



Article Evaluation of Ocean Color Algorithms to Retrieve Chlorophyll-*a* Concentration in the Mexican Pacific Ocean off the Baja California Peninsula, Mexico

Patricia Alvarado-Graef ¹, Beatriz Martín-Atienza ^{1,*}, Ramón Sosa-Ávalos ², Reginaldo Durazo ¹, and Rafael Hernández-Walls ¹

- ¹ Facultad de Ciencias Marinas, Universidad Autónoma de Baja California, Ensenada 22860, Mexico; alvaradograef@uabc.edu.mx (P.A.-G.); rdurazo@uabc.edu.mx (R.D.); rwalls@uabc.edu.mx (R.H.-W.)
- ² Centro Universitario de Investigaciones Oceanológicas, Universidad de Colima, Manzanillo 28060, Mexico; rsosa@ucol.mx
- * Correspondence: atienza@uabc.edu.mx; Tel.: +52-646-1528213

Abstract: Mathematical algorithms relate satellite data of ocean color with the surface Chlorophyll-a concentration (Chl-a), a proxy of phytoplankton biomass. These mathematical tools work best when they are adapted to the unique bio-optical properties of a particular oceanic province. Ocean color algorithms should also consider that there are significant differences between datasets derived from different sensors. Common solutions are to provide different parameters for each sensor or use merged satellite data. In this paper, we use satellite data from the Copernicus merged product suite and in situ data from the southernmost part of the California Current System to test two widely used global algorithms, OCx and CI, and a regional algorithm, CalCOFI2. The OCx algorithm yielded the most favorable results. Consequently, we regionalized it and conducted further testing, leading to significant improvements, especially in eutrophic and oligotrophic waters. The database was then separated according to (a) dynamic boundaries in the area, (b) bio-optical properties, and (c) climatic conditions (El Niño/La Niña). Regional algorithms were obtained and tested for each partition. The Chl-a retrievals for each model were tested and compared. The best fit for the data was for the regional algorithms that considered the climatic conditions (El Niño/La Niña). These results will allow for the construction of consistent regionally adapted time series and, therefore, will demonstrate the importance of El Niño/La Niña events on the bio-optical properties of the area.

Keywords: Chlorophyll-*a*; remote sensing; ocean color algorithms; El Niño–Southern Oscillation (ENSO)

1. Introduction

Oceanic ecosystems, havens for vast biodiversity, essential for climate regulation, and providers of diverse benefits to humanity, are intricate systems. To comprehend their dynamics, we must monitor oceanographic variables consistently. Long time series with adequate spatial resolution are required in order to understand these complex ecosystems. While ship-based monitoring offers precise results, its infrequent coverage and high costs limit its efficacy. Moored or remotely operated instruments are an alternative but they are spatially restricted. Satellite-derived data, on the other hand, provide expansive temporal coverage, albeit primarily for surface variables. Despite their coarser spatial resolution, they encompass a vast majority of the ocean's surface.

Satellite sensors, with their diverse optical bands, sensitivities, and overpass times [1], produce data that can vary significantly between sensors [2,3]. Such variations arise from differences in sensor characteristics and inherent uncertainties in calibration and even with processing algorithms [4]. The parameters used in the algorithms are particular to



Citation: Alvarado-Graef, P.; Martín-Atienza, B.; Sosa-Ávalos, R.; Durazo, R.; Hernández-Walls, R. Evaluation of Ocean Color Algorithms to Retrieve Chlorophyll-*a* Concentration in the Mexican Pacific Ocean off the Baja California Peninsula, Mexico. *Remote Sens.* **2024**, *16*, 1748. https://doi.org/10.3390/ rs16101748

Academic Editors: Gabriel Senay and Chung-Ru Ho

Received: 22 March 2024 Revised: 28 April 2024 Accepted: 11 May 2024 Published: 15 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). each sensor [5]. Because of all of these facts, continuous climate data records cannot be constructed [6].

To enhance the satellite data's time coverage and, therefore, provide consistent time series, the merging of data from different sensors becomes imperative [2,7,8]. Operational satellite sensor constellations, such as Copernicus, provide ocean color data as elements of an integrated, sustained observing system [8] that has better global coverage and allows for the study of phenomena of a larger timescale than what an individual sensor provides [3]. The efficacy of such merged data must be regionally validated to ensure that the algorithms used are optimal.

Bio-optical algorithms that relate ocean surface Chlorophyll-*a* concentration (Chl-*a*) with satellite data are predominantly empirical, deriving their parameters from in situ data [9]. Level-3 Chl-*a* images available on the NASA Ocean Color website employ an OCx algorithm merged with the color index algorithm (CI) [10] with different parameters for different sensors [5]. However, these global algorithms might not be regionally optimal [6]. For instance, some algorithms underestimate the values of Chl-*a* in the California Current area [6,11], while others overestimate them in the Northern Bering Sea and Chukchi Sea [12,13]. Thus, evaluating global algorithms' performance at regional levels becomes paramount, especially when using merged satellite data. Such evaluations pave the way for robust, long-term data series, enhancing our understanding and management of ocean ecosystems.

This research aims to evaluate various bio-optical algorithms for the southern part of the California Current System (CCS), offshore from the Baja California peninsula, and, if need be, to provide a regionally optimal working algorithm. The CCS, a region known for its coastal upwelling, eddies, and fronts, has been under scrutiny since 1949 via the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program focusing on the California coast, and since 1997 via the Mexican Research Program of the California Current (IMECOCAL), which centers on the region off the coast of Baja California. This region is a unique transitional zone where colder subarctic waters mingle with warmer tropical and subtropical currents [14,15]. This confluence, along with factors like circulation and water mass mixing, dictates the region's biological and chemical processes [16,17]. In particular, the mixture and the upwelling of nutrients determine the characteristics of the ecosystem [14,16–18]. Furthermore, interannual variations such as El Niño–Southern Oscillation (ENSO) strongly influence this intricate transitional ecosystem.

Our study assesses the efficacy of two global algorithms (OCx, CI) and a CCS-specific algorithm [9] using in situ Chlorophyll-*a* measurements off the Baja California Peninsula, as well as Copernicus merged satellite data. The significance of this evaluation stems from the fact that the performance of Chl-*a* algorithms for Copernicus data in this region remains untested. Copernicus, with its merged sensor data, offers enhanced spatial and temporal coverage, enabling a deeper dive into regional processes; for example, the longer-term trends modulating atmosphere–ocean interactions in Eastern Boundary Upwelling Systems triggered by decadal-scale fluctuations linked to climate forcing [19].

Post evaluation, we propose and test a regionalization tailored for this area and satellite dataset. To refine our understanding, we segmented the dataset based on hydrographic properties, bio-optical properties, and climatic events like ENSO. For each segment, specific algorithms were introduced and evaluated. Our findings facilitate the selection of the most representative model for this region using the Copernicus dataset, shedding light on the region's distinct bio-optical properties. A meticulously analyzed Chl-*a* retrieval from a consistent dataset across instruments will bolster our understanding of biogeochemical processes and the broader implications of climate change.

2. Methodology

In this section, we describe the methodology used to estimate Chl-*a* from satellite data in the IMECOCAL region. We initiate with the application of a global algorithm to the remotely sensed reflectance, $R_{rs}(\lambda)$, in order to obtain the modeled Chlorophyll concentra-

tion, mChl-*a*. Then, mChl-*a* is compared with the in situ Chlorophyll concentration data (Chl-*a*) to ascertain the algorithm's suitability for the region. Should discrepancies arise, we propose a regional adaptation of the algorithm.

The methodology unfolds in three phases. First, we describe the data, both $R_{rs}(\lambda)$ and in situ Chl-*a*. Subsequently, we describe the algorithms that transform $R_{rs}(\lambda)$ mChl-*a*. Lastly, we make an assessment of the algorithm using a comparison between the mChl-*a* and the in situ Chl-*a* data.

2.1. Remotely Sensed Reflectance Data

The satellite-derived surface reflectance data, $R_{rs}(\lambda)$, was obtained from Copernicus (https://www.copernicus.eu/ accessed on 11 September 2018). The Copernicus program has the objective of offering merged data records from multiple sensors (SeaWiFS, MODIS, MERIS, VIIRS-SNPP&JPSS1, and OLCI-S3A&S3B) using the highest quality merging processes to date. Level 3 monthly averages of $R_{rs}(\lambda)$ at 412 nm, 443 nm, 490 nm, 555 nm, and 670 nm wavebands were retrieved for the area (Figure 1), which ranges from 23°N to 32°N in latitude and 120°W to 112°W in longitude. The composite images, with a resolution of 4 by 4 km, cover the period from January 1998 to May 2016, aligning with the collection dates of the in situ Chl-*a* data.



Figure 1. General grid of IMECOCAL stations where in situ samples were collected. The line number increases from 100 toward the south on line 137. Numbers above transects indicate station numbers, which increase with distance from the coast. The blue arrows indicate the main directions of the California Current System in the region.

2.2. In Situ Chlorophyll a Concentration Data

The in situ Chl-*a* concentration data belongs to the database collected from the IME-COCAL region (Figure 1) ranging from January 1998 to April 2016 (Table 1). To obtain the in situ Chl-*a*, water was collected from the network of stations in 5 L Niskin bottles coupled in a rosette at varying depths: 1 m, 10 m, 20 m, 50 m, 100 m, 150 m, and 200 m. Two liters of water at each depth were filtered through Whatman GF/F filters at a pressure below 150 mm Hg. The filters were then placed in Histoprep tissue capsules (Fisherbrand, Fisher Scientific, Pittsburg, PA, USA) and frozen in liquid nitrogen, pending analysis. Chlorophyll

extraction was performed using acetone at 90% and at 4 °C in darkness for 24 h as recommended by Ref. [20]. Chl-*a* concentration (mg m⁻³) was determined by the fluorometric method [21,22], with a Turner Designs A10 fluorimeter for the 1998 to 2005 samples and a Turner Designs trilogy fluorimeter for 2005 to 2016. Both fluorimeters were calibrated with pure Chl-*a* (Sigma-Aldrich, Darmstadt, Germany). A more comprehensive description of the methodology can be found in Ref. [23]. For the purpose of this study, the mean in situ Chl-*a* concentration between the surface and 10 m depth data was used.



Table 1. Dates of IMECOCAL campaigns where in situ Chl-a data were obtained are marked in blue.

2.3. Description of the Algorithms

Bio-optical algorithms are designed to estimate the near-surface Chlorophyll-*a* concentration from satellite data by using a functional relationship between remotely sensed reflectance data, $R_{rs}(\lambda)$, and modeled Chlorophyll-*a* concentration, mChl-*a*. These algorithms, developed since the 1970s, predominantly use empirical equations [9]. Among the plethora of available algorithms, the NASA Ocean Color website predominantly employs a combination of the O'Reilly band ratio (OCx) and the Hu Color index algorithm (CI) [10]. In addition to these global algorithms, there are specialized algorithms tailored for the California Current System, such as CalCOFI 2 [24]. This study will evaluate the efficacy of all the aforementioned algorithms within the IMECOCAL region.

2.3.1. OCx Algorithm

The OCx algorithm is one of the primary methodologies we employed to derive mChl-a from R_{rs} data. This algorithm is characterized by a polynomial relationship expressed as:

$$\log_{10} \left(\text{mChl} - a \right) = a_0 + \sum_{i=1}^4 a_i \mathbf{F}^i, \tag{1}$$

where the a_i coefficients of this function are sensor-specific and based on global data [5,25], and F is the logarithm of the blue/green ratio of remotely sensed reflectance (\Re), as seen below:

$$\mathbf{F} = \log_{10}[\Re] = \log_{10} \left[\frac{max(Rrs(443, 490))}{Rrs(555)} \right],$$
(2)

The OCx algorithm was assessed using the parameters associated with the different sensors (SeaWiFs, MERIS, MODIS, and OCTS [25]). The optimal fit was for the SeaWiFs parameters $a_0 = 0.3272$, $a_1 = -2.9940$, $a_2 = 2.7218$, $a_3 = -1.2259$, $a_4 = -0.5683$. Consequently, these parameters were selected for this study.

2.3.2. CI algorithm

Another globally recognized algorithm explored in this study is the CI algorithm [10]. The CI algorithm employs a three-band model, utilizing reflectance values from the blue, green, and red bands (443 nm, 555 nm, and 670 nm, respectively).

$$CI = Rrs(555) - \left[Rrs(443) + \frac{555 - 443}{670 - 443} (Rrs(670) - Rrs(443)) \right],$$
 (3)

This index is subsequently related to the mChl-*a* concentration.

$$\log_{10} \left(\text{mChl} - a \right) = -0.409 + 191.6590 \text{ CI}$$
(4)

The CI algorithm is employed for retrievals below 0.15 mg m⁻³, while the OCx algorithm is used for values exceeding 0.2 mg m⁻³. For intermediate values, a weighted combination of both algorithms is utilized [25].

2.3.3. CalCOFI 2 Algorithm

Given that our research data originate from the southern part of the CCS, we also considered algorithms specifically tailored for the California Current region. The CalCOFI 2 band linear algorithm [24] determines mChl-*a* using the logarithm of the blue/green ratio of reflectance, F (as defined in Equation (2)), in the following linear relationship:

$$\log_{10} (\text{mChl} - a) = 0.444 \text{F} - 2.431 \tag{5}$$

2.4. Model Validation

To ensure the accuracy of our mChl-*a* model, we compared it against in situ Chl-*a* data. If discrepancies were observed, regional adjustments were proposed. For algorithm validation, we employed several metrics: the coefficient of determination R^2 , the adjusted coefficient of determination R^2_a , the sum of squares error (SSE), and the root mean squared logarithmic error (RMSLE). Additionally, the Akaike information criterion (AIC) was computed to compare the modeled results.

$$AIC = -2ln(L) + 2k \tag{6}$$

where L is the log-likelihood and k is the number of parameters in the model [26]. The AIC, commonly used for model selection, compares the performance of a set of models by estimating their quality based on error, data size, and the number of parameters. The smaller the value of the AIC, the better the quality of the model.

3. Results

3.1. Performance of OCx, CI and CalCOFI 2 Algorithms

The performance of two global algorithms, OCx (Equation (12)) and CI (Equation (4)), alongside the CCS-specific CalCOFI 2 (Equation (5)), was assessed for the IMECOCAL region waters (Figure 2). All algorithms were applied to the reflectance data, yielding the logarithm of the mChl-*a* concentration. The results, plotted against F (Equation (2)), are shown in Figure 2. The CI algorithm exhibited greater dispersion in mChl-*a* estimates compared to OCx and CalCOFI 2. This is attributed to the CI algorithm's reliance on blue, green, and red bands (Equation (3)), whereas OCx and CalCOFI 2 utilize only the blue and green bands. This means that the CI algorithm has one more input variable than the others; therefore, there is greater dispersion on the resulting mChl-*a*.

As waters adopt a greener hue, F values decrease as mChl-*a* values increase. Upon applying the algorithms to the IMECOCAL data (Figure 2), the primary discrepancies between the OCx and CI algorithms correspond to values of mChl-*a* higher than or equal to 0.3 mg m⁻³ (log₁₀(mChl-*a*) ~ -0.5) and a ratio of blue/green reflectance of $\Re \leq 2.5$ (F ≤ 0.4). For these reflectance ratios, OCx yields mChl-*a* values surpassing those of CI. The



difference between CI and OCx retrievals for waters with large \Re values is small, yielding very similar results for the area, though with slightly higher retrievals for the CI algorithm.

Figure 2. Results from applying OCx (blue squares), CI (yellow circles), and CalCOFI2 (pink triangles) algorithms to remote sensing reflectance data ($F = \log_{10}[\Re]$, Equation (2)) off the coast of Baja California compared to the log10 of in situ Chl-*a* data (gray dots).

The CalCOFI 2 algorithm consistently underestimated the Chlorophyll concentration across the data spectrum. This underestimation increases with the value of the blue/green reflectance ratio \Re . Therefore, for the IMECOCAL area, OCx is the algorithm that yields the best results from a qualitative point of view. A comprehensive quantitative analysis will be presented subsequently. Based on both the qualitative and quantitative scrutiny of the algorithm retrievals, Ocx is used as the basic algorithm for further comparisons, discarding the *CI* and CalCOFI 2 algorithms.

3.2. Performance of the Regionalized Algorithm versus Ocx

While the Ocx algorithm aligns well with the region (Figure 2), refining it to better match the area's unique characteristics can enhance its accuracy. A polynomial fit was thus applied to model the functional relationship between F and the in situ Chl-*a* IMECOCAL data. Polynomial functions were tried because any function can be reduced to a polynomial through Taylor series. After evaluating polynomials from the first to the fifth order, the following fourth-order function was selected based on quantitative metrics (R^2 , R_a^2 , SSE, RMSLE, and AIC).

$$\log_{10} (\text{mChl} - a) = 0.1746 - 1.9952\text{F} + 1.9992\text{F}^2 - 4.1958\text{F}^3 + 3.3837\text{F}^4$$
(7)

Figure 3 shows the contrast between the in situ Chl-*a* data and the mChl-*a* derived from the OCx (global) and the regional algorithm (Equation (7)). The regional algorithm mChl-*a* values closely mirror those of the global algorithm within the $1.5 < \Re < 4.5$ (0.18 < F < 0.65) range of blue/green reflectance ratios. For values of $\Re > 4.5$ (F > 0.65), the regional algorithm estimates are approximately 1.8 mg m⁻³ higher than the global algorithm, which is even larger than the CI retrievals (Figure 2). Conversely, for ratios $\Re < 1.5$ (F < 0.18), the regional algorithm yields lower mChl-*a*. Quantitatively, the regional algorithm consistently outperforms OCx, as evidenced by the SSE values for different F ranges. For the F < 0.18 area, the OCx sum of square error, SSE, is 24.4881(mg m⁻³)², while



Figure 3. Regional (red circles, Equation (7)) and OCx (blue squares) algorithm retrievals applied to remotely sensed reflectance data ($F = \log_{10}[\Re]$, Equation (2)) off the coast of Baja California compared to the log10 of in situ Chl-*a* data (gray dots).

3.3. Adjusting the Algorithm to Regional Properties

3.3.1. Dynamic Properties

In the vicinity of 32°N, water from the California Current (CC) converges with an intrusion of water from the Central North Pacific [27]; thus, the CC turns toward the shore and part of it veers north to incorporate into the Southern California Eddy and the Southern California Countercurrent, while the remainder continues southward along the Baja California coast [28]. Therefore, this convergence forms a persistent feature called the Ensenada Front, an ecological transition zone [27] marked by a strong gradient in Chlorophyll-*a* concentration. In the Ensenada Front, the eutrophic waters of the CC end abruptly, possibly due to the subduction of this pigment and nutrient-rich waters to the southwest [28]. The Ensenada Front separates the CC region [28,29] into a northern mesotrophic area which has been studied by the CalCOFI project, and a southern oligotrophic area studied by the IMECOCAL project. As depicted in Figure 3, the CalCOFI 2 algorithm's performance underscores the need for distinct algorithms for these areas given the notably different bio-optical properties.

The southernmost part of the CCS, adjacent to the Baja California Peninsula, serves as a transitional zone where cold water from the subarctic meets warmer water coming from the tropics and subtropics [14,15]. The location of this transitional zone fluctuates seasonally [14], influencing the state of the oceanic ecosystem because of mixing circulation and water masses, and convergence modulates the biological and chemical processes [16,17]. Previous studies [14,17] have identified two distinct provinces separated at around 28°N at Punta Eugenia (Figure 1). North of Punta Eugenia, there are subarctic waters throughout most of the year, with year-round upwelling that is most intense during spring and early summer [15]. In the southern region, coastal upwelling occurs mainly in spring and summer, while the tropical and subtropical influences are limited to summer and fall. Also, there are two cyclonic gyres north and south of Punta Eugenia [14]. Ref. [15] suggested that the separation in provinces is due to the change of sign in the wind stress curl at that latitude. This interruption in circulation has consequences for the biological properties of the area. Just as the Ensenada Front separates regions with markedly different bio-optical properties, this other dynamic boundary could have the same effect. Therefore, the dataset was separated to determine if these differences would yield an improvement in the mChl-*a* retrievals. The southern province includes lines 123 to 137 and the northern province includes lines 97 to 120, that is, north of Punta Eugenia (Figure 1). Despite expectations of differences in the distribution or values of \Re or mChl-*a*, the data dispersion was similar

across both provinces, and the polynomials fitted were very similar to the regional algorithm proposed. Variance in mChl-*a* over the northern province was 1.2583 $(\text{mg m}^{-3})^2$, while in the southern province it was 1.1840 $(\text{mg m}^{-3})^2$, with no statistical difference between the two areas.

3.3.2. Bio-Optical Properties

As with the latitudinal dynamic boundaries described before, coastal proximity is anticipated to influence Chlorophyll-a concentration due to various factors, including the continental outflow of inorganic matter carried hundreds of miles from the coast by Santa Ana winds [30,31] and upper layer mixing and thermocline shoaling originated by coastal upwelling, among others. These physical processes bring about nutrient enrichment in the euphotic zone and the subsequent increase of Chl-a. In order to explore the influence of nearshore coastal processes and given the correlation between station number and proximity to the coast (Figure 1), the station number was plotted against the green/blue ratio, \Re^{-1} (Figure 4a). A discernible pattern emerged with green/blue ratios higher than or equal to 1 nearer to the coast. This is likely due to the fact that these stations are nutrient-enriched and contain more Chl-a due to the photon absorption by phytoplanktonic pigments [32]. When the green/blue ratio decreases, the mChl-a concentration also decreases. This offshore trend suggests that the data could be categorized based on bio-optical properties (green/blue reflectance ratio) in three categories: coastal waters with a green/blue ratio of reflectance of 1 or higher, transitional waters with a green/blue ratio between 0.5 and 1, and oceanic waters with a green/blue ratio of 0.5 and lower. Figure 4b shows the locations of the stations according to these three categories.



Figure 4. (a) Variation of green/blue reflectance ratio (\Re^{-1}) with the station number (Figure 1) that increases with the distance from the coast. The stations were classified as coastal waters with a green/blue ratio of reflectance of 1 or higher (red squares, 105 stations), transitional waters with a green/blue ratio between 0.5 and 1 (green crosses, 653 stations), and oceanic waters with a green/blue ratio of 0.5 and under (blue circles, 2607 stations). (b) Station locations.

Polynomial algorithms were then fitted for each category (Figure 5), with the order determined by quantitative analysis metrics (R^2 , R_a^2 , SSE, RMSLE, and AIC)

$$\log_{10} \left(\text{mChl} - a \right) = 0.2138 - 2.6481 \text{F}$$
(8)

$$\log_{10} (\text{mChl} - a) = 0.2501 - 1.7957\text{F} - 0.4325\text{F}^2$$
(9)

$$\log_{10} (\text{mChl} - a) = 0.2786 - 2.1925\text{F} + 0.8474\text{F}^2$$
(10)



Figure 5. Retrievals of \log_{10} of mChl-*a*, using Equation (8), for coastal waters (dotted line), Equation (9) for transitional waters (solid line), and Equation (10) for oceanic waters (dashed line), compared to the \log_{10} of in situ Chl-*a* data (red squares for coastal waters, green crosses for transitional waters, and blue circles for oceanic waters).

The results of these algorithms, depicted in Figure 5, align with the general data trend and closely resemble the regional algorithm (lower mChl-*a* prediction than OCx for coastal waters (small \Re , large \Re^{-1}) and higher mChl-*a* predictions than OCx for oceanic waters (small \Re^{-1}). These algorithms consider the bio-optical properties of each area; coastal stations have a greater abundance of microphytoplankton (diatoms, dinoflagellates, silicoflagellates) associated with a larger phytoplanktonic absorption coefficient, whereas, in oceanic waters, smaller phytoplanktonic cells are found (nano and picophytoplankton) with smaller absorption coefficients [33]. They also offer the advantage of simpler inversion and reduced overfitting risk. The regional algorithm, on the other hand, provides a unified function for the entire \Re range.

3.3.3. Climatic Events

In addition to the dynamic and bio-optical properties, it is crucial to consider the impact of large-scale climatic events on the region's properties. Notably, discrepancies have been observed during ENSO events in the CCS [29]. The shifting boundaries between subarctic waters and tropical/subtropical waters during cold and warm events have been identified [17]. Such shifts, generally associated with changes in circulation and biological and chemical characteristics, necessitate an analysis of in situ data variations during these events.

To classify the data, three distinct indexes were employed: the Southern Oscillation Index (SOI, https://www.ncdc.noaa.gov/ accessed on 12 December 2018), the Oceanic

Niño Index (ONI, https://origin.cpc.ncep.noaa.gov/ accessed on 12 December 2018), and the multivariate ENSO index (MEI, https://www.esrl.noaa.gov/ accessed on 12 December 2018). Data were categorized as El Niño or La Niña based on the classification provided by any of the three indexes for the respective month and year. If none of the indexes indicated a cold or warm event for that period, the data were labeled as normal. Subsequently, a polynomial function was fitted for each condition, as depicted in Figure 6.



Figure 6. Algorithm retrievals for data classified as (**a**) La Niña, (**b**) El Niño, (**c**) or normal compared to the \log_{10} of in situ Chl-*a* data.

In waters off the Baja California Peninsula, coastal upwelling determines the variability of the phytoplankton community [34]. Therefore, when there is a disruption in these patterns, i.e., during El Niño/La Niña events, biomass and community composition experience a shift. A warm El Niño event is usually associated with weaker alongshore northwesterly winds, deepening of the seasonal thermocline, and the near-shore poleward transport of warmer, nutrient-depleted water. Therefore, a decrease in diatom abundance and an increase in the small cell communities is expected. A cold La Niña event brings stronger winds and colder, nutrient-rich water, which tend to favor diatom growth [35,36].

The extent of eutrophic and mesotrophic areas of the California Current during ENSO events has been analyzed, determining that there are important differences between the waters off of Baja California to those off the coast of California [29]. Furthermore, a reduction of the eutrophic and mesotrophic areas off the Southern California Bight associated with a decrease in upwelling was noted [29], along with an increase in the extent of the mesotrophic areas off of Baja California extending up to 700 km offshore, thus unlikely being due to upwelling but possibly to cyanobacteria. During the El Niño event of 1997–1998, Ref. [37] noted that in the region south of Punta Eugenia, tropical and subtropical waters dominated the region, while north of Punta Eugenia, there was a coastal poleward flow that displaced the core of the California Current offshore. Due to the large variability of this transitional region, both at the seasonal and interannual scales, one would not expect a global algorithm to adequately comprise the processes within this ecosystem.

The analysis for interannual warm and cold events in Figure 6 revealed that most data have high \Re values and low Chlorophyll-*a* during El Niño. Therefore, during El Niño events, waters tend to be more oligotrophic, showing the expected shift towards smaller cells [36]. This is because warmer waters make both thermocline and nutricline deeper, diminishing nutrient availability in the euphotic zone [38]. Also, different behaviors can be seen with the simplest algorithm for the normal (Equation (11)) and La Niña (Equation (12)) data and more complex for El Niño events (Equation (13)). The algorithms developed for different climatic conditions, as detailed in Equations (11)–(13), were selected based on quantitative analysis metrics (R^2 , R^2_a , SSE, RMSLE, and AIC; see Table 2 below):

 $\log_{10} (\text{mChl} - a) = 0.2962 - 2.0437\text{F} + 0.4425\text{F}^2$ (11)

$$\log_{10} (\text{mChl} - a) = 0.2337 - 2.1695\text{F} + 1.0492\text{F}^2$$
(12)

$$\log_{10} (\text{mChl} - a) = 0.3141 - 2.4323\text{F} + 2.2698\text{F}^2 - 3.1653\text{F}^3 + 2.0039\text{F}^4$$
(13)

Table 2. The coefficient of determination (R^2), the adjusted coefficient of determination (R_a^2), the sum of square error (SSE), the root mean squared logarithmic error (RMSLE), and the Akaike information criterion (AIC) for each of the tested and proposed algorithms. The R^2 and R_a^2 marked with an asterisk were outside of the expected range of [0, 1].

Error	CalCOFI 2	CI	OCx	Regional	Bio-Optical Properties	Climatic Condition
R^2	0 *	0.2640	0.4054	0.4441	0.4473	0.4585
R_a^2	0 *	0.2637	0.4046	0.4434	0.4465	0.4579
SSE	156,141.81	450.60	364.03	340.33	338.39	331.52
RMSLE	7.1715	0.3853	0.3463	0.3348	0.3339	0.3304
AIC	25,949.82	8195.45	7553.75	7349.37	7332.03	7267.75

3.4. Quantitative Comparison of Algorithms

The performance of each algorithm was evaluated using various error metrics, as outlined in Equations (6)–(8), with the results presented in Table 2. The CalCOFI 2 algorithm exhibited the highest error, confirming its unsuitability for waters off of Baja California. In contrast, the OCx algorithm outperformed the CI algorithm, as evidenced by the SSE values. The algorithm considering climatic conditions demonstrated the most promising results, with the lowest values for SSE, RMSLE, and AIC. The difference between the coefficient of determination and its adjusted value depends on the degrees of freedom and is reflected in the third decimal number, indicating that the differences are not important and, therefore, can be considered the same. The RMSLE and the SSE are the lowest for the algorithm with the climatic condition. The regional algorithm, while being an improvement over the global algorithms tested, was further enhanced when the dataset was segmented based on bio-optical properties and climatic conditions.

4. Discussion

The applicability of global algorithms to the waters off of the Baja California Peninsula was assessed. Both CI and OCx algorithms exhibit similarities for waters with mChl-*a* values less than 0.3 mg m⁻³ (log₁₀(mChl-*a*) ~ -0.5) and a blue/green reflectance ratio of $\Re \ge 2.5$ (F ≥ 0.4), as depicted in Figure 2. CI is deemed superior for retrievals below 0.15 mg m⁻³, while OCx is more effective for higher mChl-*a* concentrations (Chl-*a* ≥ 0.2 mg m⁻³) [25]. The CI algorithm includes information on reflectance at 670 nm [10], and therefore, the retrieved Chl-a values show more dispersion than the OCx. Ref. [39] stated that regionalizing the algorithms diminishes the bias, but uncertainties are still important because of the natural variation of the optical properties of phytoplankton, the presence of other optically active components, and the atmospheric correction. These authors found that removing the 443 nm band improved the performance of the Ocx algorithm. However, given that the natural data dispersion surpasses the difference between CI and OCx for mChl-*a* < 0.3 mg m⁻³, OCx was deemed more suitable for the region, leading to its selection for regionalization.

The proposed regional algorithm has a slightly better fit for the area as compared to Ocx (Table 2). However, the metrics shown in Table 2 analyze the whole of the algorithm. As shown in Figure 3, the divergence between the algorithms is in the extreme conditions of oligotrophic (high R or F, low mChl-*a*) and eutrophic waters (low \Re or *F*, high mChl-*a*). For these conditions, the regional algorithm outperforms the OCx, yielding results with a lower specific SSE for the oligotrophic and eutrophic areas. This discrepancy suggests that the IMECOCAL region's oligotrophic waters possess a higher pigment concentration than the global database used to derive OCx parameters. As Ref. [28] posits, the presence of small cells (cyanobacteria), typical of tropical and subtropical waters in warm waters off of the Baja California Peninsula, could explain this variation. While other factors like particulate matter or continental debris might influence this difference, their sporadic occurrence

means their impact is limited to specific situations and data points. As per Ref. [5], OCx is effective in eutrophic areas due to its higher mChl-*a* output. However, the IMECOCAL region's data indicates lower mChl-*a* values for smaller \Re values, leading the regional algorithm to produce reduced mChl-*a* results. This suggests that the eutrophic zone in the IMECOCAL region has a lower pigment concentration than the global database.

The regional algorithm, tailored to the unique oceanographical and optical properties of the area, offers improved predictions for both eutrophic and oligotrophic zones. The segmentation of the database resulted in models with minimal complexity (Equations (10)–(12)). Notably, a linear model was the best fit for coastal waters (Figure 6). The most optimal overall fit was achieved when the database was segmented based on climatic conditions. The El Niño condition presents a more intricate data distribution compared to La Niña and normal conditions. During El Niño events, Chlorophyll-a was lower than during the transitional and La Niña conditions. When analyzing the interannual variations associated with ENSO, ref. [40] found an increase in the mesotrophic area in the IMECOCAL region during these warm events. This coincides with the fact that for the IMECOCAL area, the dispersion in R values is less during an El Niño event. Ref. [41] found that during the cold event of 2008, the productivity rates increased. This coincides with the high Chlorophyll-a observed in the IMECOCAL data during La Niña and normal conditions during El Niño. During La Niña events, the colder waters and intensified upwelling [17] enhance nutrient availability for phytoplankton, thereby elevating mChl-a levels, predominantly in coastal and transitional stations. Such events alter the region's physical, biological, and chemical attributes, promoting microphytoplankton growth. The consequential shift in the phytoplankton community, as reflected in the algorithms, underscores the climatic conditions' significance for the region.

5. Conclusions

The Chlorophyll-*a* satellite images are one of the most used products in oceanography [39]. However, the global algorithms applied are biased because their parameters are obtained with data from all around the globe and for a particular sensor. Therefore, it is important to assess their efficiency in order to determine if they are regionally appropriate before use [39]. Off the coast of California, the global algorithms underestimate the Chlorophyll-*a* concentration by a factor of 5 [2,6]. However, for the IMECOCAL area, these algorithms retrieved mChl-a of the same order of magnitude as the in situ Chl-a data. The CalCOFI 2 algorithm, on the contrary, drastically underestimates the in situ Chl-a concentration in the IMECOCAL area, suggesting that there are significantly different bio-optical properties between these sections of the CC. The fact that the Ensenada Front works as a dynamic boundary that greatly alters the bio-optical properties in different sections of the CCS suggests that the boundary at around 28° N [14] that reduces ecological connectivity would do the same. Nonetheless, the data are not significantly different, showing that this dynamic boundary does not have a perceptible effect on the bio-optical properties of seawater in the area. The differences observed in the algorithms could be due not only to the bio-optical differences in the data per se but also to the differences in the merged satellite data. Therefore, it is important to establish accurate parameters for the specific merged satellite data.

The regional algorithm differs from the global OCx mainly in oligotrophic waters, with higher Chlorophyll-*a* values that could be due to the presence of small cells characteristic of tropical and subtropical waters. The database was then classified according to the green/blue ratio of reflectance (\Re^{-1}), which allowed for a distinction between coastal, transitional, and oceanic waters, each with a polynomial function fitted to it. Though the differences between the regional algorithm and this set of algorithms are small, the classification in coastal, transitional, and oceanic waters allows for a better parameterization for different properties compared with the combined use of CI and OCx by NASA. The weighted approach used for CI and OCx is based on the Chl-*a* value, which requires retrieval for the determination of an adequate algorithm. The use of this set allows for

a distinction of properties based on a value of \Re , that is, the classification of the data is performed before the retrieval.

The evaluation of the algorithms using merged satellite data provides a deeper understanding of the area and a better option for constructing long time series that can aid in the comprehension of the processes in the area. With the Copernicus satellite data, the climatic conditions algorithm allows for the construction of a more consistent time series for the IMECOCAL region.

Author Contributions: Conceptualization, P.A.-G. and B.M.-A.; Data curation, R.S.-Á.; Formal analysis, P.A.-G. and B.M.-A.; Investigation, P.A.-G. and R.S.-Á.; Methodology, P.A.-G., B.M.-A., R.S.-Á. and R.H.-W.; Resources, R.D.; Software, P.A.-G.; Supervision, B.M.-A.; Writing—original draft, P.A.-G.; Writing—review and editing, B.M.-A., R.S.-Á., R.D. and R.H.-W. All authors have read and agreed to the published version of the manuscript.

Funding: Patricia Alvarado-Graef received a scholarship from the SEP-PRODEP project "Productividad primaria y procesos de mesoescala en la costa de Baja California", within the 2018 call for proposals "Fortalecimiento de Cuerpos Académicos".

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Maritorena, S.; Siegel, D.A. Consistent merging of satellite ocean color data sets using a bio-optical model. *Remote Sens. Environ.* 2005, 94, 429–440. [CrossRef]
- 2. Kahru, M.; Kudela, R.M.; Anderson, C.R.; Mitchell, B.G. Optimized merger of ocean chlorophyll algorithms of MODIS-Aqua and VIIRS. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2282–2285. [CrossRef]
- 3. O'Reilly, J.E.; Werdell, P.J. Chlorophyll algorithms for ocean color sensors–OC4, OC5 & OC6. *Remote Sens. Environ.* 2019, 229, 32–47. [CrossRef] [PubMed]
- 4. McClain, C.R. A decade of satellite ocean color observations. Annu. Rev. Mar. Sci. 2009, 1, 19–42. [CrossRef]
- O'Reilly, J.E.; Maritorena, S.; O'Brien, M.C.; Siegel, D.A.; Toole, D.; Menzies, D.; Smith, R.C.; Mueller, J.L.; Mitchell, B.G.; Kahru, M.; et al. SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. In NASA Tech. Memo. 2000-206892; Hooker, S.B., Firestone, E.R., Eds.; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 2000; Volume 11, p. 49.
- 6. Kahru, M.; Kudela, R.M.; Manzano-Sarabia, M.; Mitchell, B.G. Trends in the surface chlorophyll of the California Current: Merging data from multiple ocean color satellites. *Deep-Sea Res. Part II Top. Stud. Oceanogr.* **2012**, 77–80, 89–98. [CrossRef]
- Gregg, W.; Aiken, J.; Kwiatkowska, E.; Maritorena, S.; Melin, F.; Murakami, H.; Pinnock, S.; Pottier, C. IOCCG Report Number 6, Ocean Color Data Merging. In *Report of the International Ocean-Color Coordinating Group*; Gregg, W., Ed.; International Ocean-Colour Coordinating Group: Dartmouth, NS, Canada, 2007; p. 68.
- 8. Groom, S.; Sathyendranath, S.; Ban, Y.; Bernard, S.; Brewin, R.; Brotas, V.; Brockmann, C.; Chauhan, P.; Choi, J.K.; Chuprin, A.; et al. Satellite Ocean Colour: Current Status and Future Perspective. *Front. Mar. Sci.* **2019**, *6*, 1–30. [CrossRef]
- O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. Ocean 1998, 103, 24937–24953. [CrossRef]
- 10. Hu, C.; Lee, Z.; Franz, B. Chlorophyll *a* algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *J. Geophys. Res. Ocean* **2012**, *117*, C01011. [CrossRef]
- 11. Kahru, M.; Kudela, R.M.; Anderson, C.R.; Manzano-Sarabia, M.; Mitchell, B.G. Evaluation of satellite retrievals of ocean Chlorophyll-*a* in the California Current. *Remote Sens.* **2014**, *6*, 8524–8540. [CrossRef]
- 12. Lee, S.H.; Ryu, J.; Park, J.; Lee, D.; Kwon, J.; Zhao, J.; Son, S. Improved Chlorophyll-*a* Algorithm for the Satellite Ocean Color Data in the Northern Bering Sea and Southern Chukchi Sea. *Ocean Sci. J.* **2018**, *53*, 475–485. [CrossRef]
- 13. Mustapha, S.B.; Bélanger, S.; Larouche, P. Evaluation of ocean color algorithms in the southeastern Beaufort Sea, Canadian Arctic: New parameterization using SeaWiFS, MODIS, and MERIS spectral bands. *Can. J. Remote Sens.* **2012**, *38*, 535–556. [CrossRef]
- Durazo, R.; Ramírez Manguilar, A.M.; Miranda, L.E.; Soto Mardones, L.A. Climatología de variables oceanográficas. In *Dinámica del Ecosistema Pelágico Frente a Baja California 1997–2007*; Gaxiola Castro, G., Durazo, R., Eds.; SEMARNAT: Ciudad de Mexico, Mexico; INE-SEMARNAT: Ciudad de Mexico, Mexico; CICESE: Ensenada, Baja California, Mexico; UABC: Ensenada, Baja California, Mexico, 2010; pp. 25–58 (In Spanish). Available online: https://imecocal.cicese.mx/wp/wp-content/uploads/ Durazo-Gaxiola-2010.pdf (accessed on 12 July 2023).
- 15. Durazo, R. Seasonality of the transitional region of the California Current System off Baja California. *J. Geophys. Res. Ocean* **2015**, 120, 1173–1196. [CrossRef]
- Gaxiola-Castro, G.; Durazo, R.; Lavaniegos, B.; De La Cruz-Orozco, M.E.; Millán-Núñez, E.; Soto-Mardones, L.; Cepeda-Morales, J. Pelagic ecosystem response to interanual variability off Baja California. *Cienc. Mar.* 2008, 34, 263–270. [CrossRef]

- 17. Durazo, R. Climate and upper ocean variability off Baja California, Mexico: 1997–2008. *Prog. Oceanogr.* 2009, *83*, 361–368. [CrossRef]
- Kurian, J.; Colas, F.; Capet, X.; McWilliams, J.C.; Chelton, D.B. Eddy properties in the California Current System. J. Geophys. Res. Ocean 2011, 116, C08027. [CrossRef]
- Boyce, D.G.; Lewis, M.R.; Worm, B. Global phytoplankton decline over the past century. *Nature* 2010, 466, 591–596. [CrossRef] [PubMed]
- 20. Venrick, E.L.; Hayward, T.L. Determining chlorophyll on the 1984 CalCOFI surveys. CalCOFI Rep. 1984, 25, 74.
- Holm-Hansen, O.; Lorenzen, C.J.; Holmes, R.W.; Strickland, J.D.H. Fluorometric determination of chlorophyll. *ICES J. Mar. Sci.* 1965, 30, 3–15. [CrossRef]
- 22. Yentsch, C.S.; Menzel, D.W. A method for the determination of phytoplankton, chlorophyll and phaeophytin by fluorescence. *Deep-Sea Res.* **1963**, *10*, 221–231. [CrossRef]
- De la Cruz-Orozco, M.E.; Gómez-Ocampo, E.; Miranda-Bojórquez, L.E.; Cepeda-Morales, J.; Durazo, R.; Lavaniegos, B.E.; Espinosa-Carreón, T.L.; Sosa-Avalos, R.; Aguirre-Hernández, E.; Gaxiola-Castro, G. Phytoplankton biomass and production off the Baja California Peninsula: 1997–2016. *Cienc. Mar.* 2017, 43, 217–228. [CrossRef]
- Mitchell, B.G.; Kahru, M. Algorithms for SeaWiFs standard products developed with CalCOFI bio-optical data. *CalCOFI Rep.* 1998, 39, 133–147.
- NASA. SeaWiFS Project Online; NASA: Seattle, WA, USA, 2010. Available online: https://oceancolor.gsfc.nasa.gov/SeaWiFS/ (accessed on 9 October 2019).
- 26. Burnham, K.P.; Anderson, D.R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed.; Springer: Greer, SC, USA, 2002; p. 488.
- 27. Venrick, E.L. Summer in the Ensenada Front: The distribution of phytoplankton species, July 1985 and September 1988. *J. Plankton Res.* 2000, 22, 813–841. [CrossRef]
- 28. Haury, L.R.; Venrick, E.L.; Fey, C.L.; McGowan, J.A.; Niiler, P.P. The Ensenada Front. CalCOFI Rep. 1993, 34, 69-88.
- 29. Kahru, M.; Mitchell, B.G. Influence of the 1997-98 El Niño on the surface chlorophyll in the California Current. *Geophys. Res. Lett.* **2000**, *27*, 2937–2940. [CrossRef]
- 30. Castro, R.; Parés-Sierra, A.; Marinone, S.G. Evolución y extensión de los vientos Santa Ana de febrero de 2002 en el océano, frente a California y la península de Baja California. *Cienc. Mar.* 2003, *29*, 275–281. [CrossRef]
- Sosa-Ávalos, R.; Gaxiola-Castro, G.; Durazo, R.; Mitchell, B.G. Effect of Santa Ana winds on bio-optical properties off Baja California. *Cienc. Mar.* 2005, 31, 339–348. [CrossRef]
- 32. Falkowski, P.G.; Raven, J.A. Aquatic Photosynthesis, 2nd ed.; Princeton University Press: Princeton, NJ, USA, 2007; p. 484.
- Alvarado-Graef, P.; Martín-Atienza, B.; Sosa-Avalos, R.; Durazo, R. An Ocean Color Algorithm Based on Power Functions to Retrieve Inherent Optical Properties from Remotely Sensed Data Off the Baja California Peninsula, Mexico. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 1868–1876. [CrossRef]
- 34. Gaxiola, G.; Cepeda–Morales, J.; Nájera–Martínez, S.; Espinosa–Carreón, T.L.; De la Cruz–Orozco, M.E.; Sosa–Avalos, R.; Aguirre– Hernández, E.; Cantú–Ontiveros, J.P. Biomasa y producción del fitoplancton. In *Dinámica del Ecosistema Pelágico Frente a Baja California 1997–2007*; Gaxiola-Castro, G., Durazo, R., Eds.; SEMARNAT: Ciudad de Mexico, Mexico; INE-SEMARNAT: Ciudad de Mexico, Mexico; CICESE: Ensenada, Baja California, Mexico; UABC: Ensenada, Baja California, Mexico, 2010; pp. 59–86.
- 35. Moran, X.A.G.; López–Urrutia, A.; Calvo–Díaz, A.; Li, W.K.W. Increasing importance of small phytoplankton in a warmer ocean. *Glob. Chang. Biol.* **2010**, *16*, 1137–1144. [CrossRef]
- González-Silvera, A.; Santamaría-del-Ángel, E.; Camacho-Ibar, V.; López-Calderón, J.; Santander-Cruz, J.; Mercado-Santana, A. The Effect of Cold and Warm Anomalies on Phytoplankton Pigment Composition in Waters off the Northern Baja California Peninsula (México): 2007–2016. J. Mar. Sci. Eng. 2020, 8, 533. [CrossRef]
- 37. Durazo, R.; Baumgartner, T. Evolution of oceanographic conditions off Baja California: 1997–1999. *Prog. Oceanogr.* 2002, 54, 7–31. [CrossRef]
- Mann, K.H.; Lazier, J.R. Dynamics of Marine Ecosystems. Biological-Physical Interactions in the Oceans, 3rd ed.; Blackwell Publishing: Hoboken, NJ, USA, 2006; p. 496.
- Clay, S.; Peña, A.; DeTracey, B.; Devred, E. Evaluation of Satellite-Based Algorithms to Retrieve Chlorophyll-*a* Concentration in the Canandian Atlantic and Pacific Oceans. *Remote Sens.* 2019, 11, 2609. [CrossRef]
- 40. Kahru, M.; Mitchell, B.G. Empirical chlorophyll algorithm and preliminary SeaWiFS validation for the California Current. *Int. J. Remote Sens.* **1999**, *20*, 3423–3429. [CrossRef]
- Gómez-Ocampo, E.; Gaxiola-Castro, G.; Durazo, R.; Beier, E. Effects of the 2013-2016 warm anomalies on the California Current phytoplankton. *Deep-Sea Res. Part II Top. Stud. Oceanogr.* 2018, 151, 64–76. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.