Temporal Aspects of Chlorophyll-a Presence Prediction Around Galapagos Islands

Fernando Chávez-Castrillón[1,2], Massimo Coltorti[2], Roberta Ivaldi[3], Estrella Lucena-Sánchez[4,5], and Guido Sciavicco[5]

[1] Ecuadorian Navy, Dept. of Educ. and Doct.
fchavez@armada.mil.ec
[2] University of Ferrara, Dept. of Phys. and Earth Sci.
massimo.coltorti@unife.it
[3] Hydrographic Institute of the Italian Navy
roberta_ivaldi@marina.difesa.it
[4] University of Modena and Reggio Emilia, Dept of Phys., Inf., and Math.
[5] University of Ferrara, Dept. of Math. and Comp. Sci.
{estrella.lucenasanchez|guido.sciavicco}@unife.it

Abstract. Chlorophyll-a is a specific form of chlorophyll used in oxygenic photosynthesis which has been linked to nutrient presence in sea waters, and being able to correctly determine its concentrations may turn out to be a key step in helping preventing and controlling illegal fishing activities in certain areas. In this work, we consider open access data taken from the Copernicus space program (currently used in the European Union for Earth observation and monitoring) that include several physical and biochemical variables and measurements of the ocean surrounding Galapagos Islands (Ecuador). We use such data in an attempt to build a reliable spatial temporal model that can be used to forecasting the presence of Chlorophyll-a, using a novel technique called spatial-temporal regression. Our initial results can be used to design a more complex, reliable, and implementable prediction model for real-time forecasting of Chlorophyll-a presence.

Keywords: Chlorophyll-a concentrations · Temporal regression · Illegal fishing prevention

1 Introduction

Oceanography is an important Earth science that studies the physical and biological aspects of the ocean.It requires a large amounts of data for modelling, investigating, predicting and explaining the different natural phenomena. Such data is usually provided by scientific instruments like satellites, oceanographic ships, buoys, and others. Because available data are growing in quantity and in complexity, the need is emerging for a *machine learning* approach to integrate, if not substitute, the more classical statistical approach used in oceanographic research. Among the many applications of oceanography to marine resources management, controlling and preventing illegal fishing stands out as a very important one. Illegal, unregulated, and unreported fishing is becoming more sophisticated, and, as it turn out, oceanographic conditions are predominant predictors of the seasonal variations in fishing effort [3]. Being able to automatically identifying favourable fishing zone is one possible strategy for helping illegal fishing activity monitoring and preventing, and, to this end, automatically identifying favourable oceanic conditions in a promising strategy.

Chlorophyll-a is a specific form of chlorophyll used in oxygenic photosynthesis which has been linked to nutrient presence in several different areas [7, 8, 19]. It is known that certain kind of satellite data can be used to predict the presence of Chlorophyll-a in oceanic areas [7, 9]. In this work, we consider open access data taken from the Copernicus space program, currently used in the European Union for Earth observation and monitoring, in an attempt to build a reliable *spatial-temporal* prediction model for Chlorophyll-a presence around Galapagos Islands, particularly, in the Galapagos Marine Reserve (GMR), with the purpose of creating the basis for an *implementable, cost-effective*, and *reliable* model for potential fishing area prediction to be used in illegal fishing control activities. At certain geographical point the presence of Chlorophyll-a, in combination with relevant physical, chemical, and biological variables of the same point can be thought of as a multivariate spatial-temporal series, in which the Chlorophyll-a plays the role of dependent variable. As such, *multivariate spatial-temporal regression* can be used to estimate not only the functional model, but also the temporal component for each predictor. Multivariate spatial-temporal regression is simply a multivariate regression in which the spatial-temporal component are explicitly taken into account via suitable data transformations. In its simplest form, it consists of adding suitable *lagged* data to the original ones so that the temporal history of an element plays a role in the regression; more complex techniques include automatic optimization of lags, such as in [11]. The problem considered in this paper is particularly complex from the spatial-temporal point of view. Therefore, in this work we want to first assess the expected improvement that lagged data can entail, effectively paving the way toward a more systematic exploration and optimization of the possible data transformations that take into account both the space and time components.

This paper is organized as follows. In Section 2 we give a short account of the current literature that concerns oceanographic data and learning. In Section 3 we give some practical motivations for this work and the problem we want to solve. Then, in Section 4 we describe the data that we have used, the mathematical model that we have applied, and discuss our results, before concluding.

2 Related Work

As the quantity, the complexity, and the availability of oceanographic data grows, machine learning-based approaches to their analysis are becoming ever more common [1]. Typical applications range from climate prediction, habitat modelling, and climate change analysis, to species distribution and identification, resource management, and environmental protection (see [18] for a recent review).

Examples of concrete applications include species identification [10], automatic detection and classification of ocean pollution, oil spills, alga bloom, plastic pollution [6], as well as several fishing control-related applications. Fishing control, and connected activities, in particular, are of special interest in this work.

The surveillance of illegal fishing activities and the detection of abnormal fishing vessel behaviours are critical issues for the management of marine resources. Machine learning techniques were employed for fishing gear recognition starting from *vessel monitor system* (in short, *VMS*) data to detect abnormal VMS patterns of fishing vessels in Indonesia [14]. Moreover, while the coastal fisheries in national waters are closely monitored, at least by some countries, in high seas, there is a lot of uncertainty. For the automatic control of fishing activities in high seas it is necessary to understand the general behaviour of fishing fleets, to enforce fisheries management and conservation measures worldwide. *Satellite-based automatic information systems* (in short, *S-AIS*s) are now commonly installed on the vessels and have the function to control the ships' positions, and have been proposed as a tool for monitoring the movements of fishing fleets in near real-time. Using this data, models have been developed to detect potential fishing activity from trawlers, longliners and purse seiners [17]. However, illegal fishing control is related with ocean resource and habitat management, because it affects the conservation of fishery resources, and taking the correct decisions is often a hard problem, due to the non-availability of specific data. Accordingly [18], machine learning techniques have demonstrated the potential to eliminate data gaps, predict future events, and increase the accuracy of the results. The satellite remote sensing for marine applications started in the early 1960s, with the first pictures of the Earth. Currently, satellites are being used in the indirect detection of fishes, via measuring water temperature, which is the most used environmental parameter in investigations concerning the relationship between environment and fish abundance [16]. Nevertheless, other oceanographic variables exist that can be used to increase the accuracy of the prediction. The lack of datasets that include such oceanographic variables taken directly from permanent observation stations or from ships have contributed to satellite remote sensing playing a pivotal role, considering that satellites offer the opportunity to measure and monitor multiple oceanographic variables at the same time [15]. This approach has been applied on large spatial scales with high temporal resolutions in coastal waters, but while oceanic color satellites suffer of serious limitations, such as the low spatial resolution of sensor systems, using machine learning techniques, such as artificial neural networks and support vector machines, allowed to develop Chlorophyll-a models, for example in [13].

Sea surface temperature and chlorophyll images are considered fundamental for the identification of fishing zones, which in turn is essential for illegal fishing detection. Features such as eddies, gyres, meanders, and upwelling that are indicative of fish abundance areas, can be derived from satellite information. One of the elements that indicate the quality of the water is the concentration of Chlorophyll-a, whose presence is highly correlated with the phytoplankton biomass [7]. Phytoplankton is the base of the food chain in the marine ecosys-

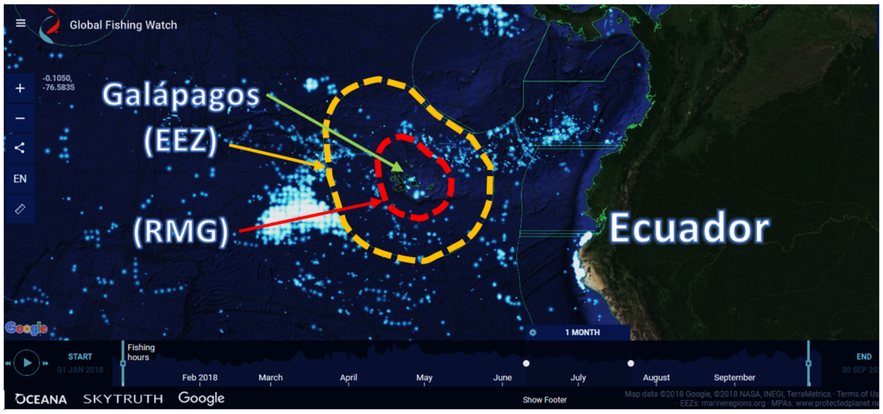

Fig. 1: An aerial view of fishing ships around Galapagos exclusive economic zone. Snapshot taken in August, 2019. Source: *http://www.globalfishingwatch.org*.

tems, and it is the main responsible for the primary production. A very recent multivariate statistical approach to predict Chlorophyll-a levels in coastal marine ecosystems, taking into account 20 variables from 64 observations is reported in [9]. The main differences between [9] and our approach are that the former uses costal data extracted from sensor, instead of high sea data from satellite, and, it does not include a spatio-temporal study of the cause-effect relationships.

3 Motivation

Galapagos Islands are located more than $1000kms$ westward of the Ecuador's continental coast. In 1998, the Government of Ecuador created the *Galapagos Marine Reserve* (in short, *GMR*) to preserve the resources of the islands. In 2001, Galapagos was declared a World Heritage by UNESCO. Due to its location, the islands receive, the influence of two currents, the so-called Humboldt's cold current and Panama's warm current from the east, and the so-called cold and deep Cromwell current from the west. These currents carry waters plenty of nutrients from the sea bottom to the surface. Because of this combination, Galapagos has extremely high productivity areas with diverse marine organisms [12], attracting various species towards the exclusive economic zone of the Galapagos and its surroundings. For this reason, it receives pressure from industrial fishing, principally from Asia, as shown in Fig. 1. Very often, the activities around the maritime limits become in illegal fishery, which is very difficult to control due to the immense Galapagos' maritime territory. The effective control of maritime spaces can only be carried out through satellite monitoring; however, an extremely expensive solution. Less expensive solutions require the use of VMSs and S-AISs systems, which are installed onboard the fishing ships, to monitor the position and fishing activities; although, these systems can be disconnected by the ships when performing illegal fishing activities.

	variable	description	unit
biochemical variables	**Chl**	total chlorophyll-a	mg/m^3
	Fe	dissolved iron	$mmol/m^3$
	NO3	nitrate	$mmol/m^3$
	O2	dissolved oxygen	$mmol/m^3$
	pH	ph	-
	PO4	phosphate	$mmol/m^3$
	Si	dissolved silicate	$mmol/m^3$
	SPCO2	surface CO2	pa
physical variables	**ST**	sea water surface temperature	$°C$
	DT	sea water -40m temperature	$°C$
	So	salinity	$1/e^3$
	Zos	sea surface height	m
	Mlotst	mixed layer depth	m
	Uo	northward sea current water velocity	m/s
	Vo	eastward sea current water velocity	m/s

Table 1: A description of the physical and the biochemical variables used in this experiment.

An alternative solution to help illegal fishing control while reducing the cost of surveillance is being able to predict the areas where fishing activity may take place. This prediction is related to oceanographic variables, chlorophyll levels and sea temperatures being two of the most important ones. It is in fact known that distribution and migration of species is strongly influenced by these two variables [19]. Therefore, in this research we try to develop a model that can predict chlorophyll levels at open sea, in an attempt to identify the areas where is most probable to find a high concentration of chlorophyll. With this information, law enforcement ships can monitor these areas more closely, looking forward to intercept ships during illegal and unregulated fishing activities.

4 Chlorophyll Prediction

Temporal regression. Given a data set A with n independent variables $A_1, \ldots,$ A_n and one observed variable B, solving a linear regression problem consists of finding $n+1$ *parameters* (or *coefficients*) $c_0, c_1, \ldots, c_n$ so that the equation:

$$B = c_0 + \sum_{i=1}^{n} c_i \cdot A_i + \epsilon, \tag{1}$$

where ϵ is a random value, is satisfied. Starting from a data set of observations:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix} \tag{2}$$

the regression problem is usually solved by suitably estimating the coefficients c_i so that, for each $1 \leq j \leq m$:

$$b_j \approx c_0 + \sum_{i=1}^{n} c_i \cdot a_{ij} + \epsilon. \tag{3}$$

There are several available, and well-known algorithms to solve such an inverse problem. The performance of such an estimation can be measured in several (standard) ways, such as *correlation, covariance, mean absolute error*, among others. When A is a multivariate time series, composed by n independent and one dependent time series, then data are temporally ordered and associated to a time-stamp:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 & t_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 & t_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m & t_m \end{bmatrix} \tag{4}$$

Using linear regression to explain B, then, entails that finding optimal coefficients for:

$$B(t) = c_0 + \sum_{i=1}^{n} c_i \cdot A_i(t) + \epsilon, \tag{5}$$

because we aim to explain B at a certain point in time t using the values $A_1(t), \dots, A_n(t)$. Equations (1) and (5) model exactly the same problem, only in the latter the temporal component is made explicit.

Lag (or *temporal*) (linear) regression consists of solving a more general equation, formulated as:

$$B(t) = c_0 + \sum_{i=1}^{n} \sum_{l=0}^{p_i} c_{i,l} \cdot A_i(t-l) + \epsilon. \tag{6}$$

In other words, we use the value of each independent variable A_i not only at time t, but also at time $t-1, t-2, \dots, t-p_i$, to explain B at time t; each $A_i(t-l)$ is associated to a coefficient $c_{i,l}$, which must be estimated, along with each *maximum lag* p_i. There are available techniques, based on standard regression algorithms, that allow one to solve the inverse problem associated to (6). The question to be addressed is how to estimate p_i, for each variable A_i. Moreover, our problem presents a further degree of complexity. Even assuming that we have solved (6), it would not be very useful for illegal fishing control, because it would be a prediction for the current day, while it may be necessary more than one day to reach the geographical points with higher Clorophill-a predictions (which are candidates for illegal fishing areas). To solve this inconvenience, we use linear temporal *forecasting* regression with a temporal range of k units:

	Mean	**Median**	**Maximum**	**Minimum**	**Variance**	**Skewness**	**Kurtosis**	**Std Dev.**
SPCO2	47.48	47.79	62.06	32.77	9.70	-0.30	0.80	3.12
O2	177.70	188.66	230.27	57.95	1630.82	-0.68	-0.64	40.38
NO3	9.96	8.19	26.88	0.20	32.98	0.63	-0.64	5.74
PO4	1.15	1.09	2.10	0.43	0.08	0.53	-0.08	0.28
Si	7.38	6.77	21.65	2.63	9.79	0.82	-0.01	3.13
pH	7.93	7.95	8.03	7.66	0.00	-1.41	2.01	0.06
Fe	9.8E-05	7.0E-05	6.4E-04	5.2E-06	7.7E-09	1.5E+00	2.2E+00	8.8E-05
Chl	0.45	0.32	2.03	0.13	0.09	1.69	2.47	0.31
Vo	-0.05	-0.03	1.11	-1.42	0.05	-0.51	1.85	0.23
Uo	-0.03	-0.03	1.26	-1.41	0.07	-0.06	1.45	0.27
Mlotst	13.30	9.00	87.30	2.90	103.16	1.83	3.77	10.16
So	34.88	34.96	36.68	32.90	0.19	-0.89	1.21	0.44
Zos	0.21	0.20	0.46	0.04	0.00	0.61	-0.01	0.06
ST	24.35	24.36	29.10	18.35	1.89	-0.02	-0.11	1.37
DT	21.19	21.99	28.18	11.41	9.87	-0.51	-0.82	3.14

Table 2: Basic statistical values of physical and biochemical variables recorded during January, 2018.

$$B(t+k) = c_0 + \sum_{i=1}^{n} \sum_{l=0}^{p_i} c_{i,l} \cdot A_i(t-l) + \epsilon. \tag{7}$$

In the particular case in which all p_is are equal, which is the case in our experiments, we denote the unique maximum lag of the problem by using p.

Data. The space program that is currently used in the European Union for Earth observation and monitoring is called Copernicus. It encompasses three complete constellations, each one with two satellites plus an additional single satellite. This system provides 150 TB of open access data every day, including fundamental measurements or estimates of several physical, chemical, and biological oceanic variables. Ocean color information of the Sentinel satellites is employed for monitoring water quality through chlorophyll-a and phytoplankton analysis; other oceanic variables such as wave height, tide, sea current, salinity, temperature, nutrient, and oxygen are used to develop hydrodynamic models to forecast the evolution of ocean variables relevant for aquaculture. Moreover, the temperature, salinity, mixed layer thickness, wind, sea currents, wave heights, mixed layer thickness, chlorophyll, phytoplankton, zooplankton, and nutrients are used to develop models related with oceanic conditions and fish's habitat spatial distribution [2]. Since our ultimate goal is to build a spatial-temporal model that explains, and therefore predicts, Clorophyll-a concentrations, phytoplankton-related values have been excluded from this analysis, as they can be considered *effects* of Clorophyll-a presence, rather than *causes*.

	Chl	**SPCO2**	**O2**	**NO3**	**PO4**	**Si**	**pH**	**Fe**	
	1	-0.15	-0.87	0.88	0.85	0.83	-0.81	0.83	**Chl**
Chl	1	1	0.22	-0.05	0.08	-0.01	0.07	-0.04	**SPCO2**
ST	0.22	1	1	-0.94	-0.89	-0.90	0.87	-0.84	**O2**
DT	-0.48	0.33	1	1	0.97	0.87	-0.88	0.87	**NO3**
Vo	-0.06	0.07	0.06	1	1	0.85	-0.90	0.83	**PO4**
Uo	-0.02	0.12	-0.04	-0.05	1	1	-0.75	0.88	**Si**
Mlotst	-0.42	0.07	0.49	0.14	0.04	1	1	-0.68	**pH**
So	-0.31	-0.43	-0.19	0.14	0.03	0.39	1	1	**Fe**
Zos	-0.11	0.52	0.58	-0.01	-0.08	0.11	-0.49	1	
	Chl	**ST**	**DT**	**Vo**	**Uo**	**Mlotst**	**So**	**Zos**	

Table 3: Correlation matrix for our variables recorded during January, 2018.

Our research area is located around the Galapagos Islands, namely between 6°N and 10°S and between 85° W and 116°W. Our data come from the data base Global Analysis Forecast-PHY-CPL-001-012, containing the values of physical variables, and the data base Global Analysis Forecast-BIO-001-028, containing biochemical variables (see [4, 5]). A summary of the considered variable can be found in Tab. 1. Our data set was constructed with daily mean values of these variables during January, 2018 (we call this data set $A_{training}$) and during February, 2018 (A_{test}), with a spatial granularity of $\frac{1}{4}$ nautical mile in every direction, for a total of 8125 geographically distinct points per day. Our data contained no missing values. Basic statistical values of all 14 physical and biochemical variables can be found in Tab. 2, and their one-to-one correlation can be found in Tab. 3. As it can be appreciated, Clorophyll-a presents a high positive correlation with nutrients (NO3, PO4, Si, Fe), and some negative correlation with oxygen, pH, temperature, and mixed layer depth. Other parameters like surface CO2, salinity, sea surface height, and sea current water velocity seem to have a low correlation.

Our strategy. Our strategy consists of the following three steps:

1. Estimate the maximum lag p using a selection of 1000 randomly chosen geographical points from the data set $A_{training}$, so that the average correlation for temporal forecasting regression with k units, with $1 \leq k \leq 5$ is sufficiently high.
2. Evaluate the decision by applying temporal forecasting regression with k units, with $1 \leq k \leq 5$, with maximum lag p using the data set $A_{training}$ with all available points recorded during January, 2018.
3. Validate the obtained model by testing it with $1 \leq k \leq 5$ on the data set A_{test} with all available points recorded during February, 2018.

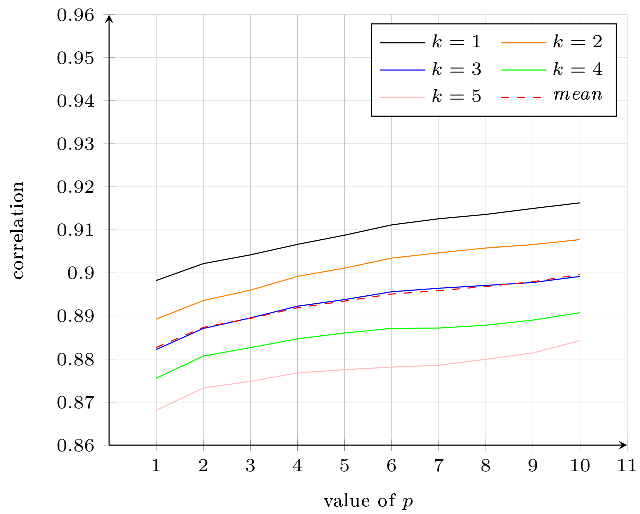

Fig. 2: Estimating the best value of p - full training mode.

Experiments. In our first experiment, 1000 random samples were selected from the $A_{training}$ to fit a multivariate regression model to estimate the levels of Clorophyll-a, and, in particular, to evaluate the effect on the predicting capabilities of our model of adding longer histories to each of our variables. Thus, we instantiate (7) with p from 1 to 10, and with k from 1 to 5. Our goal is to establish how the 10-fold cross-validation values for each value of k increases as p increases. The results can be found in Fig. 2 and Fig. 3, in which we also show the average results among the five groups. As it can be seen, the improvement is clearly visible for p from 1 to 7, and becomes less marked afterwards, but still present at least up to $p = 10$. We conclude that using $p > 10$ is not justified, and we fix the remaining experiments to the value $p = 10$.

Following our strategy, we applied a 10-days lag transformation to the data set $A_{training}$, and evaluated the correlation, the determination index, the root squared error, mean absolute error, and root mean squared error of multivariate regression in 10-folds cross-validation mode. The results are shown in Tab. 4, top. Then, we validated our approach on the set A_{test}, and the results are shown in Tab. 4, bottom.

Discussion. In the first experiment, as we have explained, we evaluated the optimal value of p. Using a random choice of geographical points we guarantee that the vale of p is not biased by local conditions, and using only a small subset of the available points we guarantee that our methodology is, in fact, applicable. As we have seen, the resulting graph has the classical *elbow*-shaped aspect, allowing one to estimate at which point adding new variables is not worth the expected improvement. Also, we can see how our results are consistent for different values

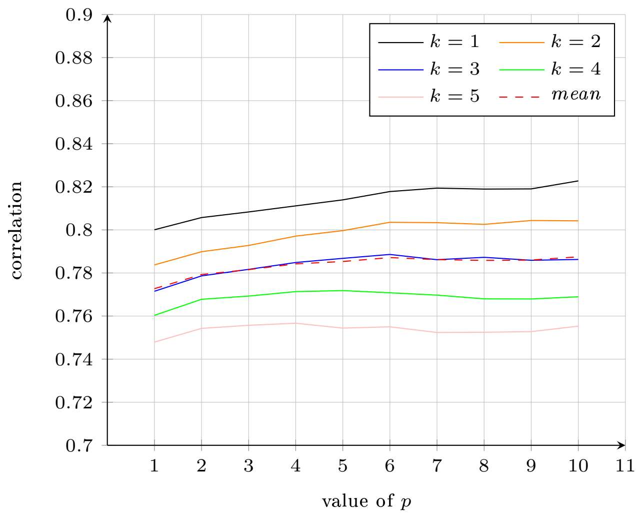

Fig. 3: Estimating the best value of p - 10-folds cross validation mode.

of k, suggesting that our approach is stable. In the second experiment, we set $p = 10$. This means that for one specific prediction we considered 10 days of history per each variable. With our training data, recorder on January, 2018, our multivariate linear regression model shows interesting correlation values ranging from 0.75 (5 days ahead prediction) to 0.82 (1 day ahead prediction). In order to establish the validity of our approach, the computed model was tested on our data set A_{test}, with values recorded during February, 2018. As it turned out, the correlation values are consistent, ranging from 0.82 to 0.87. In Fig. 4, we show how the error in the prediction is geographically distributed on a particular, random, day in February, 2018. As it can be seen, the 'bulk' of the concentration of Chlorophyll-a is correctly predicted even in the 5-days ahead model; the longer the forecast, however, the more difficult is to predict the correct values in low-level concentration areas.

5 Conclusions

In this work we have designed a first prediction model for Chlorophyll-a concentration in the waters surrounding Galapagos Islands, Ecuador. The main motivation behind this study is the design of an implementable, cost-effective system that allows naval forces to guess possible illegal fishing areas with enough time to intervene, and enough accuracy to minimize false alarms. Using open access data, we tacked this problem as a spatial-temporal multivariate linear regression problem, and we focused, to begin with, on the temporal component. While the results seem encouraging, the role of the spatial and the temporal component of this problem are yet to be identified clearly. The ultimate goal of this

	value of k	deter. ind.	r.s.m. err.	m.a. err.	max err.	corr. ind.
10-folds cv $\mathbf{A_{training}}$	1	0.8079	0.0161	0.0836	1.1830	0.8214
	2	0.7861	0.0177	0.0884	1.0985	0.8033
	3	0.7632	0.0191	0.0928	1.0539	0.7836
	4	0.7389	0.0206	0.0965	1.0314	0.7656
	5	0.7188	0.0218	0.0995	1.0751	0.7508
test mode $\mathbf{A_{test}}$	1	0.7301	0.0149	0.0863	0.6514	0.8728
	2	0.7215	0.0153	0.0887	0.6233	0.8670
	3	0.6910	0.0170	0.0937	0.6338	0.8550
	4	0.6018	0.0213	0.1025	0.7468	0.8296
	5	0.5721	0.0223	0.1037	0.7675	0.8264

Table 4: Results of Chloropylle-a forecasting. Full-training mode (top), and test mode (bottom).

project is, indeed, to build a long-term explanation/prediction model, so that naval operations can be planned in a proper way.

As future work, we want to experiment with more complex techniques of data transformation and optimization, in order to establish optimal lags for each independent variables, but also optimal spatial relationships between neighboring geographical points.

References

1. Ahmad, H.: Machine learning applications in oceanography. Aquatic Research **2**(3), 161–169 (2019)
2. ao, R.S.: Blue Book - Corpernicus for a sustaintable ocean. Mercator Ocean International (2019)
3. Cimino, M.A., Anderson, M., Schramek, T., Merrifield, S., Terrill, E.J.: Towards a fishing pressure prediction system for a western pacific EEZ. Scientific reports **9**(1), 1–10 (2019)
4. COPERNICUS: Product User Manual for Global Biogeochemical Analysis and Forecasting Product. Marine Enviroment Monitoring Service (2019), https://resources.marine.copernicus.eu/
5. COPERNICUS: Product User Manual for Global Physical Analysis and Coupled System Forecasting Product. Marine Enviroment Monitoring Service (2020), https://resources.marine.copernicus.eu/
6. Del Frate, F., Petrocchi, A., Lichtenegger, J., Calabresi, G.: Neural networks for oil spill detection using ers-sar data. IEEE Transactions on geoscience and remote sensing **38**(5), 2282–2287 (2000)
7. Desortová, B.: Relationship between Chlorophyll-α concentration and phytoplankton biomass in several reservoirs in Czechoslovakia. Internationale Revue der gesamten Hydrobiologie und Hydrographie **66**(2), 153–169 (1981)

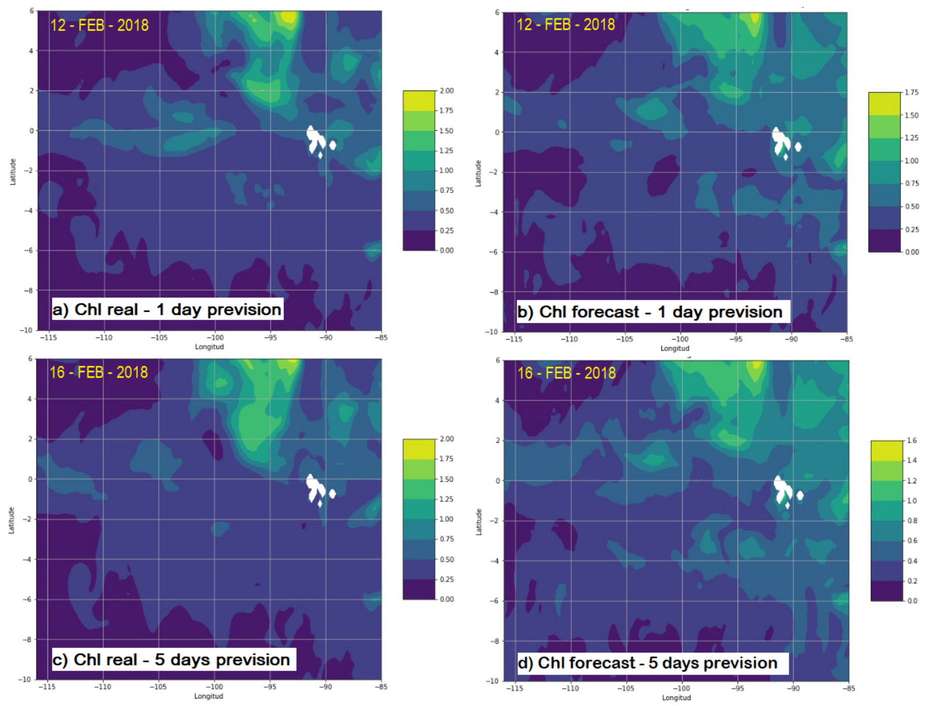

Fig. 4: Real and forecast chlorophyll map in the study area.

8. Dutta, S., Chanda, A., Akhand, A., Hazra, S.: Correlation of phytoplankton biomass (Chlorophyll-a) and nutrients with the catch per unit effort in the PFZ forecast areas of northern bay of bengal during simultaneous validation of winter fishing season. Turkish Journal of Fisheries and Aquatic Sciences **16**, 767–777 (06 2016)
9. Franklin, J.B., Sathish, T., Vinithkumar, N.V., Kirubagaran, R.: A novel approach to predict Chlorophyll-a in coastal-marine ecosystems using multiple linear regression and principal component scores. Marine Pollution Bulletin **152**, 110902 (2020)
10. Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J., Patti, C., et al.: Ipez: an expert system for the taxonomic identification of fishes based on machine learning techniques. Fisheries Research **102**(3), 240–247 (2010)
11. Jiménez, F., Kamínska, J., Lucena-Sánchez, E., Palma, J., Sciavicco, G.: Multi-objective evolutionary optimization for time series lag regression. In: Proceedings of the 6th International Conference on Time Series and Forecasting (ITISE). pp. 373 – 384 (2019)
12. Jones, P.J.: A governance analysis of the galápagos marine reserve. Marine Policy **41**, 65–71 (2013)
13. Kwon, Y.S., Baek, S.H., Lim, Y.K., Pyo, J., Ligaray, M., Park, Y., Cho, K.H.: Monitoring coastal Chlorophyll-a concentrations in coastal areas using machine learning models. Water **10**(8), 1020 (2018)

14. Marzuki, M.I., Gaspar, P., Garello, R., Kerbaol, V., Fablet, R.: Fishing gear identification from vessel-monitoring-system-based fishing vessel trajectories. IEEE Journal of Oceanic Engineering **43**(3), 689–699 (2017)
15. Monolisha, S., George, G., Platt, T.: Fisheries oceanography-established links in the eastern arabian sea (2017)
16. Santos, A.M.P.: Fisheries oceanography using satellite and airborne remote sensing methods: a review. Fisheries Research **49**(1), 1–20 (2000)
17. de Souza, E.N., Boerder, K., Matwin, S., Worm, B.: Improving fishing pattern detection from satellite ais using data mining and machine learning. Plos One **11**(7) (2016)
18. Thessen, A.: Adoption of machine learning techniques in ecology and Earth science. One Ecosystem **1**, e8621 (2016)
19. Zainuddin, M.: Skipjack tuna in relation to sea surface temperature and Chlorophyll-a concentration of Bone Bay using remotely sensed satellite data. Jurnal Ilmu dan Teknologi Kelautan Tropis **3**(1) (2011)