Emirates Journal of Food and Agriculture. 2020. 32(5): 345-357
doi: 10.9755/ejfa.2020.v32.i5.2104
http://www.ejfa.me/

RESEARCH ARTICLE

Chlorophyll-a concentration assessment using remotely sensed data over multiple years along the coasts of the United Arab Emirates

Eihab M. Fathelrahman[1]*, Khalid A. Hussein[2], Safwan Paramban[1], Timothy R. Green[3], Bruce C. Vandenberg[4]

[1]Department of Integrative Agriculture, College of Food and Agriculture, United Arab Emirates University, P.O. Box 15511, Al Ain, UAE, [2]Department of Geography and Urban Sustainability, College of Humanities and Social Science, United Arab Emirates University, P.O. Box 15511, Al Ain, UAE, [3]Water Management and Systems Research, Agricultural Research Services (ARS), USDA, 2150 Centre Ave., Bldg. D., Suite 200, Fort Collins, Colorado, 80526. U.S, [4]Center for Agricultural Resources Research, 2150 Centre Avenue, Suite 320, Fort Collins, CO 80526, U.S.A

ABSTRACT

The United Arab Emirates (UAE) recently witnessed algal/phytoplankton blooms attributed to the high concentrations of Chlorophyll-a associated with the spread and accumulation of a wide range of organisms with toxic effects that influence ecological and fishing economic activities and water desalination along coastal areas. This research explores the UAE coasts as a case study for the framework presented here. In this research, we argue that advances in satellite remote sensing and imaging of spatial and temporal data offer sufficient information to find the best-fit regression method and relationship between Chlorophyll-a concentration and a set of climatic and biological explanatory variables over time. Three functional forms of regression models were tested and analysed to reveal that the Log-Linear Model found to be the best fit providing the most statistically robust model compared to the Linear and the Generalised Least Square models. Besides, it is useful to identify the factors Sea Surface temperature, Calcite Concentration, Instantaneous Photosynthetically Available Radiation, Normalized Fluorescence Line Height, and Wind Speed that significantly influence Chlorophyll-a concentration. Research results can be beneficial to aid decision-makers in building a best-fit statistical system and models of algal blooms in the study area. The study found results to be sensitive to the study's temporal time-period length and the explanatory variables selected for the analysis.

Keywords: Chlorophyll-a concentration; Remote sensing; Water quality; Best-Fit regression model; Coastal areas

INTRODUCTION

Chlorophyll-a (Chl-a) concentration can be considered as a direct indicator to assess the biophysical characteristics of a water body. Chl-a has been useful for measuring the variety and abundance of phytoplankton and/or the algal biomass (Boyer et al. 2009). Hence, geospatial technology is one of the suitable methods for the representation of the Chl-a variability in the area (Matthews et al. 2001). The concentration of Chl-a has also been used as a key parameter to represent water quality conditions (Le et al. 2013 and Cheng et al. 2013). The spatiotemporal difference of Chl-a concentration results can be used to perform further analysis of the water quality level and help improve water resource management. The concentration of Chl-a varies spatially from place to place in coastal regions. Since conventional methods for studying water quality by field sampling are expensive, time-consuming, and spatially incomplete, remote sensing is a highly applicable alternative method for the large, medium, and small scale mapping of water quality. Satellite imageries of various sensors have been successfully applied to water quality mapping and monitoring in terms of Chl-a concentration, seas surface temperature and dissolved organic matter by using several ocean colour algorithms, such as OC3 and FLH (Fengyun 2010).

The marine ecosystems of the UAE are fragile and susceptible to algal bloom events including the harmful algal blooms (HABs), which named red tide (Al Shehhi

***Corresponding author:**
Eihab M. Fathelrahman, Department of Integrative Agriculture, College of Food and Agriculture, United Arab Emirates University, P.O. Box 15511, Al Ain, UAE. **Phone:** +971 3 713 4589. **E-mail:** eihab.fathelrahman@uaeu.ac.ae

Received: 10 January 2020; **Accepted:** 20 May 2020

et al 2014; UAE-MOEW 2015). UNESCO (2015) indicated that the algal blooms are known to have several names including phytoplankton blooms, micro-algal blooms, toxic algal blooms, or red tide (UNESCO 2015). The UNESCO recognized about three hundred species of microalgae. One-fourth of these species produce toxins. Such algal bloom events could be captured via remote sensing capabilities by measuring the Chl-a presence and accumulation over time. All such blooms could be captured via remote sensing capabilities through measuring the Chl-a presence and accumulation over time. The UNESCO also indicated that about three hundred species of microalgae were recognized as they are forming mass occurrence known as blooms. About one-fourth of these species produce toxins. The scientific community refers to these events HABs, recognizing a wide range of organisms' accumulation involved and that some species have toxic effects at low cell concentrations.

Since the 1990s, the advancement of satellite imaging technology offers efficient means for understanding the spectral signature of water quality parameters. This can be achieved by improvements of mathematical models have led to semi-empirical methods becoming an important technique for monitoring water quality remotely (Guo et al 2016). These remote sensing studies found that Chl-a estimation procedure and variables selection show the importance of considering the optical properties of the water body. For example, in ocean water, phytoplankton is generally the principal constituent and the concentrations of other components vary with Chl-a. Thus, the properties of these waters are dominantly phytoplankton and they can be observed using spectral features in the reflected light can be directly related to Chl-a concentration (Moses et al 2009).

Satellite ocean colour sensors' data, remote sensing methods, and algorithms are widely used for the detection, mapping, and monitoring of phytoplankton blooms as the earth observation provides synoptic recognition of the ocean, at both spatially and temporally dimensions (Blondeau-Patissier et al 2014). Many satellites have been developed for spatial, temporal, and radiometric resolutions of ocean colour and surface water temperatures to obtain high accuracy levels (Steissberg et al. 2005). However, most of these satellites offer data with a frequency of every 16 days, such as Landsat and ASTER (Advanced Space borne Thermal Emission and Reflection Radiometer), for a given location. Besides, Landsat does not record data at night. For example, the Moderate Resolution Imaging Spectroradiometer (MODIS) is advanced technology. As well the High-Resolution Radiometer (HRR) component overcome these temporal resolution limitations and of estimating and ocean colour and surface temperature (Lu et al 2011). They provide daily coverage of the Earth's surface at a resolution of 1×1 km with a large scan angle (Justice et al 1998). The main limitation of these sensors is their inability to penetrate clouds, which makes them available for only clear-sky conditions (Prigent et al 2003). In Lake Okeechobee, Florida, Lamon (1995) generated a regression model to predict the Chl-a levels in different zones of the lake.Temperature, total nitrogen, total phosphorous, wind velocity, and Chl-a data were used in this study. Lamon applied exploratory data analysis to determine the spatial and temporal scale of variations in Chl-a levels. Ali et al. (2013) studied the variability of water quality parameters that can help in understanding the pattern of changes and assessing the sustainability of the creek. DubaiSat-1 imagery and Spectral-based and field measured chlorophyll-a concentrations at the five stations data along the creek have been used to study the variability of chlorophyll-a as an essential algal growth indicator. Based on this study, the implications of urban development in the UAE creek influence the variability of the concentrations of chlorophyll-a, phosphates, and total nitrogen. The analysis showed a high correlation between the spectral-based chlorophyll-a concentrations on one hand and the field-based phosphates and total nitrogen concentrations on the other hand.

To assess the likely incidences of coastal algal blooms using field observations Wong *et al*. (2007) suggested a simple and practical algal bloom statistical model using vertical stability theory and framework for shallow coastal waters. The model included three-dimensional hydrodynamic circulation modules. The simulated processes included algal growth, decay, settling, and vertical turbulent mixing. This study assessed the possibility of algal bloom occurrence and evaluated the results based on field observations over four years. The model's outcome supports the explanation of the observed spatial and temporal patterns of bloom occurrences concerning the water nutrient condition and vertical turbulence and waterbody nutrient conditions.

Malek *et al*. (2011) applied statistical models to evaluate the Chl-a concentrations in tropical Putrajaya Lake, Malaysia. The study aimed to evaluate the performance of four different models, namely Recurrent Artificial Neural Network, Fuzzy Logic, Hybrid Evolutionary Algorithm (HEA) and Multiple Linear Regressions to predict the concentration of Chl-a. The authors used the root mean square Error (RMSE), correlation coefficient of determination (R^2), and area under the receiving operating characteristic (ROC) curve to assess the fitness of these models. The authors suggested such evaluations are important towards developing a trustworthy algorithm to estimate Chl-a concentration for eutrophication management of tropical lakes. They concluded that HEA produced the best performance compared to the other models. The authors suggested such evaluations are important towards developing a trustworthy algorithm to estimate Chl-a concentration for eutrophication management of tropical lakes. Malek et al. (2011) concluded that HEA produced the best performance for their case study.

Kiefer et al. (2015) used Chl-a concentration as a signal of phytoplankton abundance and trophic level at Lake Geneva, Switzerland. The authors used 11,234 satellite images from the Medium Resolution Imaging Spectrometer (MERIS) sensor on the Envisat satellite from 2002 to 2012-time period to quantify the spatial and time-based variations of Chl-a concentration. The methods used to perform the analysis included calibration of remotely sensed data and reference datasets using linear regression, Chl-a evolution, and spatiotemporal variability using means and standard deviation of the lognormal distribution, and spatial representativeness of two selected locations. The analysis of time-based evolution showed an overall decline of Chl-a concentrations in Lake Geneva but also strong spatial variability between different segments of the lake.

Rousseaux and Watson (2017) applied a biochemical model to forecast ocean Chl-a. The model was able to reproduce the dominant features of variation of the Chl-a concentration in the study area. The results indicated the potential for forecasting Chl-a concentration in this region but also highlighted various deficiencies and propulsions for enhancements to the current biogeochemical forecasting system. The authors suggested the potential outcome of their analysis to offer a fundamental basis for future applications, including the effects of El Niño events on fisheries. A semi-analytical approach of deriving Chl-a concentration was applied to the light absorption coefficient of phytoplankton. Zheng and DiGiacomo (2017) introduced a model named generalized stacked-constraints model (GSCM) to partition satellite-derived total light absorption coefficient of water into phytoplankton and non-phytoplankton components. The study shows that semi-analytical methods can provide a sufficient Chl-a product compared to reflectance-band-ratio algorithms. The GSCM to satellite ocean-colour data analysis revealed the issue of infeasible solutions when input data are subject to significant errors.

To discuss and analyse the spatial and temporal distribution of chlorophyll-a and its correlation with salinity and total suspended solids (TSS) in the seawaters at the Cirebon in West Java, Indonesia, Buditama et al. (2017) developed an appropriate model. The objective of this research was to offer a source of information for fishermen, and other relevant stakeholders to predict fertile water regions which can be used as an indicator in discovering potential areas to catch pelagic fish in the study area seawaters. Chlorophyll-a concentration, salinity, and TSS were identified in this study, using remote sensing data obtained from Landsat-8 Operational Land Imager (OLI) multi-temporal images according to dry and wet month parameters in the years 2014-2015. The results of the research showed that chlorophyll-a levels tended to be higher in wet months relative to dry months. The research found that the distribution of chlorophyll-a concentration is affected negatively by the accumulation of salinity but positively by increasing the total suspended solids.

Recently, Arabi et al. (2018) used remote sensing of water constituent concentrations of time-series data at the Wadden Dutch Sea during the years 2008-2010. The capability of the 2SeaColor model to retrieve accurate estimates, and the favourable location of the study area, which is mildly influenced by tidal phase variations, contributed to a further understanding of the long-term variability of Chl-a concentrations. The results of this study support the ongoing efforts on the Sentinel-3 Ocean and Land Colour Instrument (OLCI) calibration and validation at the study area (Arabi et al. 2018). In other previous studies, Feng et al. (2015) found the chlorophyll generally declined with increasing temperature and light. Meanwhile, Shen et al. (2018) found that the spatial distribution of high Cha concertation is negatively correlated with low Sea-Surface-Temperature (SST) in the studies area in China. Poll et al. (2013) assessed the relationship between SST and vertical density stratification, nutrient concentration and phytoplankton biomass, and composition. The authors suggested stratification temperature does not necessarily result in phytoplankton biomass standing in the region studied. Elawad (2010) studied the inter-annual Chlorophyll variations in the Red Sea and noticed that physical parameters were found to have control of phytoplankton blooms. For example, the author considered the positive effects of wind speed and solar radiation on Chl-a concentration.

All previous studies recommend inclusion of biophysical, weather, and seasonality when studying the factors that influence Chl-a presence and accumulation over time. These studies also agreed that advances in remote sensing offer a sufficient set of explanatory variables to build statistical models to study algal blooms phenomena. In this study, we modify regression models used in previous scholarly work and argue that climatic and biophysical explanatory variables derived from remote sensing data are sufficient for developing such models. We also test the sensitivity and the fitness of the developed statistical functional forms.

RESEARCH OBJECTIVES

The present research explores statistical analysis between Chl-a presence and accumulation that can be captured through remote sensing data for the study area (UAE Coastal region that covers in the Arabian Gulf and the Gulf of Oman). Such presence of Chl-a was regressed against the explanatory variable to explore the importance of such variables and find the optimum best fit statistical model that represents the Cha-a obtained data. This research tests the hypothesis that advances in remote sensing offer sufficient information to statistically fit the model that represent such a relationship between Chl-a concentration and the explanatory variables. All relevant biological, physical, and environmental influential factors across the selected period from January 2008 to December 2011 are captured, summarised, and used to study the regression relationships. The study period was selected due to several observed events of algal blooms outbreaks in the study area (e.g. Richlen et al 2010; Zhao et al 2014).

The specific objective of the study is to identify the best-fit statistical functional form model that represents the correlation between Chl-a concentration that was retrieved using remotely sensed data as a dependent variable, and the selected explanatory variables. Explanatory variables are Particulate organic carbon, Sea Surface Temperature, Calcite Concertation, Instantaneous Photo-synthetically Available Radiation, Normalized Fluorescence Line Height, and Wind Speed). The results of the study are evaluated with data from the year 2013 not used for calibration. Furthermore, the study estimates the best statistical model parameters, summarises the results, and discusses the interpretations and implications of such results. All of this information can be useful for the development of a model of Chl-a presence and accumulation.

STUDY AREA

The study area (Fig. 1) covers parts of the two gulfs, the Arabian Gulf and the Gulf of Oman. The Arabian Gulf is located between latitude 24.0º N and 30.0º N and longitude 48.0^{0}E and 57.0^{0} E. Meanwhile, the Gulf of Oman situated between 22.0^{0} N to 26.0^{0} N and 56.0^{0} E to 60.0^{0} E. The total length of the coastline of the UAE is 1318 km, which consist 650 km at the Arabian Gulf and remaining 668 km at the Gulf of Oman.

DATA AND METHODS

The MODIS-Aqua Level 2 data of the Arabian Gulf and the Gulf of Oman, obtained from the National Aeronautics and Space Administration (NASA) Goddard Space Flight Centre (NASA's Ocean colour website: https://oceancolor.gsfc.nasa.gov/), were used in this study. The data covered the period 2008 to 2011 in which algal bloom events have occurred. The dependent variable Chl-a concentration and independent variables Calcite Concertation, normalized fluorescence line-height, particulate organic carbon, instantaneous photosynthetically available radiation, and sea surface temperature were derived from the dataset. Definitions of the variables are presented in Table 1. The data were analysed using SeaDas, ENVI, and ArcMap software. The median values of the variables mentioned above were used to carry out the regression analysis. The median was chosen because it is not biased, and it eliminates the impact of outliers compared to the mean. Chl-a concentration was estimated using the Ocean Colour Index (OCI) algorithm (Hu et al. 2012). The daily data of wind speed were downloaded from the National Oceanic and Atmospheric Administration (NOAA) - National Centres for Environmental Information (https://www.ncdc.noaa.gov/cdo-web/).

The spatial resolution of the processed MODIS data is 1 km^2, and wind speed data is 0.25X0.25 degrees. The total number of days of data for the years 2008, 2009, 2010, and 2011 was 490, 120, 100, 143, and 127, respectively for the period covered. The data for the year 2013 (132 days) were used for model evaluation. The number of days varied from year to year due to the absence of algal blooms in the study area during certain days. For example, the year 2009 witnessed the least number of algal bloom days. Selected dependent variable (Chl-a concentration) and independent variables calcite concertation, normalized fluorescence line-height, particulate organic carbon, instantaneous photosynthetically available radiation, and sea surface temperature maps are illustrated in the results and discussion section of the paper. The mathematical algorithms used in this research to generate the variables are described in Appendix A. Descriptive statistics of the variables used for the base model (for the period 2008-2011 development and the model evaluation during the year 2013) are shown in Table 2 and Table 3, respectively.

Statistical models

Three statistical functional forms are used in this study a Linear Model, Semi-Log Model, and a Generalized Least Square Model (GLS). The purpose of such statistical analysis is to select the most appropriate statistical functional form that represents the "Best" fitness model to generalize on the results and offer interpretations of the relationship between the dependent variable Chl-a as an index for algal and set of explanatory/independent variables (i.e. the percentage change in the independent variables) as follow:

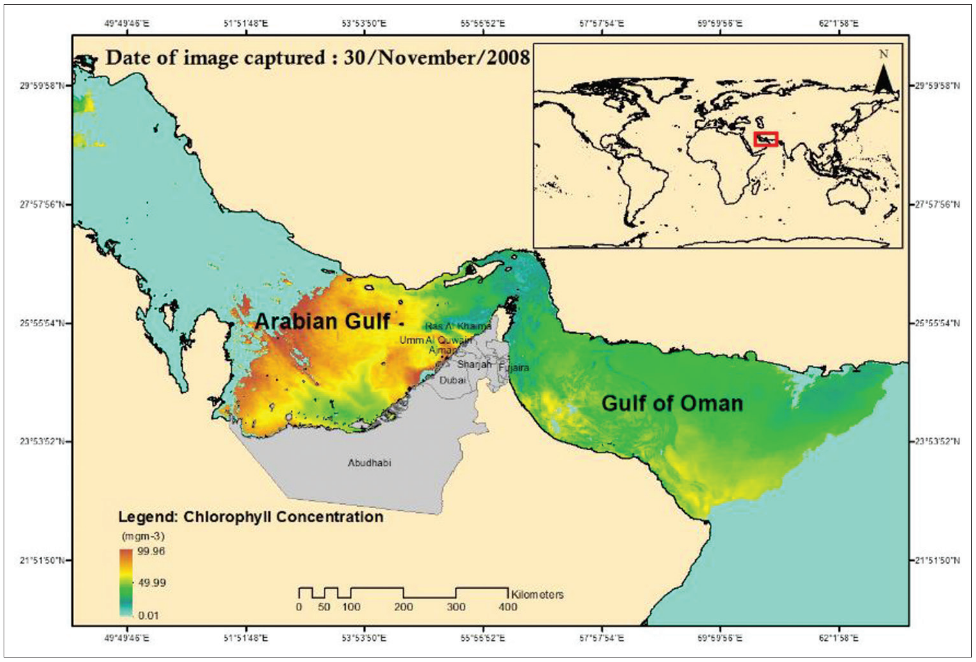

Fig 1. Research Study Area.

Table 1: Regression dependent and explanatory variables definitions

Variable	Definition	Measuring Unit
Chlorophyll-a Concentration	The amount of photosynthetic plankton present in the ocean and a water quality parameter providing food to aquatic life. (Nazeer and Nichol, 2015).	mg m^{-3}
Calcite concentration	Insoluble mineral in ocean surface waters, including dead calcareous plankton shells.	mol m^{-3}
Particulate organic carbon	A large component of organic matter in the carbon cycle, which serves as a primary food source for aquatic foods.	mg m^{-3}
Instantaneous photosynthetically available radiation	Total PAR (Photosynthetically Available Radiation) incident on the ocean surface at the time of the satellite observation (https://oceancolor.gsfc.nasa.gov/).	Einstein $m^2 s^{-1}$
Normalized fluorescence line height	Wavelength signal that describes the spectral distribution above the surface of the water body, measured by the difference between the observed NLW (678 nm) and a linearly interpolated NLW (678 nm) from two surrounding bands (https://oceancolor.gsfc.nasa.gov/). NLW stands for normalized water-leaving radiance.	$Wm^{-2} \mu m^{-1} sr^{-1}$
Sea surface temperature	A measure of the energy due to the motion of molecules at the top layer of the ocean (https://podaac.jpl.nasa.gov/)	degree
Sea surface wind speed	Movement of the atmosphere relative to the surface of the sea (https://www.ncdc.noaa.gov/dataaccessmarine-ocean-data/blended-global/blended-sea-winds).	m s^{-1}

Table 2: Dependent and explanatory variables descriptive statistics

Variables	Units	Minimum	Maximum	Mean	Standard Deviation	CV*
Chlorophyll-a	mg m^{-3}	0.096	2.9	0.8	0.5	61%
Sea surface temperature	°C	2.5	32.8	25.1	3.8	15%
Calcite concentration	mol m^{-3}	0.001	0.7	0.001	0.029	1970%
Particulate organic carbon	mg m^{-3}	0.001	431.9	168.0	76.4	45%
Instantaneous photosynthetically available radiation	Einstein $m^{-2} s^{-1}$	0.001	0.2	0.0	0.009	414%
Normalized fluorescence line height	W $m^{-2} \mu m^{-1} sr^{-1}$	0.001	0.3	0.1	0.1	42%
Seasonal value	-	0	1	0.9	0.2	24%
Wind speed	m s^{-1}	0.683	11.3	4.7	2.1	45%

*CV = Coefficient of Variation
Number of Observations: 490 for the whole period covered between 2008 and 2012

Table 3: Descriptive Statistics of the variables used for the Validation Year 2013

Variables	Units	Minimum	Maximum	Mean	Standard Deviation	CV*
Chlorophyll-a	mg m $^{-3}$	0.082	1.8	0.7	0.4	53%
Sea Surface Temperature	°C	10.2	31.2	25.3	3.5	14%
Calcite Concentration	mol m $^{-3}$	0.001	0.003	0.001	0.001	177%
Particulate Organic Carbon	mg m $^{-3}$	0.001	371.2	160.0	72.1	45%
Instantaneous Photosynthetically Available Radiation	einstein m^{-2} s^{-1}	0.001	0.2	0.004	0.022	525%
Normalized Fluorescence Line Height	W m^{-2} μm^{-1} sr^{-1}	0.002	0.3	0.1	0.1	43%
Seasonal Value	-	0	1	0.9	0.3	35%
Wind Speed	m s^{-1}	0.805	11.5	4.5	2.3	51%

*CV = Coefficient of Variation Number of Observations: 132

Linear functional form:

$$Y_i = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \mu_i \quad (1)$$

Semi-Log functional form:

$$Log\, Y_i = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \mu_i \quad (2)$$

Generalized Least Square (GLS) functional form

GLS assumes the heteroscedastic variance σ_i^2 are known to obtain the functional form, divided by σ_i :

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{01}}{\sigma_i}\right) + \beta_1\left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2\left(\frac{X_1}{\sigma_i}\right) + \beta_3\left(\frac{X_2}{\sigma_i}\right) + \beta_4\left(\frac{X_3}{\sigma_i}\right) + \beta_5\left(\frac{X_4}{\sigma_i}\right) + \beta_6\left(\frac{X_5}{\sigma_i}\right) + \beta_7\left(\frac{X_6}{\sigma_i}\right) + \beta_8\left(\frac{X_7}{\sigma_i}\right) \quad (3)$$

Where
Y_i = Chl-a concentration in mg m^{-3}
α = constant
X_1 = Sea Surface Temperature in °C
X_2 = Calcite Concentration in mol m^{-3} = 1.0 × 10^{-6} kg / m^{-3}
X_3 = Particulate Organic Carbon in mg m^{-3} = 1 mol / m^3
X_4 = Instantaneous Photosynthetically Available Radiation in Einstein m^{-2} s $^{-1}$ = 1 m^{-2} s^{-1}
Einstein is a unit defined as the energy in one mole (6.022×10^{23}) of photons
X_5 = Normalized Fluorescence Line Height in W m $^{-2}$ um $^{-1}$ sr $^{-1}$
X_6 = Seasonal Value, equals 1 during high fishing and other economic activities months in the winter from first of January to end of April each year and 0, otherwise
X_7 = Wind Speed in W m -2 um -1 sr -1 m2/sr = Watt per square meter per steradian

μ_i = regression error term not captured by the explanatory variables

RESULTS AND DISCUSSION

Table 2 shows the dependent variable Chl-a concentration and each of the explanatory variables with descriptive statistics to explore the distributions of the variables. Considering the ratio of the standard deviation to the mean value (coefficient of variation), the descriptive statistics do not infer the potential of outliers except for Organic Carbon, which is expected in the study area due to the oil export activities. The descriptive statistics for the validation year 2013 also show the distribution of dependent variables that are normally distributed with no outliers. This is shown because the coefficient of variation values are less than 100 % for all the variables in Tables 2 and 3.

As shown in Fig. 2, spatial patterns of the Chl-a concentration sensed in December over four years (2008-2011) vary substantially from year to year. No consistent pattern of the Chl-a concentration can be detected from Fig. 2. This indicates the need to use the statistical model and analysis to investigate the factors that influence the changes in the Chl-a concentration over time.

This study aims to compare the statistical outcome for three functional forms used to identify the best-fit model that represents the regression between the dependent variable (Chl-a concentration) and the explanatory variables listed in equations 1, 2, and 3 above. Table 4 below shows the results for three selected function forms/models for Chl-a in mg m^{-3} as the dependent variable against the explanatory variables (i.e. particulate organic carbon, sea surface temperature, calcite concertation, instantaneous photosynthetically available radiation, normalized fluorescence line-height, and wind speed).

Overall, the three functional forms were found to represent the data with R^2 above 0.6. However, the best-fit model was found to be the *semi-log model* with an R^2 value of 0.84 and an F-distribution of 357. These values were the highest compared to the other two models, the Linear and the Generalized Least Square (GLS) model results.

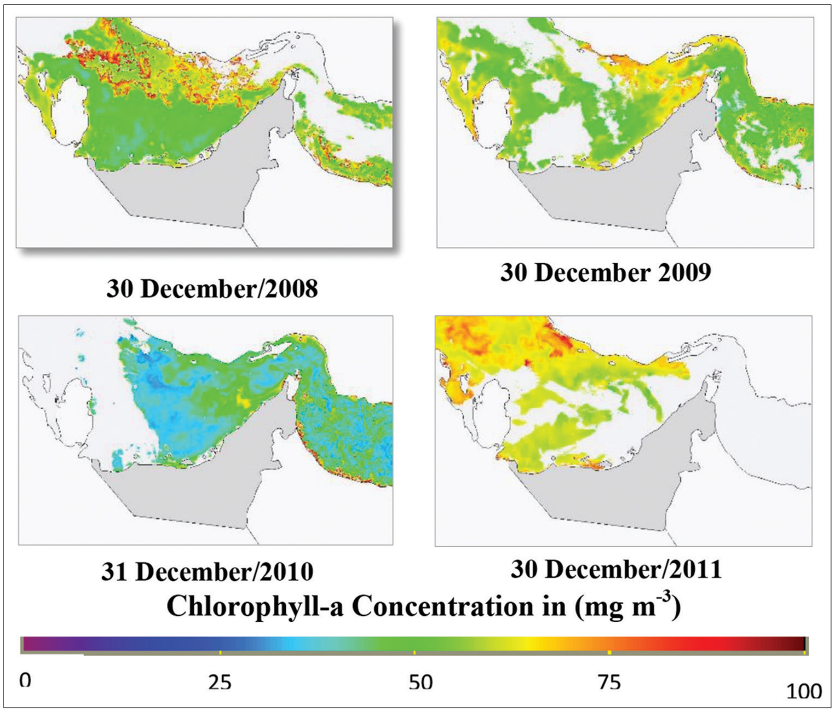

Fig 2. Dependent Variable Chlorophyll-a Concentration Spatial Distribution in mg m-3.

Table 4. Statistical results for the three functional forms models

Goodness-of-Fit Metrics	Linear	Semi-Log	Log-Log Generalised Least Squares (GLS)	Validation Model Parameters
R	0.906	0.916	0.939	0.938
R^2	0.821	0.838	0.638	0.879
Adjusted R^2	0.818	0.836	0.633	0.872
F-Test	315.795	357.274	121.263	128.711
P-Value	0.000**	0.000**	0.000**	0.000**
Dufigurerbin-watson	1.830	1.795	1.780	2.156*
Coefficients	**Linear**	**Semi-Log**	**Log-Log Generalised Least Squares (GLS)**	**Validation Models Parameters**
Sea surface temperature				
Beta	0.057	0.073	0.007	-0.035
T-Test	2.565	3.468	0.226	-0.957
P-Value	0.011*	0.001*	0.821	0.341
Calcite concentration				
Beta	0.070	0.075	0.332	-0.164
T-Test	3.563	4.040	11.055	-4.261
P-Value	0.0001**	0.0001**	0.0001**	0.0001**
Particulate organic carbon				
Beta	0.809	0.761	0.429	0.868
T-Test	35.231	34.842	14.609	19.669
P-Value	0.0001**	0.0001**	0.0001**	0.0001**
Instantaneous photosynthetically available radiation				
Beta	-0.072	-0.076	-0.067	0.077
T-Test	-2.905	-3.242	-1.988	2.230
P-Value	0.004	0.001*	0.047	0.028*
Normalized fluorescence line height				
Beta	0.175	0.243	0.346	0.223
T-Test	6.819	9.963	10.607	5.352
P-Value	0.0001**	0.0001**	0.0001**	0.0001**
Seasonal value				
Beta	-0.006	0.065	0.128	0.048
T-Test	-0.280	3.288	4.358	1.386
P-Value	0.780	0.001*	0.0001**	0.168
Wind speed				
Beta	-0.043	-0.054	-0.105	-0.016
T-Test	-2.154	-2.859	-3.750	-0.491
P-Value	0.032	0.004*	0.0001**	0.624

The parameters coefficients of the best-fit model can be interpreted statically. The dependent variable value will increase by 1% as sea surface temperature decreases by 0.035 % Calcite Concentration decreases by 0.164%, Particulate Carbon increases by 0.868%, Instantaneous Photo-synthetically Available Radiation increases by 0.077% and Normalized Fluorescence Line Height increases by 0.223%. Furthermore, the best-fit model's results showed that all explanatory variables are significant at either 99% or 95% level of confidence. The P-values of the explanatory variables in this model are less than 0.001 for sea surface temperature, less than 0.0001 for calcite concertation, less than 0.0001 for particulate organic carbon, less than 0.01 for instantaneous photosynthetically available radiation, less than 0.0001 for normalized fluorescence line-height, less than 0.001 for seasonal value, and less than 0.005 for wind speed. In other words, *a mix of climatic and biological variables* seem to have more influence on the Chl-a concentration in the study area. This is mostly because the region is witnessing significant carbon and organic matter accumulation at the UAE Arabian Gulf and the Gulf of Oman shores due to increasing economic development activities. These results indicate that both environmental and biophysical (e.g. organic matter) explanatory variables are important in building the best statistical regression model. The study included such variables based on an agreement with similar studies that included similar cases (Wong et al. (2007); Elawad (2010); Zheng and DiGiacomo (2017) and Shen et al. (2018)).

The study results are evaluated with data from the year of 2013, which shows that all variables other than wind speed are significant. The P-values of validation model parameters less than 0.0001 for calcite concentration, particulate organic carbon, and normalized fluorescence line-height, 0.028 for instantaneous photosynthetically available radiation, 0.168 for seasonal value, 0.341 for sea surface temperature, is 0.624 for wind speed (Table 4). Thus, the study results showed that climatic, biological, and fisheries activities are significant factors determining the algal concentration, using Chl-a as an indicator for such analysis.

This study has used the advances in remotely sensed data to study the statistical regression correlation between the concentration of Chl-a and factors that influence its presence and concentration. Fig. 2 shows the histograms generated for the selected regression models. These models include a) Linear Model, b) Semi Log Model, c) Generalized Least Square Model, and d) Validation Model. The histograms show a similarity of the estimated dependent variable Chl-a between the linear model and the generalized least square model due to the similarity of the models' mathematical specifications. However, as expected, the semi-Log model outcome showed a unique shape of the histogram due to its mathematical difference – see equations 1 to 3 for the mathematical specification of each model.

There is an agreement between the three statistical models on the direction (signs) and the inference of a causal relationship between the dependent variable on the one hand, and the selected explanatory variables on the other hand. However, the three models' results indicated significant variation in the magnitude of the contribution of each of the explanatory variables on the dependent variable Chl-a. In other words, the models' were sensitive to the selection of the functional form. The best regression model results, using the log-linear functional form, for the base period 2008-2011 and the validation period of the year 2013 are illustrated in Figs. 3 and 4 respectively. The model validation, interestingly, shows a similar but higher overall regression fitness compared to the Log-Linear model (the best-fit model) for the temporal study period 2008-2011. However, the validation period model shows differences in

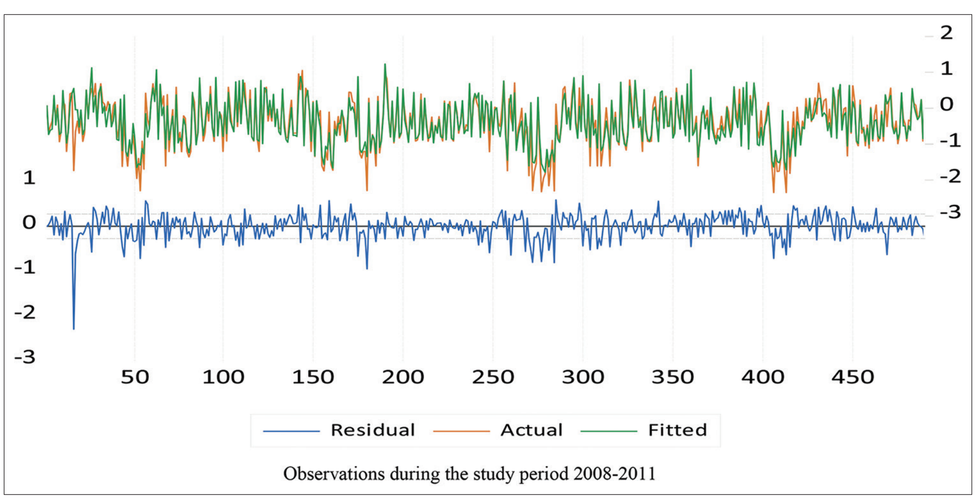

Fig 3. Time series of log-transformed Chl-a Concentration: best fit and residuals during the study period 2008-2011.

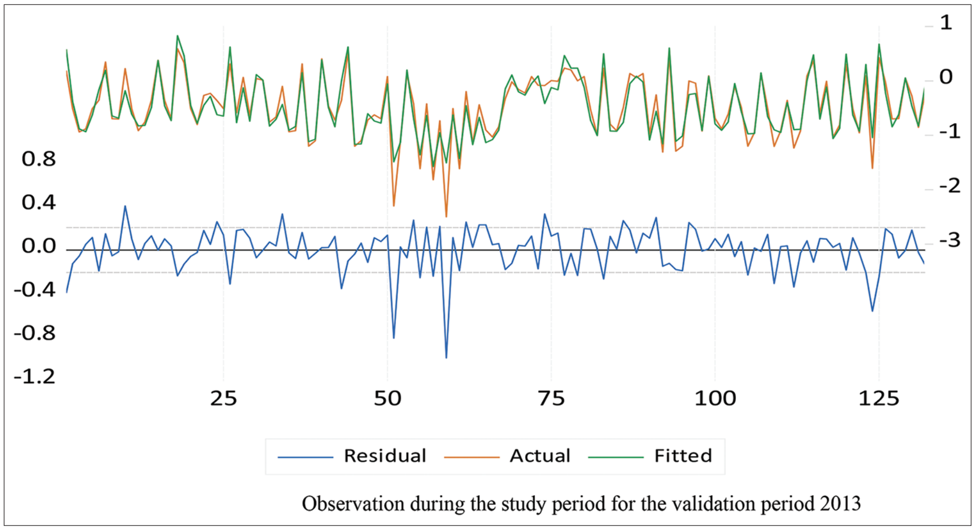

Fig 4. Time series of log-transformed Chl-a Concentration: best fit and residuals.

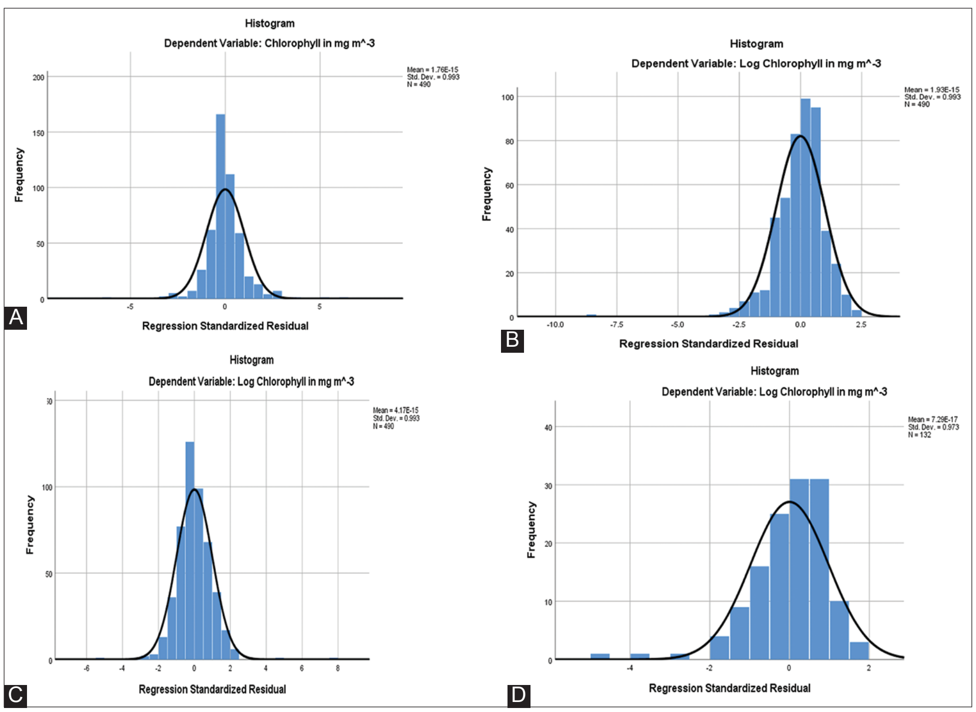

Fig 5. Histograms generated for the Models. A) Linear Model, B) Semi Log Model, C) Generalized Least Square Model, D) Validation Model.

the sign of the explanatory variables sea-surface temperature and Calcite Concentration. This indicates short time analysis against longer-time series analysis for the Chl-a concentration may produce different results for specific explanatory variables. In other words, this study found results to be sensitive to the study's temporal time-period length and the explanatory variables selected for the analysis.

The histograms in Fig. 5 illustrate the distributions of the estimated dependent variable at each type of statistical function forms, Linear Model, Semi-Log Model, Generalized Least Square Model, and the Validation Model predicted results. The results showed the best-fit model taking a normal distribution bell-shape curve. The best-fit model, Semi-Log model, is then used to derive results for the validation period in the year 2013. The results of the validation model show the overall statistical fitness of the functional form selected in this study, which is the Semi-Log model.

CONCLUSIONS

Fishing and water desalination activities have been affected negatively in the UAE by high concentrations of algal. This study quantified the concentrations of Chl-a in space and time using remote sensing data sources in the study area (Arabian Gulf and Gulf of Oman) and relevant biological, physical, and environmental influential factors. Due to the circulation of two connected water bodies, the two gulfs were considered a whole study area for this research. The changes across time and space of the dependent variable Chl-a as well as all the selected study explanatory variables were complex, not following simple patterns of autocorrelation or cross-correlation with a dominant single explanatory variable overall time. Such finding indicates the need to use multivariate statistical regression models to derive strong correlations between the study's dependent variable and the explanatory variables. The results of this study have the potential to form the base for developing a comprehensive monitoring program of Chl-a, which in turn assists the decision-makers in developing proper mitigation action plans to control chlorophyll concentration. This study offers statistical function forms for algal concertation and the possible explanatory variables that show the significance that can further be used for forecasting and perdition models.

Over the short-term, the information from this research would be useful to support the relevant community of researchers relevant to the subject and the public government sector to prepare strategic plans for such control. Such plans aim at protection and reduction of the impacts on local fisheries and coastal services organizations near the coast of the Arabian Gulf and the Gulf of Oman study identified the best-fit statistical model that represents the correlation between remotely sensed Chl-a as a dependent variable and the explanatory variables during the period 2008-2011. The Semi-Log Model provided the best fit to data among the three models tested. The high R^2 value of 0.84 and an F test value of 357.3 indicate that the overall model shows a high level of statistical confidence. The explanatory variables that were found to be significant at a 99% level of confidence are Calcite Concentration, Instantaneous Photo-synthetically Available Radiation, and Normalized Fluorescence Line Height. Meanwhile, the Sea Surface Temperature was found to be significant at a 95% level of confidence. These study results showed that both climatic and biological explanatory variables have highly significant influences on the Chl-a concentration in the study area. Furthermore, the study validation results showed that short- versus long-term temporal series produce different results in terms of signs, variables significances, and model fitness. Such regression model results can support awareness and preparedness for similar areas that may occur or at similar conditions in the future. The framework and methods of this study offer a case study for potential applications to coastal areas extending beyond the current study region. Furthermore, the model can support the expansion of the model for foresting and develop awareness and preparedness efforts in the study area. Lessons learned from this study should apply to similar environments around the world.

Authors contributions

The authors contributed equally to the research data analysis, modelling, and manuscript writing

Acknowledgment

Mr. Ahmed Taha at the United Arab Emirates University reviewed the manuscript and provided invaluable inputs.

Conflict of interest

The authors certify that there is no actual or potential conflict of interest in this article.

REFERENCES

AL Shehhi, M. R., I. Gherboudj, and H. Ghedira. 2014. An overview of historical harmful algal blooms outbreaks in the Arabian Seas. Mar. Pollut. Bull. 86: 314-324.

Ali, T. A., M. M. Mortula. and S. Atabay. 2013. Study of water quality in Dubai Creek using DubaiSat-1 multispectral imagery. In: Geo-informatics in Resource Management and Sustainable Ecosystem, Springer, Berlin, Heidelberg, pp. 200-210.

Arabi, B., M. S. Salama, M. R. Wernand, and W. Verhoef. 2018. Remote sensing of water constituent concentrations using time series of *in-situ* hyperspectral measurements in the Wadden Sea. Remote Sens. Environ. 216: 154-170.

Blondeau-Patissier, D., J. F. Gower, A. G. Dekker, S. R. Phinn. and V. E. Brando. 2014. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Prog. Oceanogr. 123: 123-144.

Buditama, G., A. Damayanti. and T. G. Pin. 2017. Identifying Distribution of Chlorophyll-a Concentration Using Landsat 8 OLI on Marine Waters Area of Cirebon, IOP Conference Series: Earth and Environmental Science, IOP Publishing, United Kingdom, p. 012040.

Campbell, J. W. 2003. Development of algorithms and strategies for monitoring chlorophyll and primary productivity. In: Coastal Ocean, Estuarine and Inland Water Ecosystems, Final Technical Report No. NAS5-96063. pp. 1-20. Available from: https://www.modis.gsfc.nasa.gov. [Last accessed on 2019 Apr 10].

Cheng, C., Y. Wei, G. Lv. and Z. Yuan. 2013. Remote estimation of chlorophyll-a concentration in turbid water using a spectral index: A case study in Taihu Lake, China. J. Appl. Remote Sens. 7: 073465.

Elawad, A. E. S. 2012. Study of Inter-Annual Variability of Chlorophyll in the Red Sea, Master's Thesis. The University of Bergen, Norway.

Feng, J., J. M. Durant, L. C. Stige, D. O. Hessen, D. Ø. Hjermann, L. Zhu, and N. C. Stenseth. 2015. Contrasting correlation patterns between environmental factors and chlorophyll levels in the global ocean. Glob. Biogeochem. Cycles. 29: 2095-2107.

Food and Agriculture Organization (FAO). 2018. Information on Fisheries Management in the United Arab Emirates. Food and Agriculture Organization, Rome. Available from: http://www.fao.org/fi/oldsite/fcp/en/are/body.htm. [Last accessed on 2019 Jan 20].

Fengyun, M. 2010. Progress in water quality monitoring based on remote sensing and GIS. In: 2010 International Conference on Challenges in Environmental Science and Computer Engineering, IEEE, New Jersey, United States, pp. 208-211.

Guo, Q., X. Wu, Q. Bing, Y. Pan, Z. Wang, Y. Fu, D. Wang. and J. Liu. 2016. Study on retrieval of chlorophyll-a concentration based on Landsat OLI imagery in the Haihe River, China. Sustainability. 8: 758.

Justice, C. O., E. Vermote, J. R. Townshend, R. Defries, D. P. Roy, D. K. Hall, V. V. Salomonson, J. L. Privette, G. Riggs, and A. Strahler. 1998. The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. IEEE Trans. Geosci. Remote Sens. 36: 1228-1249.

Kiefer, I., D. Odermatt, O. Anneville, A. Wüest. and D. Bouffard. 2015. Application of remote sensing for the optimization of *in-situ* sampling for monitoring of phytoplankton abundance in a large lake. Sci. Total Environ. 527: 493-506.

Komick, N. M., M. P. F. Costa and J. Gower. 2009. Bio-optical algorithm evaluation for MODIS for western Canada coastal waters: An exploratory approach using in situ reflectance. Remote Sens. Environ. 113: 794-804.

Lamon, E. C. 3rd. 1995. A regression model for the prediction of chlorophyll a in Lake Okeechobee, Florida. Lake Res. Manage. 11: 283-290.

Le, C., C. Hu, J. Cannizzaro, D. English, F. Muller-Karger. and Z. Lee. 2013. Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. Remote Sens. Environ. 129: 75-89.

Lu, L., V. Venus, A. Skidmore, T. Wang. and G. Luo. 2011. Estimating land-surface temperature under clouds using MSG/SEVIRI observations. Int. J. Appl. Earth Obs. Geoinfor. 13: 265-276.

Malek, S., S. M. S. Ahmad, S. K. K. Singh, P. Milow. and A. Salleh. 2011. Assessment of predictive models for chlorophyll-a concentration of a tropical lake. BMC Bioinformatics. 12: S12.

Matthews, A., A. Duncan. and R. Davison. 2001. An assessment of validation techniques for estimating chlorophyll-a concentration from airborne multispectral imagery. Int. J. Remote Sens. 22: 429-447.

Moses, W. J., A. A. Gitelson, S. Berdnikov. and V. Povazhnyy. 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data-successes and challenges. Environ. Res. Lett. 4: 045005.

National Oceanic and Atmospheric Administration. 2020. National Environmental Satellite, Data, and Information Service of the National Oceanic and Atmospheric Administration. Available from: https://www.ncdc.noaa.gov. [Last accessed on 2020 Mar 15].

National Environmental Satellite, Data, and Information Service of the National Oceanic and Atmospheric Administration. Available from: https://www.ncdc.noaa.gov/ [Last accessed on 2020 Mar 15].

Nazeer, M. and J. E. Nichol. 2015. Development and application of a remote sensing-based Chlorophyll-a concentration prediction model for complex coastal waters of Hong Kong. J. Hydrol. 532: 80-89.

Ocean Colour Web: Algorithm Description. Available from: https://www.oceancolor.gsfc.nasa.gov/atbd. [Last accessed on 2018 Feb].

Prigent, C., F. Aires. and W. Rossow. 2003. Retrieval of surface and atmospheric geophysical variables over snow-covered land from combined microwave and infrared satellite observations. J. Appl. Meteorol. 42: 368-380.

Ryan, K. and K. Ali. 2016. Application of a partial least-squares regression model to retrieve chlorophyll-a concentrations in coastal waters using hyper-spectral data. Ocean Sci. J. 51: 209-221.

Steissberg, T. E., S. J. Hook. and S. G. Schladow. 2005. Characterizing partial upwellings and surface circulation at Lake Tahoe, California-Nevada, USA with thermal infrared images. Remote Sens. Environ. 99: 2-15.

UAE-MOEW. 2015. United Arab Emirates Ministry of Climate Changes and Environment Formerly Ministry of Environment and Water (MOEW), Red Tide Information. Available from: http://www.moew.gov.ae/en/knowledge-and-statistics/red-tide.aspx. [Last accessed on 2017 Dec 10].

UNESCO. 2015. Harmful Bloom Programme. UNESCO, Paris. Available from: http://www.hab.ioc-unesco.org/index.php?option=com_content&view=featured&itemed=100001. [Last accessed on 2020 Feb 15].

Poll, W. H., G. Kulk, K. R. Timmermans, C. P. D. Brussaard, H. J. Van Der Woerd, M. J. Kehoe. and A. G. J. Buma. 2013. Phytoplankton chlorophyll a biomass, composition, and productivity along a temperature and stratification gradient in the northeast Atlantic Ocean. Biogeosciences. 10: 4227-4240.

Wong, K. T., J. H. Lee. and I. Hodgkiss. 2007. A simple model for forecast of coastal algal blooms. Estuar. Coast. Shelf Sci. 74: 175-196.

Zheng, G. and P. M. Digiacomo. 2017. Remote sensing of chlorophyll-a in coastal waters based on the light absorption coefficient of phytoplankton. Remote Sens. Environ. 201: 331-

APPENDIX A: SUPPLEMENTAL MATERIAL, ALGORITHMS USED TO GENERATE THE STUDY'S VARIABLES

Chlorophyll-a Concertation

The Ocean Colour Index (OCI) algorithm is a three-band reflectance difference executing the difference between Rrs (remote sensing reflectance) in the green band, and a reference formed linearly between Rrs in the blue and red bands. (Le et al. 2013)

$$\mathrm{OCI} = \mathrm{Rrs}(\lambda\mathrm{green}) - \left[\mathrm{Rrs}(\lambda\mathrm{blue}) + \frac{\lambda\mathrm{green} - \lambda\mathrm{blue}}{\lambda\mathrm{red} - \lambda\mathrm{blue}}\left\{\mathrm{Rrs}(\lambda\mathrm{red}) - \mathrm{Rrs}(\lambda\mathrm{blue})\right\}\right] \tag{S1a}$$

where λblue, λgreen, and λred are the instrument-specific wavelengths closest to 443, 555, and 670nm respectively. Thus, Equation (S1a) is applied as:

$$OCI = Rrs(555) - \left[Rrs(443) + \frac{555-443}{670-443}\left\{Rrs(670) - Rrs(443)\right\}\right] \tag{S1b}$$

or $$OCI = Rrs(555) - \left[Rrs(443) + 0.493\left\{Rrs(670) - Rrs(443)\right\}\right] \tag{S1c}$$

Particulate Organic Carbon (POC)

The algorithm behind the generation of particulate organic carbon (POC) in mg m^{-3} is calculated using an empirical relationship derived from *in-situ* measurements of POC and blue-to-green band ratios of remote sensing reflectance.

This algorithm depends on the availability of bands cantered at 443 in the blue region and between 547 and 565 nm in the green region.

$$POC = a\left[\frac{Rrs(443)}{Rrs(555)}\right] + b \tag{S2}$$

Where $a = 203.2$ and $b = 1.034$

Instantaneous Photosynthetically Available Radiation (IPAR)

The iPAR product represents the total PAR (Photosynthetically Available Radiation) incident on the ocean surface at the time of the satellite observation.

$$iPAR = \frac{1}{h\,c}\int_{400}^{700} \lambda E_d(\lambda, 0-)\,d\lambda \tag{S3}$$

Where h is Planck's constant, *c* is the speed of light, and $Ed(\lambda,0-)$ is spectral downwelling irradiance just below the sea surface. Ed (λ, 0-) is derived by attenuating extra-terrestrial solar irradiance at each sensor wavelength through the atmosphere using the derived atmospheric correction model. The iPAR product is reported just above the ocean surface and does not account for transmission losses through the interface.

Normalized Fluorescence Line Height (NFLH)

The normalized fluorescence line height (NFLH) is measured in mW cm^{-2} μm^{-1} sr^{-1}, calculated as the difference between the observed NLW(678) and a linearly interpolated NLW(678) from two surrounding bands (Ocean colour Web: Algorithm description).

$$NFLH = nLw(678) - \left(\frac{70}{81}\right) x\,nLw(667) - \left(\frac{11}{81}\right) x\,nLw(748) \quad (S4)$$

Sea Surface Temperature

Sea surface temperature is the temperature of the ocean surface water. It is calculated in units of °C using the 11μm and 12μm long-wave infrared bands (Ocean colour-NASA, https://oceancolor.gsfc.nasa.gov/atbd/).

$$SST_{sat} = aij_0 + aij_1 T_{11}\mu_m + aij_2(T_{11}\mu_m - T_{12}\mu_m)\, T_{sfc} + aij_3(sec(\theta - 1)\,(T_{11}\mu_m - T_1 2\mu m) + aij_4(mirror) + aij_5(\theta^*) + aij_6(\theta^2) \quad (S5)$$

Where
T11 = Brightness Temperature 11μm Channel
T12 = Brightness Temperature 12μm channel
Tsfc = Reference SST
θ =sensor zenith angle
θ* =sensor zenith angle is made negative for pixels in the first half of the scan line
Mirror = mirror side number
Coefficients a_{ij} = algorithm coefficient set for the month of year and latitude zone ij