JGR Oceans

RESEARCH ARTICLE
10.1029/2019JC015131

# Phytoplankton Diversity in the Mediterranean Sea From Satellite Data Using Self-Organizing Maps

**Roy El Hourany[1,2], Marie Abboud-Abi Saab[3], Ghaleb Faour[2], Carlos Mejia[1], Michel Crépon[1], and Sylvie Thiria[1,4]**

[1]IPSL/LOCEAN, Sorbonne Université (Université Paris VI, CNRS, IRD, MNHN), Paris, France, [2]National Center for Remote Sensing, National Council for the Scientific Research, Beirut, Lebanon, [3]National Center for Marine Sciences, National Council for the Scientific Research, Batroun, Lebanon, [4]OVSQ, Versailles Saint-Quentin-en-Yvelines University, Versailles, France

**Key Points:**
- Self-organizing maps allow accurate differentiation of phytoplankton assemblages from secondary pigments obtained from satellite observations
- Identification of phytoplankton assemblages gives insights on the dynamics of the ocean food web and its role in climate regulation
- Variability of remote sensing derived phytoplankton groups was evidenced in the Mediterranean Sea

**Correspondence to:**
R. El Hourany,
roy.elhourany@locean-ipsl.upmc.fr

**Citation:**
El Hourany, R., Abboud-Abi Saab, M., Faour, G., Mejia, C., Crépon, M., & Thiria, S. (2019). Phytoplankton diversity in the Mediterranean Sea from satellite data using self-organizing maps. *Journal of Geophysical Research: Oceans*, *124*, 5827–5843. https://doi.org/10.1029/2019JC015131

Received 6 MAR 2019
Accepted 30 JUL 2019
Accepted article online 2 AUG 2019
Published online 15 AUG 2019

**Abstract** We present a new method to identify phytoplankton functional types (PFTs) in the Mediterranean Sea from ocean color data (GlobColour data in the present study) and AVHRR sea surface temperature. The principle of the method is constituted by two very fine clustering algorithms, one mapping the relationship between the satellite data and the pigments and the other between the pigments and the PFTs. The clustering algorithms are constituted of two efficient self-organizing maps, which are neural network classifiers. We were able to identify and estimate the percentage of six PFTs: haptophytes, chlorophytes, cryptophytes, *Synechococcus*, *Prochlorococcus*, and diatoms. We found that these PFTs present a peculiar variability due to the complex physical and biogeochemical characteristics of the Mediterranean Sea: Haptophytes and chlorophytes dominate during winter and mainly in the western Mediterranean basin, while *Synechococcus* and *Prochlorococcus* dominate during summer. The dominance of diatoms was mainly observed in spring in the Balearic Sea in response to deep water convection phenomena and near the coastline and estuaries due to important continental inputs. Cryptophytes present a weak concentration in the Aegean Sea in autumn. The validation tests performed on in situ matchups showed satisfying results and proved the ability of the method to reconstruct efficiently the spatiotemporal patterns of phytoplankton groups in the Mediterranean Sea. The method can easily be applied to other oceanic regions.

**Plain Language Summary** The identification and spatiotemporal distribution of phytoplankton assemblages give powerful insights on the dynamics of the marine food web and the ocean role in climate regulation in the context of the global change. A new method to identify phytoplankton functional types from satellite observations has been developed and applied in the Mediterranean Sea. It is based on artificial intelligence clustering, the so-called self-organizing maps. The method was able to differentiate multiple phytoplankton assemblages and to provide their different pigment compositions. This approach had been validated successfully with in situ data sets and the spatiotemporal variability of the phytoplankton functional types showed a significant coherence. The method is very general and can be applied to other regions.

## 1. Introduction

The Mediterranean Sea is a semienclosed basin, covering approximately 0.8% of the world's ocean surface area. Although it has a limited geographical dimension, it is considered as one of the most complex marine environments with comparable features present in the global ocean such as a thermohaline circulation, deep water convection events, and complex geomorphology and bathymetry (Bethoux et al., 1999).

Hydrological differences along the basin cause the presence of an increasing oligotrophic gradient from west to east. It results a west to east decrease of surface chlorophyll-a concentration (Turley et al., 2000) that was observed in situ and from space (Antoine & Morel, 1995), with the Eastern Mediterranean Levantine waters exhibiting highly oligotrophic conditions (Abdel-Moati, 1990; Dowidar, 1984; Yacobi et al., 1995). However, the seasonal evolution of Chlorophyll-a (Chla in the following) distribution still follows the typical succession of temperate regions, characterized by a phytoplankton abundance increase in late winter/early spring, a decrease during the summer season and a second smaller phytoplankton bloom in autumn (Sammartino et al., 2015; Siokou-Frangou et al., 2010).

Phytoplankton community in oligotrophic areas throughout the world ocean is mainly composed by picoplankton (Yogev et al., 2011). Nevertheless, the Mediterranean phytoplankton community reveals a considerable variability over both temporal and spatial scales and large dissimilarities in phytoplankton assemblage composition (Siokou-Frangou et al., 2010). Many studies have pointed the dominance of picoplankton in the Mediterranean Sea due to its oligotrophy, but the occurrence of regional phytoplankton blooms causes the coexistence of numerous microalgal groups (Siokou-Frangou et al., 2010).

An extensive amount of observations of the phytoplankton community along the Mediterranean coastline is available. On the contrary, longitudinal data based on large-scale investigations in open sea waters are scarce in the literature (Ignatiades et al., 2009). This lack of measurements can be partly overcome by using new tools, such as remote sensing techniques. Remote sensing of surface optical properties has provided synoptic views of the abundance and distribution of sea surface constituents. Satellite ocean color sensors have been an effective platform to estimate Chla in the surface waters, providing synoptic measurements over the world ocean (Antoine et al., 1996; Behrenfeld et al., 2005; Behrenfeld & Falkowski, 1997; Longhurst et al., 1995; Westberry et al., 2008).

The spatiotemporal distribution and identification of remote sensing-derived phytoplankton groups give powerful insights on the dynamics of the marine food web and the ocean's role in climate regulation in the context of the global change (Bracher et al., 2017; Kostadinov et al., 2017). It has been for several decades recognized that detection of phytoplankton from remote sensing images was a major challenge in ocean optics. Therefore, several remote sensing algorithms have been developed to characterize the global distribution patterns of phytoplankton functional types (PFTs) or size classes (PSC; Alvain et al., 2005, 2008; Aiken et al., 2009; Brewin et al., 2010; Ciotti & Bricaud, 2006; Hirata et al., 2008; Sathyendranath et al., 2014; Uitz et al., 2006). This was done by using relationship between Chla and PSC (Bricaud et al., 2006; Hirata et al., 2008; Mouw & Yoder, 2010; Uitz et al., 2006) and by taking into account the spectral information (Alvain et al., 2005; Ben Mustapha et al., 2013; Sathyendranath et al., 2014). Some of these algorithms are based on spectral features, such as backscattering (Kostadinov et al., 2009), absorption (Bracher et al., 2009; Ciotti & Bricaud, 2006; Mouw & Yoder, 2010; Roy et al., 2013), or a hybrid between absorption and backscattering (Fujiwara et al., 2011), while other algorithms exploit second-order anomalies of water leaving radiance such as PHYSAT (Alvain et al., 2005, 2008; Ben Mustapha et al., 2013; Navarro et al., 2014, 2017). The PHYSAT method is based on the identification of specific signatures in the normalized water leaving radiance spectra measured by an ocean color sensor after removal the impact of Chla variations. Applied to oceanic Case I water, this method is designed to detect satellite pixels in which the dominant groups are nanoeukaryotes (and separately Phaeocystis-like and coccolithophores), two types of picoplankton (*Prochlorococcus*- and *Synechococcus*-like cyanobacteria) and diatoms (Alvain et al., 2008). Furthermore, a regionalized version of the PHYSAT method has been specifically developed for the Mediterranean Sea due to the peculiarities of phytoplankton assemblages and succession than can be found in this basin and its particular optical properties (Navarro et al., 2014, 2017).

We have developed an innovative approach using self-organizing maps (SOM; Kohonen, 2013) in order to evidence the relationship between satellite data measured at the ocean surface and the phytoplankton assemblage. The SOMs are unsupervised neural classifiers commonly used when dealing with environmental studies (Cavazos & Cavazos, 1999; Hewitson & Crane, 2002; Liu et al., 2006; Liu & Weisberg, 2005; Niang et al., 2006; Reusch et al., 2007; Richardson et al., 2003). In El Hourany et al. (2019), a global approach was built to estimate phytoplankton pigment concentration data using SOM (SOM-Pigments) and optical satellite data. This approach was calibrated and efficiently validated using a global database of High-performance liquid chromatography (HPLC) measured pigments collocated with GlobColour satellite data (Chla, Rrs at four wavelengths; 412, 443, 490, and 555 nm) alongside the AVHRR sea surface temperature (SST). The output of this method provided 10 types of pigment that can be associated with phytoplankton groups (Vidussi et al., 2001).

In the present work, we propose a new SOM (the so-called SOM-PFT) to identify dominant PFTs from satellite data. This algorithm is regionalized and centered on the oligotrophic, yet complex, ecosystem of the Mediterranean Sea and allowed us to track specific features of phytoplankton assemblage along with their associated biooptical properties. While observing the availability of in situ HPLC measurements in the Mediterranean Sea, the heterogeneity of the sampling locations, which are significantly more abundant in

the western basin, limits the development of a regional algorithm. Therefore, our method was based on estimating the phytoplankton pigments composition of the Mediterranean Sea from satellite data by using the SOM-Pigments developed by El Hourany et al. (2019). Then the dominant PFTs were estimated by using a dedicated SOM (the SOM-PFT), which was calibrated on the output of the SOM-Pigments. The performances were evaluated by using in situ data sampled in the Mediterranean Sea.

The paper is articulated as follows. In section 2, we briefly describe the SOM-Pigments outputs, the satellite data used to estimate the pigment concentrations, and the in situ HPLC samples used to validate the results in the Mediterranean Sea. The SOM algorithm and the calibration procedure of SOM-PFT are described in section 3. In sections 4 and 5, we present the results and the validation of the method, while discussing the spatiotemporal variability of dominant PFTs in the Mediterranean Sea.

## 2. Materials

### 2.1. The SOM-Pigments

The SOM-Pigments (El Hourany et al., 2019) allows the estimation of phytoplankton pigment concentrations in oceanic waters from satellite ocean color data and SST. The global database used to calibrate and cross validate the SOM-Pigments gathers several oceanic campaigns and published databases of HPLC measurements, which include 10 phytoplankton pigment concentrations: chlorophyll-a (Chla), Divynil-chlorophyll-a (DVChla), chlorophyll-b (Chlb), Divynil-chlorophyll-b (DVChlb), 19'Hexfucoxanthin (19HF), 19'Butfucoxanthin (19BF), fucoxanthin (Fuco), peridinin (Perid), alloxanthin (Allo), and zeaxanthin (Zea). The SOM-Pigment was calibrated on a database comprising 12,000 in situ samples, which were collocated to remote sensing reflectance data at four wavelengths (412, 443, 490, and 555 nm) and Chla GlobColour data, along with SST AVHRR data described in the following. The results of the cross-validation procedure scored a coefficient of determination ($R^2$) of 0.75 and an average root-mean-square error of 0.016 mg/m$^3$.

The SOM-Pigment allowed us to estimate the 10 phytoplankton pigment concentrations cited above from satellite data on the global scale.

The output of SOM-Pigments regroups images (4 km, gridded) of 10 pigment concentrations. Using these pigment concentrations, it is possible to determine several PFT. For example, Zea, DVChla, and DVChlb are associated with cyanobacteria (Dandonneau et al., 2004; Guillard et al., 1985). Fuco is the principal marker of diatoms (Jeffrey, 1980). For nanoplankton quantification, Allo is a pigment typical of the cryptophytes (Gieskes & Kraay, 1983); 19HF, whose concentration is related to haptophytes and chromophytes nanoflagellates (Wright & Jeffrey, 1987); 19BF a typical marker of chromophytes nanoflagellates (Wright & Jeffrey, 1987) and Chlb for chlorophytes nanoflagellates and last, we also used the Perid, which appears in dinoflagellates (Jeffrey & Hallegraeff, 1987).

### 2.2. Satellite Data

In the following, we present the satellite data used to estimate the phytoplankton pigment concentrations in the Mediterranean Sea with the SOM-Pigments.

#### 2.2.1. GlobColour Data

To extend existing time series beyond that provided by a single satellite sensor, the European Space Agency initiated the GlobColour project (http://www.globcolour.info/) to develop a satellite based ocean color data set to support global carbon-cycle research. It aims at satisfying the requirements of scientists for a long (over a decade) time series of consistently calibrated global ocean color information with the best possible spatial coverage. This has been achieved by merging data from SeaWiFS, MODIS, VIIRS, MERIS, and OLCI satellite sensors.

The GlobColour project provides a continuous data set of merged Level 3 (Mapped, 4 km) daily remote sensing reflectance (Rrs). This product is generated for each instrument, using the corresponding Level 2 data. The merged Rrs is then computed as the weighted average of all the single-sensor products. Meanwhile, the Chla was estimated via the OC5 algorithm (Gohin, 2011).

#### 2.2.2. AVHRR SST Data

Characterized by a defined annual cycle, the use of SST to calibrate SOM-Pigments permitted to better fit the relationship between the in situ HPLC and satellite data and therefore facilitated the identification of thermophile phytoplankton groups.

For that, the SST data were generated at 4-km resolution and at a daily frequency using AVHRR instruments aboard National Oceanic and Atmospheric Administration polar-orbiting satellites. Retrieval algorithms for SST from AVHRR are mainly based upon the multichannel SST algorithm (McClain et al., 1985).

### 2.3. Med HPLC Pigment Data Set (Med-Pigments)

The in situ HPLC database (Med-Pigments) used to validate the method is a compilation of different data sets and campaigns performed in the Mediterranean Sea (Table 1 and Figure 1). A total of 1,772 HPLC samples was compiled in this database, limited to the first optical depth which is about 15–35 m on average (D'Ortenzio and Ribera d'Alcalà, 2009). The spatial availability of the data is heterogeneous; more than 70% of the data were measured at a specific station, the Boussole/DYPHAMED stations anchored in the Ligurian Sea, which are dedicated to the study of biogeochemical and physical processes. Thus, Med-Pigments can lead to a limitation to develop a representative algorithm for the whole Mediterranean. We have generated a new large database of pigment composition in the Mediterranean Sea by processing the satellite data over the Mediterranean Sea with the SOM-Pigments. The in situ Med-Pigments database was used to validate the SOM-Pigments estimated data. The methodology is thoroughly described in the following section.

## 3. The Proposed Method

### 3.1. SOMs, General Concept

The SOM algorithms (Kohonen, 2013) are nonlinear unsupervised classification methods. Such methods are able to cluster vectors of a multidimensional database into classes represented by a fixed network of neurons (the SOM map). The SOMs are defined by a web-like graph, usually a rectangular grid of dimension $p \times q$ where each node is a neuron. This graph structure is used to define a discrete distance (denoted by ?) between the neurons of the map, which represents the length of the shortest path between two neurons. Moreover, the SOM enables the partition of the initial database in such a way that each cluster is associated with a neuron of the map and is represented by a synthetic multidimensional vector (the referent vector $w$). Each vector $zi$ of the initial database will be assigned to the neuron whose referent $w$ is the closest, in the sense of the Euclidean norm, and will be called the projection of the vector $v$ on the map.

The cost function can be written as

$$J^{T}_{\mathrm{SOM}}(\chi, W) = \sum_{Zi\in \mathrm{SOM}} \sum_{C\in \mathrm{SOM}} K^{T}(\delta(c,\chi(zi)))\|zi-wc\|^{2} \quad (1)$$

where $c \epsilon$ SOM indices the neurons of the SOM map, ? is the allocation function that assigns each element $zi$ of the data to its referent vector $wc$, $(c, \chi(zi))$ is the discrete distance on the map SOM between a neuron $c$ and the neuron allocated to observation $zi$, and $K^{\mathrm{T}}$ a kernel function parameterized by the temperature $T$ that weights the discrete distance on the map and decreases during the minimization process. This cost function takes into account the proper inertia of the partition of the data set and ensures that its topology is preserved. In the same way, two close neurons on the SOM map represent close data of the initial database. The estimation of the referent vectors $w$ of a SOM and the topological order is made through the training phase.

The referent vectors are obtained by a weighted average of the initial data's vectors assigned to each neuron and their neighborhood on the map.

We introduce in the following a dedicated SOM map specialized in identifying dominant PFTs on the Mediterranean Sea (SOM-PFT).

### 3.2. SOM-PFT: Calibration and Validation Procedure

Step 1: Pigment database for retrieving the PFTs in the Mediterranean Sea

The PFTs can be retrieved from the secondary pigments. Since Med-Pigments data set is sparse in time and space, we built a high coverage secondary pigment database by processing the GlobColour and AVHRR SST data on the Mediterranean Sea region with the SOM-Pigments as did El Hourany et al. (2019). Level 3 mapped 4-km daily images of SST, Chla, and Rrs at four wavelength (412, 443, 490, and 555 nm) were used as input of SOM-Pigments. We processed 7,600 daily images of GlobColour and AVHRR from 1997 up to 2018 generating more than 3 million pixels sampling the Mediterranean. The outputs are thus constituted

**Table 1**
*Available In Situ HPLC Inventory in the Mediterranean Sea*

| Cruises | Location | Period | N | % | Source |
|---|---|---|---|---|---|
| Prosope | Western basin | 9/4/1999 to 4/10/1999 | 59 | 3 | [a] |
| SODYFT | Ligurian Sea | 02/25/2002 to 12/19/2005 | 160 | 9 | |
| SESAME | Mediterranean Sea | 02/16/2008 to 10/19/2008 | 261 | 15 | |
| BOUM_bot | Mediterranean Sea | 06/21/2008 to 07/18/2008 | 64 | 4 | [b] |
| Tara_oceans | Mediterranean Sea | 9/20/2009 to 10/27/2013 | 115 | 6 | |
| BOUSSOLE | Ligurian Sea | 07/22/2001 to 11/10/2016 | 1,113 | 63 | [c] |
| Total | | | 1,772 | | |

*Note*. Dates are formatted as MM/DD/YYYY.
[a]https://doi.org/10.5194/essd-5-109-2013 [b]https://seabass.gsfc.nasa.gov/ [c]http://www.obs-vlfr.fr/Boussole/html/boussole_data/collected.php

by nine-dimensional vectors, each vector representing nine secondary pigments (DVChla, Chlb, DVChlb, 19HF, 19BF, Fuco, Perid, Allo, and Zea) on a 4-km × 4-km pixel. The performances of the SOM-Pigments in the Mediterranean Sea were check with a cross-validation algorithm and on the Boussole data which were not learned in the SOM-Pigments calibration.

We then computed for each pixel the relative values of the nine pigment concentrations (Pr: $DVChla_r$, $Chlb_r$, $DVChlb_r$, $19HF_r$, $19BF_r$, $Fuco_r$, $Perid_r$, $Allo_r$ and $Zea_r$) using the following equation:

$$P_r = \frac{P}{\sum P} \tag{2}$$

where $P$ is the estimated phytoplankton pigment concentration and $\Sigma P$ is the sum taken over the nine P:

$$\Sigma P = \text{DVChla} + \text{Chlb} + \text{DVChlb} + 19\text{HF} + 19\text{BF} + \text{Fuco} + \text{Perid} + \text{Allo} + \text{Zea}$$

Each of the 3 million pixels is now represented by its nine $P_r$; we gathered all the pixels in a specific database denoted Pr-MED.

Step 2: Identification of PFTs in the Med

We then used Pr-MED to determine specific PFT classes in the Mediterranean Sea from the 7,600 images under study. For that, we trained a dedicated SOM (SOM-PFT), which is composed of 50 × 20 = 1,000 neurons, on Pr-MED. The 1,000 neurons represent a fine description of the initial database. Each satellite pixel of Pr-MED is thus assigned to a particular neuron, which represents a grouping of similar pigment assemblage. In order to reduce the number of clusters, we then applied a hierarchical clustering algorithm (HAC) using the Ward dissimilarity, which regroups the neurons presenting statistical similarities into classes. The problem is thus to choose a small number of classes that "best" characterize specific phytoplankton pigment assemblages associated with specific PFTs. We found that six classes give a good insight on the Mediterranean biological diversity. At this stage, these classes regroup specific pigment compositions. In order to associate each HAC class with a relevant PFT, a boxplot analysis was held to visualize the distribution of each pigment relative concentration within the HAC class and identify the dominant pigment. Following this identification, each HAC class characterized by a certain pigment composition will be assigned to a PFT class based on the literature (Dandonneau et al., 2004; Gieskes & Kraay, 1983; Guillard et al., 1985; Jeffrey, 1980; Jeffrey & Hallegraeff, 1987; Wright & Jeffrey, 1987).

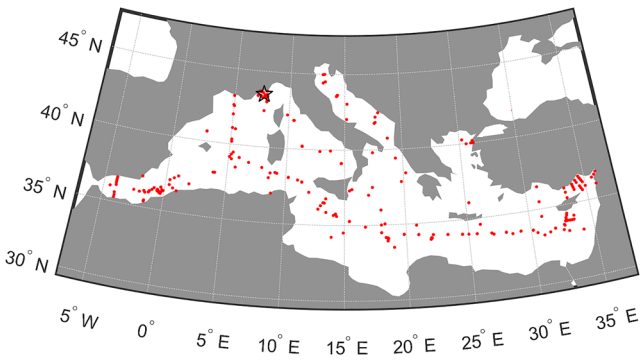


**Figure 1.** Localities of the HPLC samples in the Mediterranean Sea (Boussole station is represented with a star).

In the following, each pixel is thus assigned to a given class. Therefore, the 7,600 daily images are segmented using the six HAC classes.

Figure 2 summarizes the steps described above in a simplified flowchart.

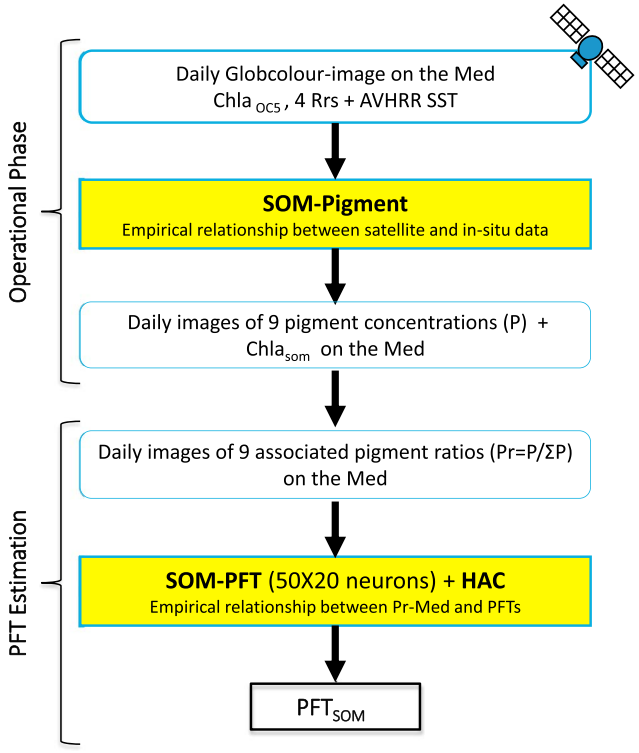


**Figure 2.** Flowchart summarizing the methodology to identify dominant PFTs from satellite data using SOM-Pigments and SOM-PFT on the Mediterranean Sea. PFT = phytoplankton functional type; SOM = self-organizing map; SST = sea surface temperature.

Step 3: Validation of the SOM-PFT using Med-Pigments

We chose to validate the SOM-PFT classification via two procedures:

1. We first used the PFT labeling proposed by Alvain et al. (2005) denoted Alvain criteria in the following. We applied the relative pigments ratios ($P_{rel}$) to identify the PFTs:

$$P_{rel} = \frac{P}{\text{Chla} + \text{DVChla}} \tag{3}$$

Alvain et al. (2005, Table 4) provided a table presenting threshold values on the different $P_{rel}$ whose combinations are related to specific phytoplankton groups (Hapto, Proc, SLC, diatoms, and dinoflagellates). Therefore, one of the tests consists in analyzing the coherence of the SOM-PFT output with the PFT given by the Alvain criteria. For that, we estimated the PFT associated with the Med-Pigments in situ measurements by using the Alvain criteria. A matchup was then defined by collocating each in situ measurement to the GlobColour and AVHRR data and by extracting the nearest pixel in a 3 × 3 pixel box surrounding the in situ measurement site. The secondary pigments of that pixel were estimated via SOM-Pigments and are presented as input to the SOM-PFT to identify the PFT. The two PFTs were then compared.

2. As a second test, we validated our PFT retrieval algorithm (SOM-Pigments → SOM-PFT + HAC) on the Boussole measurements. A 10 × 10 pixel box was extracted from GlobColour and AVHRR satellite data around the Boussole station (Figure 1) from July 2001 to November 2016. At each time, the PFT retrieval algorithm was applied on the pixel box and the ratio of each PFT class given by the SOM-PFT were calculated. The validation consists in analyzing the correlations between the time series of each PFT class ratio with the time series of the corresponding in situ $P_{rel}$. This experiment allowed us to observe and analyze the temporal coherence of this method compared to a long in situ time series.

## 4. Results

The cross validation of the SOM-Pigments showed a satisfactory result endorsing the validity of the SOM-Pigments database in the Mediterranean Sea. We found a $R^2 > 0.62$ (with a minimum $R^2 = 0.62$ for Allo concentrations and the highest $R^2 = 0.85$ for Fuco concentrations) and a low average root-mean-square error of 0.015 mg/m$^3$ (Table 2 and Figure 3).

This result allowed us to proceed with the reconstruction of the secondary pigment concentrations images on the Mediterranean as described in section 3.2, Step 1. We found a coherent spatiotemporal variability of each pigment. Figure 4 shows a monthly climatology of the pigment concentrations and the Chla OC5 (1997–2018). It can be seen that most of the diagnostic pigment variability follows that of Chla, since it is the major proxy of the total phytoplankton biomass, except DVChla, DVChlb, and Zea which present an opposed behavior.

The results of the validation of the Pr-Med data set with respect to Med-Pigments one and the spatiotemporal reconstruction of the different phytoplankton secondary pigments prove the coherence of Pr-Med data set used to train SOM-PFT. In the following, an analysis of SOM-PFT is presented along with the results of the validation experiments.

### 4.1. Analysis of the SOM-PFT

Figure 5 represents a two-dimensional map of the neurons of the SOM-PFT after the training phase. It is seen that the neurons are coherently organized with respect to the intensity of the different *Pr*. The

**Table 2**
*Statistical Results of the Cross-Validation Procedure Performed by the SOM-Pigments Using the Med HPLC Data Set*

| Pigment | $R^2$ | RMSE (mg/m$^3$) | *N* Obs |
|---|---|---|---|
| $Chla_{SOM}$ | 0.81 | 0.21 | 1,113 |
| DVChla | 0.72 | 0.007 | 663 |
| Chlb | 0.78 | 0.015 | 858 |
| DVChlb | 0.74 | 0.0005 | 79 |
| 19HF | 0.66 | 0.023 | 1,030 |
| 19BF | 0.80 | 0.006 | 1,096 |
| Fuco | 0.85 | 0.044 | 1,133 |
| Perid | 0.80 | 0.006 | 890 |
| Allo | 0.62 | 0.014 | 579 |
| Zea | 0.82 | 0.008 | 1,241 |

*Note*. RMSE = root-mean-square error; SOM = self-organizing map.

different *Pr* distribution are well distinctive on the SOM grid and poorly correlated except 19HF and 19BF images which somewhat overlaps. High $DVChla_r$ concentration ($Pr > 40\%$) occupied the upper left corner of the neuron's map, whereas high $Zea_r$ concentration ($Pr > 50\%$) were clustered at the left corner extending to the top. This means that the left corner of the SOM-PFT is specialized in Pico class detection. Whereas the $19HF_r$ and $19BF_r$ pigment ratios occupy the bottomside of the SOM-PFT with ratios >40% and >15%, respectively. High $Chlb_r$ concentration ($Pr >$ 50%) are found near the bottom right corner, and most of the $Allo_r$ information is found at the top right corner. The bottom right corner of SOM-PFT contains information on $19HF_r$, $19BF_r$, $Allo_r$, and $Chlb_r$; therefore, this SOM-PFT part is specialized in the Nano sized class detection. Whereas for the $Fuco_r$, high ratios >50% are found near the top right corner. The $Perid_r$ seems to overlap the $19HF_r$.

The six classes defined by the HAC are represented on the SOM-PFT grid in Figure 4 last panel. Figure 6 shows the boxplots of the pigment ratio composition regrouped in each class. The segmentation of the SOM-PFT into classes further highlights the specialization of each neuron to identify a specific pigment ratio dominance indicating a PFT: C1 is characterized by high $19HF_r$ and indicates dominant haptophytes nanophytoplanton (Hapto). High $19HF_r$ and $Chlb_r$ characterize C2, and therefore, this class can be attributed to chlorophytes nanophytoplanton (Chloro). C3 is characterized with high $Allo_r$ indicating a class of cryptophytes nanophytoplankton (Crypto). Meanwhile, C4 regrouped neurons with high $DVChla_r$ and $Zea_r$ indicating the dominance of *Prochlorococcus* picophytoplankton (Proc), whereas C5 is characterized with high $Zea_r$ and can be attributed to dominant *Synechococcus* picophytoplankton (Syne). Finally, C6 is highlighted by high $Fuco_r$ and denotes a dominance of diatoms microphytoplankton (Diat).

### 4.2. Validation Experiments

#### 4.2.1. Experiment 1: Validating Med-Pigments

As mentioned above, the Alvain et al. (2005) criteria (section 3.2) considers five PFTs (Hapto, Proc, Syne, Diat, and Dino), while the SOM-PFT allows to consider six PFTs whose only four are in common (Hapto, Proc, Syne, and Diat). The labeling thresholds of Alvain et al. (2005) take into account six pigment ratios: Pheopigments, Fuco, 19HF, Perid, Zea, and DVChla. To proceed with the labeling, we selected from Med-Pigments 764 in situ measurements with no missing values.

In Table 3, the confusion matrix is presented. It is obtained by comparing the results of the SOM-PFT with the Alvain et al. (2005) criteria for the Med-Pigments data set. Accordingly, 380 matchups were labeled as Hapto, 3 as Proc, 253 as Syne, 37 as diatoms, 35 as dinoflagellates, and last 56 samples were not assigned to any category. The SOM-PFT presents a performance of 70% (265 in situ vs. 380 satellite matchups) of the Hapto class, of 75.5% (191 in situ vs. 253 satellite matchups) for the Syne and finally of 94.6% (35 in situ vs. 37 matchups) for the diatoms.

Misplacements were found mostly when identifying Hapto and Syne dominated samples, where 23% (88 out of 380 matchup) of Hapto were identified as Syne, and 11.9% (30 out of 253 matchups) of Syne were assigned as Hapto by the SOM-PFT. Meanwhile, the Dino labeled samples and the samples that did not comply with any of the proposed thresholds in Alvain et al. (2005) were also presented in Table 3. We found that 54% of the Dino samples (19 out of 35) were assigned by SOM-PFT to Hapto and 28.6% (10 out of 35) to diatoms. And at last, 30.4% (17 out of 56) of the samples unidentified by Alvain et al. (2005) were assigned to Hapto by the SOM-PFT, while 44.6% (25 out of 56) were classified as Crypto or Chloro by the SOM-PFT.

#### 4.2.2. Experiment 2: Validation by Using the Boussole Time Series

In order to validate the coherence of the PFT reconstruction by the SOM-PFT, we analyzed a $10 \times 10$ pixels box around the location of the Boussole site (Figure 1) from July 2001 to November 2016. The results are given in Figures 7a–7d that show the daily estimation of the satellite PFT percentage in this box. To simplify the analysis, we regrouped three PFT classes (Hapto, Chloro, and Crypto) in a single one that represents nanoplankton size fraction (Nano).

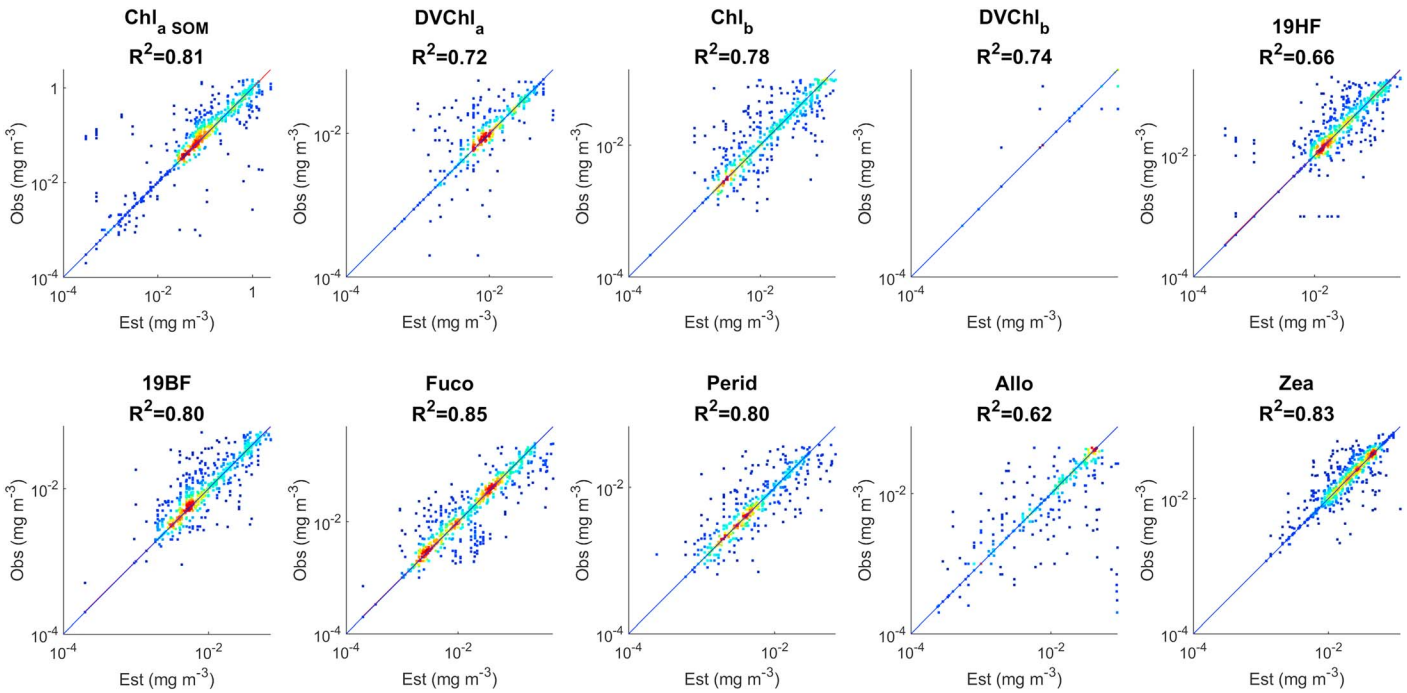


**Figure 3.** Scatter plots representing the cross-validation results of the SOM-Pigments using the Med HPLC data set. SOM = self-organizing map.

The Spearman correlation (Rs) between the satellite PFT and their corresponding in situ pigment ratios (Table 4) was estimated for diatoms, Nano, Proc, and Syne. We found a good agreement the two set of measurements. The Rs values (Table 5) are of 0.47, 0.53, 0.11, and 0.54 for diatoms, Nano, Proc, and Syne, respectively. A comparison between the in situ Chla measurement (whose concentration is maximal in winter) and satellite PFTs emphasizes the contribution level of each PFT class to Chla at the Boussole site. A relatively high Rs of 0.58 was noted for the diatoms, 0.46 for Nano, while for the Pico classes, Proc, and Syne showed a high inverse correlation of −0.65 and −0.79, respectively.

During the spring bloom period, the diatoms and the associated $Fuco_{rel}$ reach their maximum percentage (Figure 7b) when the mixed layer depth is maximum as mentioned by Marty et al. (2002). The Hapto and Chloro class distribution present maxima percentages during winter, usually around January, while Crypto are peaking in autumn (Figure 7c). The seasonal pattern of the three nanophytoplankton classes coincides with the variability of in situ $19HF_{rel}$, $Chlb_{rel}$ and $Allo_{rel}$. Meanwhile, *Prochlorococcus* show maximum values in autumn during several years, at the end of the stratification period (Figure 7d). The maxima found by the SOM-PFT are in close agreement with the maxima in the concentrations of the pigment ratio for DVChla measured in situ. These results agree with the pattern reported by Vaulot et al. (1990) and Marty et al. (2002) estimated by flow cytometry and HPLC analysis, respectively. Also coinciding with the stratification period during summer, the dominant group is *Synechococcus*, and a maximum of $Zea_{rel}$ is observed (Figure 7d).

### 4.3. Distribution of Dominant PFTs in the Mediterranean Sea

The Mediterranean Sea shows a well-marked spatiotemporal variability of the six PFT classes. Diatoms are mainly found in coastal zones, where a high occurrence of this class was found. Besides, a dominance of diatoms was also found during April in the Balearic Sea, in association with the deep convection phenomena. Simultaneously, nanophytoplankton classes (Hapto and Chloro) dominate in the Mediterranean Sea during winter and early spring season (December up to April). Meanwhile, the Pico size fraction (Syne and Proc) mostly dominate in the summer season. Figure 8 clearly indicates that the dominance of diatoms, Hapto, and Chloro in the western basin is more frequent than in the eastern basin, whereas Pico show an antagonist behavior with a quasi-dominance in the eastern Mediterranean.

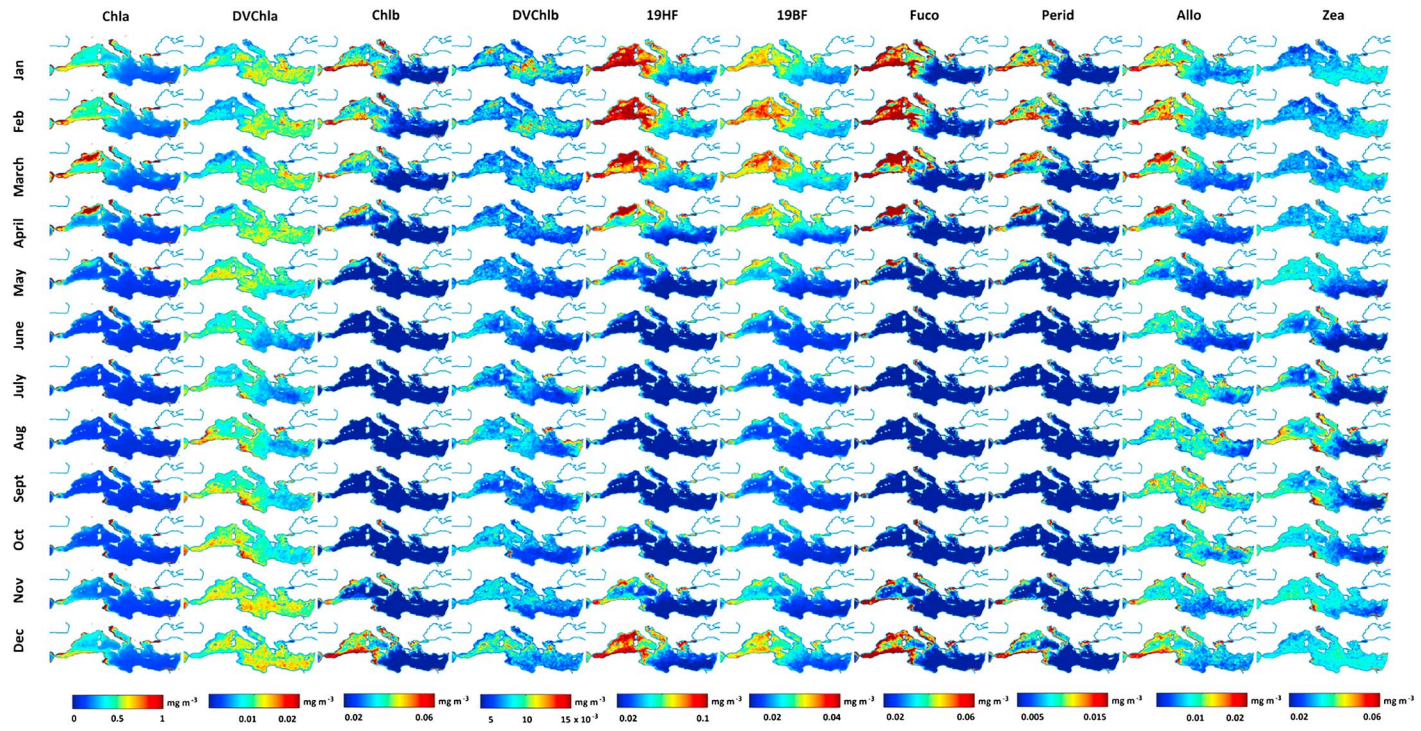


**Figure 4.** Monthly climatology images for each estimated pigment concentrations by self-organizing map-pigments along with the Chla OC5.

Let us now consider the monthly patterns of the PFTs in each subbasin (Figure 9). Diatoms present significant blooms in winter-spring in the Alboran and Balearic Seas. In the Adriatic Sea, diatoms show dominance in midsummer. Meanwhile in the open waters of the eastern Mediterranean, this class presents weak blooms.

Besides, the nanophytoplankton classes (Hapto and Chloro classes) seem to covary, showing a pronounced abundance during the November–April period. In the Balearic Sea, a sharp increase of diatoms is noted to the detriment of the nanophytoplankton classes in April, which mainly is a consequence of the convection phenomena. In the western basin, Hapto and Chloro codominate, while in the eastern basin, Chloro tend to dominate the Nano size fraction during the winter bloom. The Crypto class is absent in most of the basins, except in the Aegean Sea, where a bloom of Crypto is observed in August.

In parallel, the picophytoplankton represented by both *Synechococcus* and *Prochlorococcus* classes shows an antagonist monthly cycle. Unlike the other classes, Syne and Proc are the most abundant class during the summer season, mainly dominated by Syne. In the Adriatic and the Balearic Sea, the Syne dominated in July–September, while in the eastern part of the Mediterranean Sea, it quasi-dominated all year long. The Proc class is significantly present in the eastern Mediterranean, showing small antagonist variability in favor of the Syne.

## 5. Discussion

The present approach for identifying dominant phytoplankton groups from satellite data, which is based on the use the SOMs, is innovative and gave good performances in the validation tests. The reconstruction of the temporal variability of the PFTs at the Boussole station clearly shows that our approach efficiently reproduces the annual variability of the PFTs at least in the Ligurian Sea.

Comparing the PFTs given by the present method to these obtained by processing in situ data with the identification criteria proposed by Alvain et al. (2005), we found some differences in identifying Nano and Syne classes. These errors that are of 23% and 13%, respectively, can be associated with cumulative uncertainties during the process of the method; some errors can be due to the pigment estimation in the first step of the approach when using the SOM-Pigments (El Hourany et al., 2019), where 19HF or Zea concentrations

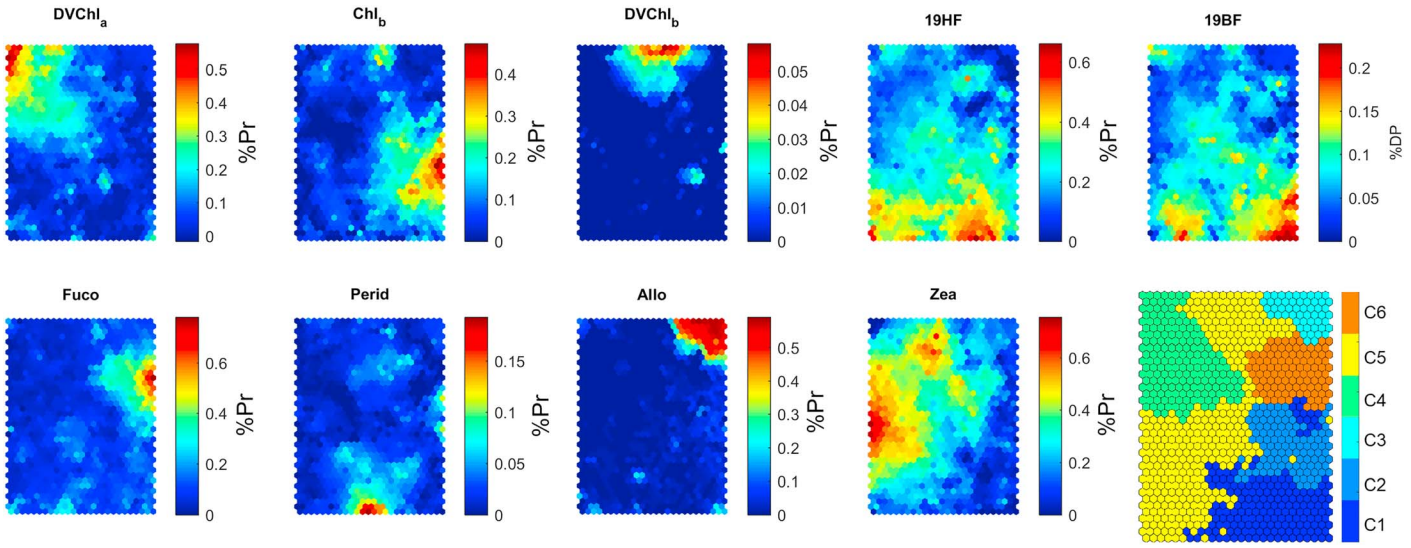


**Figure 5.** Discrete representation of each Pr on the self-organizing map-phytoplankton functional type and the hierarchical clustering in six classes (bottom right panel).

can be misestimated with high uncertainties. The use of SOM-Pigment highly depends on the quality of the satellite data (GlobColour and AVHRR) and therefore controls the pigment estimation errors. Other causes may interfere also, such as uncertainties in the thresholds of Alvain et al. (2005) and the classification criteria via SOM-PFT. Meanwhile, the SOM-PFT did not detect any Perid predominance, due to the low values of this pigments compared to that of 19HF or Zea. Therefore, we were unable to detect any dominance of dinoflagellates in the Mediterranean Sea. In parallel, the Crypto and Chloro classes were not validated using Alvain et al. (2005) thresholds. Yet an undeniable coincidence of 44.5% is noted while assigning these two classes by SOM-PFT to the in situ samples that were unidentifiable via the thresholds of Alvain et al. (2005).

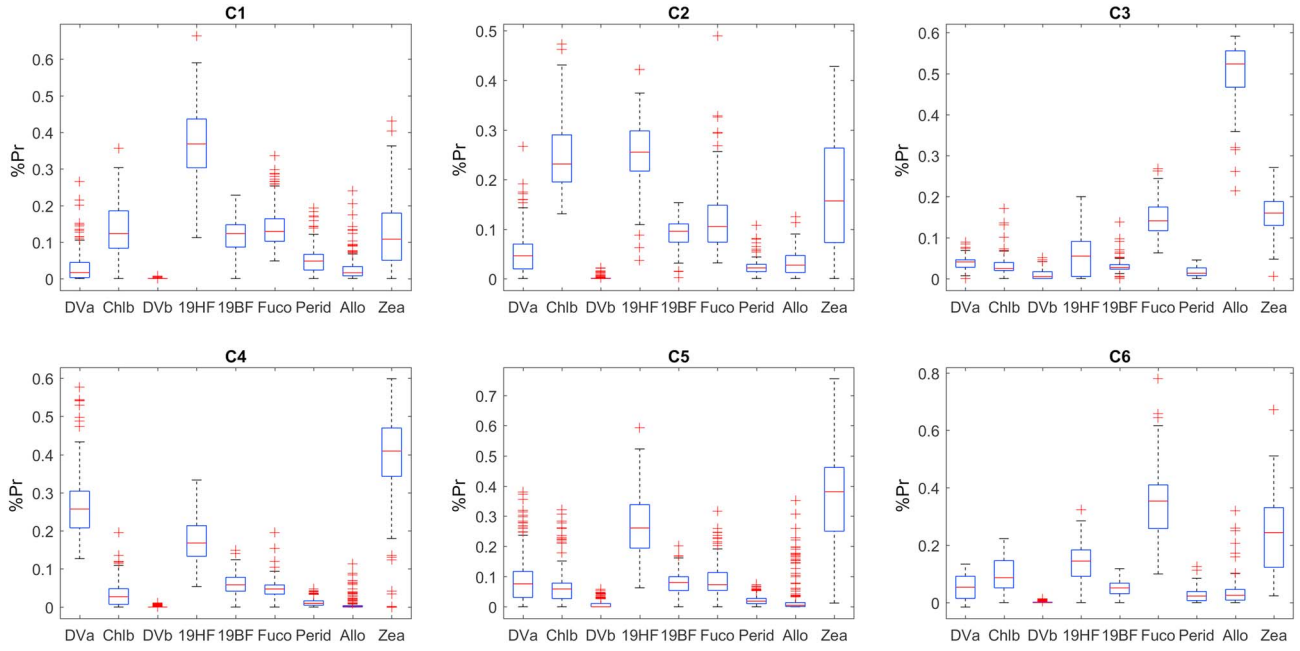


**Figure 6.** Boxplots of the nine pigment ratios for each class; C1: haptophytes (Hapto), C2: chlorophytes (Chloro), C3: cryptophytes (Crypto), C4: *Prochlorococcus* (Proc), C5: *Synechococcus* (Syne), and C6: diatoms (Diat).

**Table 3**
*Confusion Matrix Showing the Results of the Validation Test Performed by Comparing the SOM-PFT Classes With the In Situ HPLC PFT Computed With the Alvain Et Al. (2005) Criteria*

| | | Alvain et al. (2005) | | | | | |
|---|---|---|---|---|---|---|---|
| Method | | Hapto | Proc | Syne | Diat | Dino | Unidentified |
| SOM -PFT | Hapto | 70% (265) | — | 11.9% (30) | 5.4% (2) | 54.3% (19) | 30.4% (17) |
| | Proc | 1.3% (5) | 33.3% (1) | 7.5% (19) | — | 2.9% (1) | 3.6% (2) |
| | Syne | 23.2% (88) | 66.7% (2) | 75.5% (191) | — | 8.5% (3) | 14.3% (8) |
| | Diat | 4.5% (17) | — | 3.9% (10) | 94.6% (35) | 28.6% (10) | 7.1% (4) |
| | — | — | — | — | — | — | — |
| | Crypto/Chloro | 1.3% (5) | — | 1.2% (3) | — | 5.7% (2) | 44.6% (25) |
| N Obs | | 380 | 3 | 253 | 37 | 35 | 56 |

*Note*. SOM = self-organizing map; PFT = phytoplankton functional type.

Navarro et al. (2014, 2017) developed a regionalized version of PHYSAT (PHYSAT-Med), which is based on the OC-CCI merged ocean color data. Originally, PHYSAT identifies dominant phytoplankton types using normalized reflectance anomalies. PHYSAT-Med was validated using the Mediterranean in situ HPLC data set and the threshold classification method of Alvain et al. (2005). The results of this validation gave an agreement of 74% for Nano and 60% for Syne with the in situ matchups with no significant validation for diatoms detection (Navarro et al., 2014). Our results showed an agreement of 70% for dominant Nano identification, 76% for the Syne identification and an accurate recognition of diatoms (94%) by comparing our results with those given by using the in situ Med-pigment data set.

Moreover, the methodology proposed in this paper revealed not only dominant phytoplankton groups on the Mediterranean Sea but also quantitative information on major phytoplankton secondary pigments from space. Several studies were conducted to invert Chla signal into PSC or PFT abundance using secondary pigments concentration. This was based on various interpigment relationships to describe the phytoplankton variability at global scale (Hirata et al., 2011) and at the Mediterranean scale (Sammartino et al., 2015; Uitz et al., 2006, 2012, 2012). In El Hourany et al. (2019), SOM-Pigments has shown that the method reproduces accurately interpigment relationships at global scale reflecting nonlinear

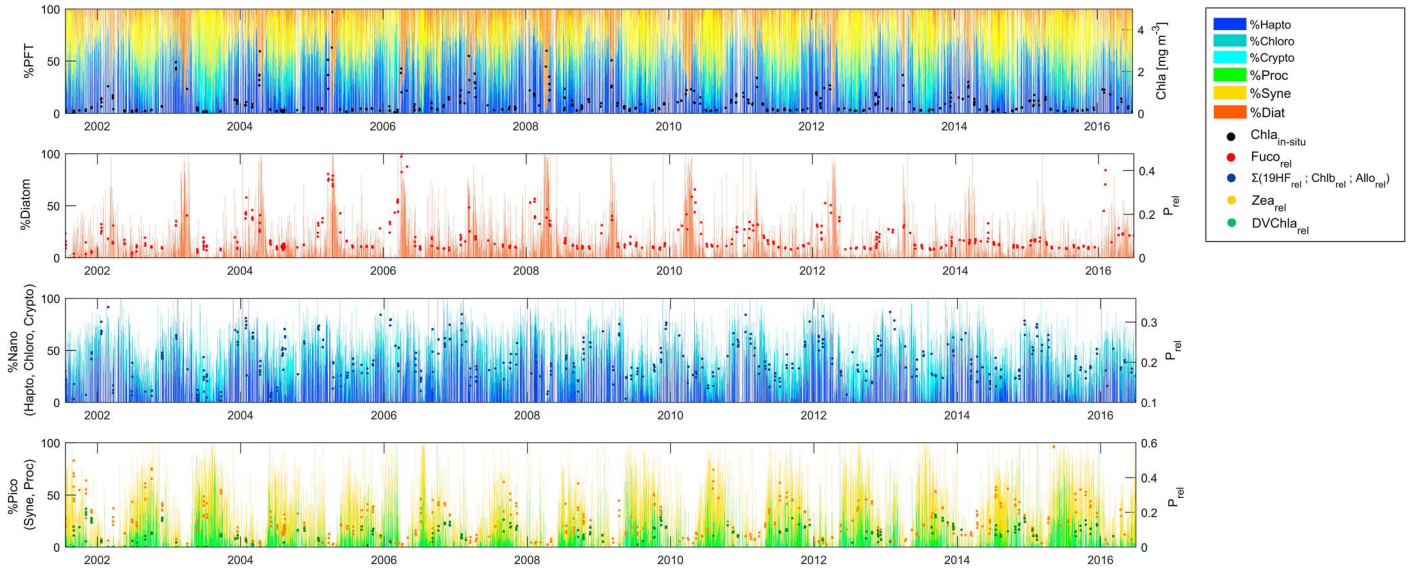


**Figure 7.** Daily percentage of dominant phytoplankton functional type computed from satellite data around the Boussole Station, Ligurian Sea, from 2001 to 2016. (a) Representation of the six phytoplankton functional type classes along with the in situ Chla (black dots), (b) diatoms (red) frequency overlapped with in situ Fuco rel (red dots), (c) Hapto (navy blue), Chloro (blue), and Crypto (cyan) frequency along with the corresponding in situ P rel (19HF rel, Chlb rel, and Allo rel; blue dots), (d) *Synechococcus* (yellow) and *Prochlorococcus* (green), overlapped with in situ Zea rel (yellow dots) and DVChla rel (green dots).

**Table 4**
*Spearman Correlation Results Performed on the Boussole Time Series Obtained by Comparing the Dominant Satellite PFT Percentage to the In Situ Pigment Ratios (Prel)*

| $R_s$ | $P_{rel}$ | $Chla_{in\ situ}$ | N Obs |
|---|---|---|---|
| Diat | 0.47* | 0.58* | 392 |
| Nano | 0.53* | 0.46* | 374 |
| Proc | 0.11 | −0.65* | 209 |
| Syne | 0.54* | −0.79* | 393 |

*Significant at the 0.001 level.

relationship between satellites derived secondary pigments and the Chla, while being consistent with the description of Hirata et al. (2011) in their analysis performed on in situ data. The results of this present study highlight the robustness of SOMs at the regional scale, capable to reproduce interpigment relationships while estimating the pigment composition from satellite data and accordingly identify dominant phytoplankton groups.

### 5.1. Phytoplankton Dynamic in the Mediterranean Sea

The oligotrophic state of the Mediterranean has been recognized for a long time (D'Ortenzio & Ribera d'Alcalà, 2009). A general decreasing gradient from west to east is typically observed in ocean color data of Chla (Barale et al., 2008; Bosc et al., 2004), to reach an ultraoligotrophic environment in the most eastern basin, the Levantine basin.

In this study, the detected winter bloom of Nano PFT classes (Hapto and Chloro classes) in most of the Mediterranean open water is induced by the nutrient availability generated by the water mixing in winter (Marty et al., 2002; Uitz et al., 2012). Such water mixing is due to the decrease of the SST and the resulting homogenization of the water column (Krom et al., 2004, 2005, 2010). The bloom stops shortly after the surface waters run out of nutrient.

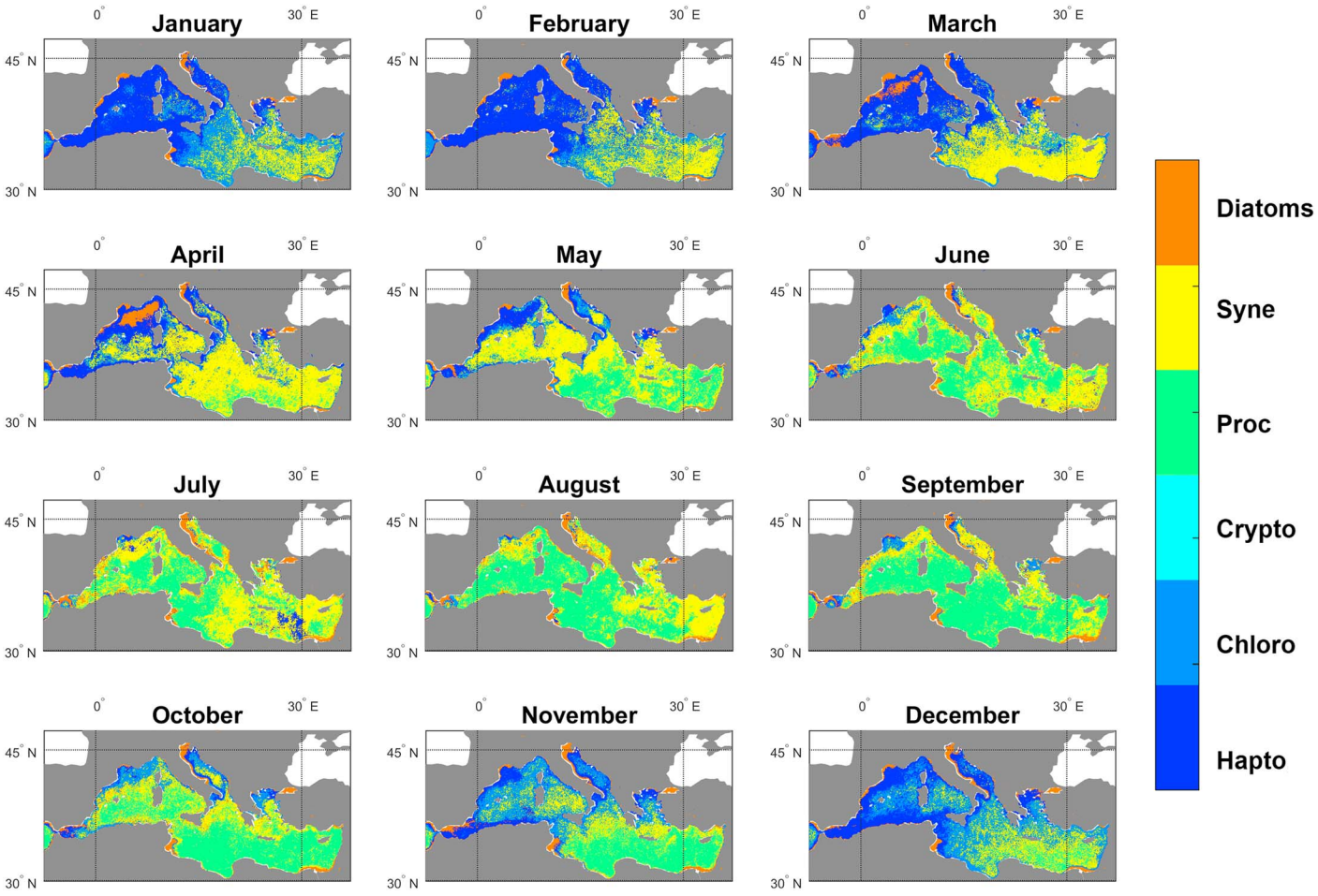


**Figure 8.** Monthly climatology images of the six dominant phytoplankton functional type classes identified using self-organizing map-phytoplankton functional type.

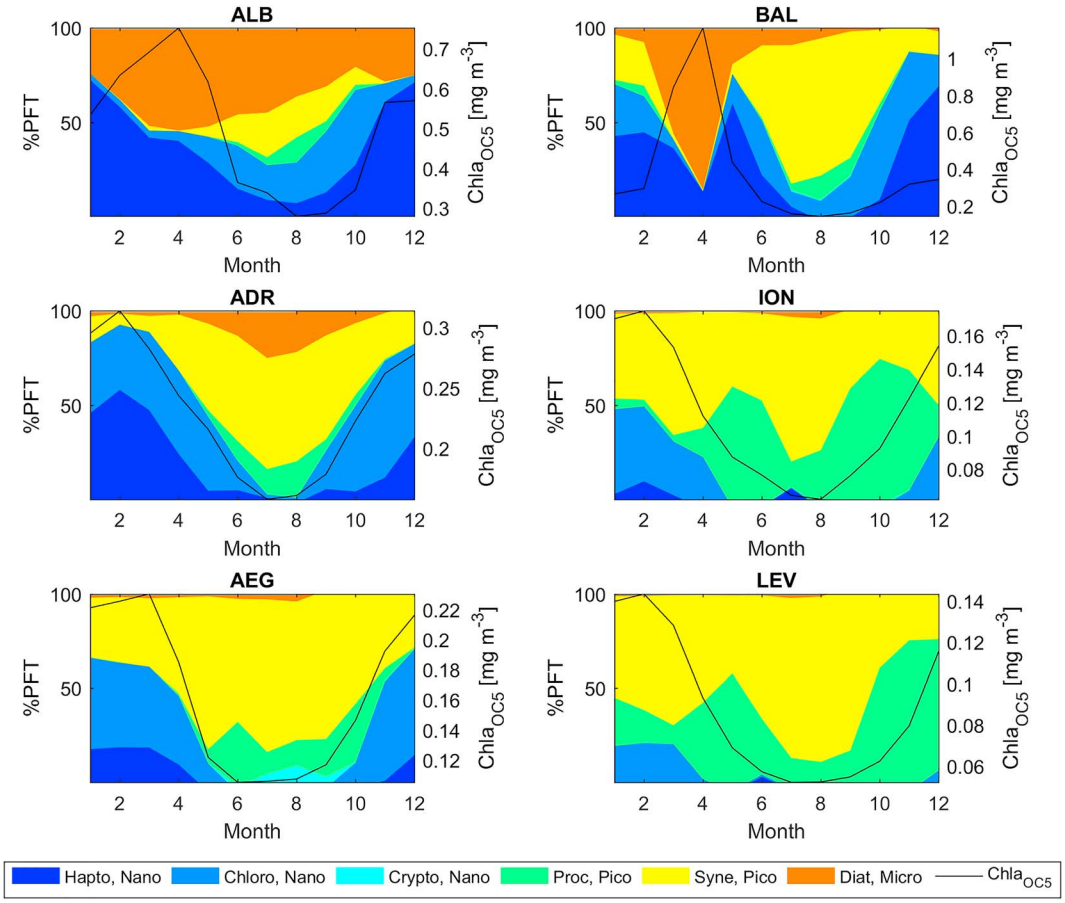


**Figure 9.** Monthly climatology of dominant phytoplankton functional type frequency alongside the Chla OC5 at the Alborian, Balearic, Adriatic, Ionian, Aegean and Levantine bassins.

Major hydrodynamic events play important roles in the mechanism of the bloom. In our study, the dominance of diatoms was observed in the northernmost part of the western Mediterranean basin. This region is characterized by the occurrence of an annual spring bloom of big phytoplankton triggered by nutrient enrichment of the euphotic zone due to deep winter convection events, which mix the whole water column (Marty et al., 2002). In addition to this pattern, the dominance of the Hapto and Chloro classes was also observed in this study all over the western basin in the winter-spring period. These waters are characterized by complex physical and chemical processes, such as water mass circulation, mesoscale hydrological structures such as anticyclonic eddies along the Algerian coast, and the intrusion of the Atlantic waters, which have the potential of alleviating the general nutrient limitation (Siokou-Frangou et al., 2010).

According to our findings, several eutrophic regions were marked by the recurrence of diatom dominance, such as the continental shelf of the Gulf of Lyon in the South of France, the Gulf of Gabes in the Tunisian coast, the Gulf of Venice in the north of the Adriatic Sea, the Nile Delta in the Levantine basin, and the northern edge of the Aegean Sea. These regions are characterized by their nutrient input originating from several sources such as the presence of a large and shallow continental shelf where the majority of the oceanic burial of organic carbon occurs and nutrients can be put in suspension by small-scale hydrodynamic phenomena (Hedges & Keil, 1995; Premuzic et al., 1982). Anthropogenic activities and riverine inputs, such as the Ebro (Eastern Spanish coast), the Rhone (Gulf of Lyon), the Po (Gulf of Venice), and the Nile (North of Egypt) also participate to nutrient enrichment. In such nutrient rich zones, the diatoms mainly develop due to the presence of essential elements for their growth including the silicon that is used to build their silica frustules. Unlike other minerals, the requirement for silicon is unique to diatoms. In the open ocean, the diatom bloom is typically ended by a shortage of silicon (Egge & Aksnes, 1992).

Moreover, the observed dynamic of the *Prochlorococcus* and *Synechococcus* classes is mainly due to their metabolic capacities, which make that these Pico are able to bloom in the summer season and then accordingly recycle dissolved organic matter in a very efficient manner under competitive circumstances such as limiting nutrients and predators (Sieburth et al., 1978). Furthermore, the abundance of Pico reaches a maximum in summer when the Chla concentration are at their lowest values. This fact can be explained by the lower photosynthetic pigment content in the Pico cells under higher irradiances and longer daylight periods during summer (Calvo-Díaz et al., 2008). During this season, primary production due to Pico exhibits a maximum value (Uitz et al., 2012). It is well known that thanks to their high surface/volume ratio, *Synechococcus* and *Prochlorococcus* can cope optimally with nutrients-impoverished environments (Le Quéré et al., 2005). The presence of *Synechococcus* in the Levantine basin has been widely reported as a dominant group where ultraoligotrophic conditions are present and particularly during summer (Siokou-Frangou et al., 2010; Uitz et al., 2012). The spatiotemporal variability of the dominant PFTs in the Mediterranean Sea observed in this study highlights the dominance of Hapto and Chloro in winter-spring seasons essentially in the western basin, whereas *Synechococcus* dominate most of the Mediterranean waters during summer. These results agree with previous studies on the Mediterranean Sea (e.g., Navarro et al., 2017; Sammartino et al., 2015; Uitz et al., 2012; Vidussi et al., 2000, 2001).

## 6. Conclusion

The clustering method we have developed was very efficient to identify the PFT from satellite measurements and to reconstruct the phytoplankton variability with significant performances. The approach was based on the optical characteristics of phytoplankton pigments and on the efficiency of the SOMs to make robust classifications. It permitted to retrieve the regional distribution of the PFT assemblage in the Mediterranean Sea and to estimate their specific pigment composition, which is innovative compared to most of the other methods inverting satellite observations. These PFTs showed a well-marked variability due to the complex physical and biogeochemical environment: Nanophytoplankton such as haptophytes and chlorophytes mainly dominated during winter in the western Mediterranean basin, while *Synechococcus* and *Prochlorococcus* dominated during summer. The dominance of diatoms was mainly observed in spring in the Balearic Sea in response to deep water convection phenomena and near the coastline and estuaries, due to important continental inputs and river discharges. Besides, we were able to observe the dynamics of several new phytoplankton types in the Mediterranean Sea, such as chlorophytes and cryptophytes. SOMs may highlight other important phytoplankton types.

The validation tests, which were performed on in situ matchups, showed satisfying performances and proved that our approach was suitable to study the phytoplankton diversity in the Mediterranean Sea. This work shows that the use of fine clustering methods such as the SOM chain we developed (SOM-Pigments → SOM-PFT + HAC) can accurately retrieve phytoplankton pigments from satellite observations and then convert pigment assemblages into well-defined phytoplankton communities, which can be identified to PFTs. The method is generic and could be applied to other oceanic regions.

**Acknowledgments**
This work was supported by the National Council for Scientific Research-Lebanon, in the frame of a doctoral fellowship, codirected between the National Center for Remote sensing-CNRSL, Lebanon; the Sorbonne Université, Faculty of Sciences-Paris VI; and the LOCEAN, France. This work has been supported by the National Center for Space Studies (CNES), France, under the TOSCA project (2018–2019) and by the Hubert Curien Progam CEDRE 2018–2019 (ALTILEV 40322Q8). This work has been supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 633211 (AtlantOS project). The different merged satellite ocean color data were obtained from the GlobColour project portal (www.globcolour.info). AVHRR Pathfinder Level 3 Daily Daytime SST Version 5.3 data set was obtained from the website (http://doi:10.7289/V52J68XX/). We acknowledge the different sources of the HPLC pigment data sets available online (https://pangaea.de/). Boussole time series data are available at the http://www.obs-vlfr.fr/Boussole/html/boussole_data/collected.php website.

## References

Abdel-Moati, A. R. (1990). Particulate organic matter in the subsurface chlorophyll maximum layer of the Southeastern Mediterranean. *Oceanologica Acta*, *13*(3), 307–315. Available at:. https://archimer.ifremer.fr/doc/00131/24250/22243.pdf Accessed: 11 February 2019

Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., & Hardman-Mountford, N. (2009). Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(15), 899–917). https://doi.org/10.1016/J.DSR2.2008.09.017

Alvain, S., Moulin, C., Dandonneau, Y., & Bréon, F. M. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Research Part I: Oceanographic Research Papers*, *52*(11), 1989–2004. https://doi.org/10.1016/j.dsr.2005.06.015

Alvain, S., Moulin, C., Dandonneau, Y., & Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochemical Cycles*, *22*, GB3001. https://doi.org/10.1029/2007GB003154

Antoine, D., André, J.-M., & Morel, A. (1996). Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global Biogeochemical Cycles*, *10*(1), 57–69. https://doi.org/10.1029/95GB02832

Antoine, D., & Morel, A. (1995). Observations, Algal pigment distribution and primary production in the eastern Mediterranean as derived from coastal zone color scanner. *Journal of Geophysical Research*, *100*(C8), 16,193–16,209. https://doi.org/10.1029/95JC00466

Barale, V., Jaquet, J. M., & Ndiaye, M. (2008). Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998-2003). *Remote Sensing of Environment*, *112*(8), 3300–3313. https://doi.org/10.1016/j.rse.2007.10.014

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*. Wiley-Blackwell, *19*, GB1006. https://doi.org/10.1029/2004GB002299

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*. Wiley-Blackwell, *42*(1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Ben Mustapha, Z., Alvain, S., Jamet, C., Loisel, H., & Dessailly, D. (2013). Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters. *Remote Sensing of Environment*, (February 2016, *146*, 97–112. https://doi.org/10.1016/j.rse.2013.08.046

Bethoux, J. P., Gentili, B., Morin, P., Nicolas, E., Pierre, C., & Ruiz-Pino, D. (1999). The Mediterranean Sea: A miniature ocean for climatic and environmental studies and a key for the climatic functioning of the North Atlantic. *Progress in Oceanography*. Pergamon, *44*(1–3), 131–146. https://doi.org/10.1016/S0079-6611(99)00023-3

Bosc, E., Bricaud, A., & Antoine, D. (2004). Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations. *18*, GB1005. https://doi.org/10.1029/2003GB002034

Bracher, A., Bouman, H. A., Brewin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., et al. (2017). Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development. *Frontiers in Marine Science*. Frontiers, *4*, 55. https://doi.org/10.3389/fmars.2017.00055

Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., & Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, Available at:, (5), 751–764. www.biogeosciences.net/6/751/2009/ (Accessed: 11 February 2019, https://doi.org/10.5194/bg-6-751-2009

Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M., & Hardman-Mountford, N. J. (2010). A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecological Modelling*. Elsevier, *221*(11), 1472–1483. https://doi.org/10.1016/J.ECOLMODEL.2010.02.014

Bricaud, A., Thiria, S., Chazottes, A., & Crépon, M. (2006). Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps View project Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps. *Applied Optics*, *45*, 8102. https://doi.org/10.1364/AO.45.008102

Calvo-Díaz, A., Morán, X. A. G., & Suárez, L. Á. (2008). Seasonality of picophytoplankton chlorophyll a and biomass in the central Cantabrian Sea, southern Bay of Biscay. *Journal of Marine Systems*, *72*, 271–281. https://doi.org/10.1016/j.jmarsys.2007.03.008

Cavazos, T., & Cavazos, T. (1999). Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. *Journal of Climate*, *12*(5), 1506–1523. https://doi.org/10.1175/1520-0442(1999)012<1506:LSCACT>2.0.CO;2

Ciotti, A. M., & Bricaud, A. (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. *Limnology and Oceanography: Methods*, *4*(7), 237–253. https://doi.org/10.4319/lom.2006.4.237

Dandonneau, Y., Deschamps, P.-Y., Nicolas, J.-M., Loisel, H., Blanchot, J., Montel, Y., et al. (2004). Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific. *Deep Sea Research Part II: Topical Studies in Oceanography*, *51*(1–3), 303–318. https://doi.org/10.1016/j.dsr2.2003.07.018

D'Ortenzio, F., & Ribera d'Alcalà, M. (2009). On the trophic regimes of the Mediterranean Sea: A satellite analysis. *Biogeosciences*, *6*, 139–148. https://doi.org/10.5194/bg-6-139-2009

Dowidar, N. M. (1984). Phytoplankton biomass and primary productivity of the south-eastern Mediterranean. *Deep Sea Research Part A. Oceanographic Research Papers*. Elsevier, *31*(6–8), 983–1000. https://doi.org/10.1016/0198-0149(84)90052-9

Egge, J. K., & Aksnes, D. L. (1992). Silicate as regulating nutrient in phytoplankton competition. *Marine Ecology Progress Series*, *83*, 281–289. Available at:. https://www.int-res.com/articles/meps/83/m083p281.pdf (Accessed: 11 February 2019, https://doi.org/10.3354/meps083281

El Hourany, R., Abboud-Abi Saab, M., Faour, G., Aumont, O., Crépon, M., & Thiria, S. (2019). Estimation of secondary phytoplankton pigments from satellite observations using self-organizing maps (SOM). *Journal of Geophysical Research: Oceans*. John Wiley & Sons, Ltd., *124*(2), 1357–1378. https://doi.org/10.1029/2018JC014450

Fujiwara, A., Hirawake, T., Suzuki, K., & Saitoh, S.-I. (2011). Remote sensing of size structure of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf region. *Biogeosciences*, *8*(12), 3567–3580. https://doi.org/10.5194/bg-8-3567-2011

Gieskes, W. W. C., & Kraay, G. W. (1983). Dominance of Cryptophyceae during the phytoplankton spring bloom in the central North Sea detected by HPLC analysis of pigments. *Marine Biology*. Springer-Verlag, *75*(2–3), 179–185. https://doi.org/10.1007/BF00406000

Gohin, F. (2011). Annual cycles of chlorophyll-a, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters. *Ocean Science*, *7*(5), 705–732. https://doi.org/10.5194/os-7-705-2011

Guillard, R. R. L., Murphy, L. S., Foss, P., & Liaaen-Jensen, S. (1985). *Synechococcus* spp. as likely zeaxanthin-dominant ultraphytoplankton in the North Atlantic. *Limnology and Oceanography*. Wiley-Blackwell, *30*(2), 412–414. https://doi.org/10.4319/lo.1985.30.2.0412

Hedges, J. I., & Keil, R. G. (1995). Sedimentary organic matter preservation: An assessment and speculative synthesis. *Marine Chemistry*. Elsevier, *49*(2–3), 81–115. https://doi.org/10.1016/0304-4203(95)00008-F

Hewitson, B. C., & Crane, R. G. (2002). Self-organizing maps: Applications to synoptic climatology. *Climate Research*. Inter-Research Science Center, *22*, 13–26. https://doi.org/10.2307/24868302

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J., & Barlow, R. G. (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment* Elsevier, *112*(6), 3153–3159. https://doi.org/10.1016/J.RSE.2008.03.011

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., et al. (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, *8*(2), 311–327. https://doi.org/10.5194/bg-8-311-2011

Ignatiades, L., Gotsis-Skretas, O., Pagou, K., & Krasakopoulou, E. (2009). Diversification of phytoplankton community structure and related parameters along a large-scale longitudinal east-west transect of the Mediterranean Sea. *Journal of Plankton Research*. Oxford University Press, *31*(4), 411–428. https://doi.org/10.1093/plankt/fbn124

Jeffrey, S. W. (1980). Algal pigment systems. In *Primary productivity in the sea*, (pp. 33–58). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-3890-1_3

Jeffrey, S. W., & Hallegraeff, G. M. (1987). Chlorophyllase distribution in ten classes of phytoplankton: a problem for chlorophyll analysis. *Marine Ecology Progress Series*. Inter-Research Science Center, *35*, 293–304. https://doi.org/10.2307/24825001

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*. Pergamon, *37*, 52–65. https://doi.org/10.1016/J.NEUNET.2012.09.018

Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J. W., et al. (2017). Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models. *Remote Sensing of Environment*. Elsevier, *190*, 162–177. https://doi.org/10.1016/J.RSE.2016.11.014

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research*. John Wiley & Sons, Ltd, *114*(C9), C09015. https://doi.org/10.1029/2009JC005303

Krom, M. D., Emeis, K. C., & Van Cappellen, P. (2010). Why is the Eastern Mediterranean phosphorus limited? *Progress in Oceanography*. Elsevier Ltd, *85*(3–4), 236–244. https://doi.org/10.1016/j.pocean.2010.03.003

Krom, M. D., Herut, B., & Mantoura, R. F. C. (2004). Nutrient budget for the Eastern Mediterranean: Implications for phosphorus limitation. *Limnology and Oceanography*, *49*(5), 1582–1592. https://doi.org/10.4319/lo.2004.49.5.1582

Krom, M. D., Woodward, E. M. S., Herut, B., Kress, N., Carbo, P., Mantoura, R. F. C., et al. (2005). Nutrient cycling in the south east Levantine basin of the eastern Mediterranean: Results from a phosphorus starved system. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *52*(22–23), 2879–2896. https://doi.org/10.1016/j.dsr2.2005.08.009

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*. John Wiley & Sons, Ltd, *11*, 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x

Liu, Y., & Weisberg, R. H. (2005). Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. *Journal of Geophysical Research*, Wiley-Blackwell, *110*(C6), C06003. https://doi.org/10.1029/2004JC002786

Liu, Y., Weisberg, R. H., & Mooers, C. N. K. (2006). Performance evaluation of the self-organizing map for feature extraction. *Journal of Geophysical Research*. Wiley-Blackwell, *111*(C5), C05018. https://doi.org/10.1029/2005JC003117

Longhurst, A., Sathyendranath, S., Platt, T., & Caverhill, C. (1995). An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*. Oxford University Press, *17*(6), 1245–1271. https://doi.org/10.1093/plankt/17.6.1245

Marty, J.-C., Chiavérini, J., Pizay, M.-D., & Avril, B. (2002). Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999). *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, *49*(11), 1965–1985. https://doi.org/10.1016/S0967-0645(02)00022-X

McClain, E. P., Pichel, W. G., & Walton, C. C. (1985). Comparative performance of AVHRR-based multichannel sea surface temperatures. *Journal of Geophysical Research*. Wiley-Blackwell, *90*(C6), 11587. https://doi.org/10.1029/JC090iC06p11587

Mouw, C. B., & Yoder, J. A. (2010). Optical determination of phytoplankton size composition from global SeaWiFS imagery. *Journal of Geophysical Research*, *115*, C12018. https://doi.org/10.1029/2010JC006337

Navarro, G., Almaraz, P., Caballero, I., Vázquez, Á., & Huertas, I. E. (2017). Reproduction of spatio-temporal patterns of major Mediterranean phytoplankton groups from remote sensing OC-CCI data. *Frontiers in Marine Science*, *4*, 1–16. https://doi.org/10.3389/fmars.2017.00246

Navarro, G., Alvain, S., Vantrepotte, V., & Huertas, I. E. (2014). Identification of dominant phytoplankton functional types in the Mediterranean Sea based on a regionalized remote sensing approach. *Remote Sensing of Environment*. Elsevier Inc., *152*, 557–575. https://doi.org/10.1016/j.rse.2014.06.029

Niang, A., Badran, F., Moulin, C., Crépon, M., & Thiria, S. (2006). Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. *Remote Sensing of Environment*. Elsevier, *100*(1), 82–94. https://doi.org/10.1016/J.RSE.2005.10.005

Premuzic, E. T., Benkovitz, C. M., Gaffney, J. S., & Walsh, J. J. (1982). The nature and distribution of organic matter in the surface sediments of world oceans and seas. *Organic Geochemistry*. Pergamon, *4*(2), 63–77. https://doi.org/10.1016/0146-6380(82)90009-2

Reusch, D. B., Alley, R. B., & Hewitson, B. C. (2007). North Atlantic climate variability from a self-organizing map perspective. *Journal of Geophysical Research*. Wiley-Blackwell,, *112*(D2), D02104. https://doi.org/10.1029/2006JD007460

Richardson, A., Risien, C., & Shillington, F. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*. Pergamon, *59*(2–3), 223–239. https://doi.org/10.1016/J.POCEAN.2003.07.006

Roy, S., Sathyendranath, S., Bouman, H., & Platt, T. (2013). The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data. *Remote Sensing of Environment*. Elsevier, *139*, 185–197. https://doi.org/10.1016/J.RSE.2013.08.004

Sammartino, M., Di Cicco, A., Marullo, S., & Santoleri, R. (2015). Phytoplankton in the Mediterranean Sea from satellite ocean colour data of SeaWiFS. *Ocean Science*, 759–778. https://doi.org/10.5194/os-11-759-2015

Sathyendranath, S., Aiken, J., Alvain, S., Barlow, R., Bouman, H., Bracher, A., et al. (2014). In S. Sathyendranath, & V. Stuart (Eds.), *Phytoplankton functional types from Space. IOCCG; 15*. Dartmouth, Nova Scotia, B2Y 4A2, Canada: International Ocean-Colour Coordinating Group. Available at:. https://epic.awi.de/id/eprint/36000/ Accessed: 9 February 2019

Sieburth, J. M., Smetacek, V., & Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. *Limnology and Oceanography*. John Wiley & Sons, Ltd, *23*(6), 1256–1263. https://doi.org/10.4319/lo.1978.23.6.1256

Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera D'Alcala, M., Vaque, D., & Zingone, A. (2010). Plankton in the open mediterranean Sea: A review. *Biogeosciences*, *7*(5), 1543–1586. https://doi.org/10.5194/bg-7-1543-2010

Turley, C., Bianchi, M., Christaki, U., Conan, P., Harris, J., Psarra, S., et al. (2000). Relationship between primary producers and bacteria in an oligotrophic sea-the Mediterranean and biogeochemical implications. *Marine Ecology Progress Series*, *193*, 11–18. https://doi.org/10.3354/meps193011

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*. John Wiley & Sons, Ltd, *111*(C8), C08005. https://doi.org/10.1029/2005JC003207

Uitz, J., Stramski, D., Gentili, B., D'Ortenzio, F., & Claustre, H. (2012). Estimates of phytoplankton class-specific and total primary production in the Mediterranean Sea from satellite ocean color observations. *Global Biogeochemical Cycles*, *26*, GB2024. https://doi.org/10.1029/2011GB004055

Vaulot, D., Partensky, F., Neveux, J., Mantoura, R. F. C., & Llewellyn, C. A. (1990). Winter presence of prochlorophytesin surface waters of the northwestern Mediterranean Sea. *Limnology and Oceanography*, *35*(5), 1156–1164. https://doi.org/10.4319/lo.1990.35.5.1156

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & Marty, J.-C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research*. Wiley-Blackwell, *106*(C9), 19,939–19,956. https://doi.org/10.1029/1999JC000308

Vidussi, F., Marty, J., & Chiave, J. (2000). Phytoplankton pigment variations during the transition from spring bloom to oligotrophy in the northwestern Mediterranean Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(3), 423–445. https://doi.org/10.1016/s0967-0637(99)00097-7

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*. Wiley-Blackwell, *22*, G82024. https://doi.org/10.1029/2007GB003078

Wright, S. W., & Jeffrey, S. W. (1987). Fucoxanthin pigment markers of marine phytoplankton analysed by HPLC and HPTLC. *Marine Ecology Progress Series*, *38*(3), 259–266. https://doi.org/10.2307/24825629

Yacobi, Y. Z., Zohary, T., Kress, N., Hecht, A., Robarts, R. D., Waiser, M., et al. (1995). Chlorophyll distribution throughout the south-eastern Mediterranean in relation to the physical structure of the water mass. *Journal of Marine Systems*, *6*(3), 179–190. https://doi.org/10.1016/0924-7963(94)00028-A

Yogev, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., et al. (2011). Is dinitrogen fixation significant in the Levantine Basin, East Mediterranean Sea? *Environmental Microbiology*, *13*(4), 854–871. https://doi.org/10.1111/j.1462-2920.2010.02402.x