Estimation of Chl-a in highly anthropized environments using machine learning and remote sensing

José G. Giménez 1, Raquel Martínez-España[2], Juan-Carlos Cano[1], José M. Cecilia[1]
[1]*Departament of Computer Engineering (DISCA)*
Universitat Politècnica de València, Valencia, Spain
[2]*Information and Communications Engineering*
Faculty of Computer Science, University of Murcia, Murcia, Spain
Email: jggimman@doctor.upv.es, raquel.m.e@um.es, {jucano,jmcecilia}@disca.upv.es

2023 19th International Conference on Intelligent Environments (IE) | 979-8-3503-1222-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/IE57519.2023.10179108

***Abstract*—Coastal lagoons are ecosystems of great socio-economic and environmental value. However, they are subject to great anthropogenic and environmental pressures, mainly due to climate change, which threatens their sustainability. High-resolution spatial and temporal monitoring systems are mandatory to (1) identify these threats, (2) understand the main problems affecting these ecosystems, and (3) predict how these ecosystems will behave in the future. In this paper, we present a monitoring system based on the European remote sensing service Copernicus that allows daily monitoring of chlorophyll-a (Chl-a) for the Mar Menor lagoon (Southeast Spain). Moreover, several machine learning (ML) models are analyzed to adapt the collected data to the particular context of the shallow and highly saline Mar Menor. The accuracy of the models are satisfactory, obtaining a global model with 0.9 value of R^2 and 0.75 mg/m^3 of mean absolute error. Also, this model is able to describe the algal bloom that provoke Chl-a peaks concentrations.**

***Index Terms*—Chlorophyll-a; Satellite Remote Sensing; Machine Learning; IoT**

I. Introduction

Remote sensing using satellite sources (SRS), based on sensors on board satellites that detect and record reflected or emitted energy, is emerging as an increasingly widespread system for monitoring different natural environments [1], [2].These systems provide high-resolution images with high-quality geometric and radiometric information. In addition, biogeophysical parameters can be obtained from the processing of these images [3]–[5], developing different products for observation of the earth, oceans, etc. [6]. ESA's Sentinels [7] and NPP VIIRS [8] products offer ever better spatial and temporal resolutions, as well as the development of new services.

SRS-based products that enable the generation of biogeochemical data are developed on a global scale and often need to be adapted to more local or regional contexts. Indeed, there can be large discrepancies between satellite-derived products and the actual state of surface parameters [9]. This implies that data generated by satellite products are not useful due to large deviations when compared to in situ measurements (i.e. ground-truth). This not only reduces their usefulness in applications but also reduces the accuracy of their derived products. Therefore, it is essential to adapt these SRS systems in local/regional contexts to improve their reliability.

In this work, different machine-learning (ML) and Deep Learning (DL) models are proposed to correct satellite data obtained from the Coopernicus ocean monitoring system. Specifically, we focus on chlorophyll-a (Chl-a) data generated from Sentinel-3. The models are applied in the regional context of the Mar Menor lagoon (Murcia, Spain); a local ecosystem at risk of collapse due to high anthropogenic pressures such as agriculture and tourism. Chl-a is an indirect measure of the lagoon health. The high mortality is attributed to the lack of oxygen in the water due to massive algal blooms. Therefore, this work focuses on chlorophyll-a (Chl-a) concentration as a critical indicator of the eutrophication status of this ecosystem. Several works have been proposed to develop SRS-based monitoring systems [10]–[13] for this variable. However, these works assume the validity of the data obtained by satellite systems, focusing mainly on one or two satellite sources, mainly Sentinel-derived products, without going into the validation process and/or coverage of all existing products for this gap. This modeling process is critical in the current unique ecosystem. The reasons are the low quantity and quality of data available in situ, the spatial mismatch between these measurements and remotely sensed pixels, the disregarding intrinsic heterogeneity of land and ocean surfaces that release nutrients into the lagoon, the shortcomings and deficiencies of theories on the scale problem in the validation.Therefore, the main contributions of this article are as follows:

- Evaluating ML and DL models accuracy focused on Chl-a concentration in Mar Menor by using remote sensing.
- Finding a global model for Mar Menor to allow us to estimate Chl-a concentration that will provide a continuous health lagoon monitoring and better dataset to Chl-a forecasting. The model is able to detect Chl-a concentrations peaks.
- Finding a cluster based on depths.

The article is structured as follows. First, we describe materials and methods carried out in this work are described in Section II. Then, we show the quantitative results offered by all SRS-

based products for monitoring the Mar Menor coastal lagoon in Section III. Finally, in section IV we present the conclusions and future work directions.

II. Materials and Methods

This section provides a full description of the materials and methods used in this work. In particular, the Mar Menor monitoring case study is briefly presented, describing the reference data used for model calibration. Moreover, the satellite sources used and the models under study are introduced.

A. Mar Menor Coastal lagoon

One of the largest coastal lagoons in Europe and the largest on the Iberian Peninsula, the Mar Menor is a 135 km^2 coastal lagoon in Murcia (Southeastern Spain). This lagoon is relatively shallow, with an average depth of 3.6 m and a maximum depth of 7 m. It is isolated from the Mediterranean Sea by a 22 km long sandy coastal barrier (called La Manga). La Manga is crossed by several gullies, which determines the semiconfined nature of its waters and gives them unique salinity and temperature characteristics (see Figure 1). In addition to its importance from an environmental point of view, the Mar Menor is also a key place for the economy of the Murcia Region. Its unique climatic conditions and abundant natural resources have attracted tourism, recreational uses, and fishing, as well as the importance of agriculture for the local economy. The basin that drains the Mar Menor is called Campo de Cartagena (CC) and is a long plain of more than 1,200 km^2 with a network of ephemeral streams that collect the scarce but intense rainfall [14].

The Mar Menor has been characterized by its oligotrophic waters and by its strong resistance to eutrophication. The lagoon's transparent waters have historically served as its key defining feature, but during the past ten years, eutrophic tendencies have emerged [15]. Inputs with high concentrations of nutrients to the lagoon have grown over the past few decades due to changes in agricultural practices in the CC and the introduction of intensively irrigated crops. Several studies [16], [17] found that flow transfer from the CC aquifer to the Mar Menor lagoon must have been significantly impacted by the entry of nitrates and other agrochemical elements from fertilization. These inputs caused a dramatic increase in pollution in the lagoon [18] and triggered an eutrophication process, resulting in water quality loss [19]. In 2016, an extreme eutrophication event resulted in sharp bloom of phytoplankton [20] that produced a significant change in the quality of the waters, with a substantial increase in turbidity, a change in the water colour, and a loss of transparency, with a drop in the depth of Secchi disc visibility to less than 1 m. These raised significant concerns among environmentalists and those working in the tourism industry, with significant socioeconomic repercussions [21]. In September 2019, Chl-a levels significantly increased, even exceeding the peak observed in 2016. This was due to surface runoff water from the drainage basin that flowed into the Mar Menor as a result of a big storm. The driving factors of primary productivity were the enormous volumes of nitrogen, phosphorus, organic matter, and sediments drained by runoff

Fig. 1: *In situ* monitoring points (ISMP) carried out by Regional goverment of Murcia (CARM) at Mar Menor.

In situ data source	CARM
Sampling starting date	2016-08-02
Sampling end date	2022-02-24
#Samples	3,251

TABLE I: In situ monitoring provide by CARM.

[22]. Under this scenario, early intervention is required to reduce nutrient and other pollutant entry into the lagoon. Thus, creating an accurate model for some environmental parameters, such as chl-a, can help authorities to research causes and take decisions related.

B. In situ data

The dataset used to analyze remote sensing data has been obtained from the Regional Government of Murcia (CARM). This entity promoted regular measurements of Mar Menor since 2016 when the degradation of the lagoon became apparent. This dataset provides almost weekly measures at different depths for twelve points spread along Mar Menor. These points are called in situ measurements points (ISMPs) and they represent the heterogeneity of the lagoon, having different characteristics such as depth, and distances to different interesting points such as sea sore and wadis. Among the data provided by CARM, it is included Chl-a, turbidity, CDOM,

	Depth(m)	**Chl-a** (mg/l)
count	6146	6146
mean	2.225	2.651
std	1.66	3.865
min	0	0.022
25%	1	0.726
50%	2	1.367
75%	4	2.964
max	6	28.112

TABLE II: Chl-a data description provided by CARM

Oxygen, salinity, and pH. This work focuses on the Chl-a as it is strongly related to anoxia episodes. Table II shows the main dataset descriptors. CARM provides filtered data so no cleaner nor outlier search is needed. Moreover, in this work the experiments manage the data by depths, which means the models will be fitted using different depths, but never mixing them. Thus, the measures taken between 0 and 1-meter depth appear in this work as depth 0, those taken between 1 and 2 meters are depth 1, and so on.

C. SRS data

SRS data is becoming popular for near real-time (NRT) monitoring systems to address natural and societal challenges such as shedding light on global and regional challenges, thereby supporting a number of initiatives (e.g. Sendai Framework, Paris Agreement, Sustainable Development Goals) [23]–[25].

This paper analyses publicly available SRS data provided by ESA satellite products. The European Copernicus Marine Service (CMS) that provides free, regular, and systematic authoritative information on the state of the Blue (physical), White (sea ice), and Green (biogeochemical) ocean, on a global and regional scale. Among several products that Copernicus offers through their different platforms, this work focuses on data from Sentinel-3 A and B. Table III describes the S3 missions. Both satellites have on board the imaging spectrometer Ocean and Land Colour Instrument (OLCI) that measures solar radiation reflected by the Earth in 21 spectral bands at a spatial resolution of 300 m. It is important to note that this work focuses on these Sentinel-3-related products as they offer daily data from the same world region around the same hour (see Table III) and having a high temporal resolution is essential for the development of a monitoring system.

Sentinel-based data are processed and split into products depending on the stage of the process. Thus, OLCI provides three different processing levels. This work focuses on Level-2 (L2) data that provides the water Chl-a concentration in mg/m^3. OLCI L2 product also provides additional data such as reflectances, Aerosol Optical Thickness (AOT), Diffuse Attenuation coefficient (Kd940), just to mention a few. Despite these parameters are not considered in this work, they are used by ESA algorithm to calculate the Chl-a data provided. Another important feature for OLCI products is the resolution. In this work, we use Water Full Resolution (WFR) product that provides the data related with water masses at 300 meters resolution.

Regarding time granularity, the CMS download tool selects files by timeless, which indicates the period from the image was acquired until it becomes public. Particularly, we focus on NTC (Non-Time Critical) files that refer to products delivered within the next 24/48 hours after data acquisition by satellite, in contrast to NRT (Near Real Time) which refers to products delivered less than 3 hours after acquisition. NTC products contain data adjusted and collated using information from other platforms thus seems more precise to use this product in the present analysis. It is also important to note that we consider the best-case scenario for data coverage provided by SRS sources; i.e., NTC data instead of NRT. Anyway, a monitoring system could provide the data in NRT and then rewrite it with its NTC data when it becomes available.

To obtain the processed Chl-a historical data from Sentinel-3 A and B, we use CREODIAS [26]; a platform that contains the data generated by Sentinel, Envisat, Lansat, and other EODATA satellites. Its design allows third-party users to prototype and builds their own value-added services and products. This service allows programmers to request the previous Sentinel products easily from Sentinel-hub. In this case, the API returns a set of netCDF4 files that contain parameters related to Chl-a, GPS coordinates, light reflectances, atmosphere light absorption, etc.

Sentinel-based products use two different algorithms to obtain Chlorophyll concentration from this raw information. We focus on the Inverse Radiative Transfer Model-Neural Network (IRTM-NN) to estimate from normalized water-leaving reflectance at OLCI bands b440 to b674, b709, θs and $\Delta\phi$ (namely $Chlor\_nn$). Unlike other chl_a estimations algorithms, the S3 NN method provides data even when some reflectances provided at level-2 are negatives. Therefore, IRTM-NN algorithm can deal with these values, something that the classical algorithm does not allow and which forces negative values to be discarded. This consideration is taken to allow us to match as many days as possible between the S3 data and the in situ data in order to obtain a larger data set to train and test the models. Similarly, we consider that the twin S3 A-B reflectances provide similar reflectance information as they use the same OLCI instrument and the same processes to obtain L2 products, so we merge the data from both sources into one dataset.

It is worth mentioning IRTM-NN relies on neural networks to address the problem as a non-linear multiple regression method that drastically reduces computational time, once the network is properly trained to fit the coefficients. Fortunately, CMS already provides the IRTM-NN model that outputs water-inherent optical properties such as total backscattering coefficient ($BBP443$), total absorption coefficient ($ATOT443$), phytoplankton absorption coefficient ($APH443$), colored Detrital and Dissolved Material absorption coefficient ($ADG443$), Chlorophyll (Chl2) concentration, Total Suspended Matter (TSM) concentration. These S3 products also provide a flag value for every pixel which indicates if valid the kind of terrain: water, land or snow. We refer the reader to [27] for insights. Anyhow, in this study all data used and associated to each ISMP are classified by S3 as water. Those invalid has been discarded.

TABLE III: Main satellite instruments features

Satellite/Instrument	Revisit time (day)	Scanning time (UTM)*	Resolution at nadir (m)	Chl Algorithm	Start Mission (year)	End Mission (year)
Sentinel-3A OLCI	1	10:10 +- 1h	∼300	Chl_nn Chl_nomc	2016	2023
Sentinel-3B OLCI	1	10:10 +- 1h	∼300	Chl_nn	2018	2025

An initial analysis shows that S3 Chl values has a middle correlation with CARM in-situ data. The correlation depends on the ISMP, that means S3 does not provide reliable results for random point. Also, in the whole sentinel dataset there is not Chl-a peaks that means its model is not able to detect algal blooms that has happened several times during the study period.

D. ML models for Chl-a forecasting

Chl-a forecasting is considered in this article as a regression problem using satellite data as input and in situ measurements as output. In order to find a specific model to fit SRS data obtained through Copernicus system to the Chl-a measurements obtained for the Mar Menor, the use of different ML models is proposed to analyze which one gives the best results. The ML algorithms selected are:

- Random Forest Regressor (RFR): Random Forest regression is a supervised learning technique, based on a set of decision trees. To avoid overfitting, RF uses the bagging technique. Thus each tree is built on different portions of the data, no single tree sees all the training data. This means that each tree is trained with different data samples for the same problem. In this way, by combining their results, some errors are compensated by others and we have a prediction that generalizes better [28].
- Decision Tree Regressor (DT): Decision trees for regression are a subtype of prediction trees applied when the response variable is continuous. In the training of a decision tree in regression, the instances, are distributed through the nodes generating the tree structure until reaching a terminal node, the condition to generate or split a node is usually the Residual Sum of Squares. When a new instance is to be predicted, the nodes of the tree are traversed according to the attributes of the new observation until a terminal node is reached. When the terminal node is reached, the output is the average of the output attribute of that node, of the training instances, [29].
- K-Nearest Neighbors Regressor (KNN): The K-nearest neighbors method is an instance-based method that tries to classify or predict a new instance x based on the k-nearest to that value. To do so, it uses different distance measures, depending on the types of attributes that make up the instance. In the case of regression prediction, the value returned is usually the average of the output attribute values of the k-nearest neighbors to the given instance. It is worth mentioning that KNN method does not require training, since the model is created at the same time as a new instance is inferred [30].
- Multi-layer Perceptron Regressor (MLP): The multilayer perceptron (MLP) is a supervised learning algorithm that learns a non-linear function approximator for classification or regression. A number of hidden layers located between the input and output layers are the real engine of this technique. MLP neurons are trained with the backpropagation learning algorithm. MLPs are designed to approximate any continuous function and can solve problems that are not linearly separable [31].
- Convolutional Neural Network (CNN): CNNs are a type of neural network that is composed of a deep learning network architecture that learns directly from data, without the need to extract features manually. In general, CNN applied in remote sensing studies is used to analyze images. However, we use another approach by using a 1D convolutional layer. We base this decision on the relationship between adjacent reflectances, that allows to CNN to find and appropriate relationship. Then a layer will find the relation similar to works that test different classical algorithms looking for this relation and their coefficients. [32].

III. Evaluation and discussion

A. Experimental set up

The dataset used in this work contains the pairing data S3-AB and ISMPs by day and coordinate. Thus, we generate several datasets by ISMP and depths containing the S3-AB reflectances and CARM Chl-a measurements. The output unit is the mg/m^3 of Chl-a, considering as inputs the remote sensing reflectances. Table IV resumes the number of instances available for each dataset.

	Depth						
ISMP	0	1	2	3	4	5	6
1	87	87	87				
2	105	105	105	105	105	98	
3	108	108	108	108	108	108	42
4	112	112	112	112	112	30	4
5	102	102	102	102	102	87	
6	103	103	103	103	103	103	46
7	108	109	109	108	106		
8	103	103	103	103	102		
9	96						
10	98	99	99	99	99		
11	104	104	105	103	54		
12	101	101	101	101	101	98	

TABLE IV: Number of instances by ISMP and depth.

To assess the models accuracy presented here we use next statistics:

- Coefficient of determination (R^2): R^2 is the proportion of the total variance of the variable explained by the regression. This measure reflects the well-fitting of a model to the variable it intends to explain, the closer to

the value 1, the better the fit of the model. This measure is calculated as:

$$R^2 = \frac{\sigma_{yp}^2}{\sigma_y^2 \cdot \sigma_p^2},$$

being $sigma_{yp}$ the variance of the actual variable y and of the predicted variable p, σ_y the variance of y (real value) and σ_p variance of p (predicted value).

- Mean Squared Error (MSE): The MSE is a measure derived from the square of the Euclidean distance, it is always a positive value that decreases as the error approaches zero. It is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - P_i),$$

being n number of instances, Y_i real value and P_i predited value.

- Mean Absolute Error (MAE): The MAE is used to quantify the accuracy of a prediction technique by comparing the predicted versus observed values. It is calculated as:

$$MAE = \frac{\sum_{i=1}^{n}(|P_i - Y_i|)}{n},$$

being n number of instances, P_i predicted value and Y_i real value.

A total of four experiments have been carried out aiming to get an appropriate model. ML techniques are initially used but also CNN model can improve the results. For the experiments, algorithms are fitted using the best parameter optimization to obtain the best fitting models.

Below, we summarize the steps follows to carry us to find the model:

- The first experiment performs a 5-fold cross-validation randomly on the ML algorithm proposed. It evaluates a local model for each ISMP and depth.
- The second experiment performs a 5-fold cross-validation randomly on ML by using all ISMPs indiscriminately to find a global model. Always, the depths are handling separately.
- The previous experiments suggest a cluster existence. Thus we repeat the previous ones excluding ISMPs to evaluate a the model by cluster.
- Finally we repeat all the experiments by using a CNN Conv1D.

B. Results and discussion

1) Creating ML local model for ISMP: This first experiment aims to train ML models for each station individually. Due to the heterogeneity of the Mar Menor, each monitoring point (i.e., ISMP) may have different characteristics that may benefit from modelling using ML methods. Thus, the main hypothesis at this section is that each ISMP requires a different ML model to be trained for this particular point. To get the best performing results we configure ML models using different parametrization, hoverwer we do not go deep into this part to do not lengthen the text and since CNN overcome them. Table V presents the best results for each ISMP and the depth where is more accurate, indicating the best ML technique. In the selection of the best results, we have taken as reference the highest R^2. In this first experiment it is important to note ML models for ISMPs 1 and 9 do not perform well, having R^2 values at 0 and large error values for both MAE and MSE. Indeed, these ISMPs are the shallowest ISMPs at Mar Menor and thus, reflectances from remote sensing introduces many errors, providing high disparity of values compared to in situ monitoring. Another relevant aspect to note is that for all ISMPs except ISMP 5, the best results are obtained at shallow depths of 0 to 1 meter or 1 to 2 meters. For ISMP 5, the best results are obtained at depths of 4 to 5 meters. This makes sense since ISMP 5 is located on the other side of the lagoon (see Figure 1). Finally, there is no single model that is better than the others for all ISMPs. The best performing techniques are MLP and RFR for 5 different ISMPs and DT and KNN for one different ISMP. Although the results of most ISMPs, excluding 1 and 9, are satisfactory, the lack of a single prediction model creates problems and complexity for the prediction system. Hence the need to look for a global model for Chl-a in the Mar Menor. Therefore, the different ML techniques for all ISMPs and all depths are tested below.

Technique	**ISMP**	**Depth**(m)	**R^2**	**MAE**	**MSE**
DT	1	0	0.0	1.526	12.919
MLP	2	0	0.863	1.141	3.473
MLP	3	0	0.762	1.132	4.171
MLP	4	1	0.808	1.1	2.75
MLP	5	4	0.808	1.1	2.750
RFR	6	1	0.864	0.932	2.88
MLP	7	0	0.79	0.943	3.966
RFR	8	0	0.877	0.84	1.589
KNN	9	0	0.0	4.682	53.934
RFR	10	0	0.738	0.919	4.481
RFR	11	1	0.861	1.047	3.1
RFR	12	1	0.887	0.965	2.823

TABLE V: The best performance model for each station and its depth wherein R^2 is max to predict Chl-a. MAE and MSE are measured in mg/m^3.

2) Creating global ML model: In this second experiment, the ISMPs are no longer considered in isolation, so a global prediction model is created for each ML technique. For these global models, a new dataset is created with data from all ISMPs and depths together. The table VI shows, for the first 4 depth levels, the results of the global models for each ML technique. This experiment discards depths greater than 4m due to the low accuracy produced also in the previous experiment.

Table VI shows that the best result obtained is with the KNN technique at depths between 1 and 2 meters for Chl-a prediction. The R^2 value is 0.6919 and the MAE and MSE are 1.1614 and 4.9211 mg/m^3 respectively. The next best performing technique is DT at the same depth with an R^2 of 0.6707. The RFR and MLP techniques are the worst performers, contrary to the local models. This is because the DT and KNN techniques are less affected by the noise introduced by ISMP 1, 5 and 9. To verify this theory of noisy data, the same experiment is performed using a 5-fold cross-validation, but without including the data from these ISMPs. As mentioned above, ISMP 1 and 9 have very shallow data and

Technique	Depth(m)	R^2	MAE	MSE
DT	0	0.553	1.249	7.481
	1	0.671	1.217	6.237
	2	0.264	1.582	10.101
	3	0.266	1.57	8.386
KNN	0	0.662	1.159	6.528
	1	0.692	1.161	4.921
	2	0.48	1.326	5.847
	3	0.409	1.519	8.036
MLP	0	0.587	1.344	6.091
	1	0.596	1.351	7.561
	2	0.578	1.397	7.224
	3	0.526	1.391	5.894
RFR	0	0.644	1.251	6.456
	1	0.631	1.148	5.562
	2	0.61	1.212	5.093
	3	0.555	1.309	4.956

TABLE VI: Results for ML at 4 depths using all stations to predict Chl-a. MAE and MSE are measured in mg/m^3.

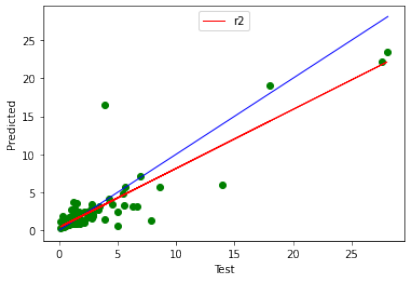
Technique	Depth(m)	R^2	MAE	MSE
DT	0	0.639	1.159	5.733
	1	0.692	1.239	5.776
	2	0.629	1.436	6.458
	3	0.488	1.45	6.417
KNN	0	0.765	1.103	5.041
	1	0.682	1.303	6.215
	2	0.558	1.402	6.882
	3	0.451	1.381	6.077
MLP	0	0.772	1.041	4.012
	1	0.75	1.267	5.543
	2	0.677	1.334	5.829
	3	0.556	1.347	5.790
RFR	0	0.833	0.964	3.647
	1	0.81	1.0783	3.969
	2	0.724	1.296	5.163
	3	0.548	1.314	5.769

TABLE VII: Models' precision at different depths to predict Chl-a in Cluster-1. MAE and MSE are measured in mg/m^3.

a small amount of data. ISMP 5 is also discarded because the best results were obtained in the local models at great depths and also, being outside the lagoon, satellite data can provide possible noise in the results.

3) Creating global model by clustering: Taking into account the above reasons, we can assume the existence of a cluster related to the maximum depth of the ISMP. Thus, we can consider two clusters initially: (1) Cluster-0 less than 2 meters deep (ISMP 1 and 9), (2) Cluster-1 more than 2 meters deep. Thus, the accuracy of the ML models for Cluster-1 improves in general, as shown in the table VII. For the DT and KNN techniques, the improvement is only about 3 or 4 hundredths of a fit in the reference to the R^2 value, and the MAE and MSE errors decrease slightly or remain the same. However, the MLP and RFR techniques obtain much better results, reaching R^2 fit values of about 0.8 and MAE and MSE errors of about 1 mg/m^3 at shallow depths of 0 to 1 meter. Figure 2 shows the dispersion graph for RFR model where this model is able to detect trend changes in Chl-a data. Actually, RFR introduces the best model able to detect these peaks.

Fig. 2: RFR dispersion plot for depth 0 at ISMP 6



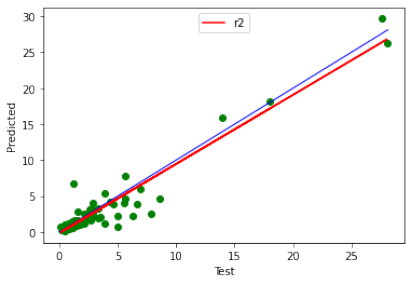
From these ML experiments we draw several conclusions. The first one is a possible cluster based on ISMP depth. Also that it possible discard ISMP 5 since it is located outside the lagoon. Finally, the global ML models fit slightly worse than local ones. However, the local models are not homogeneous and complicated the system. This fact justifies to try another techniques such as Deep Learning to find a model.

4) Using Deep Learning: Despite ML models accurate is not bad, we assume a DL model could perform better. Also, it could be work without a clustering by depth. Thus, we repeat the previous steps but using CNN Conv1D. Although this technique is mainly used to image classification or detection, we suggest another approach. To find the relation among adjacent reflectances we set a kernel windows to 4 to build the layer. Initially CNN presents a good behavior, but for isolate stations the results are not very significant compared to ML models. Similar to ML models, the best accuracy is reached using depth 0 from CARM. However, it shows a better accuracy than ML when uses all ISMPs, i.e., when CNN is training to predict a whole ISMP values using the rest of ISMPs. Tables shows again how the behavior of ISMP 1,5 and 9 have not relation to the rest. Thus, the clustering by depth seem necessary and also it justifies the exclusion of ISMP 5.

Taking in mind these results, we look for a global model trying to predict randomly whatever ISMP value into Cluster-1. Thus, training the model with 80% of data and using a cross-validation of 5 provided by the model the results are r2=0.89, MSE=1.95 mg/m^3 and MAE = 0.78 mg/m^3, which shows a good performance that outperforms the previous ML models.

Also, the figure 3 shows the CNN behaviour in Chl-a peaks in depth 0, when algal blooms phenomenons occur, that indicates this model can represent the Mar Menor Chl-a behavior.

Fig. 3: CNN dispersion plot for depth 0 at ISMP 6



ISMP	R^2	MSE	MAE
1	0.00	15.43	2.08
2	0.90	2.58	0.94
3	0.85	2.60	0.92
4	0.88	1.72	0.74
5	0.00	0.83	0.59
6	0.91	1.80	0.75
7	0.89	2.00	0.77
8	0.90	1.27	0.69
9	0.00	48.82	3.86
10	0.76	4.18	0.88
11	0.87	2.14	0.90
12	0.87	2.81	0.80

TABLE VIII: Results for CNN technique at 0 depth using a global model to predict Chl-a of whole station. MAE and MSE are measured in mg/m^3.

IV. Conclusions

Remote sensing offers a very powerful tool for monitoring large natural areas due to its ubiquity. Improvements in the instrumentation of these systems offer systems with an ever-decreasing spatial and temporal resolution, which provides the opportunity to develop warning systems for the conservation of natural areas. However, satellite systems are global and their signals need to be processed to increase their accuracy in more local/regional contexts. In this paper, different machine learning an deep learning models have been analyzed to improve the accuracy of Chl-a measurements from the reflectances of the European Sentinel-3 observation system in a particular highly anthropized context such as the Mar Menor lagoon. Our results show that Sentinel-3 reflectances data have a good relation with in-situ measurements provided by CARM. This good correlation suggests that by using remote sensing it is possible to get a good estimation of Chl-a concentration in Mar Menor. The CNN model presented represents a global model that can manage well the remote sensing information in this environment. Therefore it can provide a unique model able to describe the whole lagoon. However, there is still a lack of precision in points close to the coast or shallow, so there will be to search for a better method to estimate values in these cases. This fact seems to clash with the findings that best model results occur using surface measurements, but it is due to the reflectance light phenomenon in water and is beyond the scope of the present research. Moreover, we acknowledge that only Chl-a has been targeted and other parameters such as turbidity can be included in the model to provide a multivariate model of the lagoon. These other parameters are being studied for future work as well as test the CNN model approach on other similar environments suggest an interesting address.

Acknowledgments

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861 as well as the Ramon y Cajal Grant RYC2018-025580-I, funded by MCIN/AEI/10.13039/501100011033, " FSE invest in your future" and "ERDF A way of making Europe".

References

[1] A. M. Lechner, G. M. Foody, and D. S. Boyd, "Applications in remote sensing to forest ecology and management," *One Earth*, vol. 2, no. 5, pp. 405–412, 2020.

[2] Y. Huang, Z.-x. Chen, Y. Tao, X.-z. Huang, and X.-f. Gu, "Agricultural remote sensing big data: Management and applications," *Journal of Integrative Agriculture*, vol. 17, no. 9, pp. 1915–1931, 2018.

[3] J. Yang, P. Gong, R. Fu, M. Zhang, J. Chen, S. Liang, B. Xu, J. Shi, and R. Dickinson, "The role of satellite remote sensing in climate change studies," *Nature climate change*, vol. 3, no. 10, pp. 875–883, 2013.

[4] D. Sowmya, P. Deepa Shenoy, and K. Venugopal, "Remote sensing satellite image processing techniques for image classification: a comprehensive survey," *International Journal of Computer Applications*, vol. 161, no. 11, pp. 24–37, 2017.

[5] K. Vos, K. D. Splinter, M. D. Harley, J. A. Simmons, and I. L. Turner, "Coastsat: A google earth engine-enabled python toolkit to extract shorelines from publicly available satellite imagery," *Environmental Modelling & Software*, vol. 122, p. 104528, 2019.

[6] C. Justice, J. Townshend, E. Vermote, E. Masuoka, R. Wolfe, N. Saleous, D. Roy, and J. Morisette, "An overview of modis land data processing and product status," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 3–15, 2002.

[7] Z. Malenovskỳ, H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, "Sentinels for science: Potential of sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sensing of environment*, vol. 120, pp. 91–101, 2012.

[8] C. O. Justice, M. O. Román, I. Csiszar, E. F. Vermote, R. E. Wolfe, S. J. Hook, M. Friedl, Z. Wang, C. B. Schaaf, T. Miura, *et al.*, "Land and cryosphere products from suomi npp viirs: Overview and status," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 17, pp. 9753–9765, 2013.

[9] X. Wu, Q. Xiao, J. Wen, D. You, and A. Hueni, "Advances in quantitative remote sensing product validation: Overview and current status," *Earth-Science Reviews*, vol. 196, p. 102875, 2019.

[10] M. Erena, J. A. Domínguez, F. Aguado-Giménez, J. Soria, and S. García-Galiano, "Monitoring coastal lagoon water quality through remote sensing: The mar menor as a case study," *Water*, vol. 11, no. 7, p. 1468, 2019.

[11] I. Caballero, M. Roca, J. Santos-Echeandía, P. Bernárdez, and G. Navarro, "Use of the sentinel-2 and landsat-8 satellites for water quality monitoring: An early warning tool in the mar menor coastal lagoon," *Remote Sensing*, vol. 14, no. 12, p. 2744, 2022.

[12] D. Gómez, P. Salvador, J. Sanz, and J. L. Casanova, "A new approach to monitor water quality in the menor sea (spain) using satellite data and machine learning methods," *Environmental Pollution*, vol. 286, p. 117489, 2021.

[13] J. González-Enrique, J. J. Ruiz-Aguilar, E. Madrid Navarro, R. Martínez Álvarez Castellanos, I. Felis Enguix, J. M. Jerez, and I. J. Turias, "Deep learning approach for the prediction of the concentration of chlorophyll a in seawater, a case study in el mar menor (spain)," in *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, pp. 72–85, Springer, 2023.

[14] J. Senent-Aparicio, J. Pérez-Sánchez, J. L. García-Aróstegui, A. Bielsa-Artero, and J. C. Domingo-Pinillos, "Evaluating groundwater management sustainability under limited data availability in semiarid zones," *Water*, vol. 7, no. 8, pp. 4305–4322, 2015.

[15] H. M. Conesa and F. J. Jiménez-Cárceles, "The mar menor lagoon (se spain): A singular natural ecosystem threatened by human activities," *Marine pollution bulletin*, vol. 54, no. 7, pp. 839–849, 2007.

[16] J. Senent-Aparicio, A. López-Ballesteros, A. Nielsen, and D. Trolle, "A holistic approach for determining the hydrology of the mar menor coastal lagoon by combining hydrological & hydrodynamic models," *Journal of Hydrology*, vol. 603, p. 127150, 2021.

[17] J. C. Domingo-Pinillos, J. Senent-Aparicio, J. L. García-Aróstegui, and P. Baudron, "Long term hydrodynamic effects in a semi-arid mediterranean multilayer aquifer: Campo de cartagena in south-eastern spain," *Water*, vol. 10, no. 10, p. 1320, 2018.

[18] S. García-Ayllón, "New strategies to improve co-management in enclosed coastal seas and wetlands subjected to complex environments: Socio-economic analysis applied to an international recovery success case study after an environmental crisis," *Sustainability*, vol. 11, no. 4, p. 1039, 2019.

[19] P. Jimeno-Sáez, J. Senent-Aparicio, J. M. Cecilia, and J. Pérez-Sánchez, "Using machine-learning algorithms for eutrophication modeling: case

study of mar menor lagoon (spain)," *International Journal of Environmental Research and Public Health*, vol. 17, no. 4, p. 1189, 2020.
[20] J. M. Mercado, D. Cortés, F. Gómez-Jakobsen, C. García-Gómez, S. Ouaissa, L. Yebra, I. Ferrera, N. Valcárcel-Pérez, M. López, R. García-Muñoz, *et al.*, "Role of small-sized phytoplankton in triggering an ecosystem disruptive algal bloom in a mediterranean hypersaline coastal lagoon," *Marine Pollution Bulletin*, vol. 164, p. 111989, 2021.
[21] S. Garcia-Ayllon, "The integrated territorial investment (iti) of the mar menor as a model for the future in the comprehensive management of enclosed coastal seas," *Ocean & Coastal Management*, vol. 166, pp. 82–97, 2018.
[22] J. M. Ruiz-Fernandez, V. León, L. Marín-Guirao, F. Giménez-Casalduero, J. Alvárez-Rogel, M. Esteve-Selma, R. Gómez-Cerezo, F. Robledano-Aymerich, G. González-Barberá, and J. Martínez Fernández, "Informe de síntesis sobre el estado actual del mar menor y sus causas en relación a los contenidos de nutrientes," *Projects of Sustainability and Conservation of Mar Menor Lagoon and Its Basin; Universidad de Alicante: Alicante, Spain*, 2019.
[23] Y. Walz, A. Min, K. Dall, M. Duguru, J.-C. V. de Leon, V. Graw, O. Dubovyk, Z. Sebesvari, A. Jordaan, and J. Post, "Monitoring progress of the sendai framework using a geospatial model: The example of people affected by agricultural droughts in eastern cape, south africa," *Progress in Disaster Science*, vol. 5, p. 100062, 2020.
[24] A. Grainger and J. Kim, "Reducing global environmental uncertainties in reports of tropical forest carbon fluxes to redd+ and the paris agreement global stocktake," *Remote Sensing*, vol. 12, no. 15, p. 2369, 2020.
[25] S. M. Noe, K. Tabakova, A. Mahura, H. K. Lappalainen, M. Kosmale, J. Heilimo, R. Salzano, M. Santoro, R. Salvatori, A. Spolaor, *et al.*, "Arctic observations and sustainable development goals–contributions and examples from era-planet icupe data," *Environmental Science & Policy*, vol. 132, pp. 323–336, 2022.
[26] R. Malinowski, S. Lewiński, M. Rybicki, E. Gromny, M. Jenerowicz, M. Krupiński, A. Nowakowski, C. Wojtkowski, M. Krupiński, E. Krätzschmar, *et al.*, "Automated production of a land cover/use map of europe based on sentinel-2 imagery," *Remote Sensing*, vol. 12, no. 21, p. 3523, 2020.
[27] N. Pahlevan, B. Smith, J. Schalles, C. Binding, Z. Cao, R. Ma, K. Alikas, K. Kangro, D. Gurlin, N. Hà, *et al.*, "Seamless retrievals of chlorophyll-a from sentinel-2 (msi) and sentinel-3 (olci) in inland and coastal waters: A machine-learning approach," *Remote Sensing of Environment*, vol. 240, p. 111604, 2020.
[28] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[29] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.
[30] G. Hackeling, *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd, 2017.
[31] H. Taud and J. Mas, "Multilayer perceptron (mlp)," in *Geomatic approaches for modeling land change scenarios*, pp. 451–455, Springer, 2018.
[32] S. Harbola and V. Coors, "One dimensional convolutional neural network architectures for wind prediction," *Energy Conversion and Management*, vol. 195, pp. 70–75, 2019.