (PCA) is used to decorrelate all input dimensions to the network.

2. Atmospheric correction network.

   (a) The network is trained with simulations based on a U.S. Standard atmosphere of 50 km divided in 11 layers with 8 mixtures of maritime [52], continental [57] and H_2SO_4 [57] aerosols which

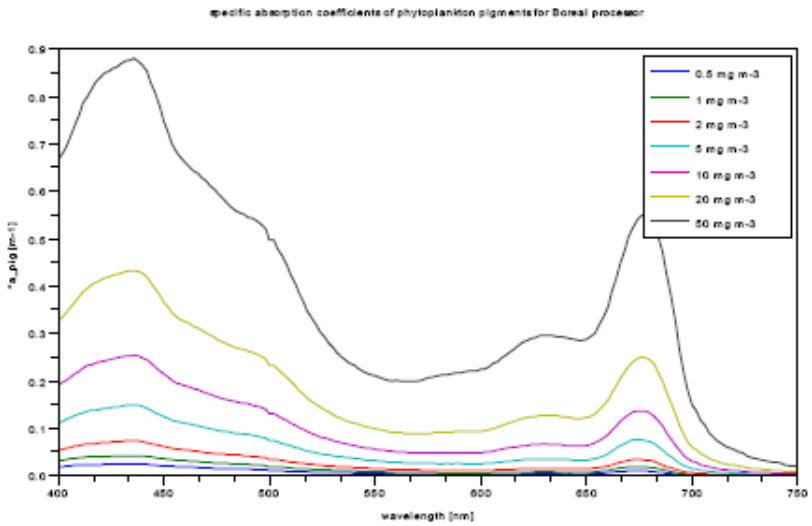


Figure 11: Absorption spectra of phytoplankton pigments of Finnish lakes [32] used in the Boreal Lakes Processor [42].

properties were derived from Mie calculations according to [2]. See Table 9

(b) A rough sea surface characterized by wind speeds of 1.5 and 7.2m/s according to [10].

(c) Residual gas absorption accounted with a modified version of the k-distribution method [5].

(d) Surface air pressure of 980, 1013 and 1040hPa were considered.

(e) A non-black ocean model, including 3 chlorophyll concentrations of 0.03, 0.3 and 3mg/m^3. Absorption coefficient of phytoplankton is computed according to [7], while its scattering coefficient is taken from [21]. The coloured dissolved organic matter is assumed to be totally absorbing according to [7]. Finally the scattering phase function of marine particles is taken from [37].

(f) Constant ozone content of 344DU.

The output of atmospheric correction algorithm is the radiance reflectances at mean sea level (MSL) for band 1-7 and 9 with additional Aerosol Optical Thickness (AOT) at 440, 550, 670 and 870 nm computed to be compared with coincident in situ measurements of AERONET from Helgoland island for validation purposes.

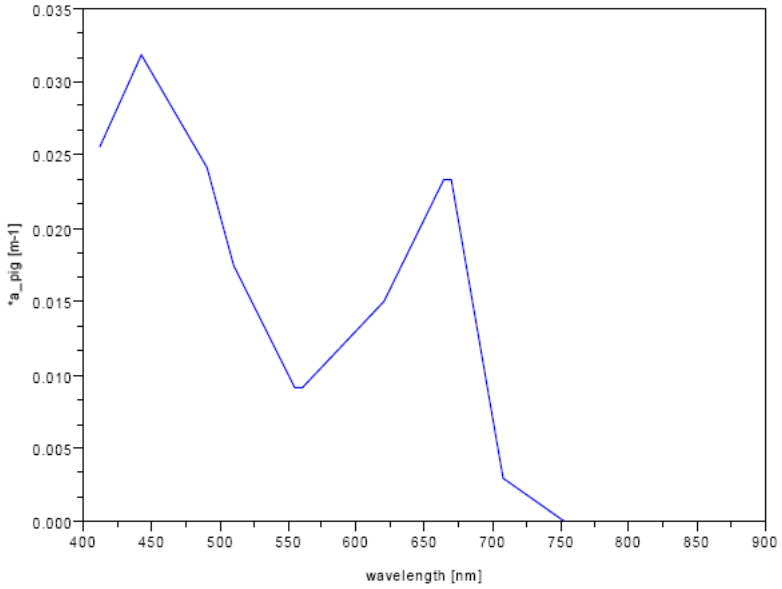


Figure 12: Absorption spectra of phytoplankton pigments of Spanish lakes [44] used in the Eutrophic Lakes Processor [42].

Aerosol Model	R(%)	W-S(%)	D-L(%)	Soot(%)	H_2SO_4 (%)	Oceanic (%)
Maritime	10	0	0	0	0	90
Continental	0	29	70	1	0	0
Stratospheric	0	0	0	0	100	0

Table 9: Composition of the different aerosols models used in the FUB/Wew atmospheric correction. Adopted from [52, 57]. Components: Rural (R), Water-Soluble (W-S), Dust-Like (D-L), Soot, H_2SO_4 and Oceanic.

Water Algorithm

Unlike in the other processors, the retrieval process of the water constituents (i.e. Chlorophyll a concentration (CHL), Total Suspended Matter concentration (TSM) and the absorption coefficient of yellow substance (YEL)) do not use the water leaving radiance reflectances estimated in the atmospheric correction algorithm explained in the previous section. Instead, an independent Neural Network for each constituent is trained with 100000 simulated vectors of the forward model, and it estimate the concentrations directly from the collected radiances at the top of the atmosphere (See Table 10).

Neural Network	Input	Hidden	Output	Function
run19_c2_080_nn.pro	19	80	12	Atm. Corr. NN
run38_c2_040_nn.pro	19	40	1	YEL NN
run39_c2_080_nn.pro	19	80	1	TSM NN
run46_c2_100_nn.pro	19	100	1	CHL NN

Table 10: Description of the number of nodes of the different 3-layer Neural Networks used in the FUB/Wew Water algorithm. YEL: Absorption coefficient of yellow substance; TSM: Total Suspended Matter concentration; CHL: Chlorophyll a concentration. The 12 output nodes of the Atmospheric correction NN are the 8 water leaving radiance reflectances and the 4 aerosol optical thicknesses.

3.2.5 ICOL processor

Improved Contrast between Ocean and Land (ICOL) is a processor developed to correct the Adjacency Effect (See Section 2.1.3). A brief description is given here, for further information read [47]. The steps followed by the algorithm are:

1. A *preparation* module transforms TOA radiance into TOA reflectance after correction of the gaseous transmittance according to [11].
2. A *Rayleigh* module corrects the Rayleigh scattering for each pixel following [53].
3. The *Rayleigh* component of the Adjacency Effect is corrected in the vicinity of land ($d < 30km$) according to [53].
4. An aerosol model is determined over the pixels. It is a Junge model implemented in *Meris Ground Segment data processing prototype* (MEGS).
5. The *Aerosol* component of the Adjacency Effect is corrected.
6. Output product is prepared by converting the reflectance again in radiances of Level 1B, the same as the input.

Some of the limitations of the current ICOL algorithm are:

1. Sun glint biases the retrieval of the surface reflectance and therefore it impacts on the adjacency effect correction.
2. Barometric pressure and aerosol model are the same in windows of $60km * 60km$, which means a relatively low resolution to evaluate the Adjacency Effect.

3. The vertical distribution of the aerosol model should be improved.

4. The adjacency effect for the aerosols it is computed for a nadir view only. Nevertheless, the algorithm applies it for off nadir views.

3.3 Flags

The flags are an output product of all processors which value shows information obtained during the processing. A non zero value for a pixel indicates that something went wrong when processing that particular pixel. They can therefore be used to exclude poor quality pixels from comparison and analyses.

In the processors studied, there exist two versions of flag products: $l1\_flags$ and $l2\_flags$. While the Level 1 flags come with MERIS data, the Level 2 flags are made by the processors and give more qualitative and reliable information.

All the processors implement $l1\_flags$ while $l2\_flags$ is not implemented by FUB/Wew processor. In Figure 13a and Figure 13b they are listed respectively.

It is important to note that when more than one flag must be applied, the output number while be the sum of those. In a binary representation, the application of a flag $N = 2^n$ means to change the value of the bit n of the flags vector from 0 to 1.

Name	Value	...	... Description
COSMETIC	1	int32	Pixel is cosmetic
DUPLICATED	2	int32	Pixel has been duplicated (filled in)
GLINT_RISK	4	int32	Pixel has glint risk
SUSPECT	8	int32	Pixel is suspect
LAND_OCEAN	16	int32	Pixel is over land, not ocean
BRIGHT	32	int32	Pixel is bright
COASTLINE	64	int32	Pixel is part of a coastline
INVALID	128	int32	Pixel is invalid

(a) l1_flags

Name	Value	Type	Description
RAD_ERR	1	int32	TOAR out of valid range
LAND	2	int32	land pixels
CLOUD_ICE	4	int32	cloud or ice
SUNGLINT	8	int32	sunglint risk
ANCIL	16	int32	missing/OOR auxiliary data
TOSA_OOR	32	int32	TOSA out of range
WLR_OOR	64	int32	WLR out of scope
SOLZEN	128	int32	large solar zenith angle
SATZEN	256	int32	large spacecraft zenith angle
ATC_OOR	512	int32	atmos. correct. out of range
CONC_OOR	1024	int32	concentration out of training range
OOTR	2048	int32	RLw out of training range
WHITECAPS	4096	int32	Whitecaps pixels
FIT_FAILED	8192	int32	fit failed
SPAREFLAG06	16384	int32	spare flag 06
SPAREFLAG07	32768	int32	spare flag 07
INVALID	65536	int32	not valid

(b) l2_flags

Figure 13: Description of the flags included in the l1_flags and l2_flags products

3.4 Comparison

3.4.1 Data

MERIS data

The MERIS data processed in this work to obtain water constituents estimations was *Level 1b Full Resolution*. Its properties are detailed in the Table 11. The specific dates were chosen taking into account the amount of clouds and the *in situ* data available (See Table 12).

In situ data

The following sources of *in situ* data have been used to analyse and compare the processors:

Name	Level 1b Full Resolution
Identifier	MER_FR__1P
Product Level	1b
Description	TOA radiance for the 15 MERIS bands at full resolution, calibrated, geo-located, annotated with Product Confidence Data, calibration data, classification flags, and environment parameters. The user will be able to order adjacent scenes covering the full MERIS swath and any number of adjacent scenes pairs leading up to the complete MERIS orbit of 17500 km, depending on the availability of the Full Resolution data. Radiometric and geometric continuity is guarantied between adjacent scenes.
File Size	2241 lines x 2241 pixels (scene)
Pixel Spacing	approximately 300 m x 300 m (along-track x across track)
Coverage	approximately 575 km x 575 km (scene), 296 x 296 km ("imagette")
Bits/Pixel	16
Unit	$10^{-3}Wm^{-2}sr^{-1}nm^{-1}$
Accuracy	From 400 to 900 nm $< 2\%$, From 900 to 1050 nm $< 5\%$

Table 11: MERIS Level 1B product characteristics. [15].

1. 3155 in situ samples of *chlorophyll a* and *turbidity* taken between May 17th, 2006 and August 13rd, 2008 at different measurement stations.

2. 5170 in situ samples of *Chlorophyll a*, *TSM*, *turbidity* and absorption at 400nm (*a400*) taken by a cruise on June 6th, 2007. The location was Länsisatama - Vuosaari (near Helsinki) and the measurements were made by Luode and the laboratory of SYKE. The instrument used was Ac9. The route followed by the instrument is shown in Figure 14.

3. 190 in situ samples taken between 2006 and 2008 of Radiances at 413nm, 441nm, 491nm, 530nm, 555nm, 668nm, 870nm, and 1019nm. All the samples belong to the Helsinki Lighthouse located at 59°56'56"N 24°55'34"E. They are provided by Aerosol Robotic Network (AERONET)

3.4.2 Tools developed

One of the main goals of the current thesis was the development of tools that help in the analysis and comparison of the MERIS processors with

Date requested	In situ data available	Images received
2006		
May 8th	X	2
May 9th	X	2
May 16th	X	0
June 13rd	X	1
June 19th	X	1
July 25th	X	1
September 11st	X	0 *Rec: 12nd Sep*
2007		
June 4th		1
June 5th		1
June 8th		1
July 2nd	X	2
July 3rd	X	3
July 16th	X	3
August 7th	X	2
August 9th	X	2
September 5th	X	2
2008		
April 20th		1
April 23th		1

Table 12: MERIS data used in the analysis.

large quantities of *in situ* and satellite data. Actually, these tools are meant to be used in future and consist of:

1. MERIS data Matlab processing tool
2. Thematic map tool
3. Data extraction tools

A brief description of each one is given here.

MERIS data Matlab processing tool

This tool was developed in Matlab to process MERIS data as automatically as possible with the four processors studied in this work (Case 2 Re-

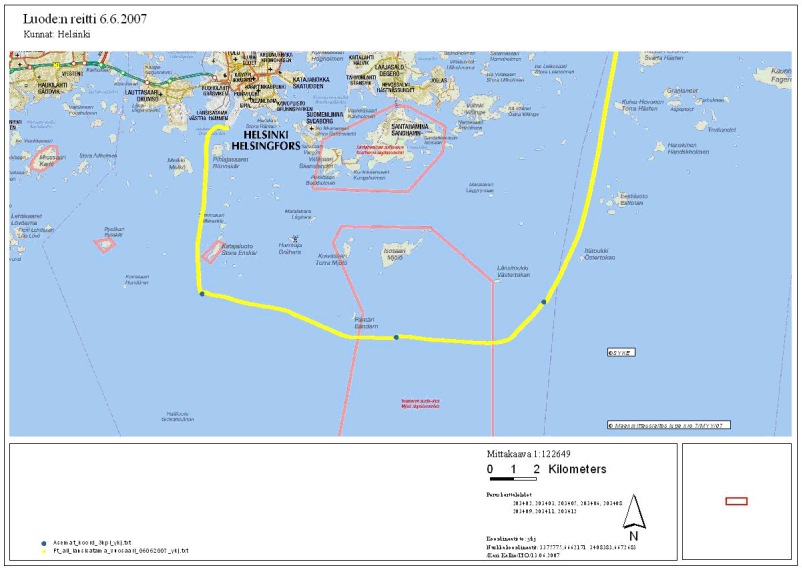


Figure 14: Flow through route followed by Luode on June 6h, 2007.

gional, Boreal Lakes, Eutrophic Lakes and FUB/Wew Water Processor). It also combines all these processors with the ICOL processor, so each input subset generates 8 different output processed with or without ICOL. For further information see the Appendix A.1.

Thematic map processor

This tool automatically generates thematic maps of the MERIS data processed with the *MERIS data Matlab processing tool* and helps to compare them by classifying the maps by date. Thus, it shows plots of 4 maps in each image with the corresponding parameter estimation map for each process with and without ICOL. Therefore, 6 different plots are created for the same date: chl a, TSM and aCDOM, with and without ICOL. An example can be seen in the Figure 15. By default, flags over clouds and land pixels are applied so that they appear in grey. For further information see the Appendix A.2.

Data extraction tool

This tool is developed to retrieve those pixels from MERIS products that match with *in situ* data provided in a text template. This tool uses the API of BEAM to compute pixel indexes from geographical coordinates and the

other way round and therefore it needs the program to be installed in the computer. This tool includes two versions:

1. Station tool (for multiple dates of in situ and MERIS data)
2. Transect tool (many data points in a single MERIS image)

Station tool: it looks for matching dates between all the MERIS products processed and the *in situ* data provided. Once a product is found, the tool checks if the geolocation of the *in situ* data is inside the borders of the map. In case it is, the tool extract the value of all the products available from a $3*3$ pixel area around the station. Although it is implemented with MATLAB, a JAVA script has been created to use the BEAM API in order to compute pixels given geolocations and vice versa. For further information see the Appendix A.3.

Transect Tool: it looks for product values in a single MERIS product at given geocordinates provided by the in situ template. It look for single pixels without averaging. Although it is implemented with MATLAB, a JAVA script has been created to use the BEAM API in the same way that the Station tool. Indeed, in this case two output text files are created with the pixels and coordinates extracted which makes the processing much faster and allow later manipulation of the data such as averaging flow trhough data that falls inside single pixels. For further information see the Appendix A.1.

3.4.3 Validation Methods

In this section, the process followed to obtain all the results presented in Section 4 are described.

1. Selection of suitable dates according to cloud cover. The website of SYKE was consulted in order to find 10-15 dates between 2006 and 2008 where previews would show clear sky. In addition, priority was given to those dates when *in situ* data from measure station was available.
2. Subsetting of the MERIS products. In order to avoid processing land areas, BEAM was used to cut the images selected. A total of 45 subsets were created including dataset from the Gulf of Finland and the Gulf of Bothnia.

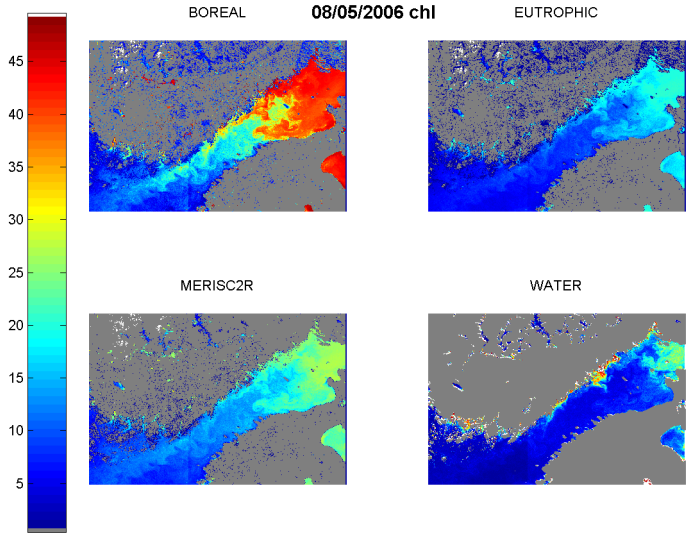


Figure 15: Plot created by the Thematic map processor.

3. Processing of the subsets using the MERIS Matlab Processing Tool. Starting from 45 original subsets, the tool processes 45 ICOL processed subsets. Next, it processes each subset with the four water processors. Therefore it creates a total of 360 subsets.

4. Thematic maps are created using the Thematic map processor. This tool creates 6 plots for each date. Those are: chlorophyll a, chlorophyll a (ICOL), TSM, TSM (ICOL), aCDOM and aCDOM (ICOL). In a single plot there are 4 maps, each one being the parameter processed with one of the processors. Thus, this tool will create a total of 270 .png files.

5. *In situ* data template is created. Depending on the source, *in situ* data provided is stored properly in a text file following the format established in the Extraction Tool.

6. Station Tool is executed. It stores in a .mat file several matrices with all the extracted products of the matching MERIS data as well as the *in situ* data that were suitable.

7. Transect Tool is executed. A single subset that matches the date of the flow through data provided is selected. The tool stores in a .mat file the extracted data.

8. Several MATLAB scripts are created to plot the data in different ways such as scatter plots, transect plots or temporal evolution plots.

9. Some quality descriptors are computed such as the Root Mean Square Error (RMSE) or the square of the coefficient of determination (R^2). The first is defined as follows:

$$RMSE = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(X_{i,meris} - X_{i,insitu})^2} \tag{81}$$

and the coefficient of determination is defined as the square of:

$$R_{xy} = \frac{\sum_{i=1}^{i=N}(X_{i,meris} - \overline{X_{meris}})(X_{i,insitu} - \overline{X_{insitu}})}{(N-1)\sigma_{X_{meris}}\sigma_{X_{insitu}}} \tag{82}$$

where N is the number of samples, $X_{i,meris}$ is a single estimated value, $X_{i,insitu}$ is a single *insitu* value and σ_x is the standard deviation of the variable X.

4 Results

In this section, the most important results obtained following the procedure described in Section 3.4.3 are presented. First, different estimated values using different processors are compared against *in situ* data. To do this, several scatter plots have been made. These can be separated into two groups:

1. *In situ* chlorophyll a (chl a) concentrations from several measurement stations as described in Section 3.4.1. Here, four scatter plots representing the results in each different processor (See Section 3.2) with and without ICOL are shown.

2. *In situ* data from flow through measurements as described in Section 3.4.1 Chl a, TSS, turbidity and absorption at 400nm. Eight scatter plots are made to compare suitable MERIS processors estimations with the measurements: four with ICOL preprocessing and four without it.

Second, the dynamics of seasonal estimation are studied and compared. These results are very important since provide a lot of information on the biological processes that take place in water. Therefore, *chl a* averaged estimations made by each of the four processors along all the year are plotted in four different graphs. Next, three dates are chosen to represent the most important phases in the chlorophyll concentrations, that is, the *spring bloom*, the *summer minimum* and the *late summer bloom*, according to the results obtained before.

Thematic maps of *chl a*, TSM and aCDOM are shown for all the processors, with and without ICOl for the three selected dates. Only thematic maps of the Gulf of Finland are shown, since showing maps from the Gulf of Bothnia would not add more information about the processors behaviour.

Moreover, transect plots are shown to compare these three parameters at the different days of the year. However, only CR2 and FUB/Wew are compared because they are the most different processors.

After this, the effects of ICOL preprocessing are analyzed. One more transect plot is presented to compare the chlorophyll a (chl a) measured the same day by all the processors with and without ICOL. Next, the effect of ICOL in estimated water-leaving reflectance is shown for every single band, in a transect fragment, next to the shore. In addition, the averaged effect in every single band along all the transect is also plotted.

And finally, water-leaving reflectance spectra are compared with *in situ* data provided by AERONET. Comparisons are made for the same three days. AERONET measurements at different hours of each day are provided. However remote sensing radiance reflectances had to be indirectly computed from radiances, which is the magnitude measured (The expression applied is the Eq. 84).

4.1 Scatter Plots

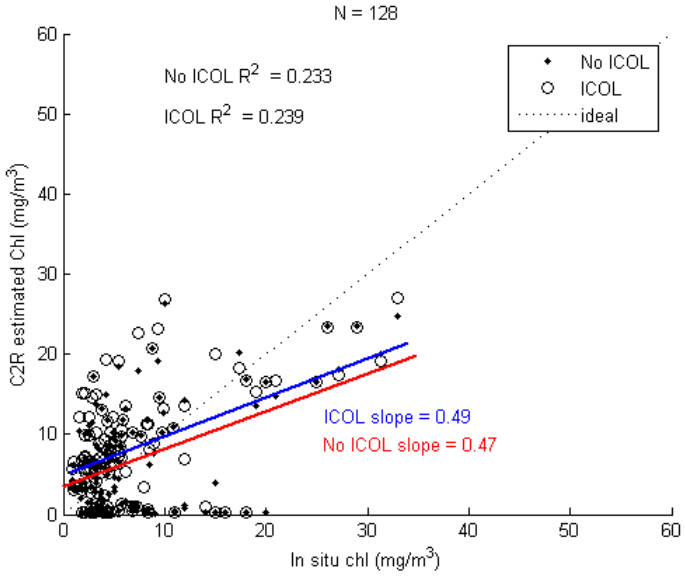
An overall summary of the results obtained with the scatter plots is presented in the Table 13. However, first of all we are going to analyse the behaviour of each processor individually.

In Figure 16a, we can see the error produced by the C2R when estimating *chl a*. A general underestimation of the processor is observed for concentrations over $20mg/m^3$ (*in situ* value), all 6 non-ICOL samples are estimated below their real value. However, in the range between $20 - 30mg/m^3$ the relative error seems to be maintained and the trend is quite correctly defined.

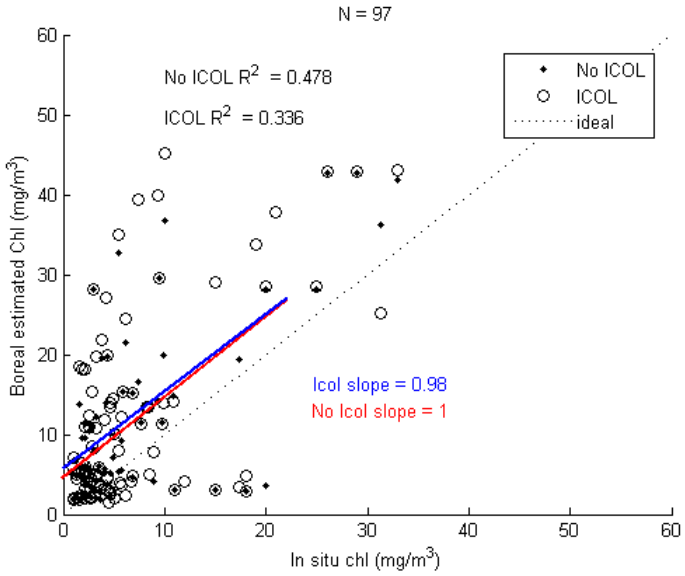
In the range between $10 - 20mg/m^3$ an interesting effect can be observed, which is that two clearly defined groups of samples are formed. There is a group where estimated values are very near to zero while the rest of values keep the behaviour of the higher range mentioned before.

Finally, in the range between $0 - 10mg/m^3$, the processor behaviour shows a less defined pattern. Estimated samples are spread through the $0 - 30mg/m^3$ range. Although the majority of points are close to the real value, the inaccuracy is still high. A group of samples at the bottom of the plot that were given almost zero estimated values while they actually range uniformly from 0 to 10 mg/m^3 can be clearly seen.

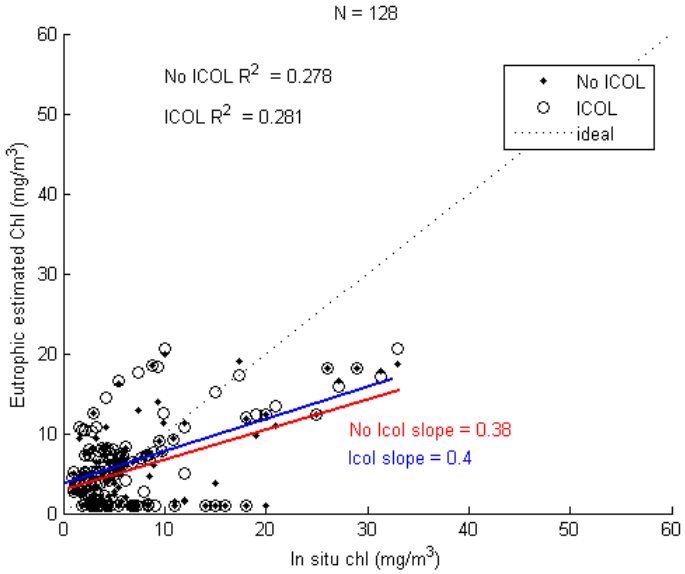
Similar correlation values, around 0.23, are obtained by the comparisons (with ICOL and without ICOL) and there are no significant changes in the general behaviour over the ranges, which can be deduced from their similar slopes (0.49 and 0.47). Only a small correction is observed in low range values where highest values are predominantly preprocessed while the lowest ones are mostly not preprocessed.



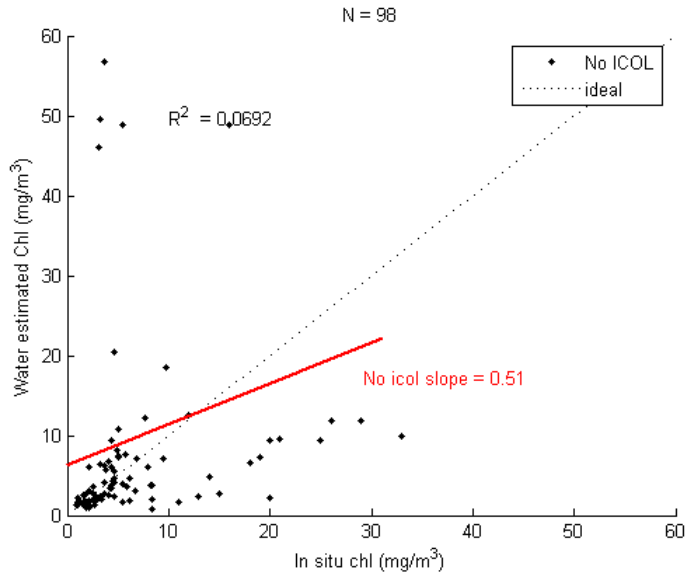
(a) Coastal Case 2 Regional Water Processor (See Section 3.2.1).



(b) Boreal Lakes Water Processor (See Section 3.2.2).



(c) Eutrophic Lakes Water Processor (See Section 3.2.3).



(d) FUB/Wew Water Processor (See Section 3.2.4).

Figure 16: Scatter plot where MERIS estimated values of Chlorophyll a (mg/m^3) are compared with *in situ* data measured between 2006 and 2008 (look at Section 3.4.1 for details) along the Baltic Sea. The MERIS data are here processed with all four processors with and without the ICOL preprocessing (See Section 3.2.5). The dotted line shows what the ideal results would be. The blue and red lines show the trend line of the estimated value and the value of the slope. N is the number of matching samples and R^2 is the square of the coefficient of determination.

Processor	R^2	RMSE (mg/m^3)	N
Coastal Case 2 Regional	0.233	6.6788	128
Coastal Case 2 Regional (ICOL)	0.239	6.7611	128
Boreal Lakes	0.478	8.9662	97
Boreal Lakes (ICOL)	0.336	10.9965	97
Eutrophic Lakes	0.278	6.0887	128
Eutrophic Lakes (ICOL)	0.281	5.7914	128
FUB/Wew Water	0.0692	18.3032	98

Table 13: Summary of the coefficient of determination R^2 and root medium squared error (RMSE), obtained for the different processors for a number of samples N when estimating Chlorophyll a (chl a)

In Figure 16b the Boreal Lakes Processor shows a quite different behaviour. Here, the general trend is an overvaluation of the estimated chlorophyll. This can be clearly noticed for ranges between $20 - 40mg/m^3$ where only one sample is underestimated while the other eight are above the ideal line. In lower ranges, the presence of almost zero estimations is much lower showing better performance than the Case 2 Regional in this particular aspect. Although the inaccuracy range for the $0 - 10mg/m^3$ measurements is higher than in the previous figure, ranging from 0 to over $45mg/m^3$, a much better result in the lower estimations can be seen.

There is a significant difference between ICOL samples and non ICOL samples in this case. At low ranges it is very clear that non ICOL estimations behave much better, while at high ranges these differences are less important. The overall correlation gives a better result for non ICOL estimations. It is interesting to notice that the slopes are almost exactly correct which denotes a very homogenous performance along the wide range of the measurements. That is systematically biased around $5mg/m^3$ over the ideal pattern though.

The Eutrophic Lakes water processor's scatter plot is shown in the Figure 16c. Since the concentrations training range (flag 1024, see Section 3.3) is too small for the measurements available (See Table 8), most of the pixels have been flagged. Hence, C2R flags were used instead.

In a similar way with the previous ones, it shows an overestimation for ICOL processed samples of low real values. In the range over $20mg/m^3$ the processor clearly underestimated the concentrations.

Three groups of samples can be distinguished in this scatter plot. The first one, is the group of samples that were given almost zero estimated concentrations in a similar way like the Case 2 Regional Processor did.

The second one, is the group of samples that show a quite consistent but too flat trend (low slope of 0.4) along the whole range. Finally, the third group is formed by the overvaluated ones at low ranges which are mostly ICOL preprocessed.

In Figure 16d, the results produced by the FUB/Wew Water Processor are shown. ICOL preprocessed results are not shown because of the anomalous behaviour they present (It is important to note that in all the comparisons made in the whole section, FUB/Wew estimations preprocessed by ICOL have been excluded from the analysis since they are out of range). Although the overall coefficient of determination is much lower (about five times) than the other processors coefficient, it estimates in a more sound way the low range concentrations. Therefore, between $0 - 10mg/m^3$ a great proportion of the samples trace a good trend around the ideal line. This trend is not maintained though for higher concentrations. An underestimation of the parameter in the range between $10 - 30mg/m^3$ can be clearly appreciated, which is reflected in the low slope obtained (0.51). Finally, a non insignificant quantity of samples is out of range with estimations over $40mg/m^3$. This fact undermines the whole coefficient of determination as mentioned before.

In figure 17, the estimated chlorophyll concentrations with all the processors are compared against *in situ* data provided by through flow measurements. Since MERIS resolution is much lower (around $300m$) than the distance between consecutive *in situ* samples, they have been properly averaged, that is, the mean value is computed for all the points geolocated within the same pixel. Thus, in this scatter plot we analyse again the general errors produced in MERIS estimations. The first impression is that a general overestimations of the chlorophyll a (chl a) is produced. While Case 2 Regional and Eutrophic Lakes show a good performance, if biased, between $6-9mg/m^3$ *in situ* concentrations, at lower concentrations $1 - 3mg/m^3$ any trend can be deduced. Indeed, in the case of the boreal lakes processor, it shows a completely vertical profile between 2 and 10 mg/m^3 at the previously mentioned range.

At this range, FUB/Wew water processor yields the best result for low concentrations, while Case 2 Regional concentrations are the most overestimated. In general, the best behaviour is produced by Eutrophic Lakes, which coefficient of determination ($R^2 = 0.568$) is well over the one achieved by the others.

ICOL processing does not improve the coefficients of determination in any case. However, a subtle correction of the Case 2 Regional processor in the range $6 - 8mg/m^3$ can be observed, bringing the estimations closer to those of the Eutrophic processor.

In TSM estimations (See Figure 18), high correlations are achieved. However, there is a general overestimation of the concentration. This is specially the case for Case 2 Regional and Eutrophic processor because of their wider range of training. However, in the range where all the processors are present, that is, between $1 - 2g/m^3$, both Boreal and FUB/Wew Water processors show very accurate results over the ideal line. When processing Boreal with ICOL, more samples are non-flagged (from N = 343, to N = 409) increasing its coefficient of determination significantly (from $R^2 = 0.343$ to $R^2 = 0.795$)

Similar results appear when comparing estimated suspended matter with measured turbidity (See Figure 19). Although they have different units, high coefficients of determination are found in all the processors. Thus, the lowest value is produced by the FUB/Wew Water processor ($R^2 = 0.746$) and the highest is given by the Case 2 Regional with ICOL ($R^2 = 0.895$).

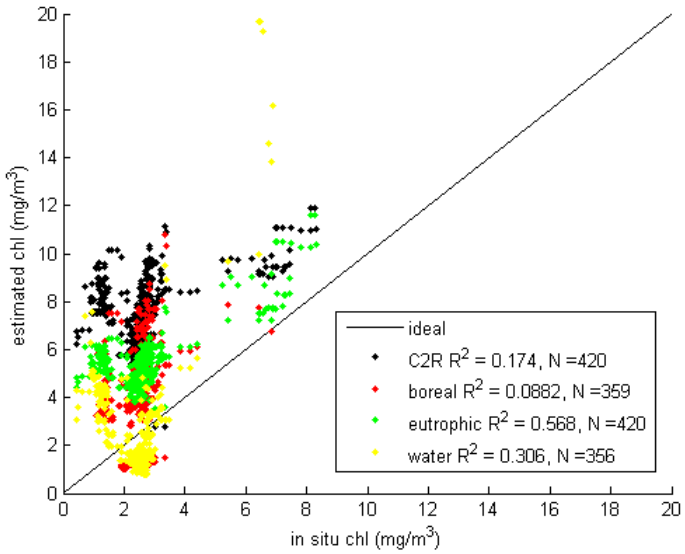
At low ranges, the same pattern shown in the case of TSM is reproduced. The highest values are given by the Eutrophic processor, while Boreal and FUB/Wew Water processor yield the lowest estimations. In the same way as before, only Case 2 Regional and Eutrophic provide results in the full range displayed.

Finally, we use the measured absorption at 400nm to compute a theoretical absorption at 442nm by means of the next expression:

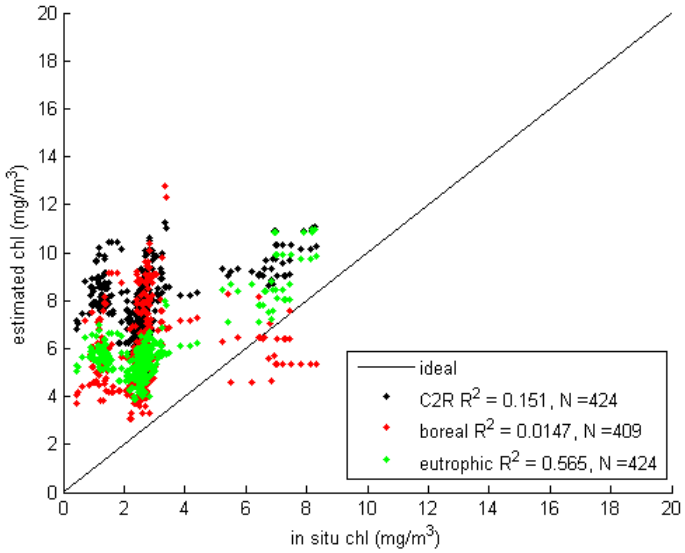
$$a(440nm) = 0.55 * a(400nm) \tag{83}$$

These values are compared with the total absorption output at 442nm provided by the MERIS processors (See Figure 20). In the case of FUB/Wew there is no output for the total absorption, thus, the aCDOM at 442 nm has been used instead. It can be observed a subtle underestimation in the case of boreal and case 2 regional processor.

It is interesting to note how ICOL improves significantly the results compressing the values around the ideal line. Thus, overestimations produced by the eutrophic lakes processors are pushed down while boreal lakes and case 2 regional underestimations are slightly pushed up. In general, compared to the others, processors seem to behave quite accurately when estimating this parameter. However a non linear trend is observed along the range of measurements.

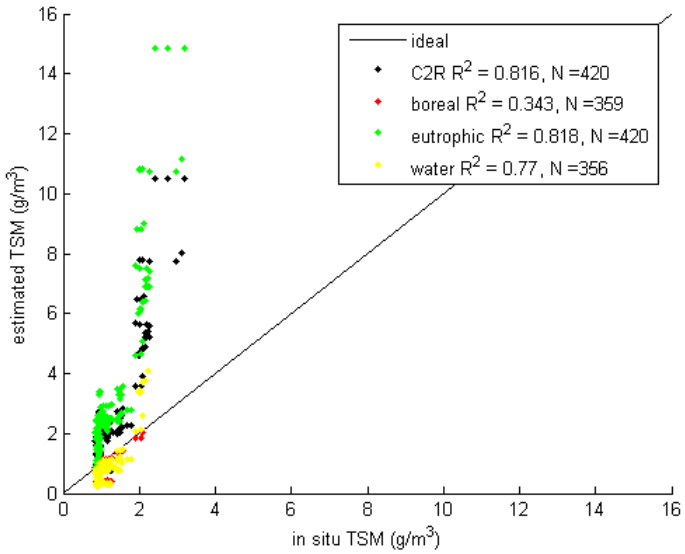


(a) Without ICOL.

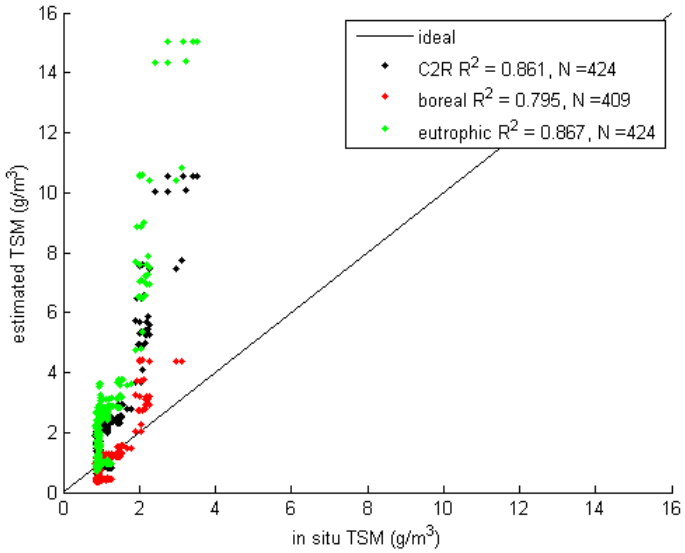


(b) With ICOL.

Figure 17: Scatter plot where MERIS estimated values of Chlorophyll a (mg/m^3) on June 5th, 2007 are compared with *in situ* data measured with the flow through instrument Ac9 on June 6th, 2007 (look at Section 3.4.1 for details) in Lansisatama-Vuosaari. The MERIS data are here processed with all the Processors (See Legend). The line shows what the ideal results would be. N is the number of matching samples shown and R^2 is the square of the coefficient of determination.

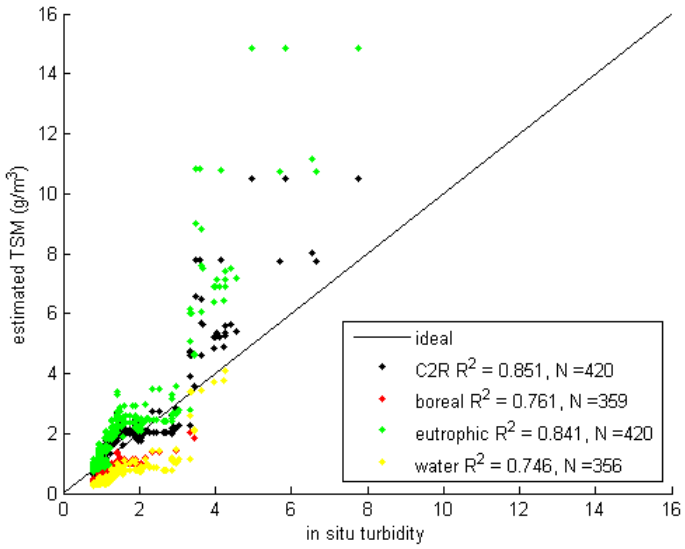


(a) Without ICOL.

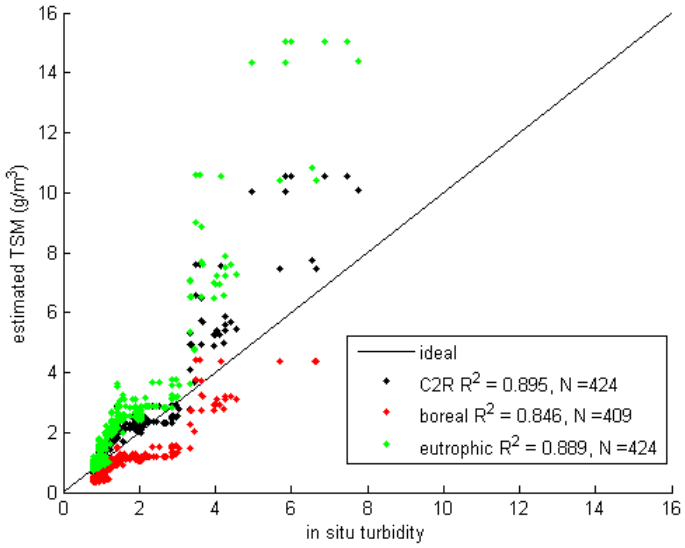


(b) With ICOL.

Figure 18: Scatter plot where MERIS estimated values of TSM (mg/m^3) on June 5th, 2007 are compared with *in situ* data (Total Suspended Solids) measured with the flow through instrument Ac9 on June 6th, 2007 (look at Section 3.4.1 for details) in Lansisatama-Vuosaari. The MERIS data are here processed with all the Processors (See Legend). The line shows what the ideal results would be. N is the number of matching samples shown and R^2 is the square of the coefficient of determination.

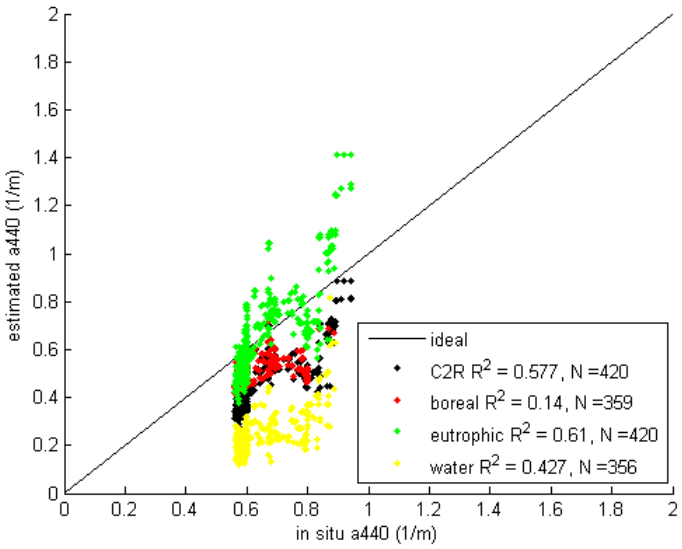


(a) Without ICOL.

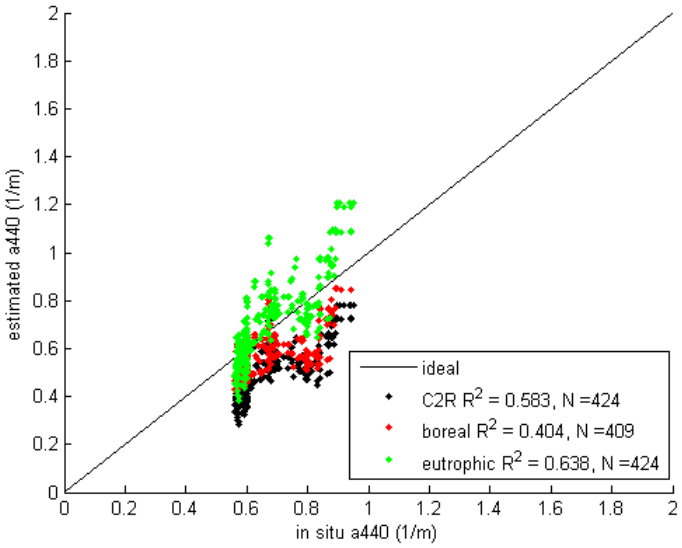


(b) With ICOL.

Figure 19: Scatter plot where MERIS estimated values of TSM (mg/m^3) on June 5th, 2007 are compared with *in situ* data (Turbidity) measured with the flow through instrument Ac9 on June 6th, 2007 (look at Section 3.4.1 for details) in Lansisatama-Vuosaari. The MERIS data are here processed with all the Processors (See Legend). The line shows what the ideal results would be ideally. N is the number of matching samples shown and R^2 is the square of the coefficient of determination.



(a) Without ICOL.



(b) With ICOL.

Figure 20: Scatter plot where MERIS estimated values of absorption at 442 nm ($1/m$) on June 5th, 2007 are compared with *in situ* data (absorption at 400nm) measured with the flow through instrument Ac9 on June 6th, 2007 (look at Section 3.4.1 for details) in Lansisatama-Vuosaari. The MERIS data are here processed with all the Processors (See Legend). The line shows what the ideal results would be. N is the number of matching samples shown and R^2 is the square of the coefficient of determination.

4.2 Evolution of the parameters

In Figure 21a the evolution of Chlorophyll concentration along the year is shown. Thus, the spring bloom, the summer minimmum and the late summer bloom can be clearly seen in the *in situ* data.

1. Spring bloom starts around the 100th Julian day and lasts until the 140th day of the year being the maximum around the day 120, which is about the end of April. The concentration raise to over $20mg/m^3$.

2. Summer minimum starts at the end of May and lasts until the end of July and the chlorophyll concentration remains under $10mg/m^3$. The minimum concentration is around $5mg/m^3$. However, it is not completely constant since it has a local maximum around the day 190.

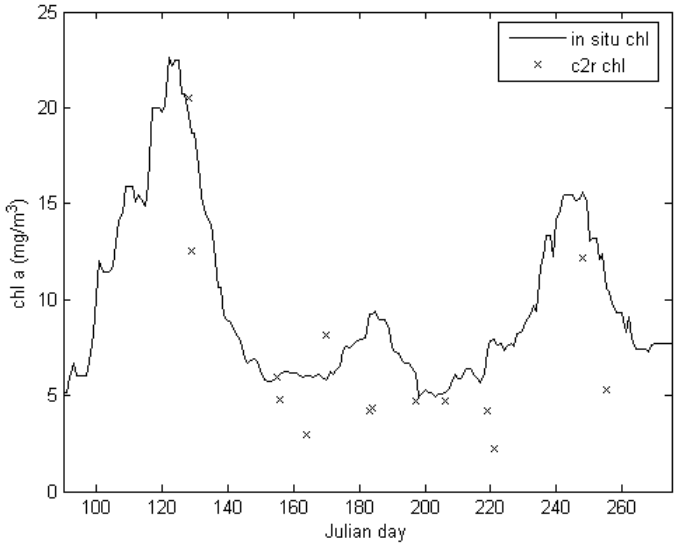
3. The late summer bloom is located between the first days of July and the end of August. Similarly to the spring bloom, it lasts around 40 days. However the concentrations are kept lower, reaching a maximum of approximately $15mg/m^3$.

As it is explained in the Figure 21a, crosses represent estimated MERIS data with the Case 2 Regional Water Processor. When there are several matches at the same location, they are averaged. Although the time sampling is not dense enough to deduce a possible interpolation of the estimated data evolution, it is clear that the estimation algorithm is sound enough to follow consistently the different highs and lows of the chlorophyll concentration. Thus, while in the spring bloom there is a sample well over $20mg/m^3$, during the minimum summer period, all the values are located under $10mg/m^3$.

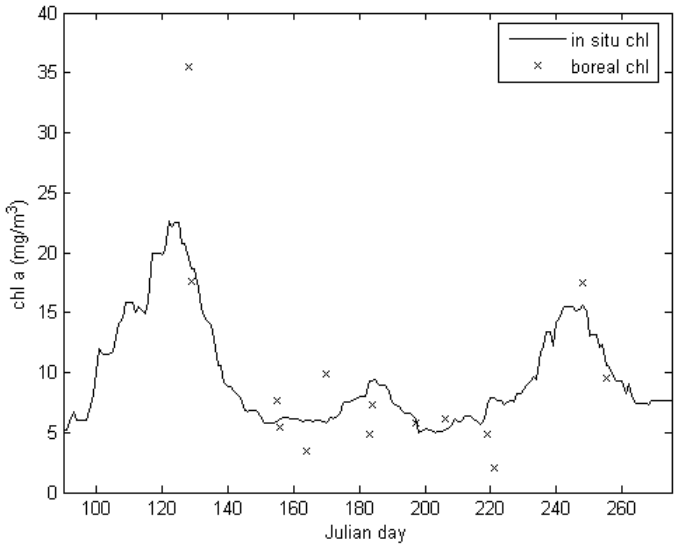
In general, a general underestimation along all the year is observed.

With the Boreal Lakes Water Processor an overestimation of the concentrations appears. However, these surpluses are specially obvious on blooms when a maximum value over $35mg/m^3$ is reached, $10mg/m^3$ well above of the real one. In the summer minimum period, estimations are fairly close to the average. Therefore, in the same way as the previous processor, it follows correctly the cyclic trends along the year.

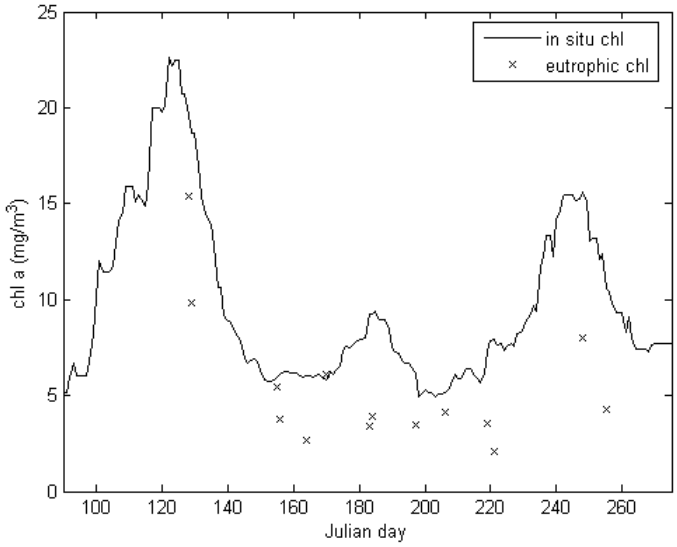
Eutrophic Lakes Water Processor behaves more similarly to the Case 2 Regional Water Processor. Thus, in general terms, concentration is underestimated both in high and low parts of the cycle. Nevertheless, the general dynamics of the parameter can be clearly observed.



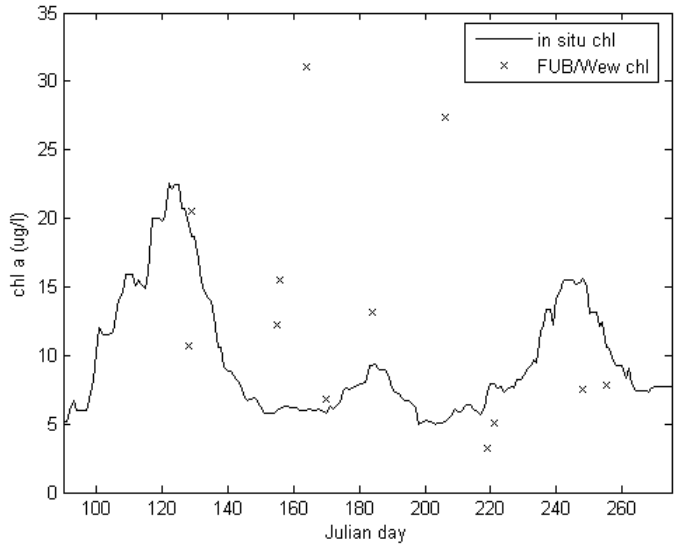
(a) Coastal Case 2 Regional Water Processor (See Section 3.2.1).



(b) Boreal Lakes Water Processor (See Section 3.2.2).



(c) Eutrophic Lakes Processor (See Section 3.2.3).



(d) FUB/Wew Water Processor (See Section 3.2.4).

Figure 21: The yearly cycle of variation of Chlorophyll a (mg/m^3) as a function of time (Julian Day) is represented in this plot from *in situ* data measured between 2006 and 2008 (look at Section 3.4.1 for details) in the Baltic Sea. A mean value has been computed when there were several measurements in the same day. In addition, a sliding window mean of ± 7 days has been applied. Finally, geographically matching MERIS estimated values of Chlorophyll a (mg/m^3) processed with all the processors under study are plotted. In the same way, the values have been averaged when there were several measurements in the same day.

Finally, FUB/Wew Water processor shows very different results. It is impossible to follow the trend since higher values are achieved during the summer minimum than in the algal bloom.

In order to analyse with more detail how the processors behave in different stages of the chlorophyll cycle, we have selected MERIS products of three dates that we consider representative of each one. These are:

1. 20th April,2008: Spring bloom.
2. 19th June,2006: Summer minimum.
3. 9th August,2007: Late summer bloom.

They are first compared using thematic maps. Right after this, a common transect path (24º01'51"E 59º54'52"N - 25º28'26"E 59º36'16"N) is plotted for each parameter in different dates.

In Figure 22, MERIS estimations for April 20th, 2008 are showed for all four processors with and without ICOL. The most clear difference between them is the average value of the chlorophyll. Thus, Boreal Lakes Processor gives the highest values (that reach up to $45mg/m^3$), and it is followed by Case 2 Regional and Eutrophic Lakes Processor, which look quite similar (with a maximum values around $20mg/m^3$, being Case 2 Regional a bit higher). Finally FUB/Wew processor shows the lowest results.

It is interesting to note however, that these differences are not uniformly distributed around all the area. The low concentrations areas (where the sea is more open) show a much more similar behaviour while the differences are large in the higher range, well inside the gulf. For instance, at the southwest part of the map, all pictures yield values around the same "blue" range (approximately between $5 - 10mg/m^3$) while the opposite occurs next to the Saint Petersburg shore, where both nominal values and even concentration distributions show different patterns.

No large differences are observed due to the effect of the ICOL preprocessing at this scale.

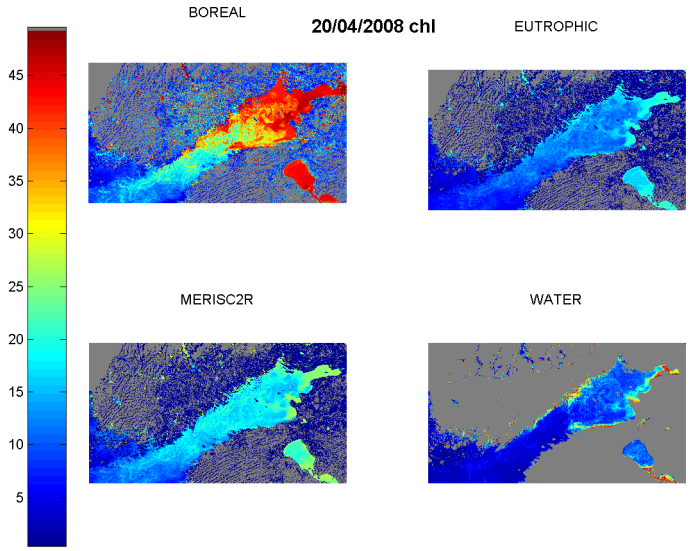
In Figure 23 the same sets of thematic maps for the TSM concentrations are presented. Here the behaviour is much more similar between processors. Higher concentrations are much more close to the land, while the rest of water present low concentration around $0 - 2g/m^3$. Hence, differences between processors are located also next to the shore. In this case, boreal lakes and case 2 regional processors show a lower effect offshore than eutrophic lakes, which maintain higher concentrations at distances far from the coast. On the other hand, FUB/Wew water processors yields a very low and homogeneous result along all the water surface. It is interesting

to note how these differences are reflected in the Lake Peipsi, at the southeast part of the map where FUB/Wew estimates a quite flat concentration while the others show a clear gradient from higher concentrations at its southern shore to lower concentration at the northern part of the lake.

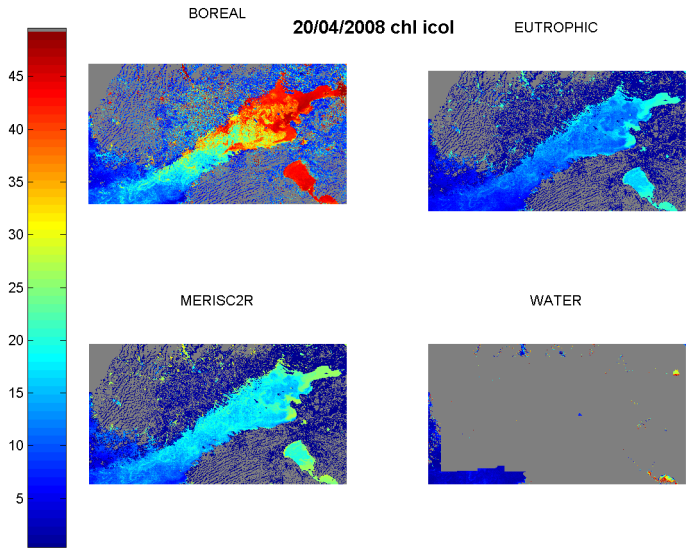
In the case of aCDOM, shown in Figure 24, the most visible differences arise when comparing those processors meant to be used in lakes which those meant to be used in coastal waters. Boreal and eutrophic lakes show a wider range of concentrations and more complex concentration distribution. This is especially clear in the case of boreal lakes, when compared for instance with the flat estimation produced by the FUB/Wew processor. Similarly to what happened with the other parameters, a common low estimation is shared in open waters, and its value is around $0.25 - 0.5m^{-1}$ in this case.

Although case 2 regional behavior is close to the FUB/Wew result, the maximum values reached in the Saint Petersburg coast are really close to those produced by the Eutrophic Lakes processor which gives to the case 2 regional processor a very interesting position between them.

In Figure 25 the summer minimum estimated concentrations of chlorophyll are displayed. The overall values are, as they should be, significantly lower than in the spring bloom. According to the previous behaviour, less differences between processors should be expected and it actually happens. However, they still arise well inside the gulf where higher concentrations are found. Unlike in the previous examples, here preprocessing with ICOL causes visible differences in the Lake Peipsi. It can be clearly seen when comparing this location with the boreal lakes processor how ICOL yields higher concentrations at the northernmost part of the lake. This effect can be seen as well when looking at Figure 26 where TSM concentrations are presented. ICOL preprocessing gives higher values to the result not only in the Lake Peipsi, but also in the Estonian coast of the Gulf of Finland. Within processors, differences are very subtle in most areas offshore but it can be still appreciated higher concentrations for eutrophic and case 2 regional processor than Boreal lakes or FUB/Wew water processors. Figures 28, 29 and 29, show the thematic maps for chlorophyll a, TSM and aCDOM, respectively, given at the summer bloom. In general, this time the processors behave in the same way that they did for the spring bloom.

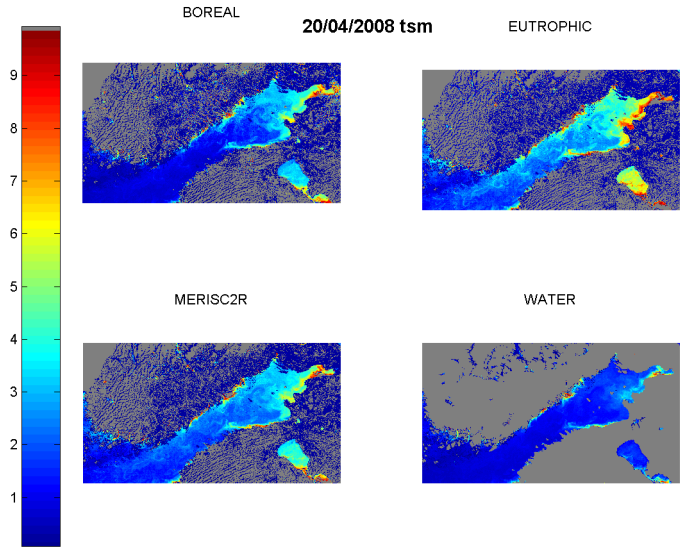


(a) Without ICOL.

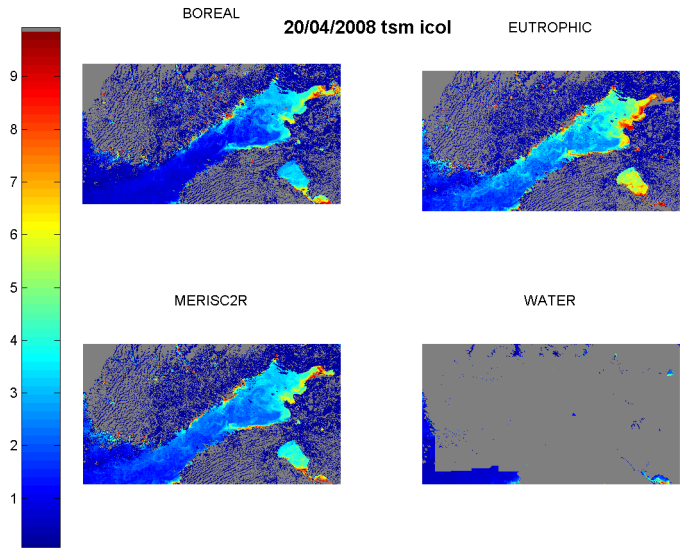


(b) With ICOL.

Figure 22: Spring bloom thematic map of the Gulf of Finland where estimated Chlorophyll a (mg/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

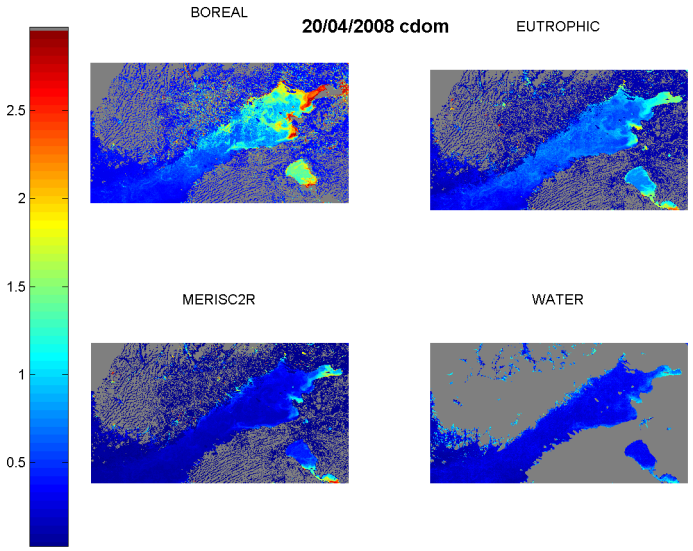


(a) Without ICOL.

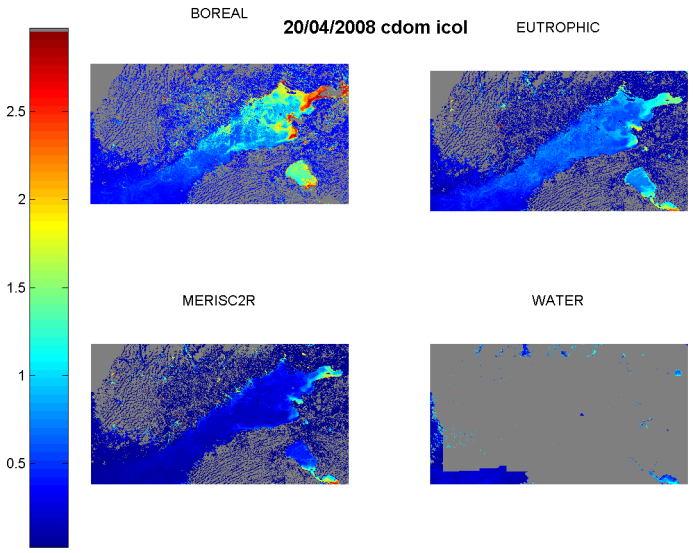


(b) With ICOL.

Figure 23: Spring bloom thematic map of the Gulf of Finland where estimated TSM (g/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

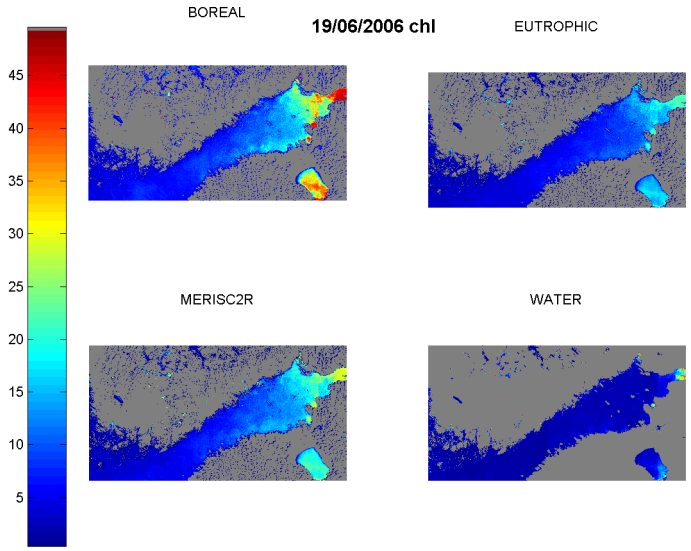


(a) Without ICOL.

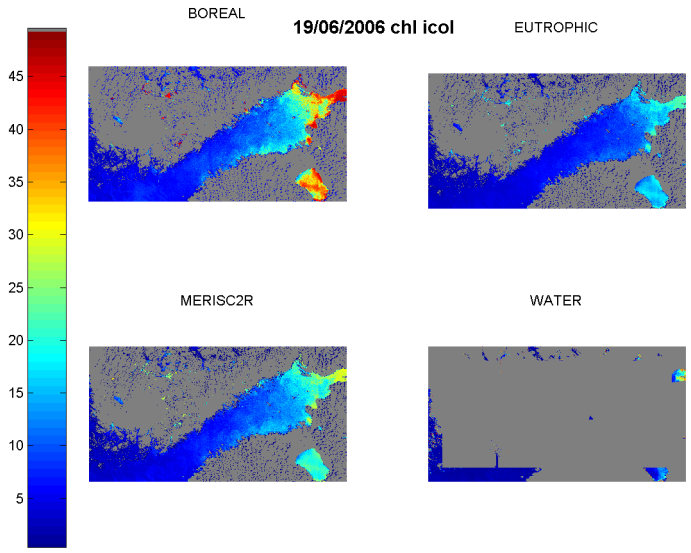


(b) With ICOL.

Figure 24: Spring bloom thematic map of the Gulf of Finland where estimated aCDOM ($1/m$) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

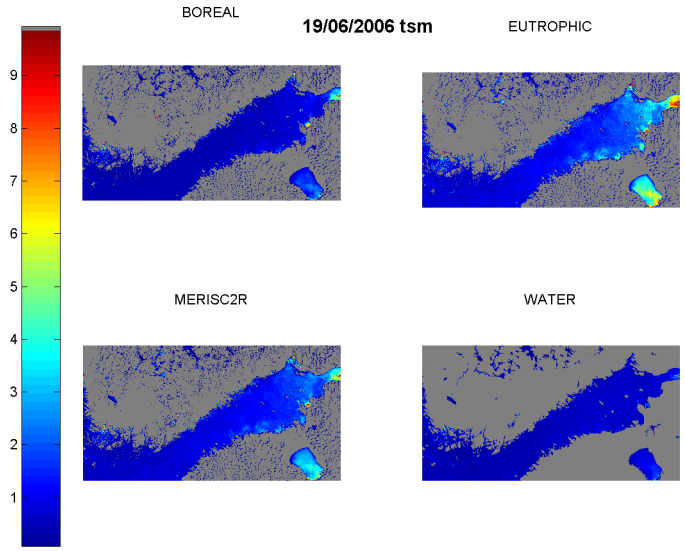


(a) Without ICOL.

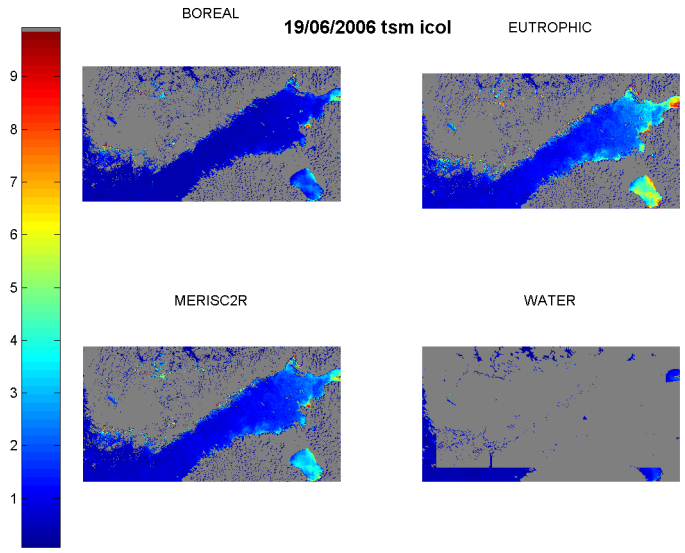


(b) With ICOL.

Figure 25: Summer minimum thematic map of the Gulf of Finland where estimated Chlorophyll a (mg/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

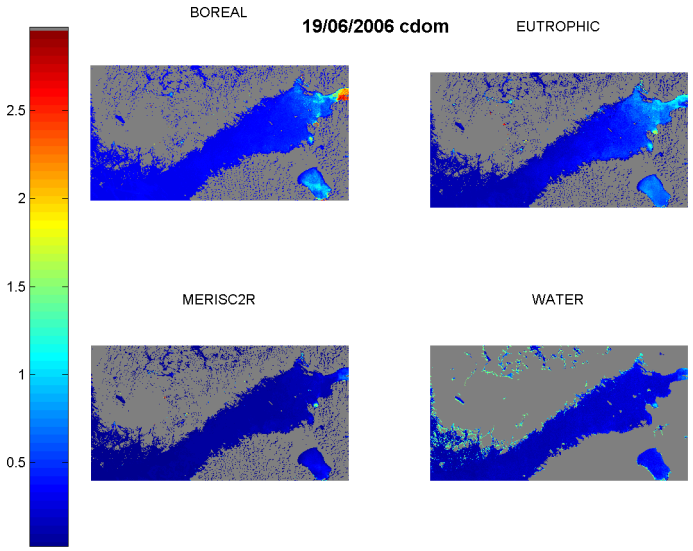


(a) Without ICOL.

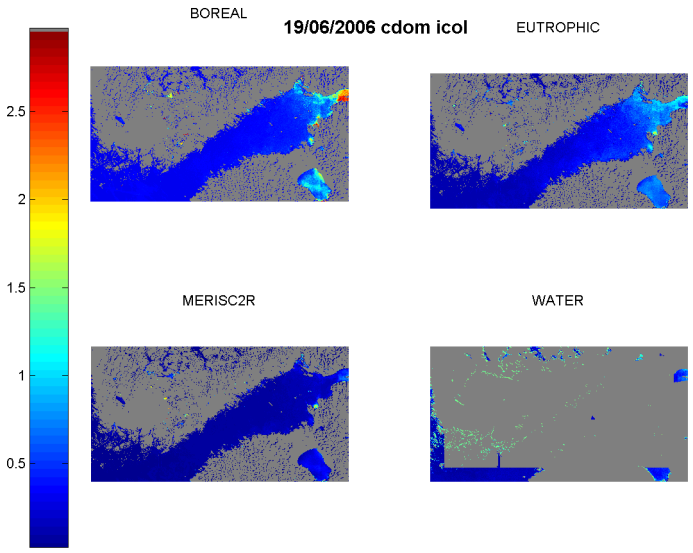


(b) With ICOL.

Figure 26: Summer minimum thematic map of the Gulf of Finland where estimated TSM (g/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

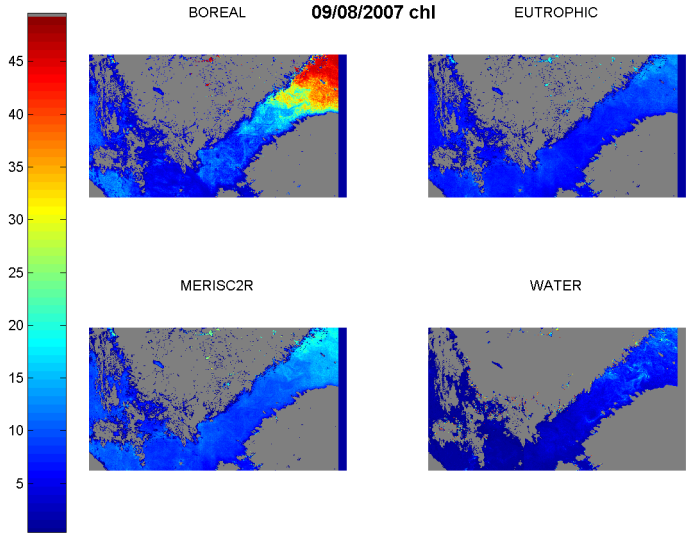


(a) Without ICOL.

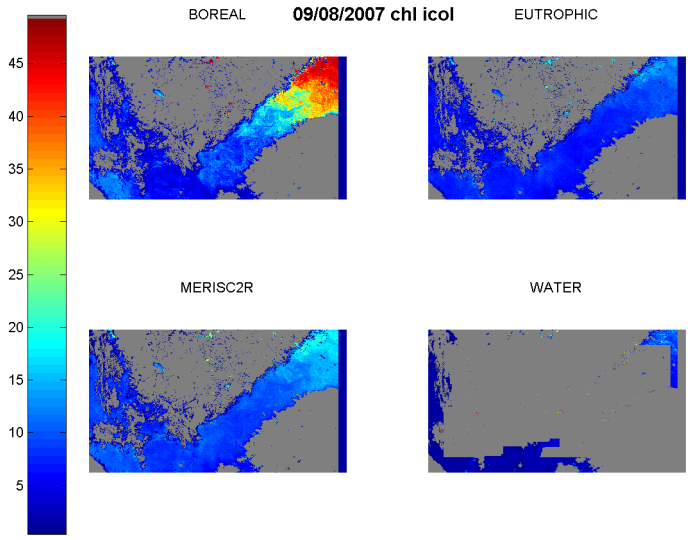


(b) With ICOL.

Figure 27: Summer minimum thematic map of the Gulf of Finland where estimated aCDOM $(1/m)$ is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

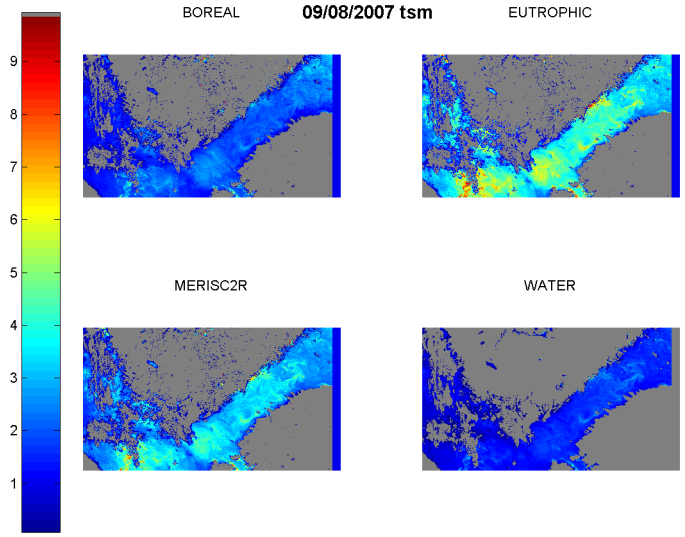


(a) Without ICOL.

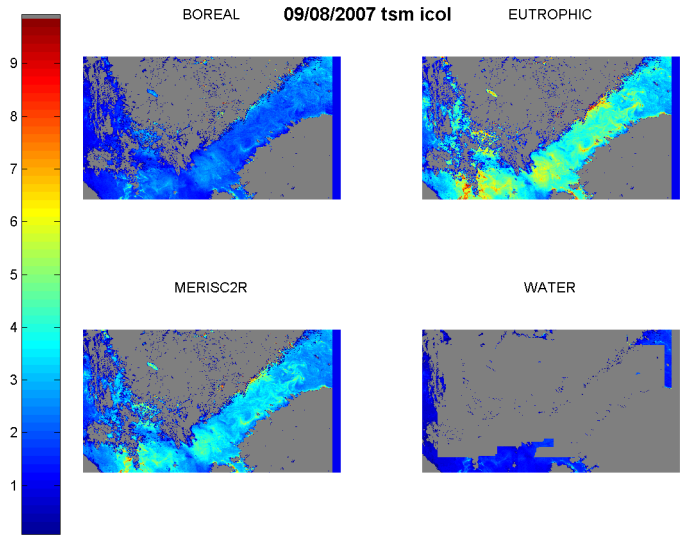


(b) With ICOL.

Figure 28: Summer bloom thematic map of the Gulf of Finland where estimated Chlorophyll a (mg/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

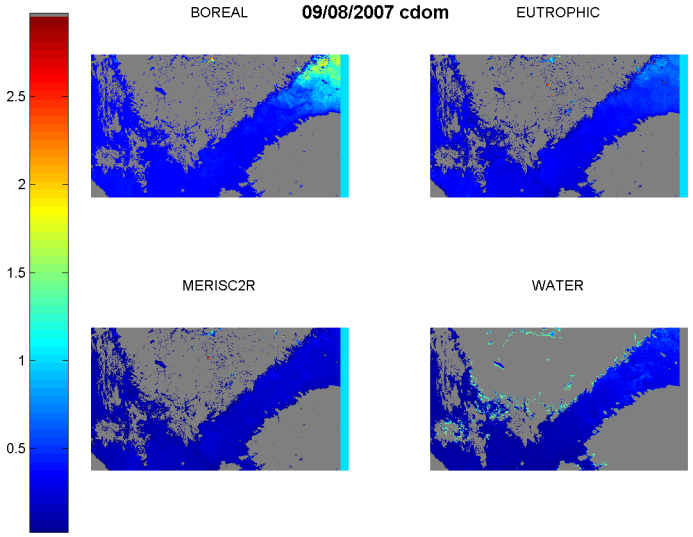


(a) Without ICOL.

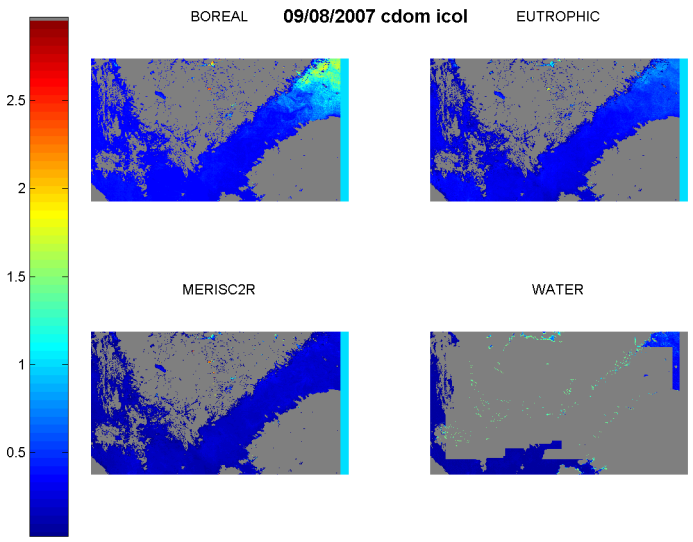


(b) With ICOL.

Figure 29: Summer bloom thematic map of the Gulf of Finland where estimated TSM (g/m^3) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.



(a) Without ICOL.



(b) With ICOL.

Figure 30: Summer bloom thematic map of the Gulf of Finland where estimated aCDOM ($1/m$) is plotted for all the processors (Coastal Case 2 Regional Water Processor (MERISC2R), Boreal Lakes Water Processor (BOREAL), Eutrophic Lakes Water Processor (EUTROPHIC) and FUB/Wew Water Processor (WATER)) without (a) or with (b) ICOL. Land flagged pixels are in grey.

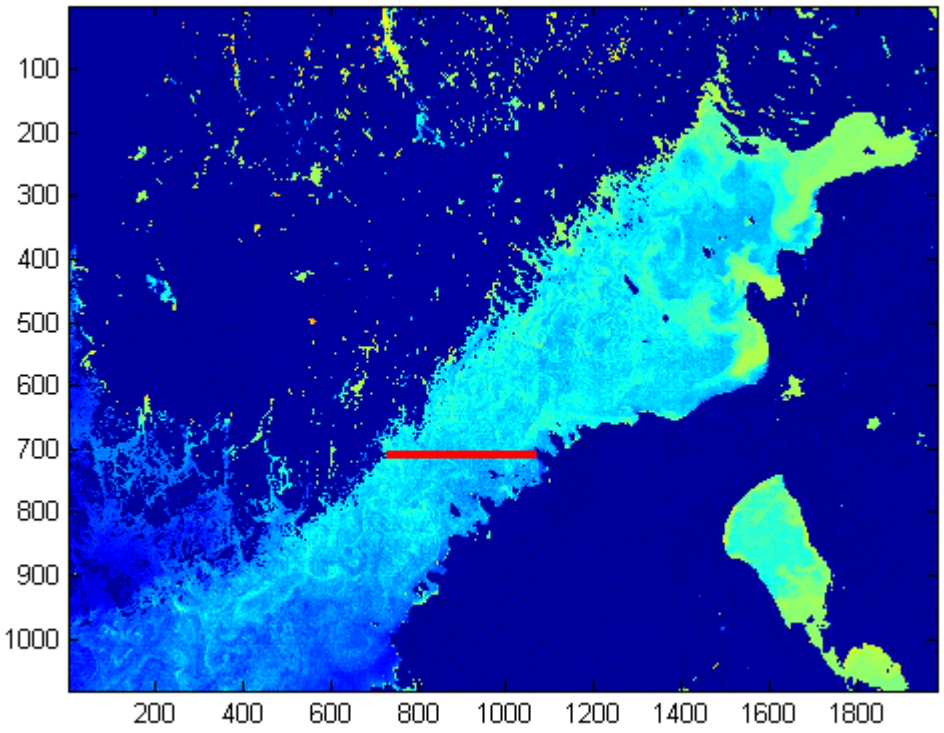


Figure 31: Colour map of the Gulf of Finland. The red line represents the transect path between 24º01'51"E 59º54'52"N - 25º28'26"E 59º36'16"N shown in Figures 32, 33 and 34. The axes units are pixels.

The differences between chlorophyll estimations are important. Again, boreal likes yields the highest concentrations, while eutrophic likes and case 2 regional are very similar. Indeed, TSM concentrations are similar between these two processors as well. However, now boreal processor behaves close to the FUB/Wew water processor, that is, giving lower concentrations. Higher concentrations of TSM than in the spring bloom are observed in general though.

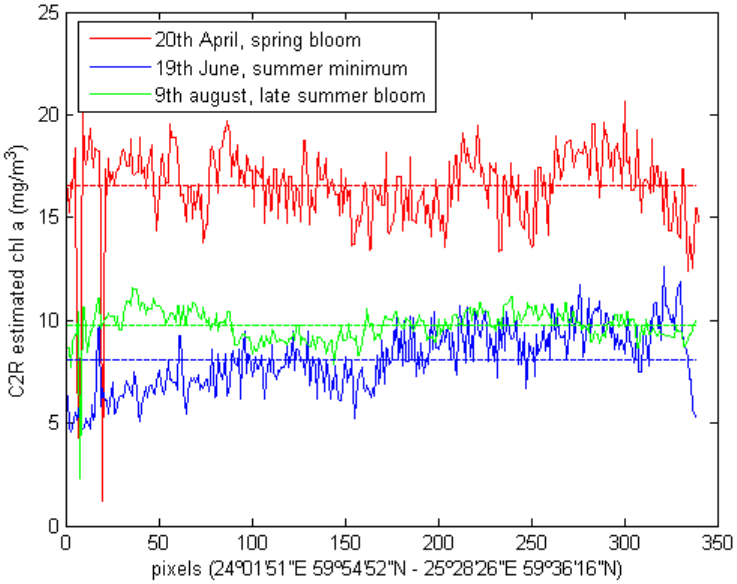
Finally, although the relative behaviour is maintained, lower absorption of aCDOM is observed. Boreal processor produces higher values (up to $1.75m^{-1}$ approximately) than the others, which are quite stable around $0.5m^{-1}$. Furthermore, the estimations are similar in those parts of the Gulf where the results are low, and the differences are more obvious at the northwest of the map, close to the Saint Petersburg coast where higher absorption rates are estimated.

In Figure 31 a colour map with a horizontal red line that crosses the Gulf of Finland from the Finnish coast to the Estonian coast can be seen. Next, the transect plots of different parameters estimated along this transect location at the three previously mentioned dates are analysed. The line has been traced in that particular location because results in this zone use to be in a central range of the overall estimations. That gives to the sample, more representative power. Only two processors will be compared here, Coastal Case 2 Regional and FUB/Wew, to specifically focus in the most different processors, leaving the differences between GKSS processors in transects for later figures.

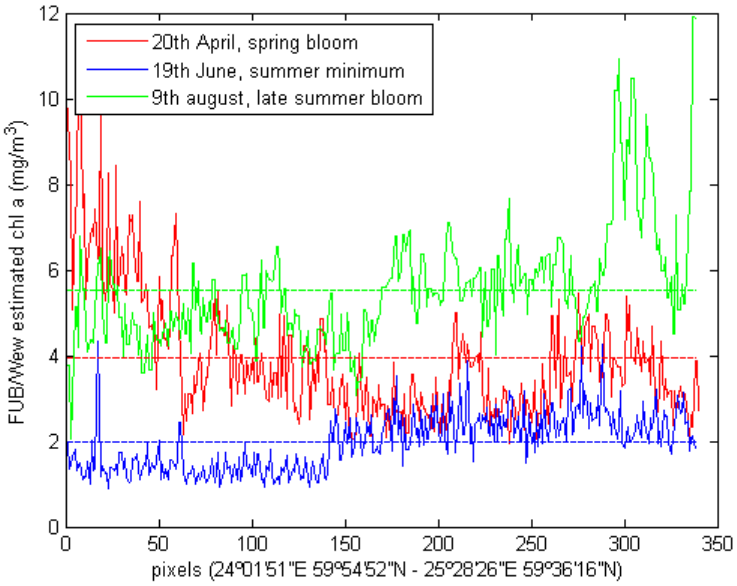
The *chl a* concentration, among the transect, is displayed in Figure 32. It can be observed that Case 2 Regional behaves as expected, that is, higher averaged values are reached in spring and summer blooms. It is interesting to note how different the results obtained with the FUB/Wew processor are. In both cases summer minimum remains at the bottom, but no comparable behaviour emerge in the other two dates. While Case 2 Regional shows a flat pattern in these two cases, FUB/Wew yields an inverse order in the averaged value among other differences such as a greater dispersion and a strange behaviour near the shores. On the other hand, results are quite similar when looking at TSM concentrations plotted in Figure 33. Although the absolute values are not even close, patterns show very common characteristics. First the summer late minimum yields higher values of TSM. Second, the higher the absolute mean is, the higher the variability of the transect is. Furthermore, low concentration are homogenous indeed, along all the transect path, while high concentration reflects significant slopes from one shore to the other, being the direction of the spring bloom slope opposite to then one of the late summer bloom.

Finally, in the case of absorption of aCDOM, important differences emerge again between processors (See Figure 34). The most obvious is now the variability. The averaged values keep the same order but the differences greatly decrease in the FUB/Wew case. In absolute terms Case 2 Regional yields much lower absorption rates.

In general terms, it can hardly be said that patterns along the three dates are somehow correlated. Some similarities might locally appear but it seems like others factors apart from the location are much more involved in the process that determines the final result.

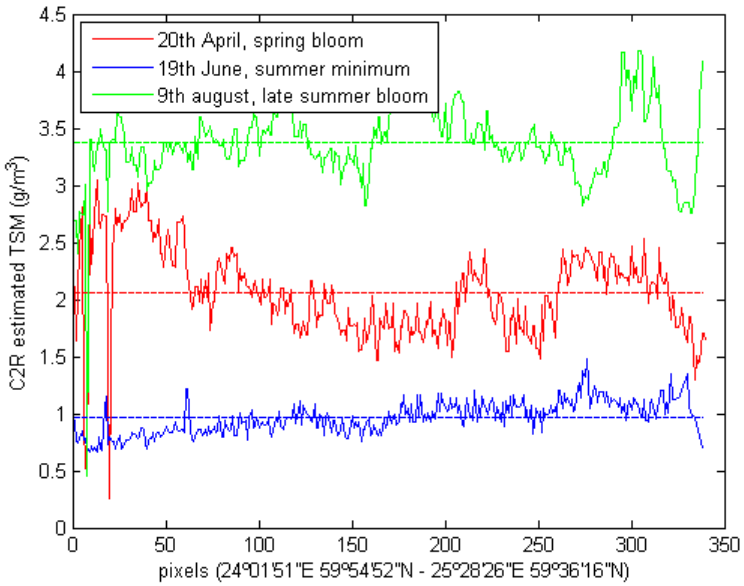


(a) Coastal Case 2 Regional Water Processor.

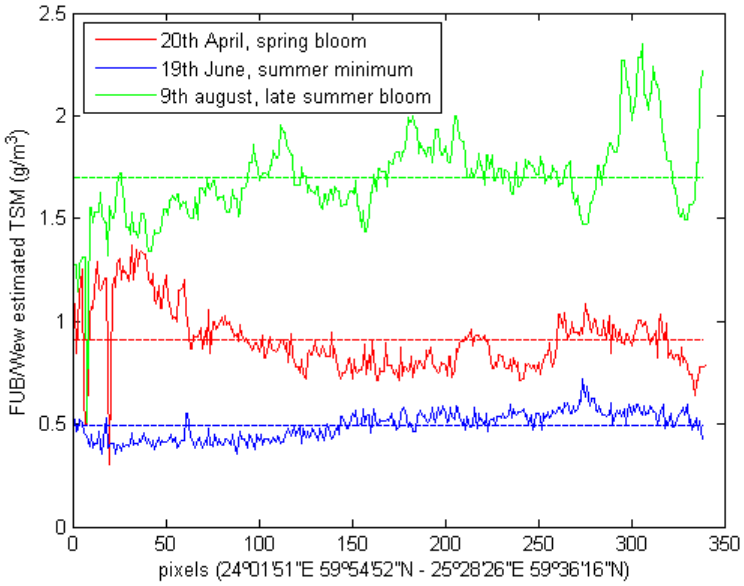


(b) FUB/Wew Water Processor.

Figure 32: Estimated Chlorophyll a (mg/m^3) values along a transect of the Baltic Sea (See Figure 31) are plotted for two processors ((a) Coastal Case 2 Regional Water Processor (MERISC2R) and (b) FUB/Wew Water Processor (WATER)) without ICOL. Three different dates are overlapped (See Legend).

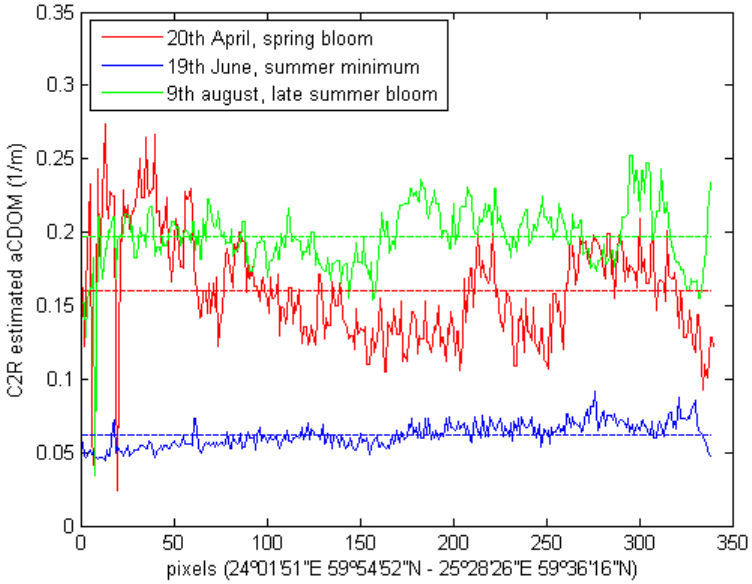


(a) Coastal Case 2 Regional Water Processor.

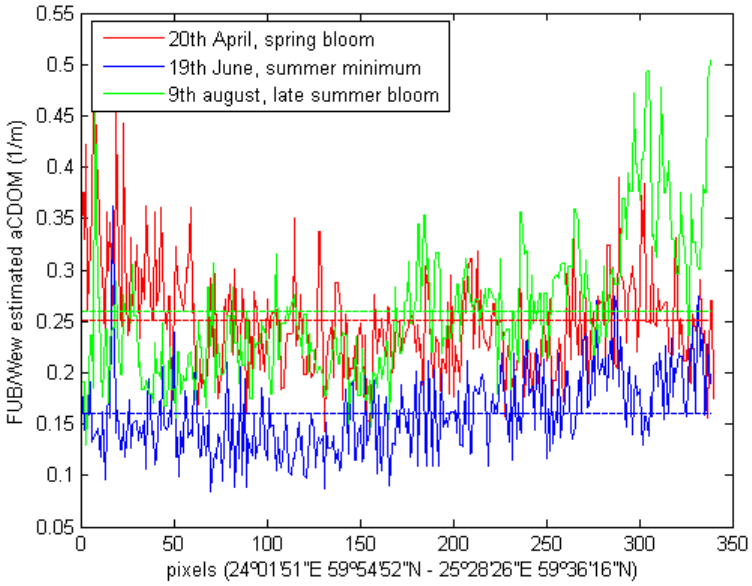


(b) FUB/Wew Water Processor.

Figure 33: Estimated TSM (g/m^3) values along a transect of the Baltic Sea (See Figure 31) are plotted for two processors ((a) Coastal Case 2 Regional Water Processor (MERISC2R) and (b) FUB/Wew Water Processor (WATER)) without ICOL. Three different dates are overlapped (See Legend).



(a) Coastal Case 2 Regional Water Processor.



(b) FUB/Wew Water Processor.

Figure 34: Estimated aCDOM $(1/m)$ values along a transect of the Baltic Sea (See Figure 31) are plotted for two processors ((a) Coastal Case 2 Regional Water Processor (MERISC2R) and (b) FUB/Wew Water Processor (WATER)) without ICOL. Three different dates are overlapped (See Legend).

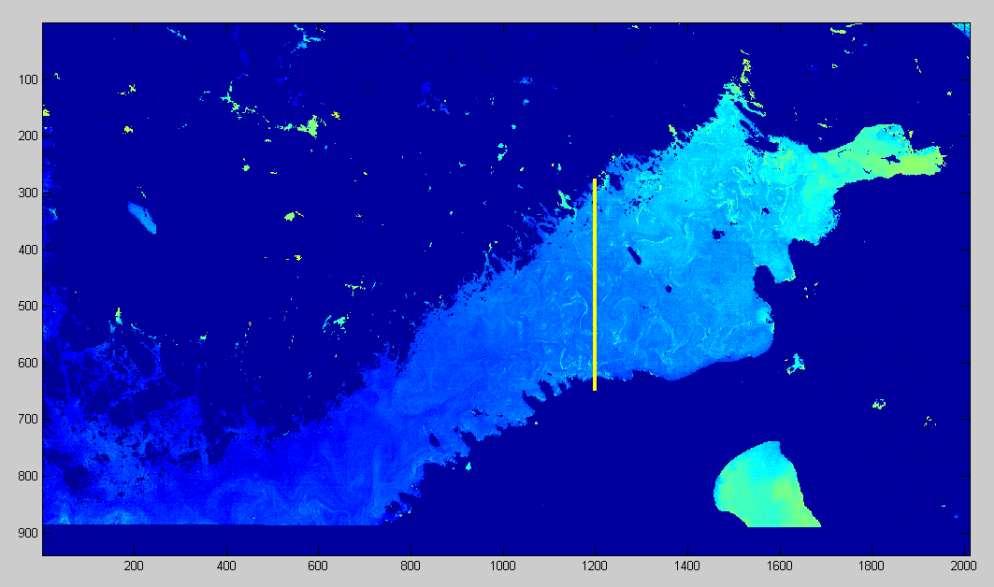
4.3 Effects of ICOL preprocessor

To deeply understand how the presence of the shore modifies the $chla$ estimations, the Figure 35 shows a common transect plot for all the processors with and without ICOL. Previous general observations like higher results of Boreal Lakes, or low estimations of the FUB/Wew processor are here obvious. However a much more interesting thing is how ICOL behaves near the shore. It can be seen how it increases the estimated values until almost a distance around 30km (100 pixels) offshore. Furthermore, it decreases the variability as can be observed at the boreal processor transect between pixels 0-100.

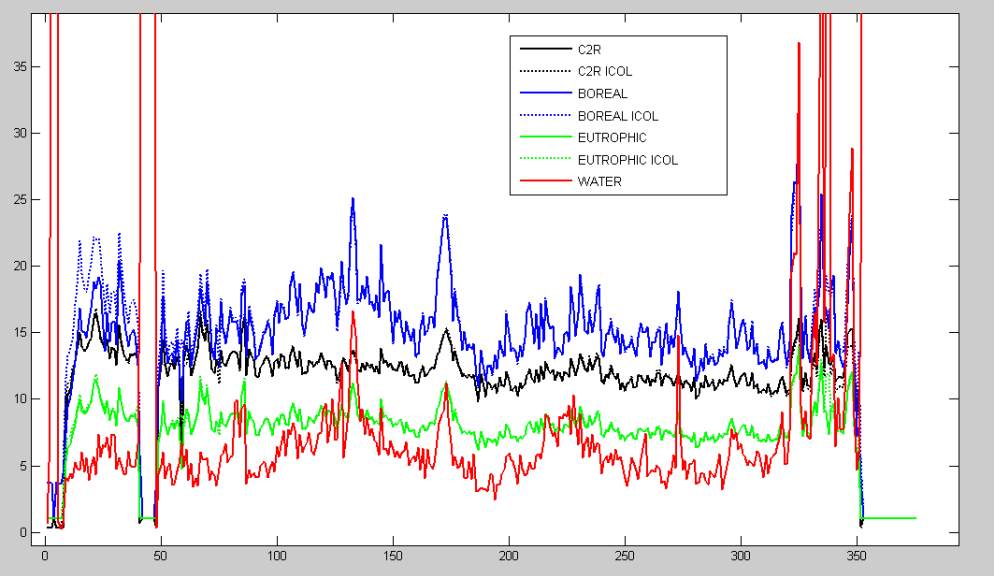
It is interesting to note also the differences between the patterns traced by the GKSS processors compared to the FUB/Wew processor pattern. Boreal, Eutrophic and Case 2 Regional show differences in absolute values (from $10mg/m^3$ to $25mg/m^3$ in the worst case) but similarities in terms of shape. On the other hand, differences arises between them and FUB/Wew such as in inland results where FUB/Wew fails to flag them correctly, resulting in out of range estimations.

This differences in the shape can be observed as well when looking at the scatter plot of Figure 36a. While Case 2 Regional and Boreal are very correlated with Eutrophic, this is not the case of FUB/Wew Water. As observed in Figure 36b, ICOL slightly affects and tends to compress the estimation to closer values. In the case of TSM, Figures 37a and 37b show that the estimations are totally correlated even between Case 2 Regional, Boreal and FUB/Wew Water, the differences being on the slope. However, the results look very different in Figure 38, where the correlation between GKSS processor and FUB/Wew processor is much lower. Actually, the differences are large for some samples that are given low aCDOM values by eutrophic processor ($0.1m^{-1}$) and uniformly distributed values by FUB/Wew along all the range ($0.1 - 1.2m^{-1}$).

In Figure 39a, the effect of the ICOL correction over the radiance reflectance near the shore can be seen. The effect therefore starts around the pixel 260 where an important compensation is performed at band 560 nm by the preprocessor. Thus, it can be deduced that the effect of ICOL extents until around $80x300m = 24km$ far from the shore. The intensity of the correction is not linear with the distance. Indeed, it follows sort of exponential shape. Furthermore, not all the wavelengths are equally corrected. In fact, as it will be seen later, the correction is proportional to the reflectance itself. Hence, the most corrected band is $560nm$ which is the green color. Beyond this band, the ICOL correction decreases, being very small at $865nm$ as well as at low wavelenghts like $412nm$.

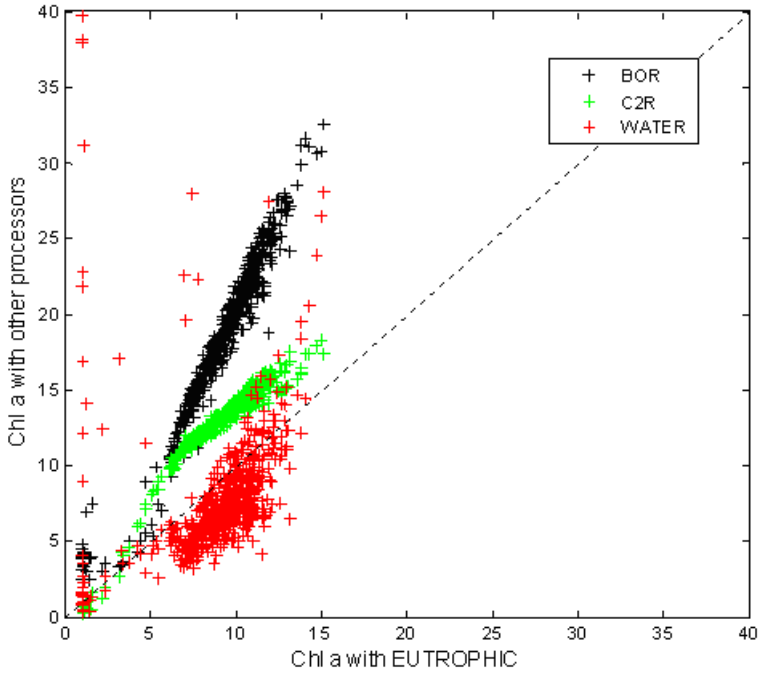


(a) Colour map showing the transect plotted below.

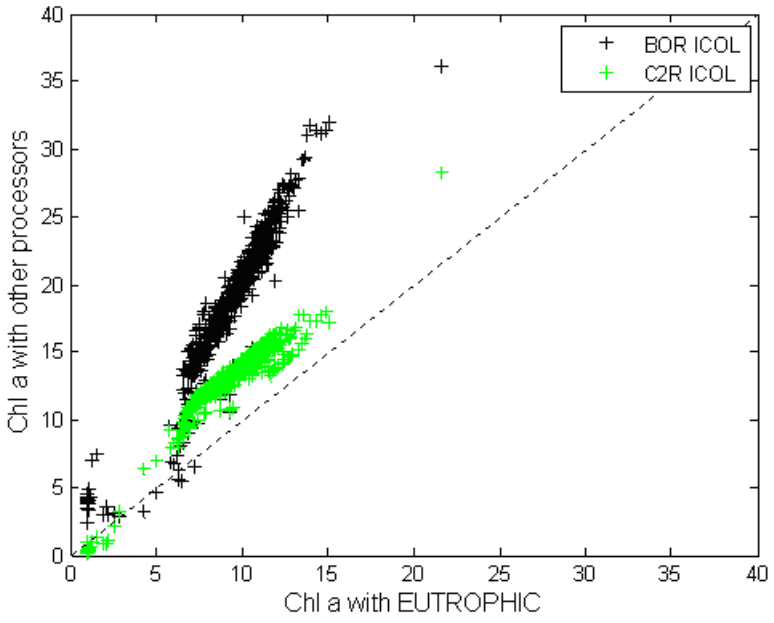


(b) Transect plot of chlorophyll a estimation for all the processors (See Legend).

Figure 35: (a) shows the zone of the Gulf of Finland on August 7th, 2007, and the vertical line defines where the transect plot (b) is located. Therefore, (b) shows chlorophyll a (mg/m^3) estimated values with all the processors with (lines) and without ICOL (dotted lines).

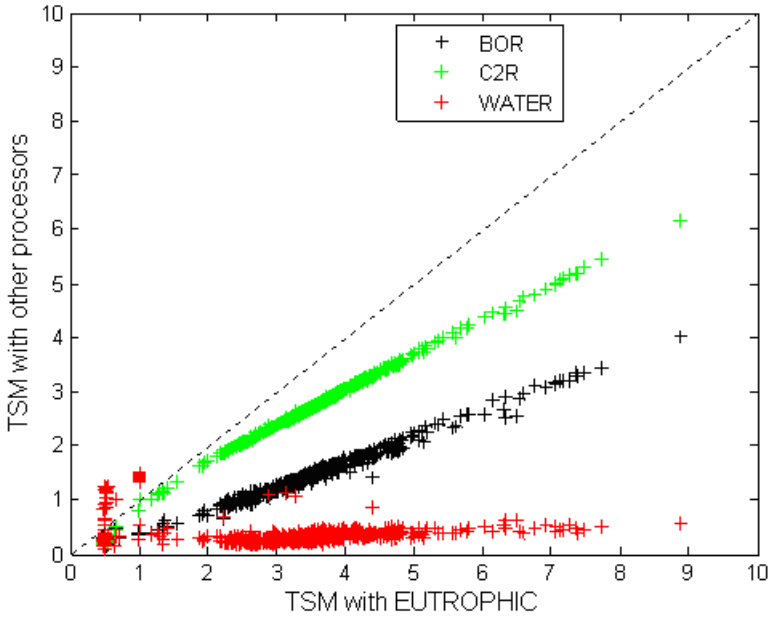


(a) Without ICOL.

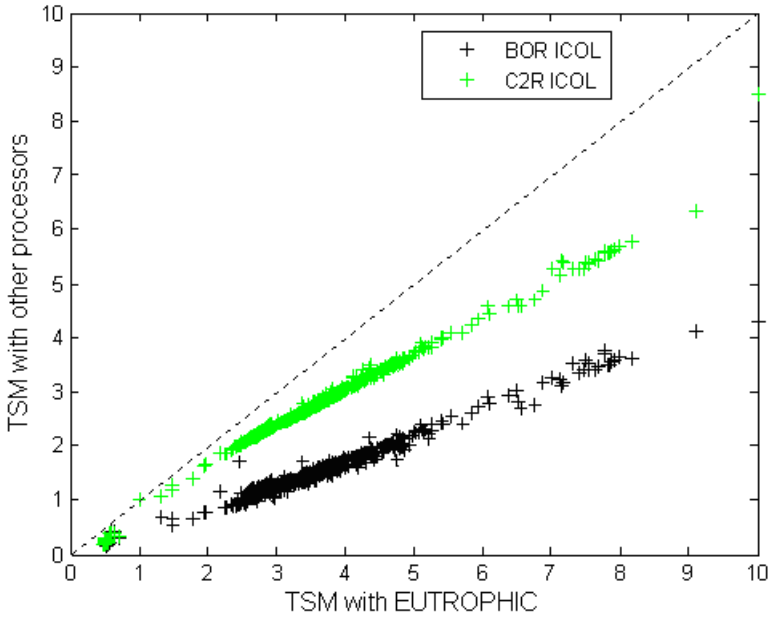


(b) With ICOL.

Figure 36: Chl a (mg/m^3)results of the EUTROPHIC processor vs. the other processors with and without ICOL on a long transect on August 7th, 2007. .

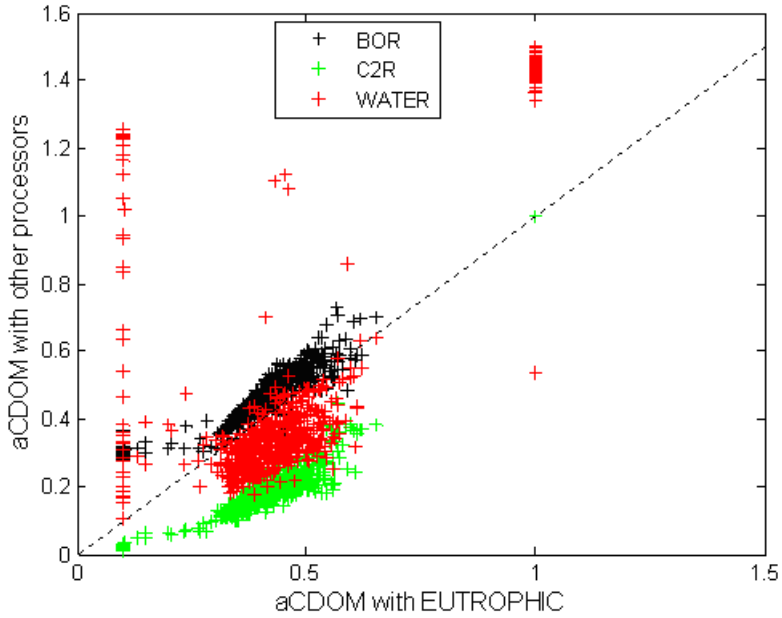


(a) Without ICOL.

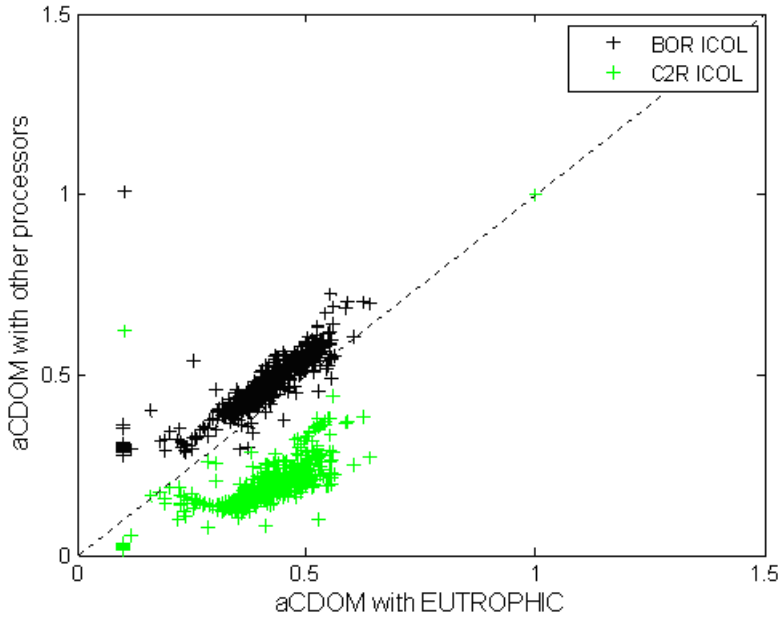


(b) With ICOL.

Figure 37: TSM (g/m^3)results of the EUTROPHIC processor vs. the other processors with and without ICOL on a long transect on August 7th, 2007.

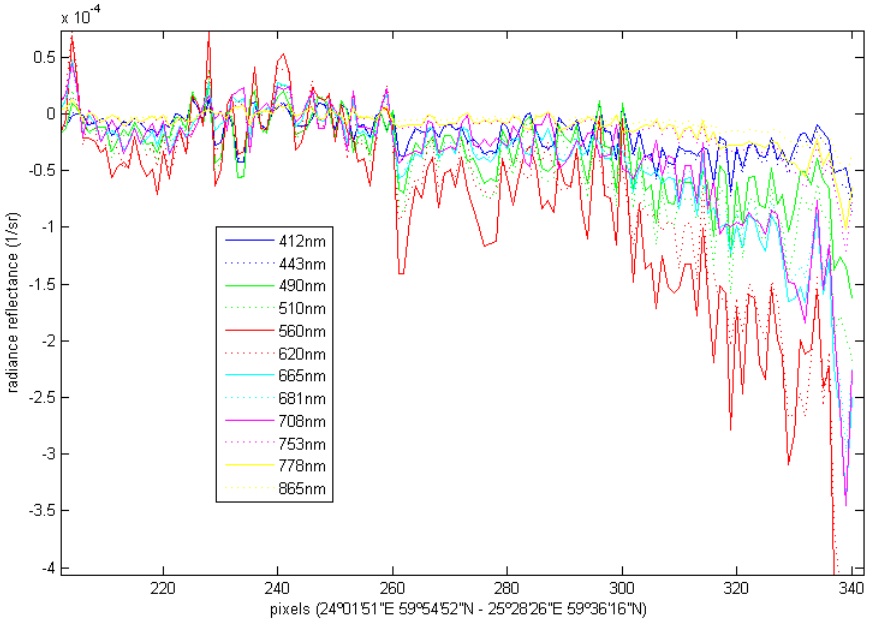


(a) Without ICOL.

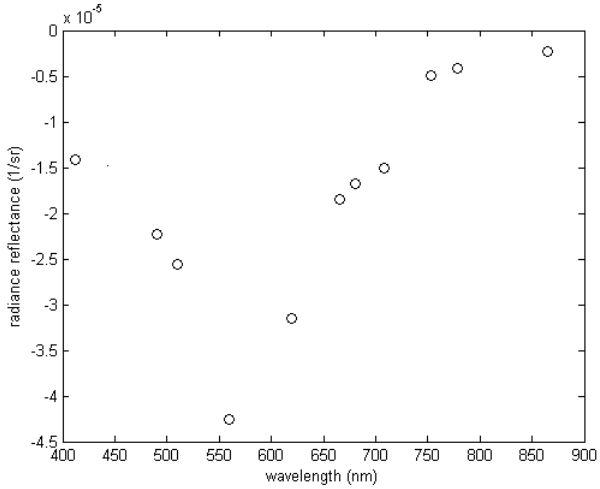


(b) With ICOL.

Figure 38: aCDOM (m^{-1}) results of the EUTROPHIC processor vs. the other processors with and without ICOL on a long transect on August 7th, 2007.



(a) Transect showing ICOL transformation of Radiance Reflectance (C2R - ICOL) at all bands near the shore.



(b) Spectrum ICOL correction (C2R -ICOL) transect mean.

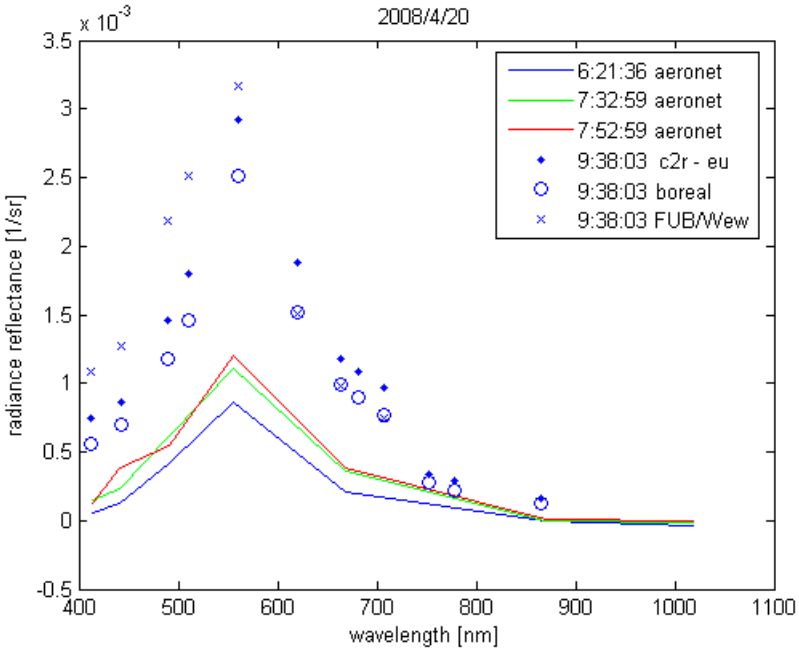
Figure 39: (a) shows the influence in the MERIS Radiance Reflectance (Processed with Coastal Case 2 Regional Water Processor) of the ICOL processor by subtracting from the original Radiance Reflectance, the ICOL corrected Radiance Reflectance. (b) is the ICOL correction spectrum pattern, where the mean of the influence in all the transect points has been taken.

The averaged correction variation depending on the wavelength can be more clearly seen in Figure 39b. Thus, the maximum correction is given at $560nm$ with a mean value over $4 * 10^{-5} sr^{-1}$. Therefore it is interesting to note that the shape is similar to the reflectances received (See Figure 40) even though the values are two orders of magnitude smaller.

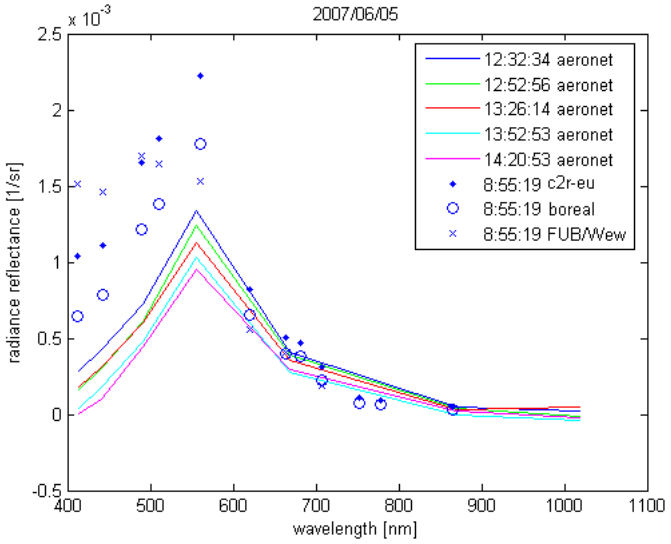
4.4 Comparison with AERONET data

Finally, results regarding radiance reflectance estimations are presented. Those are compared with AERONET *in situ* measurements located at the Helsinki Lighthouse (See 3.4.1 for more information). Spectrum representation has been chosen because the AERONET bands does not match those from MERIS, and therefore it is not possible to compute the errors directly.

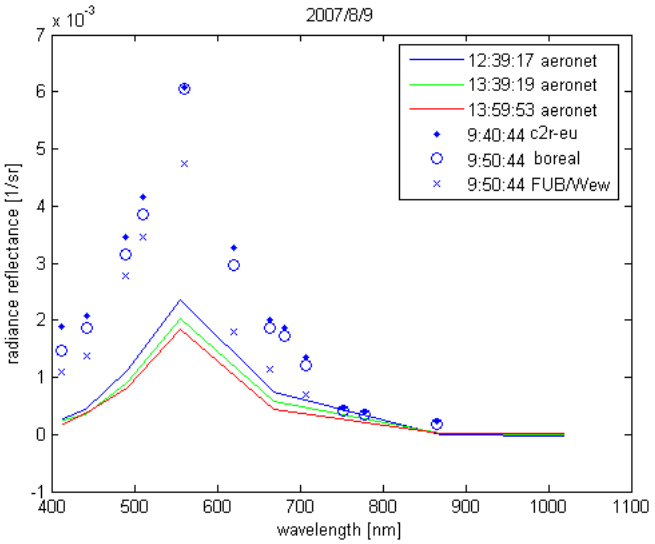
In Figure 40, the reflectances are presented at different wavelengths for the different processors (Case 2 Regional, Lakes Processors, FUB/Wew Water) estimation (at the same geolocation) and compared with the reflectances of AERONET at different hours of the day.



(a) Spring bloom: April 20th, 2008.



(b) Summer minimum: June 5th, 2007.



(c) Late summer bloom: August 9th, 2007

Figure 40: The lines show water-leaving radiance reflectance $(1/sr)$ spectrum from three different dates (chosen to be representative) at 59°56′56″N 24°55′34″E provided by AERONET Level 2 *in situ* measurements. The points are the corresponding MERIS processors estimation at the same geolocation (See Legend). In this case C2R and Eutrophic Lakes processors are shown together because they share the same atmospheric correction algorithm. Both AERONET and MERIS hours are in GMT.

The same dates were used as in previous results. Note that C2R-EU stands for Case 2 Regional and Eutrophic since they share the same atmospheric correction algorithm. Actually, radiance reflectances were calculated from radiances, which is the original data format that AERONET provides, using the following expression:

$$R_w(\lambda) = \frac{L(\lambda)}{E_0(\lambda)} \tag{84}$$

where $E_0(\lambda)$ is the extraterrestrial solar irradiance for each wavelength taken from [54]. Despite the absolute values, the shape of the spectra follow the pattern of the *in situ* measurements in most cases (an exception would be FUB/Wew at Figure 40b at low wavelengths). Therefore, results show general overestimations for low wavelengths. On the other hand, estimations over $700nm$ are much more accurate. Indeed the difference between processors estimation is lower at high wavelengths as well.

The processors behave differently compared to each other in the three plots. In Figure 40a FUB/Wew gives the maximum estimation at $560nm$, but it gives lower estimations compared to the other processors both in Figure 40b and in Figure 40c. On the other hand, Case 2 Regional yields bigger estimations than Lakes processor in all cases.

It is important to note that radiance reflectances are compared at different hours. Furthermore, even AERONET results vary within the time they are measured. These variations reach about $0.5 * 10^{-3} sr^{-1}$ at $560nm$. Therefore, the differences between *in situ* measurements and the estimation of the processors cannot be considered directly as errors.

5 Discussion

The first consideration that has to be made when evaluating the results presented in the previous section is that none of the processors can be considered as accurate. In fact, looking at the RMSE shown in Table 13 the first conclusion is that the mean error is in the order of magnitude of the measurements. At this point, a quantitative description of the results would be useless without analysing the qualitative differences between the processors and comparing them with their relative behaviour.

Similarities and differences between the processors design

Three of the processors studied in this work (Case 2 Regional, Boreal Lakes and Eutrophic Lakes) were developed by GKSS and share the same structure. Thus, the main differences can be found when comparing those with the FUB/Wew Water processor. Here are listed the most important:

1. GKSS processors use a Monte Carlo Radiative Transfer Code, while FUB/Wew processor use the Matrix Operator Method to perform the simulations. Thus, the first ones use a probabilistic method to generate simulation vectors, the second one uses a deterministic method. This is important because it can affect the distribution of the simulations along the training (See Table 8) ranges of the Neural Networks.

2. GKSS processors separate completely the atmospheric correction process and the retrieval of water constituents. On the other hand, FUB/Wew processor retrieves water constituents directly from TOA radiances. Therefore, while GKSS processors can adapt their water algorithm to different water scenarios and share a common atmospheric correction algorithm, FUB/Wew cannot do it.

3. Despite the fact that both processor families use an analytical approach to the problem, the scope of the models is different. The GKSS processors model starts with the TOSA radiances, that is, after directly correcting Rayleigh scattering and Ozone absorption, and ends with the retrieval of the Inherent Optical Properties (IOP) of water. After this, empirical relations are used to derive water concentrations (See Table 7). On the other hand, FUB/Wew processor model cover all the parts of the problem from TOA radiances to water concentrations.

The similarities are:

1. Atmospheric model used in both cases is a 50km standard atmosphere and similar characteristics are considered regarding the ozone loading, aerosol types and distribution, surface pressure, etc.

2. Both processors (GKSS and FUB/Wew) implement their models using Neural Networks.

Behaviour comparison

A detailed analysis of the differences between the processors in different scenarios has already been made in the Results Section (See 4). Therefore here we list an overview of some of the general trends observed:

1. Figure 35 shows clearly that for the same MERIS pixels, processors can be ordered from low to high estimations as follows:

   (a) FUB/WEW Water Processor.

   (b) Eutrophic Lakes Processor.

   (c) Coastal Case 2 Regional Processor.

   (d) Boreal Lakes Processor.

2. GKSS processors behave similarly at low concentrations of Chlorophyll. All of them tend to overestimate the actual value of the chlorophyll.

3. Coastal Case 2 Regional and Eutrophic Lakes processor give the most accurate chlorophyll estimations. Boreal Lakes processor tends to overestimate medium and high concentrations while FUB/Wew processor underestimate them.

4. Errors produced are not equally distributed along all the range of chlorophyll. In fact, errors produced at low concentrations are very significant in the case of GKSS processors while FUB/Wew behaves better in this zone.

5. Along the year GKSS processors can clearly follow the cycle of the chlorophyll: *spring bloom*, *summer minimum* and *late summer boom*. This capability is very important since monitoring the dynamics of waters is one of the main goals of water monitoring. On the other hand this is not the case of FUB/Wew processor.

6. Coastal Case 2 Regional and Eutrophic clearly overestimate TSM at high concentrations. However, the determination coefficients are high in both cases: 0.816 and 0.818 respectively.

7. ICOL processing does not modify significantly the results obtained compared to other sources of errors. However, it could seem surprising that in all cases it increases the radiance reflectance estimated instead of the opposite. According to [34], due to the adjacency effect of surrounding land, water looks brighter than actually is in the near infrared bands so that a correction is needed to remove that component. However, in the atmospheric correction the weights of the NN for the infrared bands are bigger than the visible bands to derive the intensity of atmospheric reflectance. That means that the adjacency effect cause an increase of the perceived atmospheric radiance, lowering proportionaly the perceived water-leaving radiance. That's why the adjacency effect try to correct this by increasing the reflectance of water.

Relationship between results and the processors implementation

It is interesting to note how can GKSS processors and FUB/WEW processor yield so different results if they are based on similar scientific models (both atmosphere and water). This fact suggests that the particular implementation of the processor can be critical. A question that therefore arises is how well Neural Networks can emulate the behavior of the models. The approach of GKSS processors and FUB/Wew processor is the opposite at this point. While the first ones try to divide the problem, that is, $R_{toa}(\lambda)$ to $R_{tosa}(\lambda)$, $R_{tosa}(\lambda)$ to $R_w(\lambda)$, $R_w(\lambda)$ to IOP's and finally IOP's to concentrations, the FUB/Wew processor tries to integrate the whole process in a single Neural Network. The general results obtained suggest that the first option is not only more flexible, but more accurate.

Indeed, fragmenting the problem can be very helpful when analysing the different sources of errors. Therefore, in those part of the process where it is feasible the use of empirical relationships or even the use of direct formulation, it can help to better control the source of errors and inaccuracies and therefore improve the results. In this way, water-leaving radiances have been compared with AERONET *in situ* measurements.

It is well known that Neural Networks behave much better when interpolating than when extrapolating. Thus, having a wide range of simulated data to train the Neural Network is a requirement. Furthermore, these data vectors should be uniformly distributed along all the range to

avoid overtraining of the Network in some regions of the range. This can be especially difficult to achieve when simulations of Radiative Transfer Models are obtained by statistical procedures like in Monte Carlo Tracing Codes. In this case, the use of a deterministic method such as the Matrix Operator Method seems to allow more control on the characteristics of the simulation vectors. It might be due to this that FUB/Wew shows a particularly good behaviour at low concentrations, that is, near one border of the training range, compared to the behaviour of the GKSS processors that are much more inaccurate and chaotic at this region (See Figure 16).

Flags

Another advantage of splitting the process as much as possible is that it allows to follow and control the value of the variables involved in it. Thus, GKSS processors can flag more accurately those pixels that get out of the training range in some stage of the process by using *l2_flags*. On the other hand, FUB/Wew only provide *l1_flags*. This feature is particularly important when dealing with a quantitave evaluation of the processors because non masked contaminated pixels (e.g. by sun glint) can originate big errors that distort the total result. However, it would be interesting to research in how the different flags behave in the different processors to evaluate its accuracy as well.

Assessment

All the reason explained in this discussion and the results obtained in Section 4 show the better behaviour of the GKSS family of processors compared to the FUB/Wew Water processor. Those are:

1. Much better RMSE (from 6.7 in the case of Coastal Case 2 Regional Processor to 18.3 in the case of FUB/Wew Water Processor (See Table 13)).
2. Much better R^2 (from 0.478 in the case of Boreal Lakes Processor to 0.0692 in the case of FUB/Wew Water Processor (See Figure 16)).
3. Much better capability to follow seasonal cycles of the chlorophyll (See Figure 21)

Between GKSS processors, the differences regards mainly to the empirical relations used to derive concentrations from IOPs and the training ranges. Thus, these differences are related to the type of water where the estimation is being made. Hence, Coastal Case 2 Regional is the most suitable

processor for the Baltic Sea. That is because of its relatively stable behaviour along the range. Eutrophic would be a good option as well but almost all values were originally flagged by this processor and thus it is not reliable.

6 Conclusions

In the present thesis, four of these processors have been described, analyzed and compared:

1. Coastal Case 2 Regional Water Processor.
2. Boreal Lakes Water Processor.
3. Eutrophic Lakes Water Processor.
4. FUB/Wew Water Processor.

As explained in Section 5, the most suitable processor is the Coastal Case 2 Regional Water Processor. Unfortunately, its estimations are far from being satisfactory: absolute errors remain in the same order of magnitude than the values estimated. Despite of their efficiency, Neural Networks can hide some of the error sources due to their opacity. Their training process has to be made therefore very accurately.

ICOL preprocessor does not provide a significant improvement of the results. In some cases it helps and in other it does not. More sophisticated algorithms with *off nadir* calculations should be developed to deal with the Adjacency Effect (See Section 3.2.5). However, it could be interesting to integrate it as a process inside the regular processors. In fact, it is possible that some processes like correction of Rayleigh scattering could be done twice when applying ICOL before the water constituents retrieval. That could explain the strange results obtained when processing ICOL corrected data with FUB/Wew (See Section 4).

From the point of view of the design, at least at the current state of the art, the approach followed by GKSS processors of splitting the estimation process in several parts is much more interesting than integrating all the model behind a Neural Network like FUB/Wew does, since it allows to evaluate the performance of each module independently. In the same way, the ability of the processors of flagging contaminated pixels is crucial to analyse them properly. FUB/Wew processor definitely should improve its flagging capabilities.

Further Research

Along the current thesis, various tools have been developed to automatise the comparison process of several processors with a large amount of MERIS products and *in situ* data. Therefore, they can be useful for further research on this field, making it faster and more efficient.

Since the beginning of this work, a new version of BEAM has been released as well as processor updates. It would be interesting to carry out a new comparison of the new version with the previous ones in order to evaluate the improvements achieved. Furthermore, it could be interesting to update the JAVA scripts that uses the BEAM's API to the new version. In fact, the tools developed can be still more flexible than they are right now. A migration to a fully JAVA programing environment could improve greatly the performance of the tools by exploiting the API libraries of BEAM. Enough documentation has been provided with the source code to fully understand it and improve its functionality.

Good and fast comparison and validation tools are essential to an agile development of new processor algorithms providing objective assessment of their functionality. Indeed, SYKE and regional environmental centers collect a considerable amount of *in situ* data from measurement stations along the Finnish coast as well as by means of through flow cruises. Thus, the Baltic Sea and specially the Finnish waters can be an ideal framework for testing different processors and take advantage of these "ground" efforts.

References

[1] Antonie D. and A. Morel, 1999, *A multiple scattering algorithm for atmospheric correction of remotely-sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrrying various aerosols including absorbing ones*. International Journal of Remote Sensing, 20, 1875-1916.

[2] Antonie D. and Morel A., 2000, *Atmospheric correction over the ocean (case 1 waters),Algorithm Theorical Basis Document 2.7*, PO-TN-MEL-GS-0005, Issue 4, Rev. 1, European Space Agency.

[3] Babin M., Morel A., Fournier-Sicre V.,Fell F. and Stramski D., 2003, *Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration*. Limnol. Oceanogr., pp. 843-859.

[4] Behnert I., Matthias V. and Doerffer R., 2007, *Aerosol climatology from ground-based measurements for the southern North Sea*, Atmospheric Res.

[5] Bennartz R., Fischer J., 2000, *A modified k-distribution approach applied to narrow band water vapour and oxygen absorption estimates in the near infrared*, Journal of Quantitative Spectroscopy & Radiative Transfer, 66, pp. 539-553.

[6] Bricaud A., Morel A., Prieur L., 1983, *Optical efficiency factors of some phytoplankters*. Limnology and Oceanography, vol. 28, no. 5, pp. 816-832.

[7] Bricaud A., Morel A., Babin M., Allali K., Claustre H., 1998, *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, pp. 31033 31044.

[8] Brockmann-Consult, *The Beam Project*, http://www.brockmann-consult.de/beam/.

[9] Chandrasekhar S., 1950, *Radiative Transfer* Clarendon Press, Oxford.

[10] Cox, C. and W. Munk, 1954, *Statistics of the sea surface derived from Sun glitter*. Journal of Marine Research, 13, 198-227.

[11] ACRI, 2006, *MERIS Level 2 Detailed Processing Model*, PO-TN-MEL-GS-0006.

[12] Elouraginia, Salem, Houda Chtiouia and Pierre H. Flamant, 2005, *Lidar remote sounding of cirrus clouds and comparison of simulated fluxes with surface and METEOSAT observations*, Atmospheric Research, Volume 73, Issues 1-2, January 2005, Pages 23-36.

[13] Eltermann, L., 1968, *UV, visible, and IR attenuation for altitudes to 50 km*. Environmental Research Paper No. 285, AFCRL-68-0153, Airforce Cambridge Research Laboratories.

[14] Environment, *Baltic Sea*, http://www.environment.fi/balticsea

[15] European Space Agency, 2002, *ENVISAT Post-Launch Products MERIS*, PO-MO-ESA-GS-1005, 5-MERIS.

[16] European Space Agency, Astrium Limited Earth Observations and Science, Astrium GmbH Earth Observations and Science, 2001, *The Envisat Mission*, Brochure.

[17] European Space Agency, 2007, *Envisat-I: Products Specifications. Volume II: Meris Products Specifications*, Doc. Ref. PO-RS-MDA-GS-2009.

[18] Fell F., Fischer J., 2001, *Numerical simulation of the light field in the atmosphere-ocean system using the matrix-operator method*, Journal of Quantitative Spectroscopy & Radiative Transfer, 69, 351-388.

[19] Fischer J., Grassl H., 1984, *Radiative transfer in an atmosphere-ocean system: an azimuthally dependent matrix-operator approach*, Applied Optics, 23, 1032-1039.

[20] Gordon H. R., Brown O. B., Jacobs M. M., 1975, *Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean*, Appl. Opt. 14, 417-427.

[21] Gordon H.R, Morel A., 1983, *Remote assessment of ocean color for interpretation of satellite visible imagery*, A review. Lectures on Coastal and Estuarine Study, Vol 4, Springer Verlag.

[22] H. R. Gordon, J. W. Brown, and R. H. Evans, *Exact Rayleigh scattering calculations for use with the Nimbus-7 coastal zone color scanner*, Appl. Opt. 27, 862-871.

[23] Gordon H. R., Wang M., 1992, *Surface-roughness considerations for atmospheric correction of ocean color sensors. II: Error in the retrieved water-leaving radiance*, Optical Society of America.

[24] Gordon H.R, 1997, *Atmospheric Correction of Ocean Color Imagery in the Earth Observing System Era*, Jour. Geophys. Res., 102D, 10781-17106.

[25] Heinemann, Thomas and Lothar Schüller, 1995, *MIE-Programmpaket für Strahlungstransportsimulaltionen mit MOMO*, Institu für Weltraumwissenschaften, Freie, Universität Berlin.

[26] S. Sathyendranath, R. P. Bukata, R. Arnone, M. D. Dowell, C. O. Davis, M. Babin, J. F. Berthon, O. V. Kopelevich, J. W. Campbell, 2000, *Colour of Case 2 Waters*, Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex Waters. IOCCG Report Number 3.

[27] Macke A, Müller J, Raschke E, 1996, *Single scattering properties of atmospheric ice crystals*. Journal of Atmospheric Sciences 53, 2813-2825.

[28] Maritorena S., Morel A., Gentili B., 1994, *Diffuse reflectance of oceanic shallow waters: influence of water depth and bottom albedo*, Lim. and Ocean.,37, 1689-1703.

[29] Mobley C.D, 1994, *Light and Water; Radiative Transfer in Natural Waters*, Academic Press, San Diego, Calif.

[30] Morel A., Prieur L., 1977, *Analysis of variations in ocean color*, Limnol. Oceanogr. 22,709-722.

[31] Morel, A. and B. Gentili, *Diffuse reflectance of oceanic waters: its dependance on sun angle as influence by molecular scattering contribution*. Applied Optics, 30, 4427-4438.

[32] Kallio K., 2008, *New SIOPs for Finnish Lakes (2008)*, Technical Note, SYKE.

[33] Kirk J.T.O., 1989, *The upwelling light stream in natural waters*, Limnol. Oceanogr., 34(8), 1410-1425.

[34] Koponen S., 2006 *Remote sensing of water quality for Finnish lakes and coastal areas* Ph. D. Thesis.

[35] Kozlyaninov M. V., Pelevin V. N., 1965, *Sur l'Utilisation de l'Approximation Unidirectionelle pour l'Étude de la propagation du Flux Lumineux en Mer*, Acad. Nauka SSR Tr. Oceanogr. Inst. 77, 73.

[36] Kröse B., Smagt P. van der, 1996, *Introduction to Neural Networks*, Eight Edition. The University of Amsterdam.

[37] Petzold T. L., 1972, *Volume scattering functions for selected ocean waters, Scripps Institution of Oceanography*, San Diego, California, pp. 25-27, 38, 64-65.

[38] Plass G.N., Kattawar G.W., 1972, Appl.Opt. 8, 455.

[39] Plass G.N., Kattawar G.W., 1973, *Matrix Operator Theory of Radiative Transfer. 1: Rayleigh Scattering*, Appl. Opt. 12, 314-329.

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

[41] Doerffer R., Schiller H., 2006, *MERIS Advanced Water Algorithm* Version 1.1., GKSS.

[42] Doerffer R., Schiller H., 2008, *MERIS Regional Coastal and Lake Case 2 Water Project Amospheric Correction ATBD*, Version 1.0. GKSS.

[43] Doerffer R., Schiller H., 2008, *MERIS Lake Water Algorithm for BEAM*, ATBD Version 1.0. GKSS.

[44] Ruiz Verdú A., 2007-2008, *Bio-optical data of Spanish Lakes*, Lake Project.

[45] Rummelhart D., McClelland J., 1986, *Parallel Distributed Processing*, MIT Press, Cambridge, Massachusetts.

[46] Santer R., Schmechtig C., 2000, *Adjacency effects on water surfaces: Primary scattering approximation and sensitivy study.* Appl. Op., Vol. 39, No. 3, pp. 361-375.

[47] Santer R., Zgolski F., 2008, *Improve Contrast between Ocean and Land*, ATBD v1.1 - MERIS Level-1C.

[48] Schiller H. and Doerffer R., 1999, *Neural Network for Emulation of an Inverse Model - Operational Derivation of Case II Water Properties from MERIS data*; Int. Journal of Remote Sensing, vol. 20, No 9, 1735 - 1746.

[49] Shiller H., 2000, *Feedforward-Backpropagation Neural Net Program ffbp 1.0*. Report GKSS 2000/37, ISSN 0344-9626.

[50] Schiller H. and Doerffer R., 2005, *Improved Determination of Coastal Water Constituent Concentrations from MERIS Data*, IEEE transactions on geoscience and remote sensing, vol 43, pp. 1585 - 1591.

[51] Schroeder, Th. 2005, *Fernerkundung von Wasserinhaltsstoffen in Kuestengeweassern mit MERIS unter Anwendung expliziter und impliziter Atmosphaerenkorrektur- verfahren*, Ph.D. Dissertation, Freie Universitatet Berlin, Berlin (Germany), 2005, http:/www.diss.fu-berlin.de/2005/78

[52] Shettle, E.P. and R.W. Fenn, 1979, *Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties*. Environmental Research Paper No. 676, AFGL-TR-79-0214, Airforce Geophysics Laboratory.

[53] Tanré D., Deroo C., Duhaut P., Herman M., Morcrette J. J., Perbos J., Deschamps P. Y., 1990, *Description of a computer code to simulate the satellite signal in the solar spectrum: 5S code.* Int. J. Rem. Sen., Vol. 11, No. 4, pp. 659-668.

[54] Thuillier G., Herse M., Simon P. S., Labs D., Mandel H., Gillotay D., Foujols T., 1998, *The visible solar spectral irradiance from 350 to 850 nm as measured by the SOLSPEC spectrometer during the Atlas I mission*, Solar Phys., 177, 41-61.

[55] Dr. Volker Liebig, 2007, *Envisat Symposium 2007: Foreword*, Montreux, Switzerland 2327 April 2007 (ESA SP-636, July 2007).

[56] Wikipedia, *Baltic Sea* http://en.wikipedia.org/wiki/Baltic_Sea.

[57] World Climate Reasearch Program, 1986, *A preliminary cloudless standard atmosphere for radiation computation. International Association for Meteorology and Atmospheric Physics*, Radiation Commission, Boulder, CO, USA, 1984, CSP-112, WMO/TD-No.24, March 1986.

[58] Zibordi Giuseppe, *Aeorosol Robotic Network (AERONET)*, Helsinki Lighthouse, giuseppe.zibordi@jrc.it

A Processing Software Instruction

The processing software package developed for the SYKE/MARCOAST comparison includes the following:

1. MERIS data Matlab processing tool
2. Thematic map tool
3. Data extraction tools

A short summary of each is presented below.

The Matlab functions and other codes have been developed and tested with BEAM version 4.2 (this version has to be installed) and Matlab 7.4. It may be necessary to adjust the paths in certain files (e.g. gpt.bat). Some comments are also included in the m-files.

A.1 MERIS data MATLAB processing tool

The purpose of this tool is to make the processing of several MERIS datasets as automatic as possible. Before processing it is necessary to save the MERIS data in the DIMAP (*.dim) format (with BEAM/VISAT). It is recommended that the user also removes from the data any unwanted areas (i.e., creates a subset of the data). This will decrease the processing time. The new datasets are placed in the same directory (called work-directory) and named so that the word "subset" appears in the beginning of the filename followed by a number starting from 0. E.g.:

subset_0_of_MER_FR__1PNEPA20060508_094650_000000982047_00294_21891_1205.N1.dim
subset_1_of_MER_FR__1PNEPA20060508_094650_000000982047_00294_21891_1206.N1.dim
...

It is strongly recommended to use exactly the same filename format as shown above since some information is gathered from the filename (e.g. the date and time).

When creating the subset, note that ICOL needs a buffer zone of 30 km on each side of the target area. So, make the subset large enough.

Also note that both the dim file and the corresponding data directory must be in the same work-directory. Additionally, both the gpt.bat and the "beam" folder (that includes auxiliary data about the processors configuration) are needed to be in the same folder as the .m files.

The main function that performs the processing (and calls the other functions) is called "startprocessors.m". The function is started with a command:

startprocessors(workdirectory)
(E.g.: *startprocessors('E:/MERIS_N1/processing')*)
The Matlab code first processes the data with ICOL (adjacency effect correction processor). NOTE: This step can take many hours per dataset if they are large. Then the makes a small correction in the ICOL processed dim files (see the beginning of the function for details).

The main function then calls a function that creates the XML and BAT files needed by the C2R, BOREAL, EUTROPHIC and FUB/WeW processors. All dim files (both ICOL preprocessed and the original file) in the work-directory are included in the processing. Finally, the DOS commands that call the relevant processors are invoked.

It is possible to comment out (with the % symbol before the command) the commands that are not wanted (in the main function). E.g., by commenting the line:
createicol(workdir);repairicol(workdir);
the ICOL processing is not included.
All result datasets are placed in the work-directory.

A.2 Thematic Maps Tool

The purpose of this Matlab tool is to automatically create thematic maps of chl a, TSM and acdom (and Z90_max, see below) from the processed data. In order for the tool to operate properly the format of the filename must be as shown in Chapter 2. The function is called with a command:
imagemaker('E:/MERIS_N1/processing')
The maps are saved as PNG files in a directory Thematicmaps (in the work-directory).

Two versions are saved. The 1st one includes only one data product per image file. E.g. the file:
a_gelbstoff_20060508_094650_BOREAL_0-3.png
includes the thematic map of acdom for date 20060508 and time 094650 processed with the BOREAL processor. The last numbers of the filename indicate the scaling used in the image (in this example the data values range from 0 (dark blue) to 3 (red)). Clouds are indicated with white and land as grey. The other flags are not included in the default processing (see also below). ICOL processing has been applied, if the word ICOL appears in the name.

Each single-product image-file also has a corresponding txt file that contains the latitude and longitude of the center point of the upper left corner pixel of the image (pixel index 0.5, 0.5).

The other image file versions are saved in directory Thematicmaps/*comparison_images* and include the results of each data product derived with the 4 processors in a single image file. The purpose of these is to make the comparison of the results easier. ICOL and non-ICOL versions are separate so the files are named as e.g.:

chl_comparison_20060508_094650.png

chl_comparison_20060508_094650_icol.png

In these images both land and clouds are grey.

The main function (imagemaker.m) includes parameters that can be changed by the user.

The parameter:

maxscale = [50 10 3];

includes the maximum value of each parameter (in the order chl a, TSM, acdom). These are used to scale the image files.

The parameters

doland = 1; %apply land flag (0 if not) doclouds = 1; %apply clouds flag (0 if not) doother = 0; %apply all flags (0 if not)

control the use of flags in the image making process. When the value is 0 the flag is not activated. When the value is 1 the flag is activated. The doother flag includes all flags (also land and cloud). This flag is not activated as default since there are cases where almost the whole image is flagged.

A separate version called *imagemaker_Z90* has been developed for making thematic maps of the parameter *Z90_max*. It operates in the same way as the original function.

A.3 Station Tool

This tool extracts all product values from the processed MERIS data that match both in time and space with the input (in situ) data given to the tool. For the time, there is an adjustable time window of +/- 2 days (changeable by the user). A 3x3 pixels square is extracted from each station (center at the coordinates of the in situ data). The in situ data that do not match any subset in time and space are ignored (not stored in the result files).

There are some parameters that have to be set by editing the script. These are the paths to the MERIS and in situ data folders and the name of the in situ file.

The tool is started with the command: extractiontool in the directory there the m-file is.

INPUT file format

The in situ data must be given to the tool as a .txt text file that contains the following in tabulated columns (only one tabulation):

1. ID number of the station
2. latitude in geographic coordinates (decimal degree units) as a float number (decimal point ".") (e.g.: 60.37603)
3. longitude in the same format (e.g.: 26.35633)
4. Finnish coordinate system: YKJP (e.g.: 6696300)
5. Finnish coordinate system: YKJI (e.g.: 3464490)
6. Julian day (number of the day in the year: 1 - 365)
7. Year (e.g.: 2008)
8. Month (1-12)
9. Day (1-31)
10. Values of water parameter one (decimal point ".")
11. Values of water parameter two
12. three
13. etc,...

The in situ data file must contain at least the first 9 columns. If some information is not available, that column must contain values 0. If Lat/Lon coordinates are not available, the YKJ coordinates must be given.

The tool will first try to use the geographic coordinates (Lat/Lon) and if their values are 0 the tool will convert the Finnish coordinates to Lat/Lon.

Similarly, there are two ways of reading the date from the text file: Julian day + year, or year+month+day. The tool will first try to use the Julian day and if that is not available it will convert the month and day to Julian days.

Examples of a line of the in situ data file are:
3047 0 0 6696300 3464490 0 2006 5 17 6.2
and:
3048 60.37603 26.35633 0 0 150 2006 0 0 4.2
OUTPUT files

The results of the extraction are stored in matrixes. The format of these matrixes is:

insitu_matrix data: N x (10+parameters), where N is the number of points that match with some MERIS subset. The first 9 columns are always the same as in the original in situ input file (if coordinate and/or date conversions have been made the new values are included in the result matrix). After those is a new column that contains the date of the MERIS data (Julian days). After that are the parameters (as they were in the original file).

MERIS data: N x (# of products) x 9, where N is as above. # of products depends on the processor of each subset. This value can be 75-76 for the C2R, BOREAL, EUTROPHIC processors and 17 for the FUB/WeW WATER processor. Finally, 9 represents the values of the 3x3 matrix in the following format:

1	2	3
4	5	6
7	8	9

Thus, the values of the center pixel are in the 5th layer of the matrix.

The name of the output matrixes is different for each processor with and without icol:

Case2Regional processor: *c2r_matrix* and *c2r_icol_matrix* Boreal processor: *boreal_matrix* and *boreal_icol_matrix* Eutrophic processor: *eutrophic_matrix* and *eutrophic_icol_matrix* Water processor: *water_matrix* and *water_icol_matrix*

The order of the output products is in the ASCII alphabetical order.

For BOREAL LAKES & C2R :

1	K_min	17	path_3	33	reflec_7	49	toa_reflec_7	65	trans_12
2	Z90_max	18	path_4	34	reflec_8	50	toa_reflec_8	66	trans_13
3	a_gelbstoff	19	path_5	35	reflec_9	51	toa_reflec_9	67	trans_2
4	a_pig	20	path_6	36	tau_550	52	tosa_reflec_1	68	trans_3
5	a_total	21	path_7	37	toa_reflec_1	53	tosa_reflec_10	69	trans_4
6	ang_443_865	22	path_8	38	toa_reflec_10	54	tosa_reflec_12	70	trans_5
7	b_tsm	23	path_9	39	toa_reflec_11	55	tosa_reflec_2	71	trans_6
8	chiSquare	24	reflec_1	40	toa_reflec_12	56	tosa_reflec_3	72	trans_7
9	chl_conc	25	reflec_10	41	toa_reflec_13	57	tosa_reflec_4	73	trans_8
10	l1_flags	26	reflec_12	42	toa_reflec_14	58	tosa_reflec_5	74	trans_9
11	l2_flags	27	reflec_13	43	toa_reflec_15	59	tosa_reflec_6	75	tsm
12	path_1	28	reflec_2	44	toa_reflec_2	60	tosa_reflec_7		
13	path_10	29	reflec_3	45	toa_reflec_3	61	tosa_reflec_8		
14	path_12	30	reflec_4	46	toa_reflec_4	62	tosa_reflec_9		
15	path_13	31	reflec_5	47	toa_reflec_5	63	trans_1		
16	path_2	32	reflec_6	48	toa_reflec_6	64	trans_10		

For EUTROPHIC:

1	K_min	17	path_2	33	reflec_6	49	toa_reflec_6	65	trans_10
2	Z90_max	18	path_3	34	reflec_7	50	toa_reflec_7	66	trans_12
3	a_btsm	19	path_4	35	reflec_8	51	toa_reflec_8	67	trans_13
4	a_gelbstoff	20	path_5	36	reflec_9	52	toa_reflec_9	68	trans_2
5	a_pig	21	path_6	37	tau_550	53	tosa_reflec_1	69	trans_3
6	a_total	22	path_7	38	toa_reflec_1	54	tosa_reflec_10	70	trans_4
7	ang_443_865	23	path_8	39	toa_reflec_10	55	tosa_reflec_12	71	trans_5
8	b_tsm	24	path_9	40	toa_reflec_11	56	tosa_reflec_2	72	trans_6
9	chiSquare	25	reflec_1	41	toa_reflec_12	57	tosa_reflec_3	73	trans_7
10	chl_conc	26	reflec_10	42	toa_reflec_13	58	tosa_reflec_4	74	trans_8
11	l1_flags	27	reflec_12	43	toa_reflec_14	59	tosa_reflec_5	75	trans_9
12	l2_flags	28	reflec_13	44	toa_reflec_15	60	tosa_reflec_6	76	tsm
13	path_1	29	reflec_2	45	toa_reflec_2	61	tosa_reflec_7		
14	path_10	30	reflec_3	46	toa_reflec_3	62	tosa_reflec_8		
15	path_12	31	reflec_4	47	toa_reflec_4	63	tosa_reflec_9		
16	path_13	32	reflec_5	48	toa_reflec_5	64	trans_1		

For FUB/WEW water:

1	aero_opt_thick_440	6	l1_flags	11	reflec_5	16	total_susp
2	aero_opt_thick_550	7	reflec_1	12	reflec_6	17	yellow_subs
3	aero_opt_thick_670	8	reflec_2	13	reflec_7		
4	aero_opt_thick_870	9	reflec_3	14	reflec_9		
5	algal_2	10	reflec_4	15	result_flags		

The products are also listed in a Matlab variables *"product_list_[processorname]"*. All result matrices are saved in a file called: *Extracted_station_data.mat*.

A.4 Transect Tool

This tool extracts from a single .dim file (given by the user as a parameter) the values of all products at the exact coordinates given by an input text file and stores them in a matrix. All coordinates are assumed to be inside the subset coverage. Additionally, it stores the in situ data provided in the text file as well.

All the parameters have to be set by editing the first lines of the script. To start the tool type: transecttool

INPUT data

The format of the in situ text file is similar to the format of the station tool but now contains only the following columns:

1. ID number
2. latitude
3. longitude
4. YKJP
5. YKJI
6. parameter #1
7. parameter #2...

In addition, the user must give the tool the path and the name of the MERIS dataset

OUTPUT files

The extracted data are stored in matrixes:

Meris_matrix: N x (#number of products) where N is the number of "in situ" points. There is one matrix for each processor (with and without ICOL)

Insitu_matrix: N x (5 + parameters) where N is the number of "in situ" points and parameters is the number of measurements.

The products are also listed in a Matlab variables "product_list_[processorname]".

All result matrices are saved in a file called: Extracted_transect_data.mat

B JAVA Class

To get the pixels that correspond to some coordinates of a subset it is needed to use the API library of the Beam program. To do this, the next java class has been compiled:

Ex1.java

```
package org.esa.beam.basics; //name of our package

//imported libraries from beam
import org.esa.beam.framework.dataio.ProductIO;
import org.esa.beam.framework.datamodel.Product;
import org.esa.beam.framework.datamodel.GeoCoding;
import org.esa.beam.framework.datamodel.GeoPos;
import org.esa.beam.framework.datamodel.PixelPos;
import org.esa.beam.util.logging.BeamLogManager;

import java.io.IOException;

public class Ex1 { // Our class
    public static void main(String[] args) {
        BeamLogManager.removeRootLoggerHandlers();
        String filePath = args[0];
        //we read our .dim file as the first argument
        Product product = null;
        try {
            product = ProductIO.readProduct(filePath, null);
            //open the .dim file product
        } catch (IOException e) {
```

```
        // error managing
            System.out.printf("Error:
            failed to open the file %s: %s\n",
            filePath,e.getMessage());
            System.exit(1);
        }
        if (product == null){
            System.out.printf("Error:
            no reader found for file %s\n",
            filePath);
            System.exit(1);
        }
        GeoCoding geoCoding = product.getGeoCoding();
        GeoPos geoPos = new GeoPos();

        float lat = new Float(args[1]);
        //we get the latitude from the second argument
        float lon = new Float(args[2]);
        // we get the longitude from the third argument
        geoPos.setLocation(lat, lon);
        //we use the geopos object to set its position
        PixelPos pixelPos = geoCoding.getPixelPos(geoPos,null);
        //here we get the pixels
        double x = pixelPos.getX();
         //we get the X pixel from the pixelPos object
        double y = pixelPos.getY();
        // the same with the Y pixel
        System.out.printf("x = %f, y = %f",x,y);
         //result is printed to be got by the script

    }
}
```

End of Ex1.java This java class is compiled as a .jar file named Beam-JavaWS.jar. To execute it, it must be called from a .bat file as follows:

Ex1.bat

```
@echo off
set BEAM4_HOME=C:/Program Files/beam-4.2
java -Xmx1024M -cp ^
```

```
"BeamJavaWS.jar;^
%BEAM4_HOME%/lib/bc-jnn-1.5.jar;^
%BEAM4_HOME%/lib/clibwrapper-jiio-1.1.jar;^
%BEAM4_HOME%/lib/jai-codec-1.1.3.jar;^
%BEAM4_HOME%/lib/jai-core-1.1.3.jar;^
%BEAM4_HOME%/lib/jai-imageio-1.1.jar;^
%BEAM4_HOME%/lib/Jama-1.0.2.jar;^
%BEAM4_HOME%/lib/jdom-1.1.jar;^
%BEAM4_HOME%/lib/mlibwrapper-jai-1.1.3.jar;^
%BEAM4_HOME%/lib/xpp3-1.1.3.3.jar;^
%BEAM4_HOME%/lib/xstream-1.3.jar;^
%BEAM4_HOME%/modules/ceres-core-0.7.jar;^
%BEAM4_HOME%/modules/beam-core-4.2.jar;^
%BEAM4_HOME%/modules/beam-envisat-reader-1.2.jar"
org.esa.beam.basics.Ex1 %*
exit /B 0
```

End of Ex1.bat

First, the Beam folder must be set. Second, the libraries must match with the beam version that the user has installed in his own computer. Finally, the main class of the .jar file must be called in the corresponding package, in this case: org.esa.beam.basics.Ex1. If the user wants to modify this class and compile it again, it is recommended to follow the tutorial from the beam website, where the setup of the workspace is explained.