Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Estimation of phytoplankton pigments from ocean-color satellite observations in the Sénégalo-Mauritanian region by using an advanced neural classifier

By

Khalil Yala[1], Ndeye Niang[2], Julien Brajard[1], Carlos Mejia[1], Mory Ouattara[2], Roy El Hourany[1], Michel Crépon[1] and Sylvie Thiria[1,3]

[1] IPSL/LOCEAN, Sorbonne Université (Université Paris6, CNRS, IRD, MNHN), 4 Place Jussieu, 75005 Paris, France
[2] CEDRIC, CNAM, 292 rue Saint Martin, 75003 Paris, France
[3] UVSQ, F-78035, Versailles, France

Corresponding author : Michel Crepon (crepon@locean-ipsl.upmc.fr)

ABSTRACT

We processed daily ocean-color satellite observations to construct a monthly climatology of phytoplankton pigment concentrations in the Senegalo-Mauritanian region. Thanks to the difficulty of the problem, we proposed a new method. It primarily consists in associating, in well-identified clusters, similar pixels in terms of ocean-color parameters and in situ pigment concentrations taken from a global ocean database. The association is carried using a new Self Organized Map (2S-SOM). Its major advantage is to allow taking into account the specificity of the optical properties of the water by adding specific weights to the different ocean color parameters and the in situ measurements. In the retrieval phase, the pigment concentration of a pixel is estimated by taking the pigment concentration values associated with the 2S-SOM cluster presenting the ocean-color satellite spectral measurements, which are the closest to those of the pixel under study according to some distance. The method was validated by using a cross-validation procedure. We focused our study on the fucoxanthin concentration, which is related to the abundance of diatoms. We showed that the fucoxanthin starts to develop in December, presents its maximum intensity in March when the upwelling intensity is maximum, extends up to the coast of Guinea in April and begins to decrease in

May. The results are in agreement with previous observations and recent in situ measurements. The method is very general and can be applied in every oceanic region.

1 - INTRODUCTION

Phytoplankton is the basis of the ocean food web and consequently drives the ocean productivity. It also plays a fundamental role in climate regulation by trapping atmospheric carbon dioxide (CO2) through gas exchanges at the sea surface and consequently lowering the rate of anthropogenic increase in the atmosphere of CO2 concentration by about 30% (*Behrenfield et al,* 2005). With the growing interest in climate change, one may ask how the different phytoplankton populations will respond to changes in ocean characteristics (temperature, salinity, acidity) and nutrient supply, which presents an important societal impact with respect to both climate and fisheries with a possible effect on fish grazing on phytoplankton.

Methods for identifying phytoplankton have greatly progressed during the last two decades. Phytoplankton was first described by microscopy. The number of samples that can be analyzed is very limited due to the necessary presence of an expert to discriminate the different taxa observed with the microscope.

Recently pigment analysis of seawater samples by high-performance liquid chromatography (HPLC) has been widely used to categorize broad phytoplankton size classes (PSC) (*Jeffreys et al*, 1997, *Brewin et al,* 2010) and even phytoplankton functional types (PFT; *Hirata et al,* 2011). HPLC enables identification of 25 to 50 pigments within a single analysis, which is much more easy and fast to conduct than microscopic observations. Each phytoplankton group (PSC and PFT) is associated with specific diagnostic pigments and a conversion formula can be derived to estimate the percentage of each group from the pigment measurements (*Vidussi et al,* 2001; *Uitz et al*, 2010). HPLC measurements are now widely used to determine phytoplankton species in situ.

The use of satellite ocean color sensor measurements has permitted to map the ocean surface at a daily frequency. Satellite sensors measure the sunlight, at several wavelengths, backscattered by the ocean. The downwelling sunlight interacts with the seawater through backscattering and absorption in such a manner that the upwelling radiation transmitted to the satellite (‘water-leaving’ reflectance) contains information related to the composition of the seawater. The light transmitted to the satellite depends on the phytoplankton cell shape (backscattering), its pigments (absorption), the dissolved matter (e.g. CDOM).

This upwelling radiation, the so-called remotely sensed reflectance $\rho_w(\lambda)$, is determined by the

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

spectral absorption ***a*** and backscattering ($\boldsymbol{b_b}$ ($\mathbf{m^{-1}}$)) coefficients of the ocean (pure water and various particulate and dissolved matters) using the simplified formulation (*Morel* and *Gentili*, 1996):

$$\rho_w(\lambda) = G\, \boldsymbol{b_b}(\lambda)/(\boldsymbol{a}(\lambda) + \boldsymbol{b_b}(\lambda)) \qquad (1)$$

where ($\boldsymbol{a}$ ($\mathbf{m^{-1}}$)) is the sum of the individual absorption coefficients of water, phytoplankton pigments, colored dissolved organic matter, and detrital particles, ($\boldsymbol{b_b}$ ($\mathbf{m^{-1}}$)) depends on the shape of the phytoplankton species. ***G*** is a parameter mainly related to the geometry of the situation (sensor and solar angles) but also to environmental parameters (wind, aerosols).

In the open ocean far from the coast (in case-1 waters), the light seen by the satellite sensor mainly contains information on phytoplankton abundance and diversity. Ocean-color measurements have been first used intensively to estimate chlorophyll-*a* concentration (*chl-a* in the following) in the surface waters of the ocean, marginal seas and lakes. (*Longhurst et al.,* 1995; *Antoine et al.,* 1996; *Behrenfeld and Falkowski*, 1997; *Behrenfeld et al.,* 2005; *Westberry et al.*, 2008).

It has been shown that it is also possible to extract additional information such as phytoplankton size-classes (PSC) by using some relationship between chlorophyll concentration and PSC (*Uitz et al.*, 2006; *Ciotti and Bricaud*, 2006; *Hirata et al.*, 2008; *Mow and Yoder,* 2010). These algorithms try to establish an algebraic relationship between the *chl-a* concentration and the PSC percentage. Some of them (*Uitz et al*, 2006; *Aiken et al.,* 2009) break-down the *chl-a* abundance into several ranges for each of which a specific relationship is computed. Others (*Brewin et al*, 2010; *Hirata et al*, 2011) are based on a continuum of *chl–a* abundance. Studies have also been done to estimate the phytoplankton groups (PFT) by taking into account spectral information (*Sathyendranath et al.,* 2004, *Alvain et al.*, 2005, 2012; *Hirata et al.*, 2011; *Ben Mustapha et al,* 2013; *Farikou et al*, 2015), which is of fundamental interest to the understanding of the phytoplankton behavior and to modeling its evolution.

Due to highly non-linear relationship linking the multispectral ocean color measurements with the pigment concentrations, we proposed a neural network clustering algorithm (2S-SOM) able to deal with multi variables related by complex relationships. The 2S-SOM algorithm is well adapted to this complex task by weighting the different inputs. The clustering algorithm was calibrated on a restricted database composed of remote sensed observations co-located with measurements taken in the global ocean.

In the present paper, we propose the retrieval of the major pigment concentrations from satellite ocean color multi-spectral sensors in the Senegalo-Mauritanian upwelling, which is an oceanic

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

region off the coast of West Africa where a strong seasonal upwelling occurs (Figure 1).

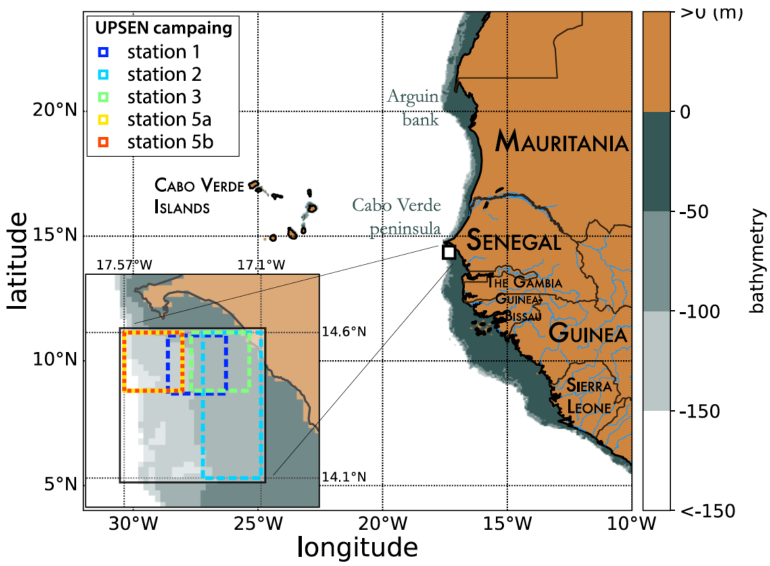

Figure 1 : *Mauritania and Senegal coastal topography. The land is in red and the ocean depth is represented in meters by the color scale on the right side of the figure. The UPSEN stations are shown in the bottom left cartoon of the figure.*

The Senegalo-Mauritanian upwelling is a one of the most productive eastern boundary upwelling system (EBUS) with strong economic impacts on fisheries in Senegal and Mauritania. Since this region has been poorly surveyed in situ, we have chosen to extract pertinent biological information from ocean-color satellite measurements. This region has been intensively studied by analysis of SeaWiFS ocean-color data and AVHRR sea-surface temperature as reported in *Demarcq* and *Faure* (2002), and more recently by *Sawadogo et al.* (2009); *Farikou et al.* 2013, 2015; *Ndoye et al*, 2014; *Capet et al*, 2017.

The paper is articulated as follows: in section 2, we present the data we used (in situ and remote sensing observations). The mathematical aspect of the clustering method (2S-SOM) is detailed in section 3. In section 4 we present the methodological results. The spatio-temporal variability of the fucoxanthin and chl-a concentration in the Senegalo-Mauritanian upwelling region are presented in section 5 as well as the results of the oceanic UPSEN campaigns. In section 6 we discuss the results and the method. A conclusion is presented in section 7.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

2- MATERIALS

In this study we used three distinct datasets: the first was used to calibrate the method, the second to conduct a climatological analysis of the Senegalo-Mauritanian upwelling region and the third was obtained during the oceanographic UPSEN campaign. These datasets are composed of satellite remote sensing observations (datasets 1, 2, 3) and in-situ measurements (datasets 1, 3).

2.1 The calibration data base (DPIG)

The calibration database (DPIG) comprises in situ pigment measurements co-located with satellite ocean-color observations done by the SeaWiFS (sea-viewing, wide-field-of-view sensor).
This learning dataset (DPIG) is composed of 515 matched satellite observations and in situ measurements made in the global ocean (mainly in the North Atlantic and the equatorial ocean; *Ben Mustapha et al*., 2014). The match-up criteria were quite severe: we used satellite pixel situated at a distance less than 20km of the in situ measurement in a time window of +/- 12h. The geographic distribution of the 515 coincident in situ and satellite measurements is shown in Fig. 2.

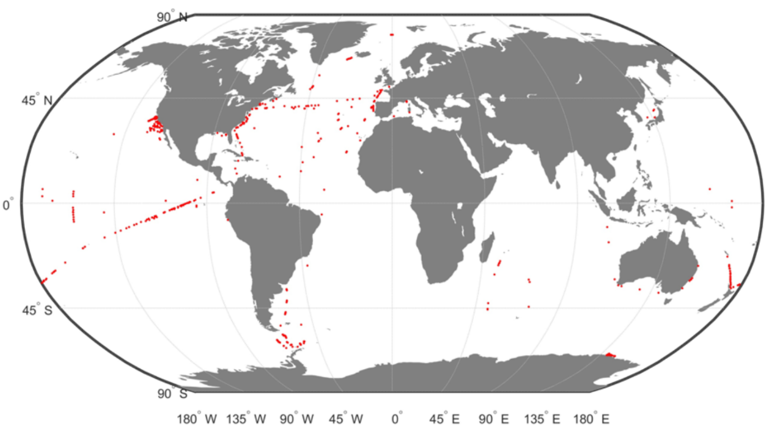

Figure 2 : *Geographic positions of the 515 in situ and satellite collocated measurements of the DPIG database.*

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Ocean Science
Discussions
Open Access

In Figure 3 we present the R^2 coefficient between the in situ *chl-a* a and the SeaWiFS *chl-a* a computed by using the OC4V4 algorithm for the DPIG collocated observations. We remark that the

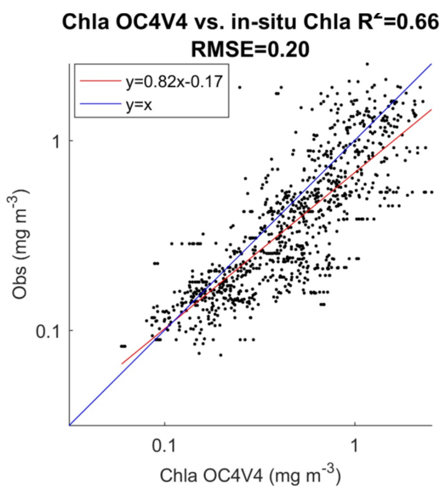

Figure 3: *Dispersion diagram of DPIG chl-a computed from the SeaWiFS observations using the OC4V4 algorithm versus in situ chl-a.*

two measurements are in good agreement at global scale. Each data of DPIG is a vector having 17 components (five ocean reflectance ($\rho_w(\lambda)$ at five wave lengths, five $Ra(\lambda)$, SeaWiFS *chl-a,* five in situ pigment ratios and an in situ *chl-a* concentration). The in situ *chl-a* a concentration ranges between 0.007 and 3. mg m^{-3} (see Table 1).

	RDIVINY A	RPERID	RFUCO	R19HF	RZEAX	CHLORO *IN SITU*
MEAN	0.1414	0.0272	0.1248	0.1859	0.1696	0.5292
STD	0.1584	0.0196	0.0971	0.0996	0.2063	0.5720
MIN	0.0037	0.0035	0.0053	0.0066	0.0027	0.007
MAX	0.8889	0.2027	0.8514	0.7654	1.5574	2.9980

Table 1 : *Pigments of the DPIG and their statistical characteristics*

The five *Ra(λ)* are defined following *Alvain et al*, 2012 :

$$Ra(\lambda) = \rho_W(\lambda)/\ \rho_{Wref}(\lambda,\ chl\text{-}a) \qquad (2)$$

where the parameter $\rho_{wref}(\lambda, chl_a)$ is an average reflectance depending on the *chl-a* concentration only which was computed according to the procedure reported in *Farikou et al*, 2015. *Ra(λ)* is a non-dimensional parameter which is independent of the *chl-a* abundance and sensitive the secondary pigments only (*Alvain et al* , 2012).

The DPIG database thus provides information on the existing links between the pigment composition and the SeaWiFS measurements. The pigment composition are defined by the pigment ratios which are non-dimensional variables of the form in the present study:

Pigment Ratio=DP/T*chl-a* (3)

Defined as the ratio of the diagnostic pigment (DP) related to the total *chl-a* (T*chl-a* = *chl-a* +Dyvinil*chl-a*) (*Alvain et al*., 2005).

The pigments of the DPIG and their statistical characteristics are given in Table 1.

2.2 The Senegalo-Mauritanian upwelling satellite data (DSAT)

Since the Senegalo-Mauritanian region has been poorly surveyed in situ, we have chosen to extract pertinent biological information from ocean-color satellite measurements. This region has been intensively studied by the analysis of SeaWiFS ocean-color data and AVHRR sea-surface temperature as reported in *Demarcq* and *Faure* (2002), *Sawadogo et al.* (2009), *Farikou et al.* (2013, 2015) and more recently by *Ndoye et al*, 2014; *Capet et al*, 2017 with in situ observations. The satellite dataset we processed to retrieve the pigment concentration consist of five $\rho_w(\lambda)$ and five *Ra(λ)* at five wavelengths (412 nm, 443nm, 490nm, 510nm and 555nm), and the SeaWiFS *chl-a* concentration observed in the Senegalo-Mauritanian upwelling region (8°N-24°N, 14°W-20°W; Figure 3) during 11 years (1998-2009) by SeaWiFS. This data set is here below denoted *DSAT*.

The satellite observations ($\rho_w(\lambda)$ and *chl-a* concentration) were provided by NASA with a resolution of nine kilometers. Due to the presence of Saharan dusts in this region, very few estimations of satellite $\rho_w(\lambda)$ and in situ *chl-a* were available, and some satellite estimations of *chl-a* could present strong over-estimations (*Gregg et al*, 2004). For this reason, we reprocessed the $\rho_w(\lambda)$ and *chl-a* data with an atmospheric correction algorithm developed specifically for Saharan dust (*Diouf et al,* 2013, http://poac.locean-ipsl.upmc.fr) in order to improve the satellite observations.

2.3 The UPSEN database

Recently, some HPLC measurements were made in the Senegalo-Mauritanian region during two oceanographic cruises (UPSEN campaigns) of the oceanographic ship "Le Suroit" from 7 to 17 March 2012 and from 5 to 26 February 2013 as reported in *Ndoye et al*, 2014; *Capet et al*, 2017. The goal was to study the dynamics and the biological variability of the Senegalo-Mauritanian upwelling. During these campaigns, in-situ HPLC measurements were carried out. We expected to be able to co-locate them with the ocean-color VIIRS (visible infra-red imaging radiometer suite) sensor observations whose wavelengths are close to those of the SeaWiFS. Unfortunately, we were only able to process satellite observations made on 21 February 2013 due to the presence of clouds and aerosols the other days. We processed the satellite observations provided by the VIIRS sensor at four wavelengths (443, 490, 510, 555 nm) for pixels in the vicinity of the ship stations (within a distance of 20km) and observed in a time window of +/- 12h, and for which the satellite *chl-a* was less than 3mg/m^3, which is the limit of validity of our method imposed by the range of *chl-a* observed in DGIP (mean of 0,52 mg/m^3). Only five stations off Cabo Verde peninsula fitted these requirements (see Figure 1 for their positions).

3 - THE PROPOSED METHOD (2S-SOM)

Classification methods were applied for retrieving geophysical parameters from large databases in several studies including weather forecasting (*Lorenz*, 1969; *Kruizinga and Murphy*, 1983), short-term climate prediction (*Van den Dool*, 1994), downscaling (*Zorita and von Storch*, 1999), reconstruction of oceanic pCO_2 (*Friedrichs and Oschlies.,* 2009), and of *chl-a* concentration under clouds (*Jouini et al*, 2013). In the present study we used a new neural network algorithm, which is an extension of the SOM algorithms (*Kohonen,* 2001)

3-1 The SOM clustering

SOM algorithms constitute powerful nonlinear unsupervised classification methods. They are unsupervised neural classifiers, which have been commonly used to solve environmental problems (*Cavazos,* 1999; *Hewitson et al,* 2002; *Richardson et al,* 2003; *Liu et al,* 2005, 2006; *Niang et al,* 2006; *Reusch et al*, 2007). SOM aims at clustering vectors of a multidimensional database (***D***) into classes represented by a fixed network of neurons (the SOM map). The self-organizing maps are defined as an undirected graph, usually a rectangular grid of dimension $p \times q$. This graph structure is used to define a discrete distance (denoted by δ) between the neurons of the map which presents the shortest path between two neurons. Moreover, SOM enables the partition of $\boldsymbol{D}$ in which each cluster is associated with a neuron of the map and is represented by a prototype that is a synthetic multidimensional vector (the referent vector $\boldsymbol{w}$). Each vector z_i of $\boldsymbol{D}$ will be assigned to the neuron whose referent $\boldsymbol{w}$ is the closest, in the sense of the Euclidean Norm (EN), and will be called the projection of the vector z_i on the map. A fundamental property of a SOM is the topological ordering provided at the end of the clustering phase: two close neurons on the map represent data that are close in the data space. The estimation of the referent vectors $\boldsymbol{w}$ of a SOM and the topological order is achieved through a minimization process in which the referent vectors $\boldsymbol{w}$ are estimated from a learning data set (The DPIG data base in the present case). The cost function is of the form:

$$J_{SOM}^{T}(\chi, W) = \sum_{zi\in D} \sum_{c\in SOM} K^{T}\left(\delta\left(c, \chi(z_i)\right)\right) \|z_i - w_c\|^2 \qquad (4)$$

where c ∈ SOM indices the neurons of the SOM map, χ is the allocation function that assigns each element z_i of DPIG to its referent vector $w_{\chi(z_i)}$ and $\delta\left(c, \chi(z_i)\right)$ is the discrete distance on the SOM between a neuron c and the neuron allocated to observation z_i and K^T a kernel function parameterized by T (where T stands for “temperature” in the scientific literature dedicated to SOM) that weights the discrete distance on the map and decreases during the minimization process. T acts as a regularization term.

This cost function takes into account the proper inertia of the partition of the data set $\boldsymbol{D}$ and ensures that its topology is preserved.

SOMs have frequently been used in the context of completing missing data (*Jouini et al*, 2013), so the projected vectors z_i may have missing components. Under these conditions, the distance between a vector $z_i \in \boldsymbol{D}$ and the referent vectors $\boldsymbol{w}$ of the map is the Euclidean distance that considers only the existing components (the Truncated Distance or *TD* hereinafter).

3-2 The 2S-SOM Classifier

In the present case, we used 2S-SOM, which is a modified version of SOM, very useful in the case of a large number of variables. It automatically structures the variables having some common characters into conceptually meaningful and homogeneous blocks. The 2S-SOM takes advantage of this structuration of $\boldsymbol{D}$ and variables into different blocks, which permits an automatic weighting of the influence of each block and consequently of each variable. Due to its capacity to weight the different parameters, 2S-SOM is able to deal with a large quantity of parameters, choosing those that are the most significant for the classification and neutralizing those which are less significant. This is done by using a more complex cost function that introduces a set of new parameters estimated during the

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

learning phase along with the referent vectors. For a neuron *c*, we define the weights α_{cb} of each block *b* and the weights β_{cbj} of the variables *j* in this block ***b***. The new cost function is

$$J^{T}_{2S-SOM}(\chi, W, \alpha, \beta) = \sum_{c}\left(\sum_{b=1}^{B}\left(\sum_{zi \in D} \alpha_{cb}\, K^{T}\left(\delta\left(c, \chi(z_i)\right)\right) d_{\beta_{cb}}(i) + J_{cb}\right) + I_c\right) \quad (5)$$

with $d_{\beta_{cb}}(i) = \sum_{j=1}^{P_b} \beta_{cbj} \left\| z^{j}_{ib} - w^{j}_{ib} \right\|^2$ (6)

where c ∈ 2S-SOM indices the neurons of the 2S-SOM map, P_b is the number of variables in the block ***b***, and ***B*** is the number of blocks, under the constraints:

$$\sum_{b=1}^{B} \alpha_{cb} = 1; \alpha_{cb} \in [0,1] \forall c \in [2S-SOM] \quad (7)$$

and

$$\sum_{j=1}^{P_b} \beta_{cbj} = 1; \beta_{cbj} \in [0,1] \forall c \in [2S-SOM]; \forall b \quad (8)$$

I_c and J_{cb} are used to regularize the weights α and η. They are defined as negative entropies weighted by μ for the blocks and η for the variables of each blocks:

$$I_c = \mu \sum_{b=1}^{P_b} \alpha_{cb} log(\alpha_{cb})$$

and $J_{cb} = \eta \sum_{j=1}^{B} \beta_{cbj} log(\alpha_{cbj})$

The penalty coefficients μ and η are two hyper-parameters that are determined at the end of the learning phase depending on the problem under study. The 2S-SOM allows, during the learning phase, an automatic weighting of the influence of each block and of each variable. After the learning phase, the influence of each group and of each variable is different for each neuron, which permits determination of a better classification that identifies the relevant variables for the different PSCs.

The learning procedure comprises three distinct phases, the third being iterated in order to choose the two hyper-parameters:

- *First :* a standard SOM map (SOM_{init}) is learned using the $J^{T}_{SOM}(\chi, W)$ cost function
- *Second* : the different values of (μ , η) are sampled
- *Third* : for each pair of (μ , η)
- SOM_{init} is used, as the initial condition of a new learning phase.
- All the parameters are estimated at the same time using $J^{T}_{2S-SOM}(\chi, W, \alpha, \beta)$.
- A 2S-SOM(μ , η) is determined

The best SOM(μ , η) is chosen according to the problem under consideration.

The 2S-SOM algorithm is available on : https://github.com/carmman/2S-SOM_versionCM

3.3 The calibration phase

The vectors of DPIG defined in section 2 can be decomposed in four blocks. The essence of the decomposition of the components of the DPIG database vectors in blocks is that each of the 17 components of the DPIG vectors gathered information with a different physical influence in the classification phase. We therefore consider different blocks (four in the present case) composed of variables having similar physical significance. The composition of each block is done as follows:

First Block (B1) comprises the five pigment in-situ concentration ratios (divinyl chlorophyll-a, peridinin, fucoxanthin, 19'hexanoyloxyfucoxanthin, zeaxanthin concentration ratios). The pigment ratios are defined in Eq. 3.

Second Block (B2) comprises the water-leaving reflectance $\rho_w(\lambda)$ at the five SeaWiFS wavelengths

Third Block (B3) comprises the five $Ra(\lambda)$,

Fourth Block (B4) comprises two variables: The in situ and the SeaWiFS *chl-a* concentrations.

At the end of the calibration phase, each element $\boldsymbol{z_i}$ of the dataset DPIG is associated with a referent $\boldsymbol{w_k}$, whose components are partitioned into four blocks, which is the closest $\boldsymbol{w_k} \in \boldsymbol{W}$ in term of the weighted assignment function used in 2S-SOM. In the present study, the 2S-SOM map is represented by a two-dimensional (9x18=162) grid that represents the partition of the DPIG dataset into different classes. Each class provided by the 2S-SOM is associated with a so-called referent vector $\boldsymbol{w_k}$ with $k \in \{1.....162\}$. The size of the map has been heuristically determined.

3.4 The Pigment retrieval

In the second, phase which is an operating phase, we estimate the pigment concentration ratios of a pixel P_j from its satellite ocean-color sensor observations only. The 11 ocean color satellite observations (5 $\rho_w(\lambda)$, 5 $Ra(\lambda)$, and *chl-a*) of pixel P_j are projected onto the 2S-SOM using the Truncated Euclidian Distance (section 3.1). We select the neuron k associated with a referent vector whose the 11 ocean-color parameters are the closest to those observed by the satellite sensor. The pigment ratios of P_j are those associated with the neuron k. At the end of the assignment phase, each pixel k of a satellite image is associated with a referent vector $\boldsymbol{w_k}$, which has 6 pigment concentration ratios among its 17 components. The flowcharts of the method (2S-SOM learning and pigment retrieval) are presented in Figure 4.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Ocean Science
Discussions
Open Access

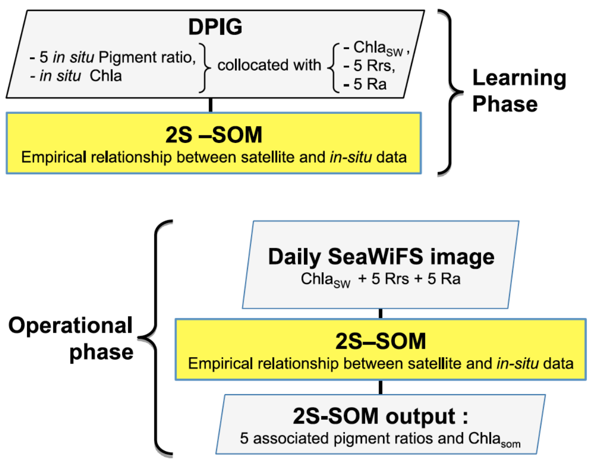

Figure 4 : *Flowchart of the method: top panel - Learning phase; bottom panel – operational phase and pigment retrieval.*

4 - METHODOLOGICAL RESULTS

4-1 Statistical validation of the method

The validation of the method was focused on the retrieval of the fucoxanthin ratio, which is a characteristic of diatoms, but the same procedure could be applied to any pigment. The hyper-parameters μ and η were optimized in order to retrieve that ratio. Due to the small amount of data in the DPIG, we estimated the accuracy of the fucoxanthin retrieval by a cross-validation procedure, which is a powerful procedure in statistics. The principle is the following: we iterated the procedure for one realization of the learning phase 30 times. The procedure was the following:

i=1 …. 30

1. determination at random of a learning dataset L_i (90% of DPIG) and a test dataset T_i (10% of DPIG)
2. training of a 2S-SOM map M_i using L_i (see section 3.2 and 3.3).
3. Validation using T_i according to the procedure described in section 3.4
4. Estimation of the RMSE$_i$ and $R^2{}_i$ on T_i between the estimated and observed fucoxanthin ratios

end

Computation of the mean RMSE and R^2 $\left(R^2, \text{RMSE} = \frac{1}{30}\sum_{i=1}^{I=30} R^2 i, \text{RMSE } i\right)$

The flowchart of the cross-validation procedure is presented in Figure 5.

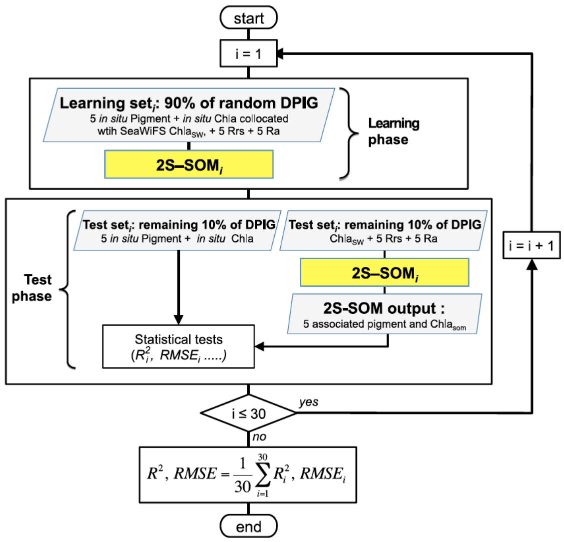

Figure 5 : *Flowchart of the cross-validation procedure for 30 partitions of the DPIG database.*

Statistical parameters (R^2 coefficients, RMSE and P-values) of the cross validation between the DPIG in situ pigments and the pigments given by the 2S-SOM averaged for the 30 2S-SOM realizations, which are presented in table 2, show the good performance of the method.

	R^2	RMSE [MG M^{-3}]	PVAL
CHLA SOM	0.84	0.22	0.001
DVCHLA	0.60	0.02	0.001
FUCO	0.87	0.02	0.001
PERID	0.81	0.01	0.001

Table 2 : *Statistical parameters (R2 coefficients, RMSE and P-values) of the cross validation between the DPIG in situ pigments and the pigments given by the 2S-SOM averaged for the 30 2S-SOM realizations*

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Ocean Science
Discussions
Open Access

4-2 Analysis of the topology of the 2S-SOM

As explained in sections 3-2 and 3-3, the referent vector components ($\mathbf{w}_k \in R^{17}$), which are estimated during the learning phase, are partitioned in four blocks B1, B2, B3 and B4. The hyper parameters μ and η are tuned in order to favor the accuracy of the retrieval of the fucoxanthin ratio. We recall that all the pigment ratios are estimated during the calibration phase, but in the present paper attention was focused on the fucoxanthin ratio when selecting the two hyper parameters. In Figure 6, we

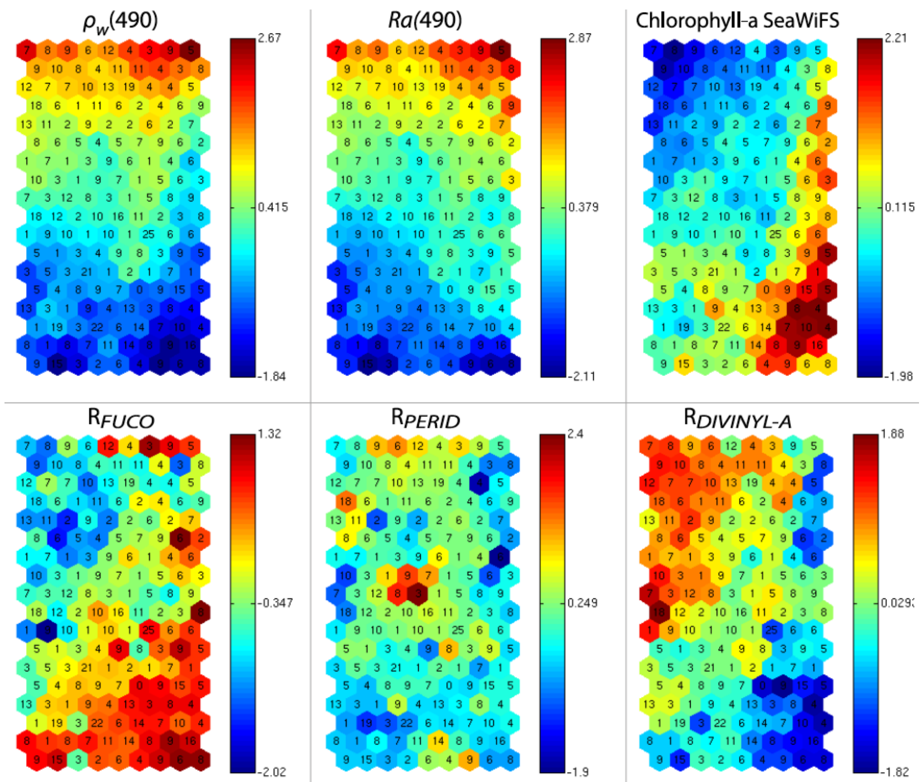

Figure 6 : *2S-SOM Map. From left to right and top to bottom, values of the referent vectors for $\rho_w(490)$, Ra(490), SeaWiFS chl-a, fucoxanthin, divinyl, peridinin. The number in each neuron indicates the amount of DPIG data captured at the end of the learning phase, the values indicated by the color bars are centered-reduced and non-dimensional values.*

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

present six of the referent vector components of the 2S-SOM map. These components are $\rho_w(490)$, *Ra(490)*, SeaWiFS *chl-a*, and the ratios of fucoxanthin, which is a specific diatom pigment. They exhibit a coherent topological order, the components having close values being close together on the topological map. The remaining eleven components (not shown) exhibit the same coherent topological order. One can observe a very good topological order for the fucoxanthin ratio that was favored by the determination of the hyperparameters μ and η. Moreover, the region in the 2S-SOM that characterizes the diatoms (high fucoxantin ratio and high *chl-a*) seems to be well evidenced. Another important remark is that the value of each component presents a large range of variation of the same order as the range of variation found in the DPIG variables. It means that the SOM map has captured most of the variability of the dataset.

Figure 6 shows a strong link between the values of the referent vectors for fucoxanthin and *chl-a* (high fucoxanthin and *chl-a* values, at the bottom right of the 2S-SOM) while fucoxanthin is high and *chl-a* low for the referent vectors at the bottom left of the 2S-SOM. Additional information will be provided by the *Ra(490)* values when the fucoxanthin is less closely linked to the chlorophyll.

Besides, for each neuron, the 2S-SOM provides a weight for each block (α_{cb}) and each variable (β_{cbj}). For a given neuron the weights (α_{cb}) of the blocks are normalized, their sum being 1. A value of 1 for one block (and therefore a value of 0 for the other blocks) indicates that the data in the neuron are gathered with respect to that block only because there is too much noise in the variables in the other blocks. By examining the weights on the map one can see which block most influences the link between the satellite measurements and the pigment ratios.

In Figure 7, we present the α_{cb} values estimated during the learning phase of the 4 blocks (B1, B2, B3, B4). For some area on the map, only the blocks related to the reflectance and the reflectance ratio are used for the definition of the neuron, while the weights for the two other blocks (pigments and *chl-a*) are null, indicating that for these neurons, in situ observations and SeaWiFS *chl-a* are more noisy. For these cases, the clustering assembled the data that are similar with respect to the reflectance only.

5 - GEOPHYSICAL RESULTS

In the present study, we apply the 2S-SOM (section 3), which explicitly makes a weighted use of the data according to their specificity (ocean-color signals or in situ observations) to retrieve the fucoxanthin concentration from remote sensed data in the Senegalo-Mauritanian upwelling region

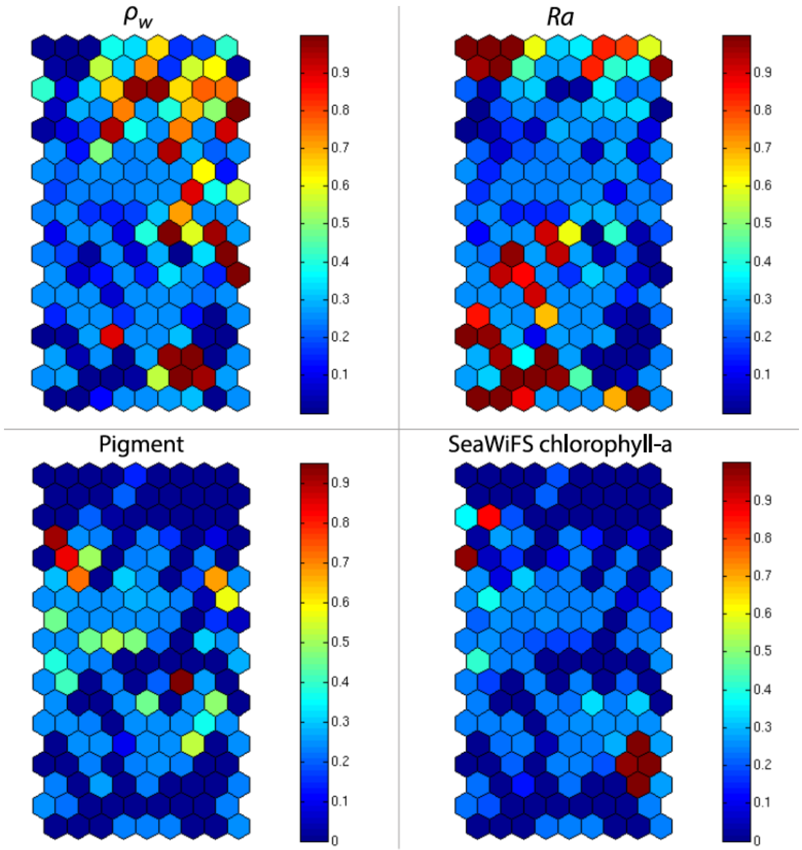

Figure 7: *2S-SOM map. Weights of the four block parameters (α_{cb}) determined at the end of the learning phase; from left to right and top to bottom: ρ_w, Ra, Pigment, SeaWifs chl-a. The color bars show the % of the weight estimated by 2S-SOM, a value of 1 or 0 indicating that the data in the neuron are assembled with respect to that block only.*

where in situ measurements are lacking. According to the good results of the cross validation method as shown in section 4.1, we expect that the 2S-SOM will provide pertinent results in a region which has been poorly surveyed.

5-1 The pigment estimation from SeaWiFS observations in the Sénégalo-Mauritanian upwelling region

We decoded the DSAT database (section 2-3) using the 2S-SOM for 11 years (1998-2009) of SeaWiFS data observed in the Senegalo-Mauritanian upwelling region (8°N-24°N, 14°W-20°W). This study was done according to the retrieval phase described in section 3.4. For each day, we projected the 11 SeaWiFS observations (5 $\rho_w(\lambda)$, 5 $Ra(\lambda)$ and *chl-a*) of each pixel P_j on the 2S-SOM. At the end of the assignment phase, each pixel of a satellite image was associated with 6 pigment concentration ratios. The underlying assumption is that the link between the remote sensing information and the pigment ratios of a pixel is this provided by the selected referent $\boldsymbol{w_k}$. Thanks to the topological order provided by the 2S-SOM, we expect that the best neurons chosen during the retrieval would give accurate concentration ratios. In Figures 8, 10 and 11 we present the fucoxanthin concentration ratio restitution for three different days and the associated SeaWiFS Chlorophyll images (1 and 6 January and 28 February 2003). Due to the limited size of the DPIG, the

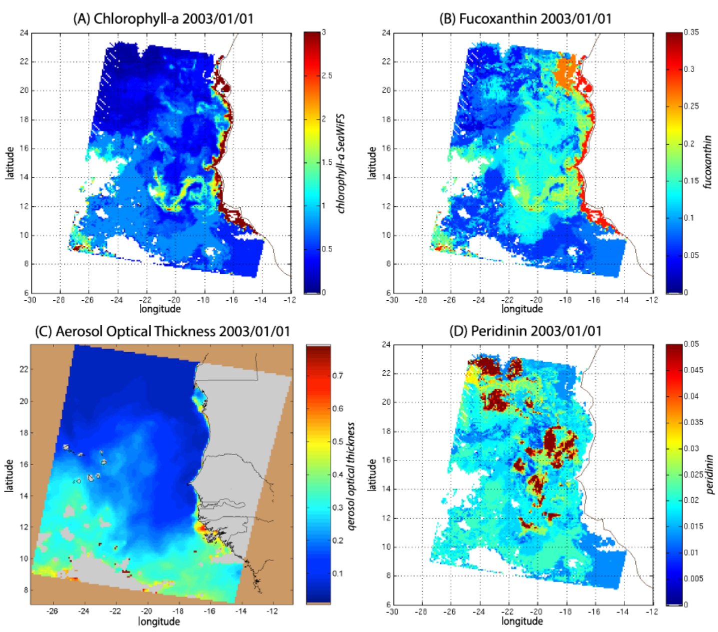

Figure 8 : *A) chl-a concentration, (B) fucoxanthin ratio, (C) aerosol optical thickness, (D) peridinin for 1 January 2003. Panels (B) and (D) show that a second-order information was retrieved, which is correlated with the chl-a concentration (A) but not equivalent. The aerosol optical thickness (C) does not seem to contaminate the estimated parameters (fucoxanthin and peridinin ratios).*

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

range of the ratio learned for the fucoxanthin is between 0.3% and 20% with a mean of 10% and the *chl-a* content is between 0.5 mg m^{-3} and 3mg m^{-3}. The statistical estimator we used cannot

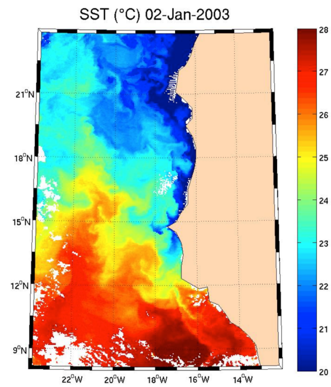

Figure 9 : *SST for 2 January 2003. Note the well-marked upwelling (cold temperature) north of 13°N.*

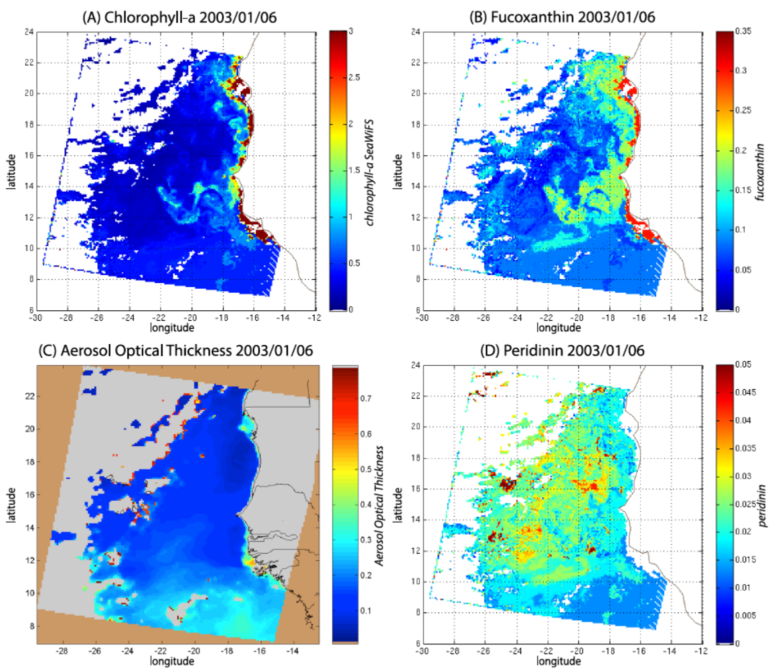

Figure 10 : *(A) chl-a concentration, (B) fucoxanthin ratio, (C) aerosol optical thickness, (D) peridinin for 6 January 2003. Panels (B) and (D) show that a second-order information was retrieved, which is correlated with the chl-a concentration (A) but is not equivalent. It is found that the aerosol optical thickness (C) does not contaminate the estimated parameters (fucoxanthin and peridinin ratios).*

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

extrapolate what has not been learned and for that raison we flagged the pixels in the SeaWiFS images that have a *chl-a* concentration greater than 3mg m^{-3}.

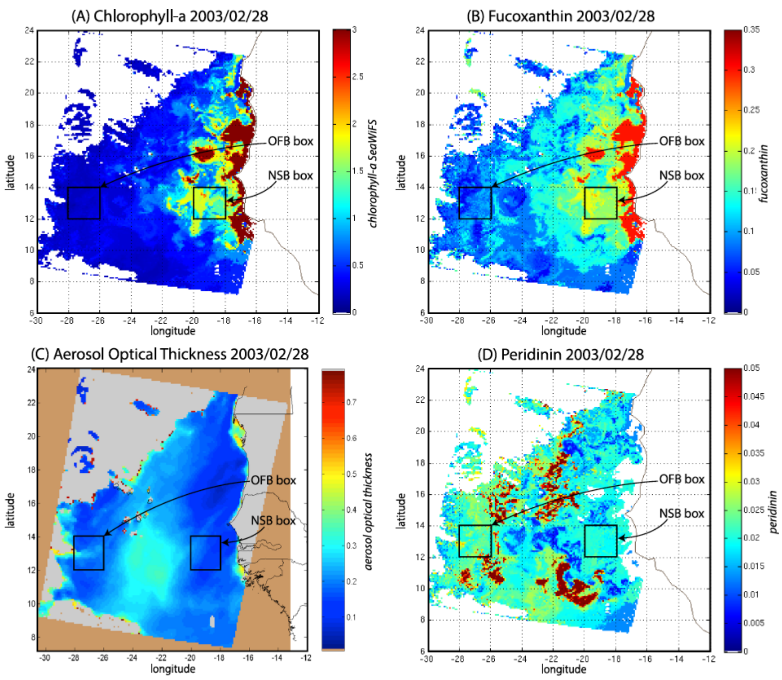

Figure 11 : *(A) chl-a concentration, (B) fucoxanthin ratio, (C) aerosol optical thickness, (D) Peridinin for 28 February 2003. Panels (B) and (D) show that a second order information was retrieved, which is correlated with the chl-a concentration (A) but is not equivalent. It is found that the aerosol optical thickness (C) does not contaminate the estimated parameters (fucoxanthin and peridinin ratios).*

Regarding the images obtained for 1 January 2003 in the Senegalo-Mauritanian region (Fig 8A, B, C, D), we observe that the *chl-a* (Fig 8A) is very high at the coast and decreases offshore in accordance with the upwelling intensity as shown in the SST image (Fig 9). Moreover, we observed a persistent well-marked *chl-a* pattern south of the Cap Vert peninsula in form of a “W“. Except in the southern part of the region, the AOT (Aerosol Optical Thickness) is low, which means that the atmospheric correction of the reflectance is quite small, which gives confidence in the ocean-color data products. The fucoxanthin concentration is maximum at the coast and decreases offshore as does the *chl-a* concentration, in agreement with the works of *Uitz et al.,* (2006, 2010). Fucoxanthin presents coherent spatial patterns. Peridinin concentration is somewhat complementary

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

to that of fucoxanthin, with the low fucoxanthin concentration area corresponding to high peridinin concentration area (northern part of Figs 8B, D). This behavior is also observed in Figure 10 (6 January 2003) and in Figure 11 (28 February, 2003) endorsing the analysis shown in Figure 8.

For 28 February, we selected two square box regions (Fig. 11), one near the coast (NSB, long [-20°, -18°], lat [12°,14°]) and the other about 800 km offshore (OFB, long [-28°, -26°], lat [12°,14°]). NSB waters correspond to upwelling waters while OFB waters correspond to oligotrophic waters. We projected the eleven ocean color parameters of the NSB and OFB pixels on the 2S-SOM map.

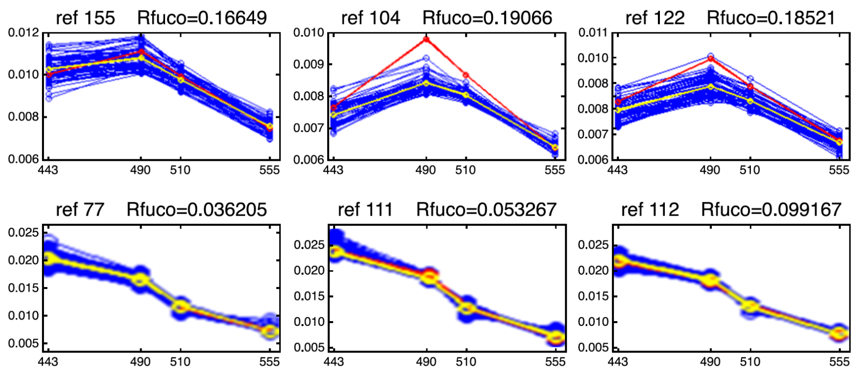

Figure 12 : *Reflectance spectra (in blue) of six referent vectors (red) selected during the decoding of 28 February: top line, in the NSB region (long. [-20°, -18°], lat. [12°, 14°]); bottom line, in the OFB region (long. [-28°, -26°], lat. [12°, 14°]). The reflectance spectra of the captured pixels are in blue and those of the w in yellow.*

Figure 12 presents the reflectance spectra (in blue) captured by three neurons of the 2S-SOM corresponding to pixels located in the NSB region (*top line*) and those captured by three neurons corresponding to pixels located in the OFB region (*bottom line*). The reflectance spectra of the associated referent vectors ***w*** are in yellow. The satellite reflectance spectra match the referent vector spectra; moreover the fucoxanthin ratio varies inversely with the mean value of the spectrum: the higher the fucoxanthin ratio, the smaller the mean value of the spectrum. The pigment concentration is greater near the coast.

We note a strong difference between the shape and the intensity of the near-shore (NSB) and offshore (OFB) spectra. The OFB spectra present mean values higher than those of the NSB spectra.

This is due to the fact that NSB spectra were observed in a region where diatoms are abundant, as shown by the high value of fucoxanthin concentration in this region (Figs 8, 10, and 11), which is a proxy for diatoms along with higher *chl-a* concentration. In Figure 12, we note the lower values of the coastal spectra at 443 nm, which can be interpreted as a predominant effect of spectral absorption by phytoplankton pigments and CDOM.

The different spectra are close together in the OFB region and more disperse in the NSB region. This can be explained by the fact that the OFB region corresponds to Case-1 waters while the NSB region waters are close to Case-2 waters and are influenced by the variability of near shore process like turbidity, or presence of dissolved matters.

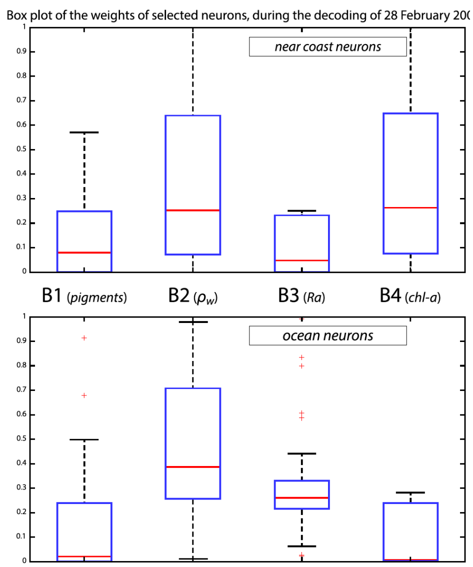

Figure 13 *Box plot of the weights of the selected neurons during the decoding of the 28 February data. From left to right, weights of blocks B1, B2, B3, B4. Top panel, in the OFB region (long. [-20°, -18°], lat. [12°, 14°]); bottom panel, in the NSB region (long. [-28°, -26°], lat. [12°, 14°]).*

We analyzed the weights of the blocks for all the neurons selected in the analysis of the costal (NSB) and offshore (OFB) boxes. Figure 13 presents the box plot of the weight a_{cb} corresponding to the neurons belonging to the four blocks (B1, B2, B3, B4), with the constrain that the sum of the weights of a neuron is 1; a weight larger than 0.25 indicates the predominance of a block in the learning for the classification (see section 3.5). It is clear that the weights for pixels near the coast (Fig 13, top

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

panel) are different from those for offshore pixels (Fig. 13, bottom panel). As already mentioned in section 4.3 and also shown in Figure 7, the weights of the 2S-SOM play a significant role in the 2S-SOM topology and consequently in the pigment retrieval. The weights of blocks B1 and B4 that take into account the influence of the pigment ratios and the chlorophyll content in the retrieval are very low for the offshore (OFB) oligotrohic region and more important for the coastal (NSB) region. The weights of the blocks B2 and B3, which take into account the influence of the reflectances ($\rho_w(\lambda)$, $Ra(\lambda)$), dominate for the offshore regions. In coastal waters the weights of all the blocks are used, with a smaller influence of B3, which is associated with R_a. This shows the automatic adaptation of the 2S-SOM to the environment in order to optimize the clustering efficiency with respect to a classical SOM.

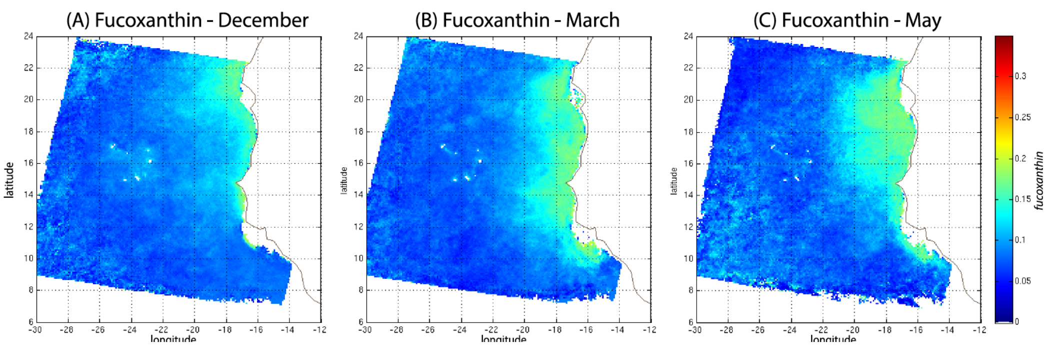

Figure 14 : *Monthly fucoxanthin concentration averaged for an 11- years (1998-2009) for December (A), March (B) and May (C).*

In order to study the seasonal variability of the fucoxanthin concentration with some statistical confidence in the Senegalo-Mauritanian upwelling region, we constructed a monthly climatology for an 11-year period (1998–2009) of the SeaWiFS observations by summing the daily pixels of the month under study. The resulting climatology is presented in Figure 14 for December (Fig. 14a), March (Fig. 14b), and May (Fig. 14c). The fucoxanthin concentration, and consequently the associated diatoms, presents a well-marked seasonality. Fucoxanthin starts to develop in December North of 19°N, presents its maximum intensity in March when the upwelling intensity is maximum, extends up to the coast of Guinea (12°N) in April and begins to decrease in May where it is observed north of Cabo Verde peninsula (15°N) in agreement with the observations reported by *Farikou et al*, (2015) and *Demarcq and Faure*, (2000).

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Ocean Science
Discussions

Figure 15 shows the fucoxanthin (in green) and the *chl-a* (in blue) concentrations computed from satellite observations for an 11-year period of SeaWiFS observations in the NSB region. There is a good correlation in phase between these two variables but not in amplitude (a good coincidence of peak occurrence but weak correlation in peak amplitude) showing that the relationship between

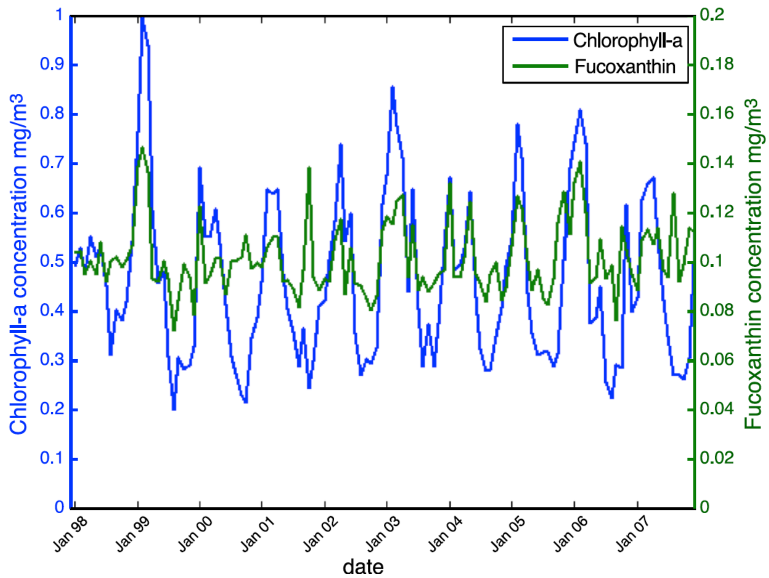

Figure 15 : . *chl-a (in blue) and fucoxanthin (in green) concentrations for near-shore pixels (in the NSB region).*

fucoxanthin and *chl-a* is complex as mentioned by *Uitz et al*, (2006). In particular, there is a weak peak in fucoxanthin in October 2001, which is not correlated with a *chl-a* peak.

5-2 Analysis of the UPSEN campaigns

Figure 16 shows, for the UPSEN stations 1, 2, 3, 5a and 5b (see figure 1 for their geographical position), the averaged in-situ spectrum (in blue), the 2S-SOM spectrum (in red) which is the spectrum of the 2S-SOM neuron captured by the collocated satellite VIIRS sensor observations, the spectrum (in black) of the learning database (DPIG) captured by that neuron that is the closest to the in-situ spectrum. These three spectra are close together showing the good functioning of the 2S-SOM.

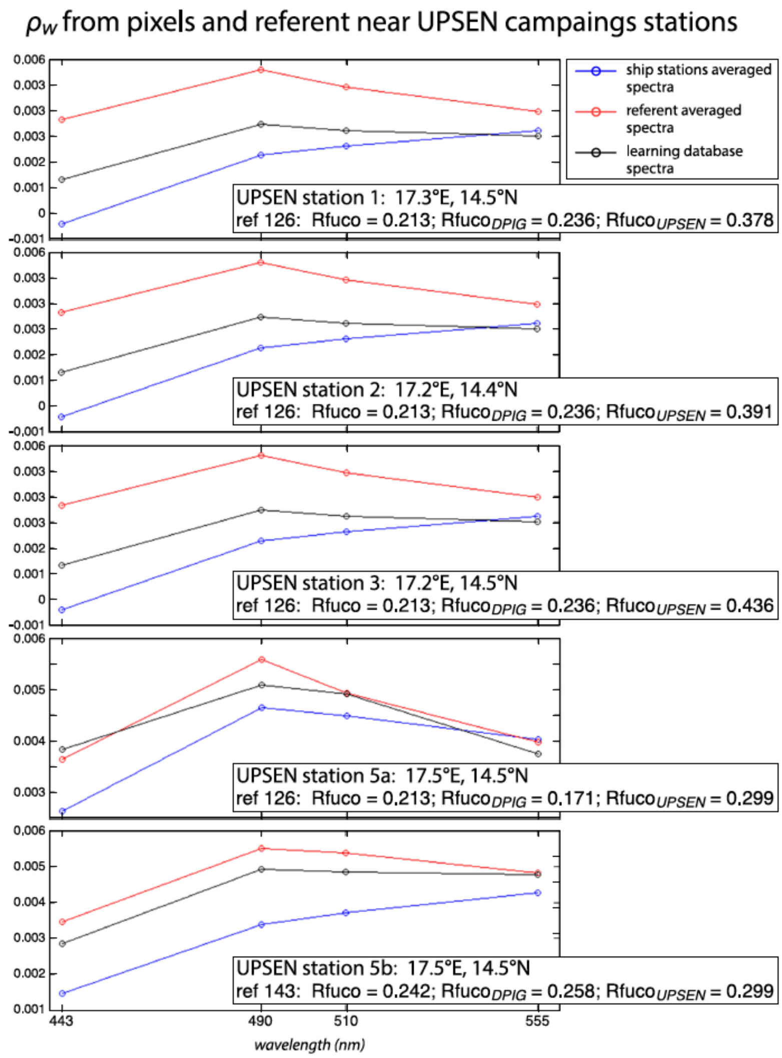

Figure 16 : *For ship stations 1, 2, 3, 5a and 5b, the averaged spectrum of the in situ spectra of the UPSEN station is shown in blue; the spectrum of the referent vector of the 2S-SOM neuron, which has captured the satellite observations which are the closest to the UPSEN station is shown in red; the spectrum of the learning database (DGIP) captured by the neuron that is the closest to the averaged satellite spectra is shown in black.*

Their shapes are close to these observed in the NSB region (Figure 12) but their intensity is lower meaning that their waters are more absorbing than the NSB waters due to a higher pigment

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

concentration. In fact, the UPSEN stations were located near the coast (figure 1) in the Hann bight south off the Cap Verde peninsula, which is very rich in phytoplankton pigments. In table 3, we present the fucoxanthin ratios associated with the referent vectors (Rfuco$_{2S\text{-}SOM}$), the closest DPIG fucoxanthin-ratios captured by the neuron of the referents and the fucoxanthin-ratios measured during the UPSEN campaign. We note that the fucoxanthin ratios of the in-situ measurements are in the range of the DPIG (see table 1), which allows a good functioning of the 2S-SOM. The pigment ratios obtained from ocean-color observations through 2S-SOM are close to pigment concentrations measured at the ship stations, which confirms the validity of the method we have developed. We remark that the best 2S-SOM estimate of fucoxanthin ratio with respect to the UPSEN in-situ measurement is given at station 5b which is the farthest off the coast. These results endorse the climatological study of the Senegalo-Mauritanian upwelling region we have done with the 2S-SOM (section 5.1).

UPSEN Station		Referent N°	Rfuco *2S-SOM*	Rfuco *DPIG*	Rfuco *UPSEN*
Stat 1	17.3 E 14.5 N	126	0.213	0.236	0.378
Stat 2	17.2 E 14.4 N	126	0.213	0.236	0.391
Stat 3	17.2 E 14.5 N	126	0.213	0.236	0.436
Stat 5a	17.5 E 14.5 N	126	0.213	0.171	0.299
Stat 5b	17.5 E 14.5 N	143	0.242	0.258	0.295

Table 3 : *For ship stations 1, 2, 3, 5a and 5b of the UPSEN campaigns, we show the referent captured by the VIIRS observations, the fucoxanthin-ratio associated with this referent (Rfuco-2S-SOM) the fucoxanthin-ratio of the closest DPIG fucoxanthin-ratio captured by the neuron of the referent and the fucoxanthin-ratio measured in situ during the UPSEN campaign*

The 2S-SOM method gives pigment concentrations that are close to those obtained by in situ observations. It could be applied to a large variety of other parameters in the context of studying and managing the planet Earth. The major constraint to obtaining accurate results would be to deal with a learning data set that statistically reflects all the situations encountered in the observations processed. Due to its construction, the method cannot be used to find values beyond the range of the learning data set.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

6 - DISCUSSION

Machine learning methods are powerful methods to invert satellite signals as soon as we have adequate database to support the calibration. Several technics have been used for retrieving biological information from ocean color satellite observations. First people employed MLPs, which are a class of neural networks suitable to model transfer function (*Thiria* et al, 1993). *Gross* et al, (2000, 2004) retrieved *chl-a* concentration from SeaWiFS, *Bricaud* et al, (2006) modeled the absorption spectrum with MLPs, *Raitsos* et al, 2008 and *Palacz* et al, 2013 introduced additional environmental variables in their MLPs such as SST in the retrieval of PSC/PFT from SeaWiFS, which improved the skill of the inversion. Another suitable procedure was to embed NN in a variational inversion, which is very efficient way when a direct model exists (*Jamet* et al, 2005; *Brajard* et al, 2006a,b; *Badran* et al, 2008). Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations were conducted by *Chazottes* et al, (2006, 2007), by using a SOM.

In the present study, due to the fact that the learning dataset we used is quite small (515 elements) which makes MLPs and classical supervised learning methods unusable, we decided to use an unsupervised neural network classification method which is an extension of the SOM, method well adapted to deal with small database whose elements are very inhomogeneous; we cluster available satellite ocean-color reflectance at five wavelengths and their derived products, such as chlorophyll concentration, and the associated in situ pigment ratios.

The major points of this study are the following:

- The clustering was carried out by developing a new neural classifier, the so-called 2S-SOM, which presents several advantages with respect to the classical SOM. As in the SOM, we defined clusters that assemble vectors, which are close together in terms of a specified distance. This classifier was learned from a worldwide database (DPIG) whose vectors are ocean-color parameters observed by satellite multi-spectral sensors and associated pigment concentrations measured in situ. In the operational phase, SeaWiFS images are decoded allowing the estimation of the pigment concentration ratios. The major advantage of 2S-SOM with respect to the classical SOM is to cluster variables having similar physical significance in blocks having specific weights. The weights attributed to the four blocks are computed during the learning phase and vary with the quality of the variables and with respect to the location on the ocean (near the coast or in deep ocean). This permits to modulate the variable influence in the cost function, which makes the clustering more informative than this provided by the SOM. For offshore waters, the block decomposition allowed us to show that more influence is given to the reflectance ratios $Ra(\lambda)$ and

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

less to the *chl-a* and pigment concentrations; on the contrary near the coast the weights indicate a more active use of the pigment composition and the *chl-a* concentration. The resulting 2S-SOM clustering therefore takes into account at best the information that belongs to the specific water content.

- The 2S-SOM decomposes the DPIG into a large number of significant ocean-color classes allowing reproduction of the different possible situations encountered in the dataset we analyze. Besides, we assume that the relationship between the pigment concentration and the remote sensed ocean-color observations is independent of the location, which is justifiable since the relationship depends on the optical properties of ocean waters through well-defined physical laws which are region-independent. This also endorses the fact that we use a global database to retrieve pigments in a definite region. On the contrary the different phytoplankton species vary from one region to another making the relationship between pigment ratio and phytoplankton species strongly depending on the region, which justifies the fact we focused our study on the pigment retrieval rather than the PSC or PFT as mentioned above. Moreover, most of the recent phytoplankton in situ identifications have been made using pigment measurements with the HPLC method (*Hirata et al*, 2011). It is therefore more natural to retrieve the pigment concentrations, which is the quantity we measured, than the associated PSC or PFT, which are estimated from the pigment observations through complex non-linear and region-dependent algorithms (*Uitz et al*, 2006). Due to the characteristics of the DPIG, the method has the ability to retrieve pigment concentration patterns over a large range (0.02 – 2 mg m^{-3}).
- We were able to analyze the pigment concentration in the Senegalo-Mauritanian region by processing satellite ocean color observations with the 2S-SOM. We found an important seasonal signal of fucoxanthin concentration with a maximum occurring in March. We evidenced a large offshore gradient of fucoxanthin concentrations, the near shore waters being richer than the offshore ones. We showed that the offshore region waters correspond to Case-1 waters while the near shore waters are close to Case-2 waters and are influenced by the variability of near shore process like turbidity, or presence of dissolved matters. The UPSEN measurements show that the pigment ratios of the Senegalo-Mauritanian region are in the range of the DPIG database used to calibrate the method, which justifies the use of the 2S-SOM algorithm to investigate this region.
- We used daily satellite observations to construct a monthly climatology of pigment concentrations of the Senegalo-Mauritanian upwelling region, which has been poorly surveyed by oceanic cruises. Due to the highly non-linear character of the algorithms for determining the pigment concentrations from satellite measurements, it is more rigorous mathematically to apply these algorithms to daily satellite data and to average this daily estimate for the climatology period under study than to

estimate them from the satellite data climatology as many authors have done (*Uitz et al., 2010*; *Hirata et al.,* 2011). We found that Fucoxanthin starts to develop in December North of 19°N, presents its maximum intensity in March when the upwelling intensity is maximum, extends up to the coast of Guinea (12°N) in April and begins to decrease in May

Another important aspect of our study concerns the validity of our results. The 2S-SOM method has been validated by focusing the retrieval accuracy on the fucoxanthin ratio by using a cross-validation procedure. These results were qualitatively confirmed by two other independent studies.

- We first applied a cross validation procedure (see section 4.1), which is powerful technique for validating models (*Kohavi,* 1995; *Varma* and *Simon*, 2006). We learned 30 different 2S-SOM using 30 different learning dataset determined at random from the DPIG dataset (each learning dataset representing 90% of DPIG) and 30 test datasets (10% of DPIG). By averaging the results, we found that the 2S-SOM method retrieves the fucoxanthin concentration with a good score (see the statistical parameters in table 2) which confirms the pertinence of the method.
- We then found that our fucoxanthin climatology is in agreement with in situ observations of phytoplankton reported in *Blasco et al*. (1980) in March to May 1974 off the coast of Senegal during the JOINT I experiment. These authors analyzed 740 water samples collected with Niskin bottles at 136 stations extending along a line at 21°40'N (in the northern part of the studied region) from 0 to 100 km offshore. The samples were taken at several depths (mostly at 100, 50, 30, 15, 5 m). Phytoplankton cells were counted and identified by the Utermohl inverted microscope technique (*Blasco,* 1977). They found that diatoms reach their maximum concentration in April–May and are the most abundant group in that period, whereas the other cells predominate in March. Similar microscope observations have been reported in the ocean area south of Dakar by *A. Dia* (1985) during several ship surveys in February–March 1982–1983.
- Our method is also in agreement with the monthly eleven years climatology presented in *Farikou et al,* (2015) who used a modified PHYSAT method to retrieve the *PFT* in the Senegalo-Mauritanian region.
- The pigment concentrations provided by the 2S-SOM from the VIIRS sensor observations are in qualitative agreement with the in-situ measurements done at five stations during the two UPSEN campaigns in 2012 and 2013, showing that the method is able to function in waters where the pigment concentrations are quite high (fucoxanthin ratios of the order 0.4).

7 - CONCLUSION

We developed a new neural network clustering method, the so-called 2S-SOM algorithm to retrieve phytoplankton pigment concentration from satellite ocean color multi spectral sensors. The 2S-SOM algorithm is a SOM specifically designed to deal with a large number of heterogeneous components such as optical and chemical measurements. The major advantage of 2S-SOM with respect to the classical SOM is to cluster variables having similar significance in blocks having specific weights. The weights attributed to the blocks during the learning phase vary with the quality of the variables. This permits to modulate the variable influence in the cost function, which makes the clustering more informative than this provided by the SOM. Moreover, the 2S-SOM method is efficient and rapid as soon as the calibration is done since it uses elementary algebraic operations only. The 2S-SOM method is like a piecewise regression that takes advantage of the unsupervised classification of the SOM. We decomposed the DPIG database into a quite large number of partitions (9x8=162) when comparing our study to other studies (*Uitz et al*, 2006, 2012). The validity of the method has been controlled through a cross validation procedure and confirmed by three qualitative studies. Statistical parameters (R^2 coefficients, RMSE and P-values) of the cross-validation between the DPIG in situ pigments and the pigments given by the 2S-SOM averaged for the 30 2S-SOM realizations which are presented in table 2, show the good performance of the method. It must be noticed that the performance mainly depends on the size of the learning set used to calibrate the 2S-SOM. This set must include all the situations encountered in the pigment retrieval. The larger the learning set, the better the method performs. Due to its generic character and its flexibility, the method could be used to determine a large variety of variables measure with satellite remote sensing observations.

The method was applied to study the seasonal variability of the fucoxanthin concentration in Senegalo-Mauritanian upwelling region. We showed a large offshore gradient of fucoxanthin, the higher concentration being situated near the shore. We were able to construct a monthly climatology for an 11-year period (1998–2009) of the SeaWiFS observations by summing the daily pixels of the month under study in a region which was poorly surveyed by oceanic cruises. The fucoxanthin concentration, and consequently the associated diatoms, presents a well-marked seasonality (Figure 10). Fucoxanthin starts to develop in December North of 19°N, presents its maximum intensity in March when the upwelling intensity is maximum, extends up to the coast of Guinea (12°N) in April and begins to decrease in May where it is observed north of Cabo Verde peninsula (15°N) in agreement with the observations reported by *Farikou et al,* (2015) and *Demarcq and Faure*, (2000).

The UPSEN campaign results endorse the validity of the study of the Senegalo-Mauritanian upwelling region done with the 2S-SOM.

Acknowledgments

The study was supported by the projects CNES-TOSCA 2013-2014 and 2014-2015. The water-leaving reflectances were obtained from the SeaWiFS daily reflectances, ρobsTOAw(λ), provided by the NASA/GSFC/DAAC observed at the top of the atmosphere (TOA) and processed with the SOM-NV algorithm (Diouf et al., 2013) from 1998 to 2010. They are available at the web site: http://poacc.locean-ipsl.upmc.fr/. The DPIG data base was kindly provided by Dr. S. Alvain. We thank Dr. Alban Lazar and Dr. E. Machu for providing in situ data measured during the UPSEN experiments and stimulating discussions for their interpretation. We also thank Ray Griffiths for editing the manuscript.

References

Alvain S, Moulin C., Dandonneau Y. and Breon F. M. : Remote sensing of phytoplankton groups in case-1 waters from global SeaWiFS imagery. Deep-Sea Res. Part1, vol **52** (11), pp 1989-2004, 2005.

Alvain S. Loisel H. and Dessailly D. : Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton group detection. Optics Express, vol **20** (2), 2012.

Antoine D., André J. M. , Morel A. : Oceanic primary production : Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. Global Biogeochem Cy. vol **10**, pp 57-69, 1996.

Badran F., Berrada M. , Brajard J., Crepon M. , Sorror C., Thiria S., Hermand J.P. , Meyer M., Perichon L., Asch M. : Inversion of satellite ocean colour imagery and geoacoustic characterization of seabed properties : Variational data inversion using a semi-automatic adjoint approach J. Marine Systems, vol 69, pp 126-136, 2008

Behrenfeld M. J., Boss E., Siegel D.A., Shea D.M. : Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem. Cy. vol 19**,** GB1006, doi:10.1029/2004GB002299, 2005

Behrenfeld M. J., and Falkowski P.G. : Photosynthetic rates derived from satellite base chlorophyll concentration. Limnol. Oceanogr, vol **42,** pp 1-20, 1997

Ben Mustapha Z. S., Alvain S. , Jamet C., Loisel H. and Desailly D. : Automatic water leaving radiance anomalies from global SeaWiFS imagery: application to the detection of phytoplankton groups in open waters. Remote Sens. Environ., vol 146, pp 97-112, 2014.

Blasco D. : Red tide in the upwelling region of Baja California. Limnol. Oceanogr. vol 22, pp 255-263, 1977

Blasco D., Estrada M. and Jones B. : Relationship between the phytoplankton distribution and composition and the hydrography in the northwest African upwelling region, near Cabo Corbeiro. Deep-Sea Res. , vol 27A, pp 799-821, 1980.

Bracher A., Bouman HA, Brewin RJW, Bricaud A, Brotas V, Ciotti AM, Clementson L, Devred E, Di Cicco A, Dutkiewicz S, Hardman-Mountford NJ, Hickman AE, Hieronymi M, Hirata T, Losa SN, Mouw CB, Organelli E, Raitsos DE, Uitz J, Vogt M and Wolanin A : Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development. Front. Mar. Sci. 4:55. doi: 10.3389/fmars.2017.00055, 2017

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Brajard J., Jamet C., Moulin C. and Thiria S. : Atmospheric correction and oceanic constituents retrieval with a neuro-variational method. Neural Networks, Vol 19(2), p178-185, 2006

Brajard J., Jamet C., Moulin C. and Thiria S : Neurovariational inversion of ocean color images. J. Atmos. Space Res. Vol 38, n 2, pp 2169-2175, 2006

Brewin R. J. W., Hardman-Mountford N. J., Lavender S. J., Raitsos D. E., Hirata T., Uitz J., et al. : An inter-comparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. Remote Sens. Environ. 115, 325–339. doi: 10.1016/j.rse.2010.09.004, 2011

Brewin R. J. W., Sathyendranath S., Hirata, T., Lavender, S.J., Barciela, R., Hardman-Montford, N.J : A three-component model of phytoplankton size class for the Atlantic Ocean. Ecol. Model. vol **22,** pp 1472-1483, 2010.

Bricaud A., Mejia C. , Blondeau Patissier D. , Claustre H., Crepon M. and Thiria S. : Retrieval of pigment concentrations and size structure of algal populations from absorption spectra using multilayered perceptrons. Applied Optics Mars 2007 vol 46 n°8., 2006

Capet X., Estrade, P., Machu, E., Ndoye, S. et al. : On the Dynamics of the Southern Senegal Upwelling Center: Observed Variability from Synoptic to Superinertial Scales : J. Phys. Oceanogr. vol **47** (1), pp 155-180, 2017

Cavazos T. : Using Self-Organizing Maps to Investigate Extreme Climate Events: An Application to Wintertime Precipitation in the Balkans. J. Climate, vol **13**, 1718–1732, 2000.

Chazotte A., Crepon M., Bricaud A., Ras J. and Thiria S. : Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method. Applied Optics, 46 (18), 3790-3799, 2007

Chazottes A., Bricaud A., Crepon M. and Thiria S. : Statistical analysis of a data base of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps. Appl. Opt. 45, 8102-8115, 2006

Ciotti A. and Bricaud A. : 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. Limnol. Oceangr. Methods, vol **4**, pp 237-253, 2006.

Demarcq H. and Faure V. : Coastal upwelling and associated retention indices from satellite SST. Application to Octopus vulgaris recruitment. Oceanografica Acta, vol **23**, pp 391-407, 2000.

Dia A. Biomasse et biologie du phytoplancton le long de la petite côte sénégalaise et relations avec l'hydrologie. Rapport interne N°44 du CRODT, Réf: 0C000798, 1981-1982. On line on the web site:http://www.sist.sn/gsdl/collect/publi/index/assoc/HASH2127.dir/doc.pdf

Diouf D., Niang A., Brajard J., Crepon M. and Thiria S. : Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method. Remote Sens. Environ. **vol 130***,* pp 74-86, 2013.

Farikou O., Sawadogo S., Niang A., Brajard J., Mejia C., Crépon M. and Thiria S. : Multivariate analysis of the Sénégalo-Mauritanian area by merging satellite remote sensing ocean color and SST observations. J. Environ. Earth Sci. vol **5** (12), pp 756-768, 2013

Farikou O., Sawadogo S., Niang A., Diouf D., Brajard J., Mejia C., Dandonneau Y., Gasc G., Crepon M., and Thiria S. : Inferring the seasonal evolution of phytoplankton groups in the Sénégalo-Mauritanian upwelling region from satellite ocean-color spectral measurements, J. Geophys. Res. Oceans, vol **120**, pp 6581-6601, 2015.

Gordon H. R. : Atmospheric correction of ocean color imagery in the Earth Observing System era. J. Geophys. Re. Atmospheres, vol **102**(D14), pp 17081-17106, 1997.

Hewitson B.C. and Crane R. G. : Sef organizing maps : application to synoptic climatology. Climate research, vol **22**, pp 13-26, 2002

Friedrich T. and Oschlies A. : Basin-scale pCO2 maps estimated from ARGO float data : A model study, J. Geophys. Res., vol **114**, C10012, doi: 10. 1029/2009JC005322, 2009.

Gross L., Frouin R., Dupouy C., Andre J. M. and Thiria S. : Reducing biological variability in the retrieval of chlorophyl_a concentration from spectral marine reflectance. Applied Optics, Vol. 43 Issue 20 pp. 4041, 2004

Gross L., Thiria S., Frouin R., Mitchell B.G : Artificial neural networks for modeling transfer function between marine reflectance and phytoplankton pigment concentration J. Geophys. Res. Vol 105,no.C2, pp3483-3949, february 15, 2000

Jamet C., Thiria S., Moullin C., Crepon M. : Use of a neural inversion for retrieving Oceanic and Atmospheric constituents for Ocean Color imagery : a feasability study. doi:10.1175/JTECH1688.1, J. Atmos. Ocean. Techno. :/ Vol. 22, No. 4, pp. 460–475, 2005

Hirata T. , Aiken J., Hardman-Mountford N., Smyth T. J. and Barlow R.G. : An absorption model to determine phytoplankton size classes from satellite ocean color, Remote Sens. Environ. vol **112,** pp 3153-3159, 2008.

Hirata T. , Hardman-Mountford N.J., Brewin R.J.W., Aiken J., Barlow R., Suzuki K., Isada T., Howell E., Hashioka T., Noguchi-Aita M. and Yamanaka Y. : Synoptic relationships between surface chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types. Biogeosciences, vol **8** (2): pp 311-327, 2011.

Ocean Sci. Discuss., https://doi.org/10.5194/os-2019-11
Manuscript under review for journal Ocean Sci.
Discussion started: 20 March 2019

Jeffreys S.W. and Vesk M. : Phytoplankton Pigment in Oceanography : Guidelines to Modern Methods, UNESCO, Paris, ed S. W. Jeffery, R.F.C. Mantoura and S. W. Wright, Introduction to marine phytoplankton and their pigment signatures, pp 33-84, 1997.

Jouini M., Lévy M. , Crépon M. and Thiria S. : Reconstruction of ocean color images under clouds using a neuronal classification method. Remote Sens. Environ. vol **131**, pp 232-246, 2013

Kohavi R. : A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann ed.. **2** (12): pp 1137–1143, 1995.

Kohonen T : Self-organizing maps (3rd ed.). Springer, Berlin Heidelberg New York. 2001

Kruizinga S. and Murphy A : Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. Mon. Wea. Rev., vol **111,** pp 2244–2254, 1983.

Liu Y. and Weisberg R. H. : Patterns of ocean current variability on the West Florida Shelf using the self-organizing map, J. Geophys. Res., **110,** C06003, doi:10.1029/2004JC002786, 2005

Liu Y., Weisberg R. H., and He R. : Sea surface temperature patterns on the West Florida Shelf using growing hierarchical self-organizing maps, J. Atmos. Oceanic Technol., vol **23**(2), pp 325– 338, 2006

Longhurst A. R., Sathyendranath S., Platt T., Caverhill C. : An estimation of global primary production in the ocean from satellite radiometer data. J. Plank. Res. vol **17**, pp 1245-1271, 1995

Lorenz E. N : Atmospheric predictability as revealed by naturally occurring analogs. J. Atmos. Sci., vol 26, pp 639–646, 1969

Morel A. and Gentili G. : Diffuse reflectance of oceanic waters. III. Implication of bidirectionality for the remote-sensing problem. Appl. Opt. vol 35, pp 4850-4862, 1996.

Mouw C. B. and Yoder J. A. : Optical determination of phytoplankton size composition from global SeaWiFS imagery. J. Geophys. Res. vol **115**, C12018, doi:10.1029/2010JC006337, 2010.

Ndoye S. , Capet X., Estrade P., Sow B., Dagorne D., Lazar A., Gaye A. and Brehmer P. : SST patterns and dynamics of the southern Senegal-Gambia upwelling center. J. Geophys. Res. Oceans, vol 119, pp 8315–8335. 2014

Niang A., Badran F., Moulin C., Crépon M. and Thiria S. : Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. Remote Sens. Environ. vol 100, pp 82-94, 2006.

Palacz A. P., St. John, M. A., Brewin, R. J.W., Hirata, T., and Gregg,W.W. : Distribution of phytoplankton functional types in high-nitrate low-chlorophyll waters in a new diagnostic ecological indicator model. Biogeosciences 10, 7553–7574. doi: 10.5194/bg-10-7553, 2013.

Raitsos D. E., Lavender, S. J., Maravelias, C. D., Haralambous, J., Richardson, A. J., and Reid, P. C. : Identifying phytoplankton functional groups from space: an ecological approach. Limnol. Oceanogr. 53, 605–613. doi: 10.4319/lo.2008.53.2.0605, 2008

Reusch D. B., Alley, R. B., and Hewitson, B. C : North Atlantic climate variability from a self-organizing map perspective, J. Geophys. Res., vol **112**, D02104, doi:10.1029/2006JD007460, 2007.

Sathyendranath S., Watts S., L., Devred E., Platt T., Caverhill C. M., and Maass H. : Discrimination of diatom from other phytoplankton using ocean-colour data, Mar. Ecol. Prog. Ser., vol 272, pp 59–68, 2004.

Uitz J., Claustre H., Morel A. and. Hooker S.B : Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res. **111,** C08005, doi:10:1029/2005JC003207. 2006

Uitz J., Claustre H., Gentili B. and Stramski D. : Phytoplankton class-specific primary production in the world's ocean: seasonal and interannual variability from satellite observations. Global Biogeochem. Cycles, vol **24**, GB 3016, doi:10:1029/2009GB003680, 2010

Van den Dool H. : Searching for analogs, how long must we wait? Tellus, vol **46A**, pp 314–324, 1994.

Varma, S., Simon, R. : Bias in error estimation when using cross-validation for model selection; BMC Bioinformatics. vol **7**. PMC 1397873. PMID 16504092. doi:10.1186/1471-2105-7-91, 2006

Vidussi F., Claustre H., Manca B. B., Luchetta A. and Marty J. C. : Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean sea during winter. J. Geophys. Res., vol 106, pp 19,939-19,956, 2001.

Westberry T., Behrenfeld M.J., Siegel D. A. and Boss E.: Carbon-based productivity modeling with vertically resolved photoacclimatation. Global Biogeochem. Cycles, vol **22**, *GB2024*, DOI:10.1029/2007GB003078, 2008

Zorita E. and Von Storch H. : The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. Journal of Climate, vol **12,** pp 2474-2489, 1999.