Satellite Imagery in e-Environment

Marko Haralović[1], Mirna Lovrić[2], Marin Maletić [3]
1,3: University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
2: University of Zagreb, Faculty of Science, Department of mathematics, Zagreb, Croatia
e-mails: marko.haralovic@fer.hr, mirnlovr.math@pmf.hr, marin.maletic@fer.hr

***Abstract* - The traditional methods of estimating water quality rely on manual sampling and measuring of specific water parameters. Taking in consideration the fact that these methods are time-consuming and expensive, the idea, and necessity, of satellite remote sensing to replace in-situ measurements was born. We fixed our focus on Copernicus Sentinel satellites to estimate four water parameters of interest: chlorophyll-a concentration, water temperature, turbidity and dissolved oxygen, all equally important in estimating the trophic state of water bodies We propose methods for automation of satellite data extraction through Python and algorithms for estimating chlorophyll-a concentration and temperature of the Adriatic Sea. We present various equations and mathematical relations found and document the best acquired results of existing optimized algorithms. The major impediment is the state of atmosphere and the Satellite precision, and it is felicitous to only the surface level of the water analysis. All of this can be used in places that require more frequent water quality testing, like fish farms, with the advantage of cutting financial and time costs of managing them.**

***Index Terms* – Chl-a, OC4Me, SeaWiFs, remote sensing, Satellite monitoring, Sentinel-2, Sentinel-3, water parameter extraction algorithms**

I. Introduction

Satellite images are images of Earth collected by satellites orbiting around the globe. With the variety of the open-source databases of the imagery, the selected database for this paper was Creodias. Creodias is an environment that serves as a source of the processed Earth Observation (EO) data. It is formed out of Copernicus Sentinel data and services, among others. Copernicus is the European Union's Earth observation program that offers satellite Earth Observation and in-situ data.[1] 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 (about 984.25 ft) x 300m (about 984.25 ft) resolution (Sentinel3 OLCI). The visualization and image overview were done through SNAP, or the Sentinel Application Platform, which is 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. There are many types of satellites and a vital part for remote sensing is proper satellite selection. The two mainly used and observed satellites in this paper were Sentinel-2 and Sentinel-3. Sentinel in general 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. Polar-orbiting satellites carry multiple instruments, including optical imagers. Sentinel 2 has 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 [2]. 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 [3]. OLCI was built for water monitoring and has well placed spectral bands for that purpose. However, its rather low spatial resolution and large imaging area allows the study of only bigger seas and oceans [4].

The need can be put forth as this: remote sensing had not been incorporated in Croatia, nor is it a regular data retrieval method in scientific world. An exploration in that terms must be done and is beyond necessary, since the pragmatical and economic benefits are obvious and immediate. The work in this paper had a real life worth and desirability, therefore the relevance could be assessed based upon it. Developing a fully functional, proven and empirically tested algorithms would be beneficial for the

scientists and for the sea parameters calculations, which would in that case be done remotely, using the satellite imagery.

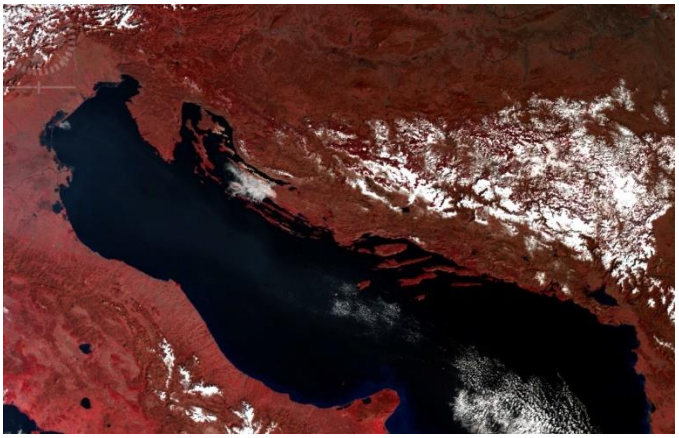


Fig. 1. Satellite image opened through SNAP toolbox

II. WATER QUALITY AND SENSING

A. Water parameters considered

After the description of the Satellites and concerning themes, the question remains how and what exactly should be considered for the remote sensing analysis. Therefore, the imposing question emerges: what the parameters of relevance are? Considering real life needs, supported by paper research and analysis, it was agreed on four parameters: dissolved oxygen, sea temperature, chlorophyll-a concentration, and turbidity. The overview of them follows.

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 CO_2 and H_2O into O_2 [5,6]. In addition, Chl-a is the main indicator of phytoplankton biomass [7,8,9] and can be used to determine the water clarity [10].

With the other two parameters 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. Turbidity is caused by particles suspended or dissolved in water that scatter light making the water appear cloudy or murky. Particulate matter can include sediment - especially clay and silt, fine organic and inorganic matter, soluble colored organic compounds, algae, and other microscopic organisms. Turbidity is measured using specialized optical equipment in a laboratory. The unit of measurement is called a Nephelometric Turbidity Unit (NTU). Measuring water transparency and Total Suspended Solids (TSS) also can be used to predict turbidity values. Secchi disks in lakes provide a simple and low-cost method for measuring water clarity. [11] Secchi depth is strongly influenced by the three optically significant constituents named before (Chl-a, TSM, and CDOM) and it corresponds to 10% of the surface light [12]. A slight comparison was drawn with the results, and it remains for future work to be considering mathematical codependency with parameters like turbidity, dissolved oxygen, etc.

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. Still, due to it being only surface temperature, the need for algorithm solution remains a field of interest.

B. Satellites used

Sentinel satellites are equipped with sensors and embedded systems which are used for spectral light analysis. Using many algorithms, atmospheric corrections and sensors, the data is calculated upon radiance and irradiance of the sunlight. Irradiance is the downwelling radiation from the sun, whereas radiance is the upwelling radiation from the Earth to the sensor. Since every satellite mission uses the light radiance and reflectance for the analysis, a problem to overcome is how does the state of atmosphere affect the remote sensing and could it be surmounted. The answer to the questions 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. [13] 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.[14]).

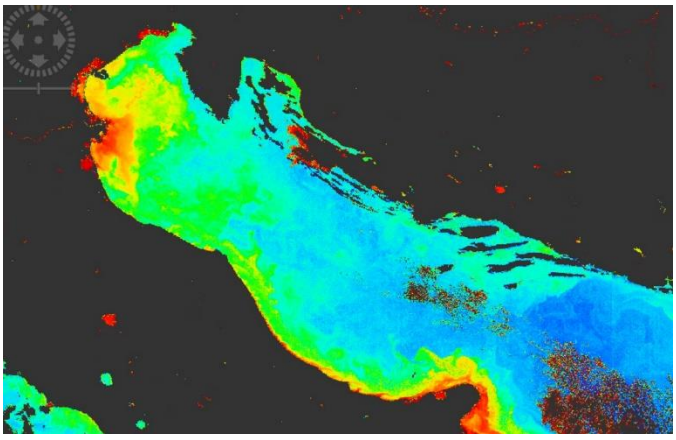


Fig. 2. Neural network Chlorophyll concentration estimation

Knowing the physical characteristics of the light, one could assume that depending on the reflectance, specific spectral wavelength will be observed which provide sufficient information about some characteristic of the water. Therefore,

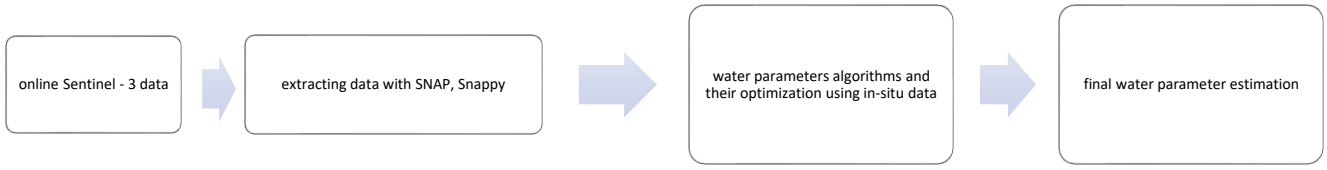


Fig. 3. Workflow diagram

every Sentinel has bands which represent different light wavelengths, distinct colors defined on the spectrum of light.

Satellite information about a band is providing us with how much radiance and reflectance of a specific length light has been detected. Moreover, sentinels provide masks; specifically designed pixel detection, which is parameterized based upon the field of interest.

III. METHODOLOGY

This work's data needs were met by using Creodias database, where the satellite images have been downloaded from. The exact dates and coordinates in the Adriatic Sea were used, based upon the in-situ measurements that were provided for the work. Considering the sheer enormous size of the collective data, an ubuntu virtual machine was set up where, using Putty, all data was transmitted. Further steps required data extraction from the sentinel images. Since for an acceptable algorithm proving and optimization a lot of data is needed, and with the amount of in-situ data available, automating the data extraction process became a necessity, given that loading and extracting data from SNAP is slow. Using Python coding language, the data about bands, masks and filters was extracted from every image using Snappy, 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. Looping through a few hundred downloaded images and extracting data needed, a dataset was created for later algorithm testing. The development of the algorithms preceded, for the analysis to be made on the dataset. Upon extraction 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. The main source of otherwise scientifically depleted field was work on Chinese sea turbidity calculation, notably high in its level.

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.

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, 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 around 500 measurements could have been used for the initial algorithm calibration and assessment, with the depth of 0 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 (R^2) gives an estimate of the proportion of variance explained by the model. The R^2 is only a valid estimator if its significance is high. A significance level of 95% ($p < 0.05$) was used here.

A. Algorithm optimization overview, correlations

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].

$$y = \beta 0 + \beta 1 x + \varepsilon.$$

Polynomial regression [18,19] is a type of regression analyzed in the n-th 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 several 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 [20]. Model of polynomial [21,22] is:

$$y = \beta 0 + \beta 1x + \beta 2x\ 2 + \cdots + \beta hx\ h + \varepsilon,$$

where h is named the polynomial degree [23,24].

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. 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.

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.

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.

R-squared (R^2) 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. It tells whether the data and model are biased, not if the model is correct.

The methodology of the solution and a brief complete overview follows.

B. Approach

Creodias data was filtered by date, coordinates, and cloud coverage. All the images covered 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 dependent on the data considered: OLCI for chl-a and SLSTR for sea temperature. SNAP was used for product evaluation and visualization. Band math was implemented through SNAP functions, color filtering was selected for the visual representation of the results. Automated image data extraction 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 were shown. An algorithm for the turbidity was found incorrect. The dissolved oxygen calculation was in the end left out of this study.

C. Chlorophyll algorithms

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.

Table 1. Chlorophyll concentration algorithms

Algorithm	Type	Equation/ C	Band Ratio (R), Coefficients (a)
POLDER	cubic	10(a0+a1* R+a2*R2 +a3*R3)	R = log(Rrs443/Rrs565) a = [0.438, -2.114, 0.916, -0.851]
CalCOFI two-band linear	power	10(a0+a1* R)	R = log(Rrs490/Rrs555) a = [0.444, -2.431]
CalCOFI two-band cubic	cubic	10(a0+a1* R+a2*R2 +a3*R3)	R =log (Rrs490/Rrs555) a = [0.450, -2.860, 0.996, -0.3674]
CalCOFI three-band	multiple regress-ion	exp(a0+a1 *R1+a2* R2)	R = ln (Rrs490/Rrs555) R2 = ln (Rrs510/Rrs555) a = [1.025, -1.622, -1.238]
CalCOFI four-band	multiple regress-ion	exp(a0+a1 *R1+a2* R2)	R = ln (Rrs443/Rrs555) R2 = ln (Rrs412/Rrs510) a = [0.753, -2.583, 1.389]
Morel-1	power	10(a0+a1* R)	R = log (Rrs443/Rrs555) a = [0.2492, - 1.76]
Morel-2	power	exp(a0+a1 *R)	R = ln (Rrs490/Rrs555) a = [1.077835, - 2.542605]
Morel-3	cubic	10(a0+a1* R+a2*R2 +a3*R3)	R = log (Rrs443/Rrs555) a = [0.20766, -1.82878, 0.75885, -0.73979]
Morel-4	cubic	10(a0+a1* R+a2*R2 +a3*R3)	R = log (Rrs490/Rrs555) a = [1.03117, -2.40134, 0.3219897, -0.291066]
SeaWiFs	max band ratio, multiple regress-ion	10(a0+a1* R+a2*R2 +a3*R3 +a4*R4))	R= log (max (Rrs443/Rrs555, Rrs490/Rrs555, Rrs510/Rrs555)), a= [¸0.3272, -2.9940, 2.7218, -1.2259, -0.5683]
OC4ME	max band ratio, polyno-mial	10(a0+a1* R+a2*R2 +a3*R3 +a4*R4))	R=log10(max (Oa443, Oa490, Oa510)/Oa555) a= [0.3272, -2.9940, -3.249491, 2.7218, -1.2259, -0.5683]
OCI	band differe-nce	10(a0+a1* R)	R=Rrs555 -0.5*(Rrs443 -Rrs670) a= [-0.4909, 191.659]

Out of the above written algorithms, after every was coded and tuned through the available data, due to being robust, efficient, precise and the most accurate, three have been proved to be prone enough for the consideration; OC4Me, SeaWiFs and CalCOFI four band. The description follows.

OC4Me Chlorophyll

Chlorophyll concentration is defined by the "OC4Me" Maximum Band Ratio (MBR) semi-analytical algorithm, developed by Morel et al [28] (cf. O'Reilly et al. [29], for a more general description of such algorithms). It is the latest version of MERIS pigment index algorithm, which is fully described in the MERIS ATBD 2.9 and in Morel et al [30]. OC4Me is a polynomial based on the use of a semi-analytical model, itself based on the analysis of AOPs measured. [31][32] It is expressed as:

$$log_{10}(Chl) = \sum_{k=0}^{4} (A_k * (log_{10}(R_j^i))^k) \quad (1)$$

R 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 Algorithm

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 equation, 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 like CalCOFI three-band ratio.

SeaWiFs algorithm

SeaWiFS images are received by OSC HRPT stations, which create level 1A images. Level 1 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. [33] 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 bands considered were wavelengths of 443,489,510 and 555, respectively. The other major distinction with the OC4Me were the coefficients inquired.

D. Temperature algorithms

For the estimation of sea temperature, a different satellite images were needed, captured by a different sensor, more accurately a SLTSR sensor with a RBT product type. They were also acquired from the Creodias database. The algorithms found, all used bands S7_BT_in, S8_BT_in and S9_BT_in of which only two were used, considering which gave better results. In Table 2., an overview of the algorithms found is given, where T4 and T5 are bands S7 and S8, or S8 and S9.

Table 2. Temperature algorithms (Kelvin) [35]

Author	Equation
McClain et al. (1983)	Ts = 1.035*T4 + 3.046*(T4 – T5) – 10.943
Price (1984)	Ts = T4 + 3.33*(T4 – T5)
Becker and Li (1990)	Ts = 1.274 + T4 + 2.63*(T4 – T5)
Prata and Platt (1991)	Ts = T4 + 2.45*(T4 – T5)
Sobrino et al. (1993)	Ts = 1.06*(T4 – T5) + 0.46*(T4 – T5)*2
Sobrino et al. (1993)	Ts = T4 + 1.8*(T4 – T5)

E. Turbidity algorithms

Turbidity as a parameter was researched in the last week of the project duration due to the time consumption of the parameters above reviewed. Algorithms found were not satisfactory enough thus leaving it for future projects.

Table 3. Turbidity equations modeled with Linear and Polynomial Regression [15]

Regression model equation	Band combination (=x)
367.82x2 - 976.42x + 649.13	B2/B3
971.47x2-1468x + 55.84	B3/B2
725.32x2- 858.52x + 255.91	B4/B3
118.8x2- 401.92x + 341.62	B2/B4
387.41x2 - 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

IV. RESEARCH RESULTS

Table 4. Chlorophyll concentration algorithms results, RMSE, deviation, correlation

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	NULL	NULL
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

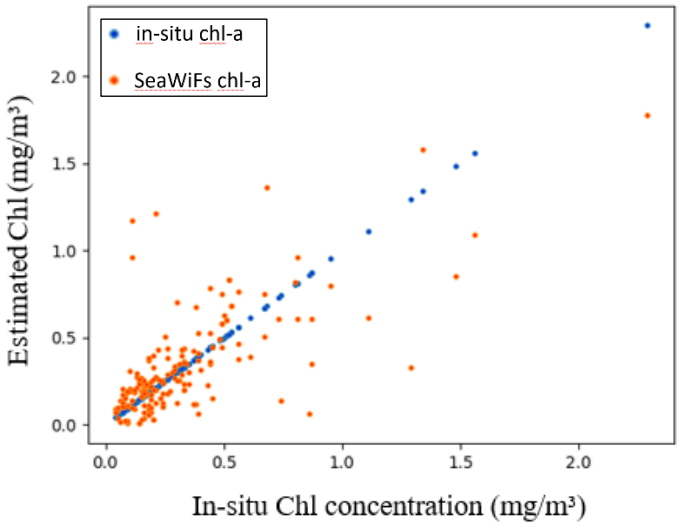


Fig. 3. Graphical visualization of SeaWiFs CHL algorithm vs. in-situ data

Table 5. Temperature algorithms results, RMSE, deviation, correlation

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

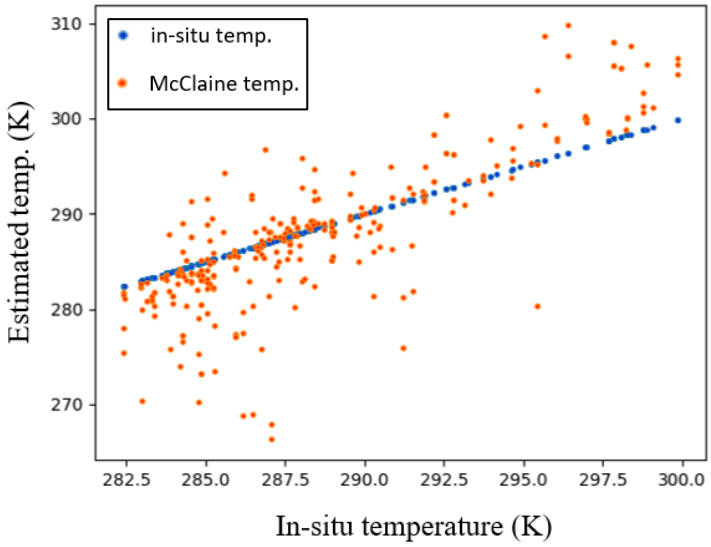


Fig. 4. Graphical visualization of McClain temperature algorithm vs. in-situ data

V. DISCUSSION

A. Chl-a

SeaWiFs algorithm performed the best out of those tested. As seen on Fig. 3., the estimated chlorophyll concentration (orange) is clustered around the (blue) in-situ values, with some estimations over- and under-estimating, while some estimations were correct to the second decimal place. We can also see that quite a few estimations deviate a lot from the in-situ measurements. Taking into consideration that the Croatia's coastal waters have quite low chlorophyll concentrations, these errors are considerable. These immense errors can be attributed to the low resolutions of Sentinel-3 images, clouds and atmospheric corrections. SeaWiFs has been 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 too low and no dependency could be outlined out of the model. The model is therefore inadequate and deficient under the conditions that the subpar R squared means the model developed cannot be used for the analysis.

The polynomic regression model stood out as incorrect, but it must be mentioned that most of the results were close to the real values, with a few bumps deviating as high as seen in the table. A negative R squared means that the appropriate model provides worse results than 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 is 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.

B. Sea surface temperature

Algorithms [1 to 5 in Table 5.] have a mean error of 2.77 Celsius/Kelvin degrees and above, which, pragmatically, is not usable for the temperature analysis of any water. Surely the parameters of the algorithms must be optimized to fit, otherwise no positive aspect 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.

C. Turbidity

As previously said, this parameter was researched with a lack of time, and in this paper, only raw, existing, and nonoptimized algorithms were presented. Therefore, turbidity results were deemed incorrect, wrong, and not useful. The formulas were otherwise proven to be correct in the areas with the high concentration of suspended matter. The Adriatic Sea has low concentrations of suspended matter; therefore, we have concluded that the existing algorithms and formulas depend on several factors other than just the parameter observed, which if considered, only suggest that the adapted algorithm must be introduced.

V.CONCLUSION AND FUTURE WORK

This work should be considered a preview of work to come. Without doubt, the algorithms could have been more accurate. But most of the problems are present 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 these conditions.

The extensive search of the proper equation would result in much fewer work mistakes and it would propel 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 not applicable for 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 also be 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 300 meters, if the in-situ data is collected near the shore, pixel resolution and values is compromised and ‘polluted’, so it would be of an 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 have to use more than one parameter at the time to be characterized as adaptable.

Due to spatial resolution of Sentinel-3 (300 meters), for Adriatic Coast, which is smaller in size, and since our case study was coastal area, conclusion is that even though Sentinel-3 satellites are used for water analysis, for this case better solution would be Sentinel-2 satellites with resolutions of 10, 20 and 60 meters. The neural network algorithm and OC4ME algorithm operated by the Sentinel projects are similar to the values we get for the parameters. Moreover, we are able to predict the values of the parameters corresponding to in-situ data correctly in most cases, with absolute error lower than 5 percent, but the problem is pixel pollution, a state that occurs when a cloud or any other atmospheric “noise” (“pollution) is detected by the satellites or if the pixel we observe is too close to the shore, when the reflectance is corrupted due to higher values that could have been prescribed to the impartial observer of sensor, that catches not only the sea surface, but the land as well.

ACKNOWLEDGEMENT

Our expression of gratitude goes to the Ericsson Nikola Tesla, not only for accepting us on this year’s Summer Camp, but also for all the support and consideration given while working there. Special thanks go to our team’s mentor Goran Kopčak who guided us through the process of research with his supervision and advice.

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-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).
[4] 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.
[5] 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]
[6] 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]
[7] 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]
[8] Wozniak, M.; Bradtke, K.M.; Krezel, A. Comparison of satellite chlorophyll an algorithm for the Baltic Sea. J. Appl. Remote Sens. 2014, 8. [CROSS:REF]
[9] 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]
[10] 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]

[11] Turbidity: Description, Impact on Water quality, Sources, Measures – A General Overview, Minnesota Pollution Control Agency, March 2008.
[12] Wetzel, R.G. Limnology. Lake and River Ecosystems, 3rd ed.; Academic Press: San Diego, CA, USA, 2001., [Cross:ref]
[13] Chavez PS (1988) An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens Environ 24(3):459–479,CrossRef
[14] Copernicus, https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm ,(accessed on 12 September 2022).
[15] 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
[16] 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
[17] 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]
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] Ridge Regression Condence Machine, Ilia Nouretdinov,Tom Melluish, Volodya Vovk, [CROSS:REF]
[26] Efficient learning machines,Theories, concepts and applications for engineers and system designers, Marietteb Awad,Rahul Khanna [CROSS:REF]
[27] Random Forests,LEO BREIMAN Statistics Department, University of California, Berkeley, CA 94720, Machine Learning, 45, 5–32, 2001, [CROSS:REF]
[28] 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.
[29] 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.
[30] [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.
[31] Bricaud, A., Morel, A., Babin, M., Allali, K., & Claustre, H. (1998). Variations of light absorption by suspended particles with chlorophyll concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. Journal of Geophysical Research-Oceans, 103, 31033-31044
[32] Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research-Oceans, 106, 7163-7180
[33] ANALYSIS OF SEAWIFS OCEAN COLOR ALGORITHMS FOR LAKE ERIE, Christopher Everett Wells, B.S. The Ohio State University, 2005 [CROSS:REF]
[34] 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]
[35] Algorithm for Estimating the Sea Surface Temperature (SST) through Sentinel-3's SLSTR sensor, Dosapati, Abhishek, 2018.

VI. WORK DIVISION

1) **Research:** Marko Haralović, Mirna Lovrić, Marin Maletić
2) **Algorithms:** Marko Haralović, Mirna Lovrić
3) **Data acquisition:** Mirna Lovrić, Marin Maletić
4) **Python code:** Marin Maletić, Marko Haralović
5) **Final report:**
   Abstract: Marin Maletić
   Chapter 1-4: Marko Haralović
   Chapter 5: Marko Haralović, Marin Maletić

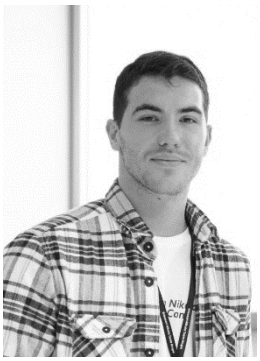
Marko Haralović was born on the 11th of the August 2002 in Karlovac, Croatia. He is a student at the Faculty of Engineering and Computing in Zagreb, Croatia and had finished his first undergraduate year at the college. He became the student at FER in 2021.

In 2022 he had a brief internship in UVI Play where he was an application developer and web content manager. Furthermore, in the Summer of 2022 he enrolled to the Ericsson Nikola Tesla d.o.o Summer Camp where he was included in the project named Satellite Imagery in e-Environment. His research and job interest are machine learning, model optimizations, deep learning, computer modeling, computer analysis, computer science used in economics, DCF analysis, mathematical modelling, visual data analysis.

Mr. Haralović's awards and honors are NCVVO diploma for state competition, XV. Gymnasium diploma of excellence, and STEM scholarship. He is also an active member of IEEE XPLORE scientific community.

Mirna Lovrić was born in Požega, Croatia, in 2000. She finished Gymnasium in Požega and received the bachelor's degree in mathematics at the Faculty of Science, University of Zagreb. She is currently pursuing the master's degree in applied mathematics at University of Zagreb, Faculty of Science.

In the Summer of 2022, she enrolled to the Ericsson Nikola Tesla Summer Camp. Her research interest includes numerical mathematics, computer simulation, computer vision and bioinformatics.



Marin Maletić, born in Zagreb, Croatia, 1999., received the B.S. degree in electrical engineering and information technology from Faculty of Electrical Engineering and Computing, University of Zagreb, and is currently pursuing the master's degree in information and communication technology. The final thesis, titled "Semantic image segmentation for application in environment exploration", was mentored by Stjepan Bogdan, PhD, in Zagreb (2022.).
In the Summer of 2022, he enrolled to the Ericsson Nikola Tesla Summer Camp where he was included in the project "Satellite Imagery in e-Environment".
His research interests vary and include automation and robotics, computer vision, artificial intelligence and machine learning.
Mr. Maletić, once part of the Croatia's national rowing team, is the head coach of the faculty rowing team and the main organizer of such events and competitions. He also enjoys singing and playing guitar.