Summer Camp 2022

Satellite Imagery in e-Environment

Prepared by: *Haralović Marko. University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb*
Lovrić Mirna, University of Zagreb, Faculty of Science, Zagreb, Croatia
Maletić Marin, University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

Mentors: *Goran Kopčak, m.sc. (Ericsson Nikola Tesla),*
Veronika Ozretić Vidak (Ericsson Nikola Tesla)
Filip Wolf (Ericsson Nikola Tesla)

Table of contents

1.Summary

Remote sensing retrieval of water parameters is ultimately too unstable to detect the level of any concentracion of the parameters of concern. It must be added that the summary of the work we considered relevant is too specific and not applicable for a general usage. Therefore, the algorithms we developed and put into use were inadequate to consistently calculate any results, to the extent that reliability of them was put at stake. Paper provides deep insight into the satellite technology and its usage, how the data from the satellites could be handled, what some indicators mean and how does retrieval of the data from the satellites is useful for the procession and approximation of the real parameters. The ground for the future scope of work has been forged, an insightful overview has been given and the theoretical background provided for a future work that will follow. Main issues are atmospheric correction algorithms unsustainability, satellite projects imprecisions and general approximation and limitation of the remote sensing algorithms. Codependency between the parameters was noticed and is yet to be explored, concluding thought is to make an overall fitting scheme which for this work is helpful and valuable.

2. Abstract

The main goal of the project is to analyze the sea water parameters using satellite imagery fetched from Creodias database and from the Copernicus Open-Source hub. The water analysis could be divided into two intuitive sections, first being freshwater sources, which mainly happen to be inland waters, and second being oceans and/or seas. Freshwater parameters will be of no importance for the project and the analysis, since the project focus has shifted to the seas, where four parameters are considered: sea level temperature, chlorophyll, dissolved oxygen, and turbidity. The reason behind the project lies in the simplicity of further analysis of the water quality which could be used for any purpose and would be available to whomever it may concern. The data collected directly from the field of the Adriatic Sea will be evaluated and upon the results for water clarity, the water will be labeled accordingly. The initial premise was to use Python programming language for the analysis of the Sentinel-3 data imagery that would be extracted from the creodias database. We aim to fetch the data from the exact locations for which there are in-situ measurements from the SNAP Windows app, which would be used for the analysis of satellite imagery. Developed algorithms will be used for comparison of the satellite imagery data with in-situ measurements and calibrated accordingly. The analysis other publicly disclosable algorithms had been considered and analyzed. Various equations and mathematical relations were observed and implemented, and the work provides the results for the best available equations from other works. Conclusion sums up to be that the remote sensing can not be performed with the current knowledge and imperfection of the Satellite imagery as it is, therefore a calibration must be done; the major impediment is the state of atmosphere and the Satellite precision and it's felicitous to only the surface level of the water analysis.

KEYWORDS: Satellite monitoring, Creodias, Sentinel-2, Sentinel-3, OC4Me,CHL, remote sensing

3.Introduction

Satellite images are images of Earth collected by imaging satellites operated by governments and businesses around the world.

There are five types of resolution when discussing satellite imagery in remote sensing: spatial, spectral, temporal, radiometric and geometric:

-spatial resolution is defined as the pixel size of an image representing the size of the surface area (i.e. m^2) being measured on the ground.

-spectral resolution is defined by the wavelength interval size and number of intervals that the sensor is measuring.

-temporal resolution is defined by the amount of time (e.g. days) that passes between imagery collection periods for a given surface location.

-radiometric resolution is defined as the ability of an imaging system to record many levels of brightness and to the effective bit-depth of the sensor.

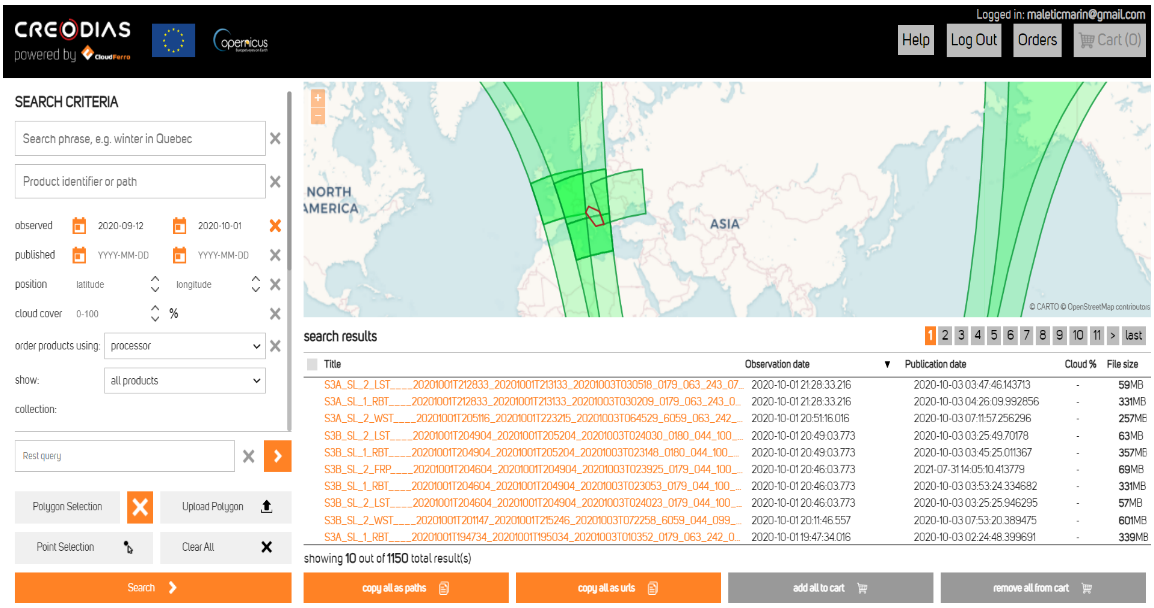
-geometric resolution refers to the satellite sensor's ability to effectively image a portion of the Earth's surface in a single pixel.

Creodias is an environment that serves as a source of processed Earth Observation (EO) data. It is formed out of Copernicus Sentinel data and services, Envisat data, Landsat and other metadata.

EO data provides very comprehensive spatial coverage, with relevant multi-spectral optical data available from 10-60 x 10-60 m resolution (e.g., Sentinel-2 MSI) to 300m x 300m resolution (Sentinel3 OLCI).

Copernicus is the European Union's Earth observation programme that offers satellite Earth Observation and in-situ data. Open database consists of airborne,ground-based and seaborne information available for wide publicity. [1]

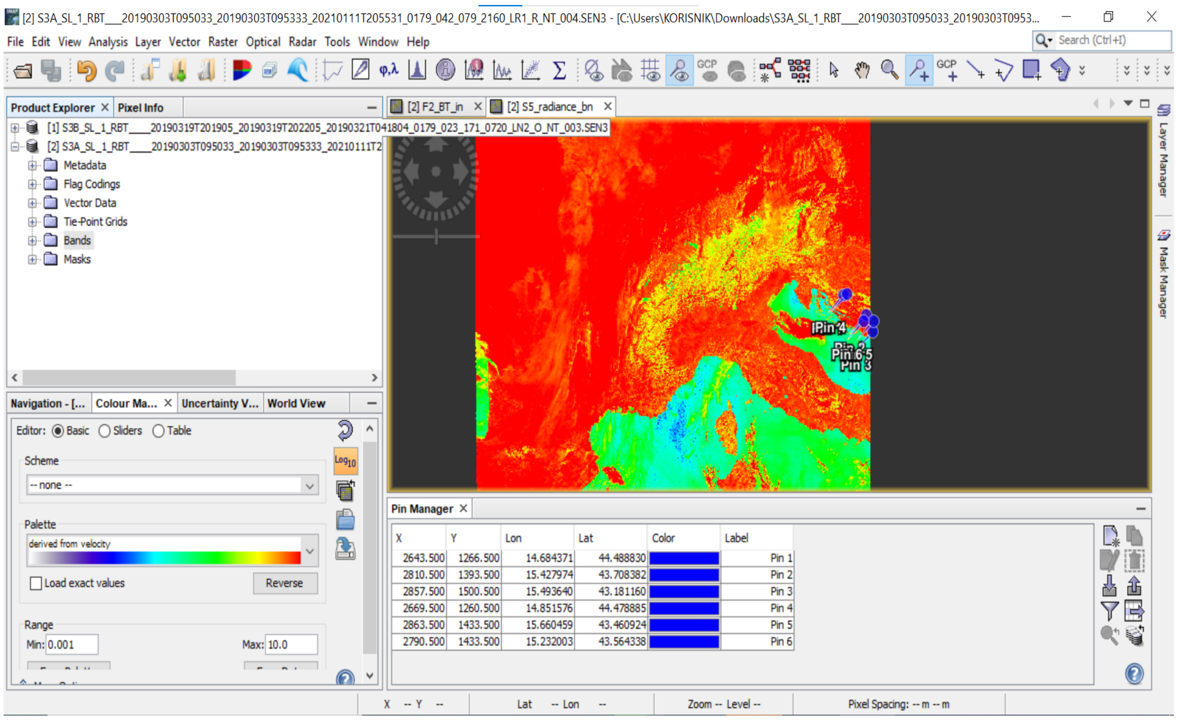
Image 1: Creodias database and data selection:



Copernicus Open Access Hub, previously known as Sentinels Scientific Data Hub, the portal provides access to Sentinel data through an interactive graphical user interface. The portal will also provide access to data produced by future Sentinel missions when available.

SNAP, or the Sentinel Application Platform is in short a collection of executable tools and Application Programming Interfaces which have been developed to facilitate the utilization, viewing and processing of a variety of remotely sensed data. It includes tools for all common remote sensing satellites.

Image 2: SNAP interface, with a product open and a color filter put on.



Sentinel is a polar-orbiting, multispectral high-resolution imaging mission for land monitoring to provide, for example, imagery of vegetation, soil and water cover, inland waterways and coastal areas. There are three types of Sentinel satellites, of which being 1,2 and 3 respectively.

Sentinel-1 for the objective has a Land and Ocean monitoring, it is composed of a pair of polar-orbiting satellites operating day and night, and performs Radar imaging to acquire the data no matter the weather conditions.

Sentinel 2 is a satellite mission responsible for land monitoring, with it being composed out of two polar-orbiting satellites providing high-resolution optical imagery, with vegetation, soil and coastal areas being the main observance goal.

Sentinel-3 is a satellite mission with the primary objective being marine observation, and a sea-surface and land temperature, ocean and land color study. It is composed out of three satellites, with a radar altimeter being the misssion’s primary device. Polar-orbiting satellites carry multiple instruments, including optical imagers.

There are several different satellites, but for this work are of no broader importance, so the list goes for the general knowledge; Sentinel-4, Sentinel-5, Sentinel-5P, Sentinel-6.

There are two different Earth observation missions launched by the European Space Agency Copernicus program [2] that are suitable for lake studies. One of them is Sentinel-2, a land monitoring constellation of two satellites (A and B, launched in 2015 and 2017, respectively). Both have the Multispectral Instrument (MSI) onboard, which offers high-resolution optical imagery at 10 m, 20 m, and 60 m spatial resolution, depending on the spectral band. MSI samples in 13 spectral bands. This mission provides global coverage every 5 days [3]. Although MSI has great spatial resolution that is suitable for smaller lakes, it lacks a band in critical wavelengths, such as the Chl-a absorption peak at 665 nm.

The other suitable mission is Sentinel-3, which is also a constellation of two satellites (A and B, launched 2016 and 2018, respectively). Sentinel-3 has a medium resolution (300 m) Ocean and Land Color Instrument (OLCI) onboard for marine and land research. It has 21 spectral bands and provides global coverage (at the equator) every two days [4]. OLCI was built for water monitoring and has well placed spectral bands for that purpose. However, its rather low spatial resolution allows the study of only about 1000 of the largest lakes on Earth [5] out of 117 million [6].

The MultiSpectral Instrument (MSI) onboard Sentinel-2 (S2) with high resolution (up to 10 m) has shown suitability for detecting cyanobacterial blooms and retrieval of chl a concentration in subalpine lakes [7]. However, further development for atmospheric correction (AC) algorithms and the need for the larger validation data over optically complex waters is necessary [8].

The S2 mission was originally designed for monitoring land cover changes and is composed of two identical satellites—S2A was launched in 2015 and S2B in 2017. The S2 mission includes satellites S2C and S2D as well, which are planned to be sent into orbit in the next decade.

The MSI sensor measures in 13 spectral bands from 443 to 2190 nm with spatial resolution 10, 20, 60 m and with a 12-bit radiometric resolution [9].

Irradiance is the downwelling radiation from the sun. Radiance is the upwelling radiation from the Earth to the sensor.

Since every satellite mission uses the light radiance and reflectance for the analysis and monitoring of the Earth, the questions of a problem come to be the following:

How does the radiance and irradiance of the light affect the monitoring?

How does the state of atmosphere affect the remote sensing and could it be surmounted?

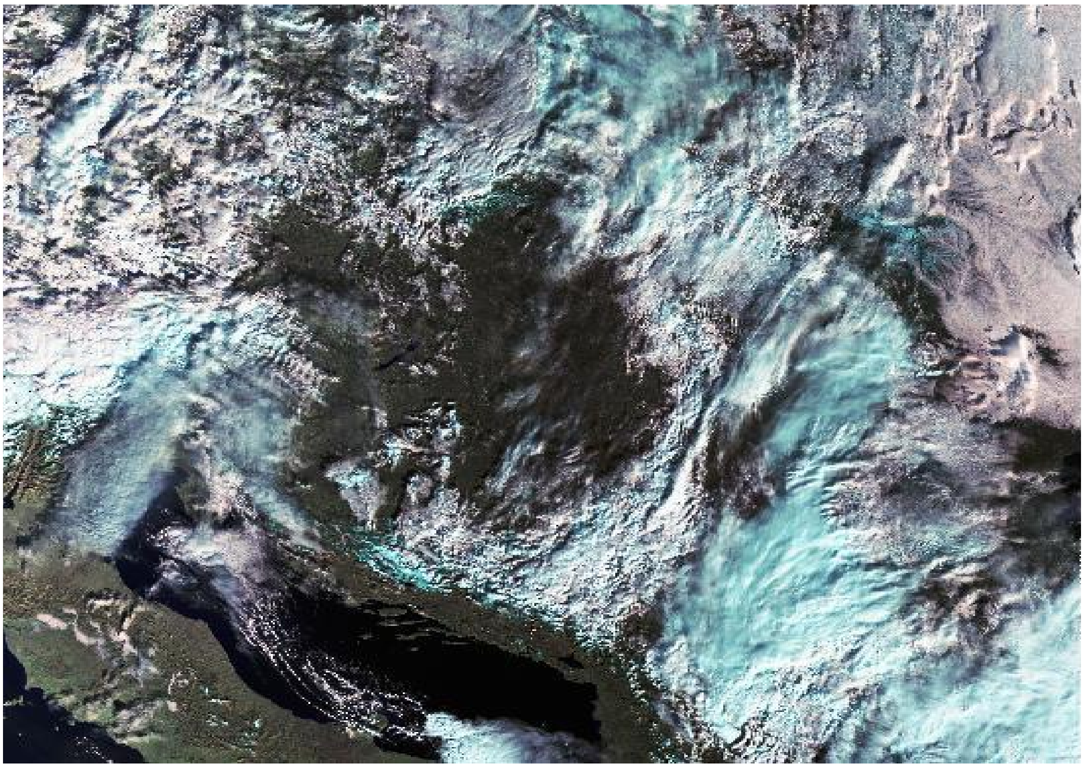
The answer to the question is that atmospheric correction is applied, as it removes the scattering and absorption effects from the atmosphere using scene.specific atmospheric data, the dark object subtraction as a technique that assumes small to none surface reflectance, and radiative transfer models. [10]

SENTINEL-2 Atmospheric Correction (S2AC) is based on an algorithm proposed in Atmospheric/Topographic Correction for Satellite Imagery. The method performs atmospheric correction based on the LIBRADTRAN radiative transfer model.[11]

Scene classification algorithm enables generation of a classification map which includes four different classes for clouds (including cirrus) and six different classifications for shadows, cloud shadows, vegetation, soils/deserts, water and snow.[11]

The Sentinel-3 mission uses a cloud shadow detection algorithm.

Image 3: image of a surface of interest with clouds, with AC applied, but still polluted.



The aim of this project, stated before, is to analyze sea water parameters. Therefore, the imposing question emerges for the tedious reader: what are the parameters of relevance. A vast discussion, followed with further paper search and analysis, proved the next four parameters to be of a key significance for the work: turbidity, dissolved oxygen, sea level temperature and chlorophyll concentracion.

Satellite remote sensing has been used as a valuable tool for supporting the implementation of the WFD, by deriving phytoplankton and cyanobacterial pigments such as chl a and phycocyanin (PC), total suspended matter (TSM), colored dissolved organic matter (CDOM), and the spectral attenuation coefficient (Kd).

Since DO is a non-optically active parameter, it cannot be directly quantified by spectral analysis. Therefore, estimating DO with remote sensing is not as direct as it was for chlorophyll-a. In [34] it was shown that water temperature and DO are highly correlated by using Pearson correlation analysis. They also tried to estimate in-situ DO concentration as a linear combination of six satellite-derived parameters by applying a multiple regression analysis. The parameters that they used were extracted from the MODIS-Aqua and the VIIRS sensors, and those parameters were SST, TSS, and chlorophyll-a for present and previous month, where TSS (total suspended sediments) is represented with the following formula ($R_{rs}(\lambda)$ is remote sensing reflectance at wavelength λ):

$$Log(TSS) = 0.649 + 25.623 \cdot [R_{rs}(555) + R_{rs}(670)] - 0.646 \cdot \frac{R_{rs}(490)}{R_{rs}(555)}$$

The results showed a strong correlation ($R^2 = 0.801$, $p < 0.001$) of DO observed in situ with a linear combination of SST, SST in one month prior (SST_{m-1}) and chlorophyll -a in one month prior ($Chl\text{-}a_{m-1}$). On the very end, the formula they got was:

$$DO = -0.131 \cdot SST - 0.132 \cdot SST_{m-1} + 0.066 \cdot Chl - a_{m-1} + 12.343$$

In [35] authors used Sentinel-2 level 2A images of Little Miami River for 2020 that were atmospherically corrected by Sen2Cor. They represented the DO by using spectral water indices such as spectral bands, Automated Water Extraction Index (no shadow) ($AWEI_{nsh}$) and the Sentinel Water Mask (SWM), where $AWEI_{nsh}$ and SWM are given with the following formulas:

$$AWEI_{nsh} = 4 \cdot (Green - SWIR1) - (0.25 \cdot NIR + 2.75 \cdot SWIR2)$$

$$SWM = \frac{(Band\ 2 + Band\ 3)}{(Band\ 8 + Band\ 11)}$$

where the bands for Sentinel are band 3 for Green, band 8 for NIR, band 11 for SWIR1 and band 12 for SWIR2. Next, they trained two machine learning algorithms – Random Forest (RF) and Support Vector Machine (SVM) – to predict DO concentrations using spectral predictors mentioned above. In the end they established that RF and SVM are effective in estimating DO concentrations with bands 5 (B5, vegetation red edge) and 8 (B8, NIR) as important predictors. Also, another often chosen predictor was band 3 (B3, green).

Chlorophyll a (chl a) is the main pigment in phytoplankton, which is known as one of the key parameters of the WFD which indicates the trophic status of water. Through photosynthesis, the phytoplankton converts CO2 and H2O into O2 and is responsible for primary production in the water column [36,37]. In addition, chl a is the main indicator of phytoplankton biomass [38,39,40] and can be used to determine the water clarity [41].

With itself being self explanatory, the description of turbidity is as follows; sea turbidity is a measure of the amount of cloudiness or haziness in sea water caused by individual particles that are too small to be seen without magnification. Highly turbid ocean waters are those with many scattering particles in them. [12]

The only other parameter of consideration in the project, in the taken position, was the Secci depth. Secchi depth refers to the depth at which a disk lowered into the water can no longer be seen from the surface. Secchi depth is related to water clarity and is a measure of how deep light can penetrate the water. Secchi depth (SD) is an approximate evaluation of the water transparency with a Secchi disk; it is essentially a function of the reflection of the light from its surface. SD is strongly influenced by the three optically significant constituents named before (Chl-a, TSM, and CDOM) and generally, it corresponds to 10% of the surface light [13].

SST (Sea Surface Temperature) could be extracted from the satellite data in near real time. Sentinel 3 has an embedded system that analyzes surface height, significant wave height, and wind speed to approximately calculate the sea level temperature. In that regard, calculation is highly precise and the data could be extracted from the SNAP data.

Every Sentinel has bands. Specifically, 12 of them, which are representing different light wavelengths, therefore different colors defined on the spectrum of light. Band symbolizes a color, and satellite information about a band is actually providing us of how much of radiance and reflectance of a specific length light has been detected.

Moreover, sentinels provide us with the masks; specifically designed pixel detection, which is parameterized based upon the field of interest. Sentinel satellites use a mask of X resolution and the constituent spectral band of Y wavelengths to detect a potentially outlying value of a value, that could be anything connected to the land and/or sea parameters.

The work had an real life worth and desirability, therefore the relevance could be assessed based upon it, if the criteria has been met: developing a fully functional, proven and empirically tested algorithms that would be beneficial for the scientists and for the sea parameters calculations, which would in that case be done remotely, using the satellite imagery.

This work's goals were met by using Creodias database, where the data has been extracted from. The exact dates and coordinates were used, based upon the real in-situ measurements that were provided for the work. After the data extraction, since there had been over 4000 data form the site, a ubuntu virtual machine was set up, where using Putty all data was transmitted, with the reason being the sheer enormous size of the data. The data was also preprocessed in the SNAP application, where many calculations were made.

Image 4: SNAP API where the data is selected and filtered through the band math: in this case CHL filter was applied.

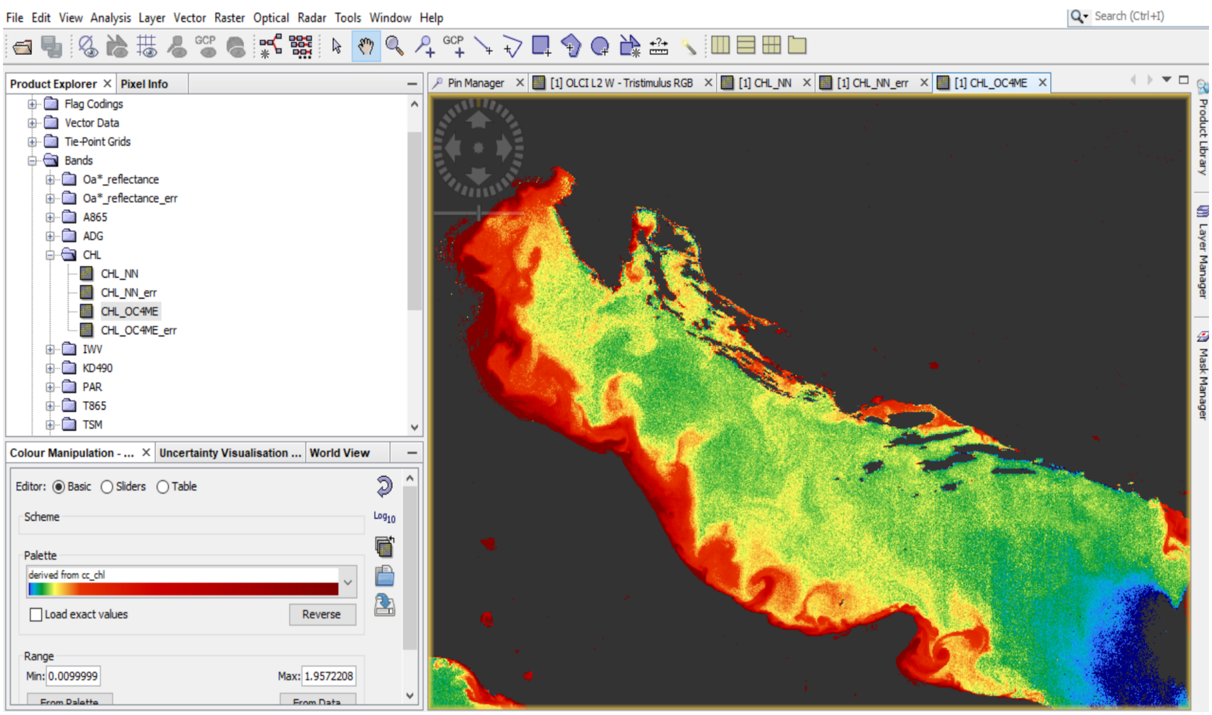
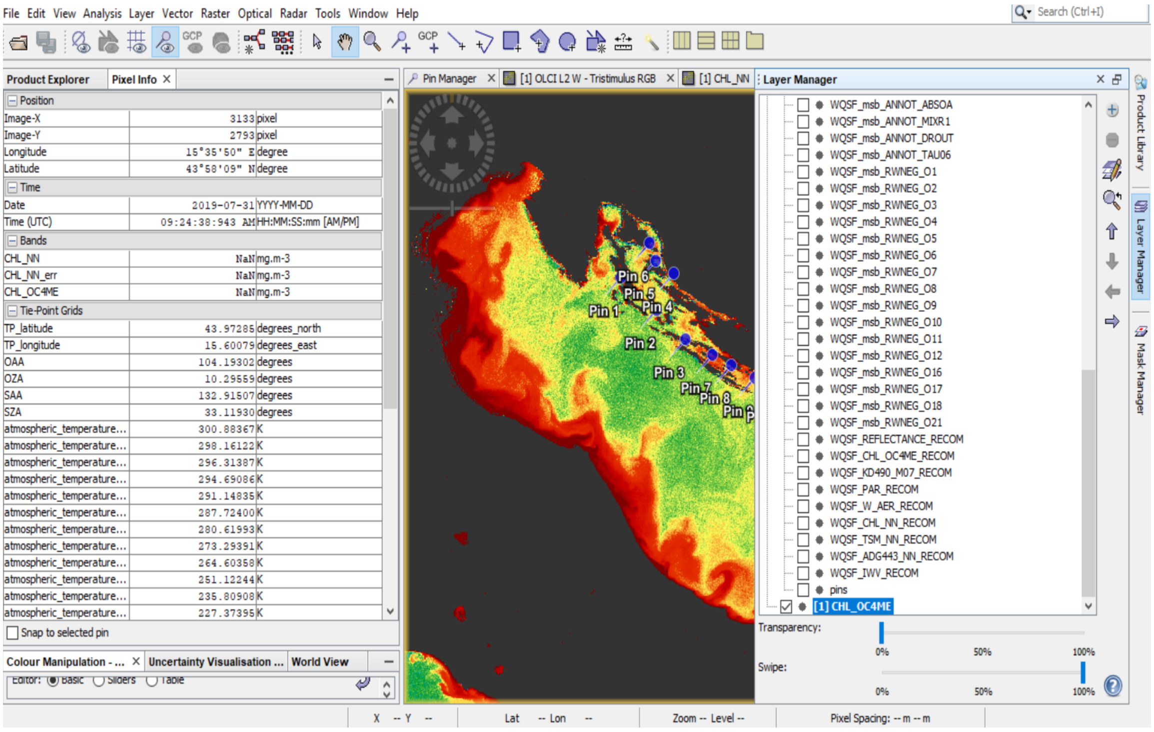


Image 5: pin selection, filtering and data tabeling, using embedded system. Specific coordinates were selected, as to be reliable and in line with the in-situ measurements.



Further steps required data extraction from the sentinel images; using Python coding language, the data about bands, masks and filters was extracted from every image. The development of the algorithms preceded, for the analysis to be possibly made on the dataset. Upon extraction using SnapPy (SnapPy is a Python library package for studying the topology and geometry of 3-manifolds, with a focus on hyperbolic structures. It is used as a bridge between Java and Python, since the SNAP app was developed in Java), the data was pulled through algorithmic calculations. Firstly, chlorophyll was examined. Approximately 20 different band ratios were examined, out of which three algorithms stood out as partially and mostly correct: OC4Me, SeaWiFs and CalCOFI four bands. Upon calculating data, reconsideration of the algorithms, data extraction and comparison to the real in-situ data, the algorithmic optimization took place. Optimization was performed on the three aforementioned algorithms, where numerous models were examined.

Sea temperature algorithms were less numerous, although precise in execution. The data preprocessing and extraction was the same, differing points came to be nothing else than the algorithm calculation.

Turbidity calculation was therefore the same, except for the formulas. To note, TBD was a parameter not available to compare to the SNAP, whilst the others were accessible.

Dissolved oxygen was considered as a parameter, but the complexity of the calculation was deemed to be too high for any proper calculation based on the accessible data and formulas, which served as a theoretical background, later proven through the analysis itself.

The band we considered are presented in the next image.

Image 6: A flowchart overview of the work.

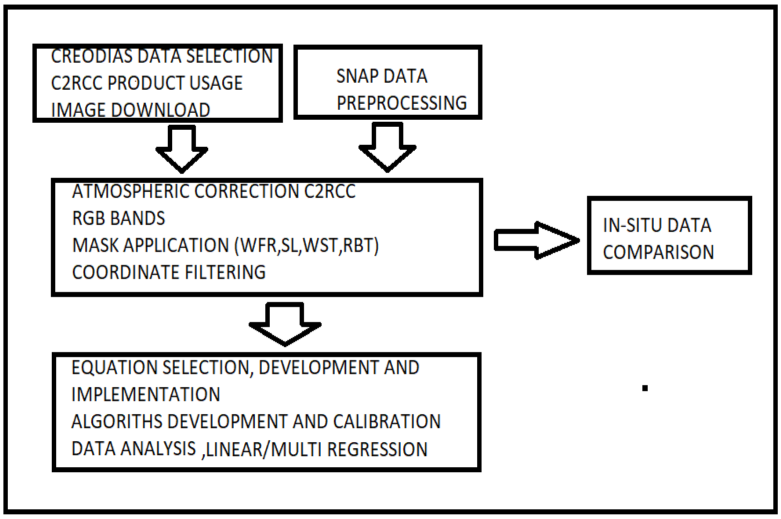


Image 7: bands and bands' resolutions

Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Ultra Blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Noticeably, the sentinel-3 has a spectral precision insufficient to dissect the minor difference between the depths of the data given, with the depths being 0,6,15 and 30 meters. Furthermore, the sentinel sensors are incapable to measure the data of non-surface area; to put it in other words, the capability of the measurement through the satellites remains on the surface level, being improbably fit for any other depth than that of 0 meters, which signifies water level. Therefore, of the 4000 in situ data collected from the Adriatic Sea, only 1000 measurements could have been used for the initial algorithm calibration and assessment, with the depth of 5 meters left as only other considerable choice of depth, for which the machine must had been calibrated, according to the data collected and calculated results.

Four evaluation metrics were used to characterize the performance of the models. The Coefficient of Determination (r2) gives an estimate of the proportion of variance explained by the model. The r2 is only a valid estimator if its significance is high. A significance level of 95% ($p < 0.05$) was used here.

3.1.1. C2RCC

The C2RCC (Case 2 Regional CoastColour processor) is developed for optically complex Case 2 waters, which use a large database of simulated ρw and TOA radiances. It is based on neural network technology and has been trained in extreme ranges of scattering and absorption properties. The C2RCC outputs results of ρw, IOPs, chl a, and TSM and provides the possibility to add additional background information such as salinity, elevation, ozone, temperature, and air pressure. [42].

3.1.2. Sen2Cor

Sen2Cor is an AC processor for vegetation applications and for scene classification, which is designed for S2 MSI data to generate L2 products. Sen2Cor relies on a large database of look-up tables and atmospheric radiative transfer model and is able to classify scenes into 12 classes (clouds, cloud shadows, vegetation, snow, water, cirrus, etc.). It outputs bottom-of-atmosphere (BOA) reflectance images, with aerosol optical thickness, water vapor, scene classification and quality indicators, such as cloud and snow probability. [43].

3.2. Algorithm optimization overview, correlations

3.2.1. Linear regression [21]

First, regression analyzes are usually used for forecasting and prediction, in which their application has major overlaps with the area of machine learning. Second, regression analysis can be used in some cases to determine causal relations between the independent and dependent variables.Simple Linear Regression is a case model with a single independent variable [22]. Simple Linear regression defines the dependence of the variable. $y = \beta 0 + \beta 1x + \varepsilon$. Simple regression distinguishes the influence of independent variables from the interaction of dependent variables[23].

3.2.2.Polynomial regression [21]

Polynomial regression [24, 25] is a type of regression analyzed in the nth degree of polynomial modeling of the relationship between independent and dependent variables. MLR is a statistical technique to predict the result of an answer variable, using a number of explanatory variables. The object of (MLR) is to model the linear relationship between the independent variables x and dependent variable y that will be analyzed. Polynomial regression is a special case of MLR in which the polynomial equation of data blends in with curvilinear interplay of the dependent and independent variables [26]. Model of polynomial [27, 28] is: $y = \beta 0 + \beta 1x + \beta 2x\ 2 + \cdots + \beta hx\ h + \varepsilon$ Where h is named the polynomial degree [29, 30].

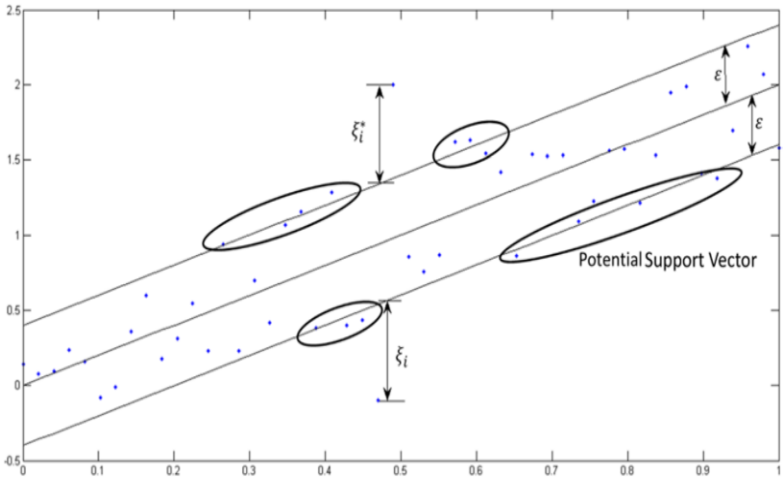
3.2.3.Ridge regression [31]

Ridge regression is a popular parameter estimation method used to address the collinearity problem frequently arising in multiple linear regression. The formulation of the ridge methodology is reviewed and properties of the ridge estimates capsulated. In particular, four rationales leading to a regression estimator of the ridge form are summarized. Algebraic properties of the ridge regression coefficients can be calculated, which elucidate the behavior of a ridge trace for small values of the ridge parameter (i.e., close to the least squares solution) and for large values of the ridge parameter. Further properties involving coefficient sign changes and rates-of-change, as functions of the ridge parameter, would be useful for specific correlation structures among the independent variables.

3.2.4.SVR (Regression model of SVM) [32]

SVMs solve binary classification problems by formulating them as convex optimization problems.The optimization problem entails finding the maximum margin separating the hyperplane, while correctly classifying as many training points as possible. SVMs represent this optimal hyperplane with support vectors. The sparse solution and good generalization of the SVM lend themselves to adaptation to regression problems. SVM generalization to SVR is accomplished by introducing an e-insensitive region around the function, called the e-tube. This tube reformulates the optimization problem to find the tube that best approximates the continuous-valued function, while balancing model complexity and prediction error. More specifically, SVR is formulated as an optimization problem by first defining a convex e-insensitive loss function to be minimized and finding the flattest tube that contains most of the training instances. Hence, a multiobjective function is constructed from the loss function and the geometrical properties of the tube. Then, the convex optimization, which has a unique solution, is solved, using appropriate numerical optimization algorithms. The hyperplane is represented in terms of support vectors, which are training samples that lie outside the boundary of the tube.

Image 5: Visual representation of a SVR model (one dimensional, linear SVR)



3.2.5. Random forests [33]

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. For random forests, an upper bound can be derived for the generalization error in terms of two parameters that are measures of how accurate the individual classifiers are and of the dependence between them. The interplay between these two gives the foundation for understanding the workings of random forests.The simplest random forest with random features is formed by selecting at random, at each node, a small group of input variables to split on. Grow the tree using CART methodology to maximum size and do not prune. Denote this procedure by Forest-RI. The size F of the group is fixed. Two values of F were tried. The first used only one randomly selected variable, i.e., F = 1. The second took F to be the first integer less than log2 M + 1, where M is the number of inputs.

3.2.6. R^2

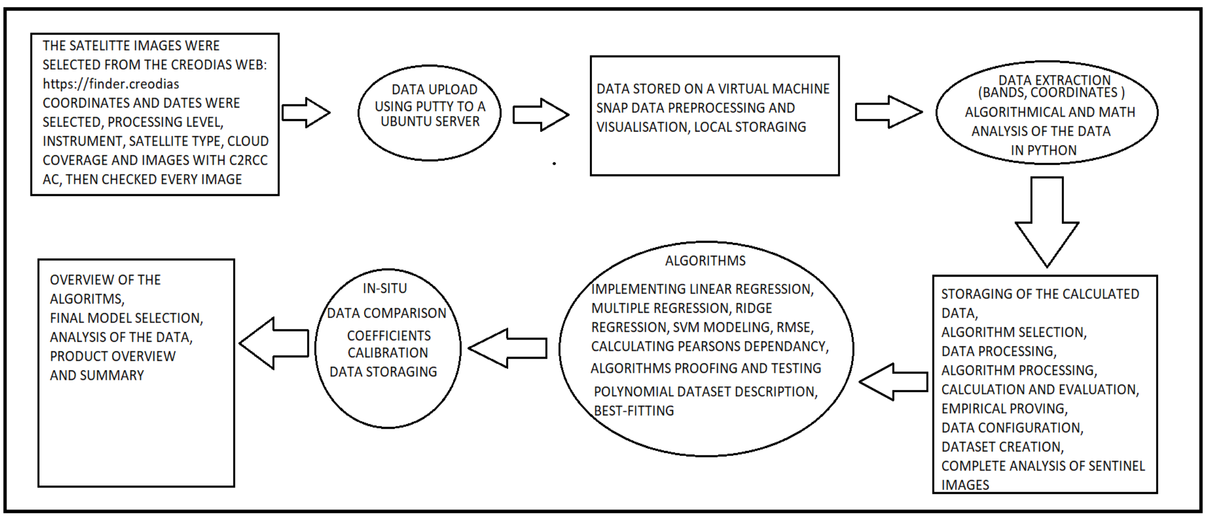
R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Whereas correlation explains the strength of the relationship between an independent and dependent variable, R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the R2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs. Ultimately it tells you whether the data and model are biased, not if the model is correct.

4.Solution description

4.1. Architecture

The implementation of the solution and a brief complete overview follows.
Creodias data was filtered by date, coordinates and cloud coverage. All of the images covered somewhat with clouds were not considered. That was the first filter applied. The satellite used was Sentinel 3, on non time critical level, Level 1A product, and the name of instrument varied dependant of the data considered: OLCI, WFR,WST and RBT were instruments of the interest.

Image 6: workflow of the project.



SNAP was used for the product evaluation and visualization. Band math was implemented through SNAP functions, color filtering was selected for the visual representation of the results. Via Putty, images were sent to the Virtual machine on Ubuntu.

Code for image openery was programmed, the data of interest was extracted (mostly bands for a specific coordinate, CHL algorithm of the Sentinel-3 data, NN data). Different formulas were considered, coded and implemented. Results were collected and compared. The excessively incorrect data was removed. Three chlorophyll algorithms were found satisfyingly correct and therefore developed and optimized. Others were deprived of optimization and later disconnected from the rest. A couple of algorithms for the temperature was shown. An algorithm for the turbidity was found seemingly correct. The data was processed through Python 3.6.8 with a specific SnapPy library for decompressing the data. The data was compared to the real in-situ measurements.

The dissolved oxygen calculation was in the end left out of this study.Definition of the parameters was given; hereby the equations will be presented upon which we will calculate the data and compare it to the in-situ data.

For CHL, models of linear regression, multi band regression and polynomial distribution were considered.

FIGURE 1. CHL equations.

Parameter	Algorithm	Type	Result Equation	Band Ratio (R),Coefficients (a)
CHL	POLDER	cubic	C = $10^{(a0+a1*R+a2*R^2+a3*R^3)}$	R = log (Rrs443/Rrs565) a = [0.438, -2.114, 0.916, -0.851]
CHL	CalCOFI two-band linear	power	C = $10^{(a0+a1*R)}$	R = log (Rrs490/Rrs555) a = [0.444, -2.431]
CHL	CalCOFI two-band cubic	cubic	C = $10^{(a0+a1*R+a2*R^2+a3*R^3)}$	R =log (Rrs490/Rrs555) a = [0.450, -2.860, 0.996, -0.3674]
CHL	CalCOFI three-band	multiple regression	C =$exp^{(a0+a1*R1+a2*R2)}$	R = ln (Rrs490/Rrs555) R2 = ln

				(Rrs510/Rrs555) a = [1.025, -1.622, - 1.238]
CHL	CalCOFI four-band	multiple regression	$C = exp^{(a0+a1*R1+a2*R2)}$	R = ln (Rrs443/Rrs555) R2 = ln (Rrs412/Rrs510) a = [0.753, -2.583, 1.389]
CHL	Morel-1	power	C = $10^{(a0+a1*R)}$	R = log (Rrs443/Rrs555) a = [0.2492, - 1.76]
CHL	Morel-2	power	$C = exp^{(a0+a1*R)}$	R = ln (Rrs490/Rrs555) a = [1.077835, -2.542605]
CHL	Morel-3	cubic	C = $10^{(a0+a1*R+a2*R^2+a3*R^3)}$	R = log (Rrs443/Rrs555) a = [0.20766, -1.82878, 0.75885, -0.73979]
CHL	Morel-4	cubic	C = $10^{(a0+a1*R+a2*R^2+a3*R^3)}$	R = log (Rrs490/Rrs555) a = [1.03117, -2.40134, 0.3219897, -0.291066]
CHL	SeaWiFs	max band ratio, multiple regression	C = $10^{(a0+a1*R+a2*R^2+a3*R^3 +a4*R^4))}$	R= log(max (Rrs443/Rrs555, Rrs490/Rrs555, Rrs510/Rrs555)), a=[¸0.3272, -2.9940, 2.7218, -1.2259,-0.5683]
CHL	OC4ME	max band ratio, polynomial algorithm	C = $10^{(a0+a1*R+a2*R^2+a3*R^3 +a4*R^4))}$	R=log10(max (Oa443,Oa490, Oa510)/Oa555) a=[0.3272,-2.994

				0,-3.249491, 2.7218,-1.2259 ,-0.5683]
CHL	OCI	band difference	C = C = $10^{(a0+a1*R)}$	R=Rrs555 - 0.5*(Rrs443 -Rrs670) a=[-0.4909, 191.659]
CHL	Globe color	polynomial	C = $10^{(a0+a1*R+a2*R^2+a3*R^3+a4*R^4))}$	R=log10(max (Oa443/Oa555, Oa490/Oa555, Oa510/Oa555)) a=[0.3205, -2.9139, 8.7428, -16.1811,9.0051]
CHL	MODIS Aqua	multiple regression	C = $10^{(a0+a1*R+a2*R^2+a3*R^3)}$	R=log10(max(Oa443/Oa547, Oa490/Oa547)), a=[0.2424, -2.7423,1.8017, 0.0015, -1.2280]
CHL	OC medium level CHL	band ratio, regression	chl = alfa *R_CHL + (1-alfa)*(chl_colour_index) alfa=(chl_colour_index-0.25)/(0.3-0.25) chl_colour_index=Oa555-0.5*(Oa443+Oa670) R_CHL= $10^{(a0+a1*R+a2*R^2+a3*R^3+a4*R^4))}$	R= log10(max(Oa443,Oa490, Oa510)/Oa555) a=[0.3272, -2.9940, 2.7218, -1.2259,-0.5683] A= -0.4909 B= 191.659

Out of the above written algorithms, after every was coded and tun through the available data, only three have been proved to be prone enough for the consideration ; OC4Me, SeaWiFs and CalCOFI four band. The description follows.

OC4Me Chlorophyll

Chlorophyll concentration, Chl, is defined by the "OC4Me" Maximum Band Ratio (MBR) semi-analytical algorithm, developed by Morel et al [16] (cf. OReilly et al. [17], for a more general description of such algorithms). It is the latest version of the MERIS pigment index algorithm, which is fully described in the MERIS ATBD 2.9 and in Morel et al [18]. A brief reminder is provided here. OC4Me is a polynomial based on the use of a semi-analytical model, itself based on the analysis of AOPs measured in situ over the past decades in various oceanic regions. [19][20]
It is expressed as:

$$log_{10}(Chl) = \sum_{k=0}^{4} (A_k * (log_{10}(R_j^i))^k)$$

is the ratio of irradiance-reflectance of band i, among 443, 490 and 510 nm, over that of band j at 560 nm. The band for the numerator is selected so that the ratio is maximized. The above computation is embedded within an iterative loop including the computation of the irradiance-reflectance from the directional water leaving reflectance that requires a bi-directionality correction relying on the Chl estimate. Convergence is reached when the difference between two successive Chl estimates are below a threshold.

CalCOFI Algorithms

The CalCOFI algorithms are derived from CalCOFI data [Mitchell and Kahru, 1998].
The CalCOFI two-band relates C to Rrs490/Rrs555 using a power equation.The CalCOFI two- band cubic is a third-order polynomial equation using Rrs490/ Rrs555. The CalCOFI three-band, a multiple regression equa- tion, has similarities with the OCTS-P algorithm and uses the Rrs490/Rrs555 and Rrs510/Rrs555 band ratios. The functional form of the CalCOFI four-band equation is similar to Cal- COFI three-band ratio.

SeaWiFs algorithms

SeaWiFS images are received by OSC HRPT stations, which create level la images. Level 1 a image data are raw, and all spacecraft and instrument telemetry are retained in raw form as in the Level 0 data. Algorithms for chlorophyll-a concentration detection using imagery from the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) have been previously widely used and have shown some good and accurate results in those papers.
The SeaWiFS sensor is unique in that the spectral bands are sensitive to fluctuations in ocean color that are due to pigment changes caused by variations of phytoplankton, changes in suspended matter, and changes in organic carbon, among others. [44]
A logical cause of the algorithms is believably justified, and a model had been created.

Noticeable similarity to the OC4Me algorithm only adds to the case that similar mathematical equations and ratio of the bands could effectively be used for the CHL calculations. The dissimilarities are the bands used, as well as the relations between them. The considered bands were that of wavelengths of 443,489,510 and 555 respectively. The other major distinction with the OC4Me were the coefficients inquired.

Figure 2: Turbidity equations of based on Chinese algorithms [14]

X	General form	Best band combination
$X1$	$R_{rc}(\lambda_1)$	$R_{rc}(\lambda_1) = B6$
$X2$	$R_{rc}(\lambda_1) - R_{rc}(\lambda_2)$	$R_{rc}(\lambda_1) = B6$, $R_{rc}(\lambda_2) = B4$
$X3$	$R_{rc}(\lambda_1) / R_{rc}(\lambda_2)$	$R_{rc}(\lambda_1) = B6$, $R_{rc}(\lambda_2) = B4$
$X4$	$\frac{R_{rc}(\lambda_1)-R_{rc}(\lambda_2)}{R_{rc}(\lambda_1)+R_{rc}(\lambda_2)}$	$R_{rc}(\lambda_1) = B6$, $R_{rc}(\lambda_2) = B4$
$X5$	$\frac{R_{rc}(\lambda_1)+R_{rc}(\lambda_2)}{R_{rc}(\lambda_1) / R_{rc}(\lambda_2)}$	$R_{rc}(\lambda_1) = B3$, $R_{rc}(\lambda_2) = B6$
$X6$	$\frac{R_{rc}(\lambda_2)-R_{rc}(\lambda_3)}{R_{rc}(\lambda_1) / R_{rc}(\lambda_2)}$	$R_{rc}(\lambda_1) = B3$, $R_{rc}(\lambda_2) = B6$, $R_{rc}(\lambda_3) = B4$
$X7$	$\frac{R_{rc}(\lambda_2)+R_{rc}(\lambda_3)+R_{rc}(\lambda_4)}{R_{rc}(\lambda_1) / R_{rc}(\lambda_2)}$	$R_{rc}(\lambda_1) = B3$, $R_{rc}(\lambda_2) = B6$, $R_{rc}(\lambda_3) = B4$, $R_{rc}(\lambda_4) = B5$
$X8$	$\frac{R_{rc}(\lambda_2)-R_{rc}(\lambda_3)+R_{rc}(\lambda_4)}{R_{rc}(\lambda_1) / R_{rc}(\lambda_2)}$	$R_{rc}(\lambda_1) = B3$, $R_{rc}(\lambda_2) = B6$, $R_{rc}(\lambda_3) = B4$, $R_{rc}(\lambda_4) = B5$

Figure 3: Turbidity equations modeled with Linear and Polynomial Regression [15]

Regression model equation for estimating turbidity	Band combination (=x)
$367.82x^2 - 976.42x + 649.13$	B2/B3
$971.47x^2 - 1468x + 55.84$	B3/B2

$725.32x^2$ - 858.52x + 255.91	B4/B3
$118.8x^2$ - 401.92x + 341.62	B2/B4
$387.41x^2$ - 1103x + 786.37	B1 + (B1/B2)
20.981x - 8.901	B3/B2
102.56x - 5.5003	B3+B4
90.319x - 10.775	B2+B3+B4
20.254*ln(x) + 46.009	B2+B3
14.735*ln(x) + 30.802	B2+B3+B4
0.4329 - B1*54.6776 + B3*42.4338	NONE
0.4532 - B1 + 56.9454 + B3*43.5723	NONE

4.2. Functional specification

The goal of the project was to develop a functional algorithm that would use satellite imagery for the water analysis.

Algorithms would have to consider various parameters, and out of the groupage of them, the 4 that were selected as the most relevant and useful for the initial calibration and systematization were CHL, dissolved oxygen, turbidity and sea temperature.

The observed sea was the Adriatic Sea.

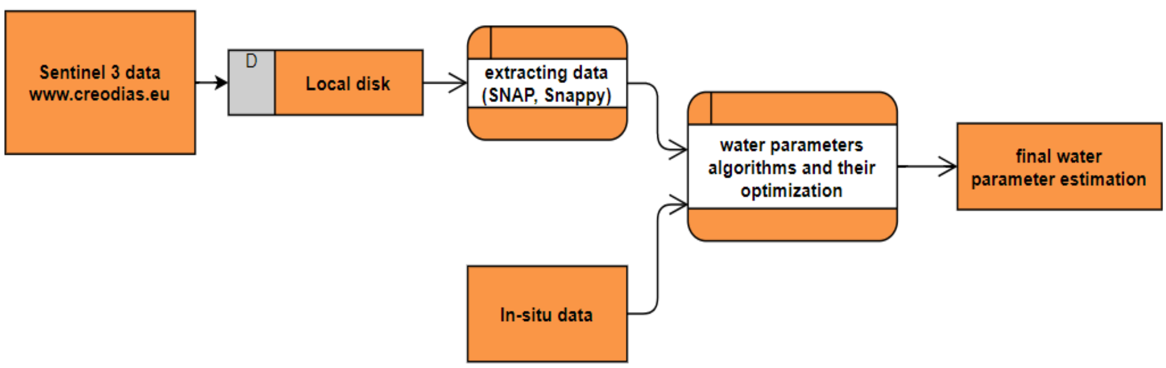
The dissolved oxygen was dropped out of the analysis due to limited time for the development.

All the code was deployed to the GitHub repository, for which the link to is given in the subsection *Source Code Organization.* The accessibility of the code lies on the permission of Ericsson Nikola Tesla, who claims all the responsibility and rights for this project, therefore the algorithms developed are under the jurisdiction of ENT groupation.

The code developed does the following:

- data extraction: a path to the locally or virtually downloaded data must be given, alongside with the latitude and longitude of a spot observed (in terms of location). Upon writing the two down in the code, all the bands could be extracted,as well as the pixel info. To note, to know which data is reachable, one must go in the documentation of the Sentinel-3 and product of Sentinel-3 used, as the extraction of the bands is name related.
- data processing: all the data of interest can be scored, used and manipulated within the functions, and upon completion, you would be eligible to extract it. A wide and extensive search for the 'right' formulas had been done, a variety of them was used for the calculation of the parameters of consideration. The functions do differ since they have been using different band rations for the calculation of the parameters. An exploration of that sort is valuable if not for the accuracy, then for the given insight into how to adapt your algorithm and how to make a rational mathematical function that would be declared as reasonable, useful and for whose implementation a slight moderation would be sufficient to generate any sort of a correct answer.
- data comparison: the in-situ data was given, the coordinates for every measurements, alongside the data of extraction. were used to get the exact images of the place on specific data from the Creodias database. The data was stored in an Excel file, so an excel reading python script was developed. Then, the comparison between the two values took place, RMSE , correlation (Spearmans) and a standard deviation, as well as the errors were calculated.
- algorithm optimization: we explored and made few optimizations of the parameters of the algorithms, of which being Linear Regression, Polynomial Regression, Ridge Regression, SVM (SVR) Model and Random Forests.All the code is available on the GitHub repository. Adaptation of the model was made, coefficients were perfected to a point, and the final equation model was done.

4.3. Information flows



Creodias, as a satellite database, was used for the data gathering, selecting and extracting.

Upon extraction, satellite images were downloaded to the local disks of the laptops (simultaneously put on a virtual machine and stored there for the common and joint usage).
Using SNAP desktop programme, using Python's library Snappy and a code generated in it, the necessary data, which would be bands, masks and pixel information needed for the analysis, was extracted and used in a code, run through and the results were calculated upon the extraction.
When the final results were calculated, a writing function generated excel table where the results were visualized and compared to the in-situ data, which was at the moment already given. In-situ data was selected based on the depth (near surface data was only relevant for the analysis).
Upon comparison of the two, an excel reading code was made for the data extraction, subsequently algorithm optimization was used for the parameter optimization.RMSE, R^2 and a correlation (in this case, Spearman's was used) were calculated and stored as the final results.
Newly adapted parameters were adapted to fit the best for the accuracy to be as high as possible, and it all was shown in the results.
In the end, final estimation came, as the observed functions were optimized for the Adriatic Sea case.
The final concluding took place and the accuracy of the developed algorithms were discussed.
All the code was pushed to the GitHub repository and is internally available.

4.4. Results and overview

Figure 4: temperature algorithms results, correlation, RMSE, deviation

Algorithm	Mean error/K	Standard deviation/K	Mean R^2	R^2 standard deviation
McClain	4.56	27.35	NULL	NULL
Price	3.04	26.5	NULL	NULL
Becker Li	2.77	26.35	NULL	NULL
Prata Platt	4.31	26.32	NULL	NULL
Sobrino	3.24	26.22	NULL	NULL
Linear Regression	-0.2	0.69	-0.9	1.53
Ridge Regression	-0.2	0.68	-0.9	1.54

SVR	-0.56	0.39	0.49	0.14
Polynomial Regression	0.44	9.8	-517	4396
Random Forests	-0.1	0.26	0.68	0.03

Polynomial regression model stood out as completely incorrect. but it must be mentioned that the majority of the results had been closer to the real values, with a few bumps deviating as high as seen in the table. The algorithm was inconsistent and ultimately useless and erroneous.
A negative R squared means that the appropriate model provides worse results that the horizontal line. It can be concluded that the developed model does not fit and follow the trend of the data.A negative R squared is only possible with linear regression when either the intercept or the slope are constrained so that the "best-fit" line (given the constraint) fits worse than a horizontal line. With nonlinear regression, the R squared can be negative whenever the best-fit model (given the chosen equation, and its constraints, if any) fits the data worse than a horizontal line.
With that being said, Linear and Ridge regression are not applicable and are seemingly incorrect approximations.
Algorithms [1 to 5 in the figure 4] have a mean error of 2.77 Celsius/Kelvin degrees and above, which is pragmatically not usable for the temperature analysis of any water. Surely the parameters of the algorithms must be optimized to fit, otherwise no silver lining of such algorithms could be seen.
With SVR being just under 0.5 mark, Random Forest is the only developed optimization model appropriate for its purpose. R squared being 0.68 and standard deviation 0.03, the model is labeled as fittable.

Figure 5: CHL calculation, SeaWiFs model

Algorithm	Mean error	Standard deviation	Mean R^2	R^2 standard deviation
SeaWiFs	0.07	0.41	NULL	NULL
OC4Me	0.34	0.76	NULL	NULL
CalCOFI_4b	0.27	0.56		
Linear Regression	0.01	0.05	0.01	0.83
Ridge Regression	0.0	0.05	0.2	0.05
SVR	0.04	0.04	0.32	0.11

Polynomial Regression	0.0	0.09	-3.71	37.91
Random Forests	0.01	0.04	0.19	0.31

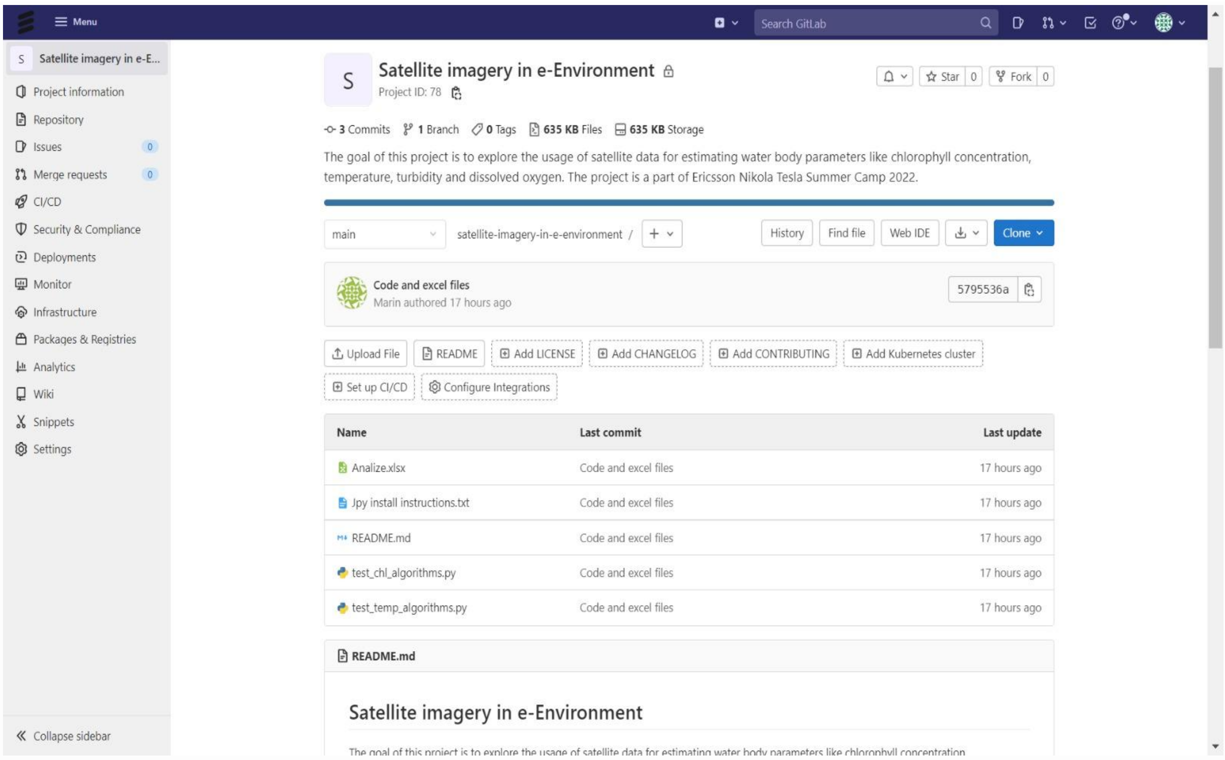
SeaWiFs has been relatively under the threshold of 0.5 standard deviation, with the Random forest algorithm being the prominently best, but the concern is that the R squared values are so low that no dependency is outlined out of the model and the model is therefore inadequate and deficient under the conditions that the subpar R squared means the model developed can not be used for the analysis.

5.Source code organization

Code was pushed to the GitHub repository. It can be found on the following link:

http://192.168.209.11/sc22_satellite/satellite-imagery-in-e-environment.git

Image 7: the Git repository snippet,a complete overview of the treelayout of it.



6.Future work

This work should be considered a preview of work to come. Without doubts, the algorithms could have been more accurate. But, the majority of the problem is due to incapability to solve the problems with the pixel resolution, with the cloud coverage and the amount of light dissipated on the third source entities that are not the sea surface.

It is of utmost importance to have an insight into how much of an error is caused by the fact that the satellite images could compromise the state of the atmosphere. This was noticed because of persistent atmospheric effects, which indicate that either sun glint, light haze, or cirrus clouds were present, or potentially a combination of all of these conditions.

The extensive search of the proper equation would result in much fewer work mistakes and it would prophel the accuracy to a more satisfying level. It is true that all the work previously done has been nothing but a specific case study, incomplete, slightly inaccurate and absolutely not applicable for a wide usage. Therefore, the next study must result in an algorithm that would be able to evaporate all the imperfections of the satellite images as such.

What should be also done is a wider exploration of various parameters, mutual codependency, interchangeability and variance of the results, as it would be beneficial when adopting any sort of an algorithm for a wider use. It must be known that the seas in general have different amounts of suspended matter, therefore the algorithms must be adapted to that fact. Additionally, since the satellites have been proven to be incorrect when measuring values for the depth different of that of 0 meters,or should it be put, they are adoptable and valuable to use only for the surface values, an exploration should be done in the regard to the questions: could it be possible to develop a correction for the depth? Furthermore, since the precision of the Sentinel-3 is around 100 meters, if the in-situ data is collected near the shore, pixel resolution and values is compromised and ‘polluted’, so it would be of a importance to declare the proper distance of the shore where the images should be taken, which products and masks should be applied.

Since the standard deviation of any of the paper’s algorithms is high, and the connection between the results and data marginal, a correction of the band relations, of the bands and the equation should be introduced in the future work. To debunk the blurred correlations between the parameters, a multi dependency should be explored. The reason is clear: an overall, general algorithm would likely have to use more than a parameter at the time to be characterized as adaptable.

7. Tips and tricks

Several issues were faced due to snappy's incompatibility with other installments: Python version 3.6.8 was required, snappy was added to path, JPY had to be added to the path as well. Another issue was similarity with the aforementioned other Python library, therefore no code could have been run on a couple of laptops.
Suggestion goes out to read the official documentation and to go over the user manual.
The correct import could be checked by typing from snappy import ProductIO,and if a mistake is reported, program can not be run, since the above described problem occurred. A reinstallment is required and another import of the libraries.

There are numerous satellite imagery sources available for the documentation, analysis, filtering,etc. Many globally available databases are in the reach, but very few have been sufficient enough for the project, in terms of data availability, filters, many did not have Sentinel-3 images and specific filters. Therefore, from the original idea of using Copernicus database, alongside with the SciHub, the focus has shifted to the Creodias, where the vast majority of the data could have been reached and downloaded all the time. With SciHub, one might have a problem reaching the documents and images at a given time, since they are often offline;meaning, cannot be accessed instantly and it requires some time for the images to be available. No similar problems were noticed using Creodias.

Cloud coverage happens to be a significant problem, or a major threat for any analysis, since no matter the AC algorithm is applied, or any kind of filter used, many images for the specific coordinates and dates, which was obviously the case and is the case when comparing any Sea parameters' data to the in situ measurements, clouds could have been seen on the images and imposed insurmountable problem for the analysis and algorithmical calculation, as it polluted the pixel info of any given image. Therefore, proper and deeper analysis followed with search of a solution would be necessary, as for this work manual selection had to be applied when selecting the images, which proved to be time consuming and virtually unnecessary, but still presented a problem to solve.

No other depth than that of a surface level should be considered in an approach to this analysis; satellite imagery in constrained to the near surface level and accuracy significantly drops when reaching lower depths. Therefore, this paper has been constructed based on the results from the comparison between in-situ measurements from the depth of 0 meters with the algorithmical results from the equations previously described.An additional space will probably be required for any more extensive research, that would consist of a numerous data from the field, since for every data and specific coordinates, a SNAP image must be downloaded, thatfore a bigger data storage should be prepared. Also to note, it is important to know that the procedure of selecting, dissecting and transmitting the SNAP images takes some time which should be reserved in advance. The approximation that would be given in regard to the time required to do so would probably be days.

8. Installation and run instructions

In order for the proper completion of the project, several installments and prerequisites were obligatory. The whole project setup was on Windows 10 laptops.

SNAP v9.0.0 was used for the data visualization and imagery processing, the main download was used for the Windows 64 bit architecture.

The Ubuntu Subsystem was used for SNAP imagery storage, since every SNAP image was of a size ranging from 140 up to 700MB, therefore for the 4200 data and belonging images, a virtual machine with extended storage had to be used.

Putty was downloaded to perform the upload from the SNAP to the Ubuntu server, with a specific security key given. A session was created and accessed from WinSCP, wherefrom all the data was uploaded.

WinSCP v5.21 was thatfore used.

The C2RCC operators are built-in into SNAP.

Python script was made and run in Miniconda3 for Windows 64. But, for the code that we programmed in python, in order for us to be able to open the images from SNAP, to extract data needed for the calculation from it and for storing it, since SNAP was written in Java, a so-called bridge was needed. Therefore, a library called SnapPy was used and installed, as well as Pandas, JPY, Traceback, NumPy and math.

Snappy is a python library for the snappy compression library from Google. Similarly, SnapPy is a program for studying the topology and geometry of 3-manifolds, with a focus on hyperbolic structures. Version 3.0.3 was installed.

All the code was written in VisualStudioCode, locally run in terminal and through Windows PowerShell.

Python version 3.6.8 was used.

Our product requires Python 3.6.8 and belonging libraries imported in order to be run.

All the code was pushed to a GitHub repository of Ericsson Nikola Tesla, where a repository was created.

9. User manual

ReadME of the project follows:

Image 8: first part of the ReadMe

```
# Satellite imagery in e-Environment

The goal of this project is to explore the usage of satellite data for estimating water body parameters like chlorophyll concentration,
temperature, turbidity and dissolved oxygen. The project is a part of Ericsson Nikola Tesla Summer Camp 2022.

## Installation

* First step is to follow instructions on installing Java JDK and jpy in the  contained Jpy install instructions text file.

* The easiest way to configure your Python installation for the usage of SNAP-Python (snappy) interface is to do it during the
installation of [SNAP](https://step.esa.int/main/download/snap-download/). Within the installer you can simply activate a checkbox and
select the path to the python executable. It is suggested to use either Python 3.5 or 3.6. For higher Python version it needs a manual
build of jpy.
Manual configuration instructions can be found [here](https://senbox.atlassian.net/wiki/spaces/SNAP/pages/50855941/Configure+Python+to+use
+the+SNAP-Python+snappy+interface)


Snappy installation success can be tested if the following code runs:

```
python36
from snappy import ProductIO
```


* Use the package manager [pip](https://pip.pypa.io/en/stable/) to install:

```
pip install numpy
pip install pandas
```

Image 9: second part of the projects ReadMe

```
## Setup
To run this project, download the code locally:

```
$ cd ../<dest_folder>
$ git clone
$ cd satellite-imagery-in-e-environment

```

## Data
Satellite images used are Sentinel 3 images from European Union's Earth observation programme, Copernicus. They can be downloaded from
[EUMETSAT](https://coda.eumetsat.int/#/home) or [CREODIAS](https://finder.creodias.eu/#).

In-situ data used to test algorithms accuracy was accuired in different months of 2016. - 2019. and can be found in contained file
'Analize.xslx'.

## Authors and acknowledgment
The work on this project was done by Marko Haralović, Mirna Lovrić and Marin Maletić with the help of Filip Wolf and mentors Goran Kopčak
and Veronika Vidak-Ozretić, employers of Ericsson Nikola Tesla.

## License
This project ownership is claimed by [Ericsson Nikola Tesla](https://www.ericsson.hr/).

## Project status
Project will be continued after the end of Summer Camp.
```
```

References

[1] Klein, T.; Nilsson, M.; Persson, A.; Håkansson, B. From Open Data to Open Analyses—New Opportunities for Environmental Applications? Environments 2017, 4, 32.

[2] Copernicus. Available online: www.copernicus.eu (accessed on 12 September 2022).

[3] ESA Sentinel Online, Sentinel-2. Available online: sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed on 12 September 2022).

[4] ESA Sentinel Online, Sentinel-3. Available online: sentinel.esa.int/web/sentinel/missions/sentinel-3 (accessed on 12 October 2019). 20. Globolakes. Global Observatory of Lake Responses to Environmental Change. Available online: www. globolakes.ac.uk (accessed on 12 September 2022).

[5] Alikas, K.; Krazer, S. Improved retrieval of Secchi depth for optically complex waters using remote sensing data. Ecol. Indic. 2017, 77, 218–227. [CrossRef]

[6] Verpoorter, C.; Kutser, T.; Seekell, D.; Tranvik, L. A Global Inventory of Lakes Based on High-Resolution Satellite Imagery. Geophys. Res. Let. 2014, 41, 6396–6402.

[7] Bresciani, M.; Cazzaniga, I.; Austoni, M.; Sforzi, T.; Buzzi, F.; Morabito, G.; Giardino, C. Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. Hydrobiologia 2018, 824, 197–214.

[8] Grendaite, D.; Stoneviˇcius, E. Chlorophyll-a concentration retrieval in eutrophic lakes in Lithuania from ˙ Sentinel-2 data. Geol. Geogr. 2018, 4, 15–28.

[9] European Space Agency Sentinel-2 Web Page. Available online: https://sentinel.esa.int/web/sentinel/ missions/sentinel-2 (accessed on 12 September 2022).

[10] Chavez PS (1988) An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens Environ 24(3):459–479,CrossRef

[11] Copernicus, https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm ,(accessed on 12 September 2022).

[12] The influence of increasing water turbidity on the sea surface temperature in the Baltic Sea: A model sensitivity study,Journal of Marine Systems , Volume 88, Issue 2, November 2011, Pages 323-331,[CROSS:REF]

[13] Wetzel, R.G. Limnology. Lake and River Ecosystems, 3rd ed.; Academic Press: San Diego, CA, USA, 2001. , [Cross:ref]

[14] Innovative GOCI algorithms to derive turbidity in highly turbid waters: a case study in the Zhejiang coastal area,Zhongfeng Qiu, Lufei Zheng,Yan Zhou, Deyong Sun, Shengqiang Wang,and Wei Wu [CROSS:REF]

[15] Modelling Reservoir Chlorophyll-a, TSS, and Turbidity Using Sentinel-2A MSI and Landsat-8 OLI Satellite Sensors with Empirical Multivariate Regression,Yashon O. Ouma , Kimutai Noor, and Kipkemoi Herbert [CROSS:REF]

[16] Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., et al. (2007).Optical properties of the "clearest" natural waters.Limnology and Oceanography, 52, 217-229.

[17] O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel,D. A., Carder, K. L., Garver, A., et al. (1998). Ocean color algorithms for SeaWiFS. Journal of Geophysical Research, 103, 24937-24953.

[18] [Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007).Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach.Remote Sensing of Environment, 111, 69-88.

[19] 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-Oceans, 103, 31033-31044

[20] Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research-Oceans, 106, 7163-7180

[21] A Review on Linear Regression Comprehensive in Machine Learning, Dastan Hussen Maulud, * , Adnan Mohsin Abdulazeez , Journal of Applied Science and Technology Trends Vol. 01, No. 04, pp. 140 –147, (2020), [CROSS:REF]

[22] Abdulazeez, A., Salim, B., Zeebaree, D., & Doghramachi, D. (2020). Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol.. Maulud et al. / Journal of Applied Science and Technology Trends Vol. 01, No. 04, pp. 140 –147, (2020) 146

[23] M. S. Acharya, A. Armaan, and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5.

[24] M. C. Roziqin, A. Basuki, and T. Harsono, "A comparison of montecarlo linear and dynamic polynomial regression in predicting dengue fever case," in 2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC), 2016, pp. 213-218.

[25] A. K. Prasad, M. Ahadi, B. S. Thakur, and S. Roy, "Accurate polynomial chaos expansion for variability analysis using optimal design of experiments," in 2015 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), 2015, pp. 1-4.

[26] Y. Chen, P. He, W. Chen, and F. Zhao, "A polynomial regression method based on Trans-dimensional Markov Chain Monte Carlo," in 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018, pp. 1781- 1786.

[27] G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Color correction using root-polynomial regression," IEEE Transactions on Image Processing, vol. 24, pp. 1460-1470, 2015.

[28] N. N. Mohammed and A. M. Abdulazeez, "Evaluation of partitioning around medoids algorithm with various distances on microarray data," in 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2017, pp. 1011-1016.

[29] H. Jie and G. Zheng, "Calibration of Torque Error of Permanent Magnet Synchronous Motor Base on Polynomial Linear Regression Model," in IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 318-323.

[30] H. Niu, Q. Lu, and C. Wang, "Color correction based on histogram matching and polynomial regression for image stitching," in 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), 2018, pp. 257-261.

[31] Ridge Regression Condence Machine, Ilia Nouretdinov,Tom Melluish, Volodya Vovk, [CROSS:REF]

[32] Efficient learning machines,Theories, concepts and applications for engineers and system designers, Marietteb Awad,Rahul Khanna [CROSS:REF]

[33] Random Forests,LEO BREIMAN Statistics Department, University of California, Berkeley, CA 94720, Machine Learning, 45, 5–32, 2001, [CROSS:REF]

[34] https://www.sciencedirect.com/science/article/pii/S0160412019327291 , accessed on 14th September 2022

[35] https://reader.elsevier.com/reader/sd/pii/S2352938522000672?token=7620DBD04E074BAC063AD8A963073BCAEA04F8A942438ADA9846E8F26DE8FE8BA7A93BA9ABB454B3F3A288308C41D923&originRegion=eu-west-1&originCreation=20220901071731 , accessed on 14th September 2022

[36] Matthews, M.W. A current review of empirical procedures of remote sensing in Inland and near-coastal transitional waters. Int. J. Remote Sens. 2011, 32, 6855–6899. [CROSS:REF]

[37] Duan, H.; Zhang, Y.; Zhang, B.; Song, K.; Wang, Z. Assessment of chlorophyll-a concentration and trophic state for lake chagan using landsat TM and field spectral data. Environ. Monit. Assess. 2007, 129, 295–308. [CROSS:REF]

[38] Moses, W.J.; Gitelson, A.A.; Berdnikov, S.; Povazhnyy, V. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data—Successes and challenges. Environ. Res. Lett. 2009, 4. [CROSS:REF]

[39] Wozniak, M.; Bradtke, K.M.; Krezel, A. Comparison of satellite chlorophyll a algorithms for the Baltic Sea. J. Appl. Remote Sens. 2014, 8. [CROSS:REF]

[40] Zhang, Y.; Ma, R.; Duan, H.; Loiselle, S.; Xu, J. A spectral decomposition algorithm for estimating chlorophyll-a concentrations in Lake Taihu, China. Remote Sens. 2014, 6, 5090–5106. [CROSS:REF]

[41] Salem, S.I.; Higa, H.; Kim, H.; Kobayashi, H.; Oki, K.; Oki, T. Assessment of chlorophyll-a algorithms considering different trophic statuses and optimal bands. Sensors (Switzerland) 2017, 17, 1746. [CROSS:REF]

[42] Brockmann, C.; Doerffer, R.; Marco, P.; Stelzer, K.; Embacher, S.; Ruescas, A. Evolution Of The C2RCC Neural Network For Sentinel 2 and 3 For The Retrieval of Ocean. In Proceedings of the conference held Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016

[43] Uwe, M.-W.; Jerome, L.; Rudolf, R.; Ferran, G.; Marc, N. Sentinel-2 Level 2a Prototype Processor: Architecture, Algorithms and First Results. In Proceedings of the conference held on ESA Living Planet Symposium, Edinburgh, UK, 9–13 September 2013; pp. 3–10

[44] ANALYSIS OF SEAWIFS OCEAN COLOR ALGORITHMS FOR LAKE ERIE, Christopher Everett Wells, B.S. The Ohio State University , 2005 [CROSS:REF]

Appendix A: List of acronyms and abbreviations

AC	Atmospheric Correction
ACOLITE	Atmospheric Correction for OLI Terra
C2RCC	Case 2 Regional CoastColour
CHL	Chlorophyll
CDOM	Chromophoric dissolved organic matter
DO	Dissolved oxygen
HPLC	High performance Liquid Chromatography
Kd	Light attenuation coefficient
MBR	Maximum Band Ratio
MCI	Maximum Chlorophyll Index
MERIS	Moderate Resolution Imaging Spectroradiometer
MSI	MultiSpectral Instrument
NIR	Near-Infra Red
NN	Neural Net
OLCI	Ocean and Land Colour Instrument
OC	Ocean Colour
OC4Me	Ocean Colour 4 Meris
OCI	Ocean Colour Index
OLI	Operational Land Imager
RBT	Radiance and Brightness Temperature
RMSE	Root Mean Squared Error
SD	Secci Depth
S2	Sentinel 2
S3	Sentinel 3
SR	Spectral Reflectance
TSM	Total Suspended Matter
TSI	Trophic State Index
TBD	Turbidity
WFR	Water Full Resolution
LW	Water Leaving Radiance
SST	Sea Surface Temperature