



SIMBA

Systeme Intégré de Modélisation de la
Baleine noire de l'Atlantique

(Integrated modelling system for the North Atlantic Right Whale)

Chlorophyll-a Algorithm Theoretical Based Document (ATBD)

Milestone #3


Date: March 31, 2022

Customer: CSA

Project no.: 9F040-190633/004



VERSION AND SIGNATURE

Prepared by	Version	Date	Modifications
Christopher Ilori	1.0	12/2021	Initial version
Christopher Ilori, Thomas Jaegler	2.0	02/2022	Revised version
Verification by			Modifications
Simon Bélanger	1.0	12/2021	Comments and suggestion
Christiane Dufresne	2.1	03/2022	Revision and edition
Carlos Araújo; Yanqun Pan	2.2	03/2022	Final revision
Final Authorization			Signature
Simon Bélanger	2.3	03/2022	



APPLICABLE AND REFERENCE DOCUMENTS

ID	Description	Reference
ATBD-Chla	Technical document: Chla	M3-T2.1.4
ATBD-Pheno	Technical document: phenology metrics	M3-T2.3.5
OCR-R	Ocean Colour Radiometry Processing Report	M3-T2.7.1
MS3-PR	Milestone#3 Progress Report	M3-PR



TABLE OF CONTENTS

VERSION AND SIGNATURE	I
APPLICABLE AND REFERENCE DOCUMENTS	II
TABLE OF CONTENTS.....	III
ACRONYMS	IV
LIST OF FIGURES	II
LIST OF TABLES	III
1. INTRODUCTION	4
1.1. AIM OF THE DOCUMENT	4
1.2. SCOPE OF THE DOCUMENT	5
1.3. RATIONALE	5
2. METHODOLOGY.....	7
2.1. DESCRIPTION OF DATA.....	7
2.2. PRINCIPAL COMPONENT ANALYSIS	10
2.3. APPLYING THE PCA ALGORITHM TO THE MATCHUP DATASET	10
2.4. IMPLEMENTATION OF THE MODEL ON SATELLITE IMAGES.....	13
3. PERFORMANCE EVALUATION	16
4. CONCLUSIONS	21
5. REFERENCES	22
APPENDIX A	23
APPENDIX B.....	25
APPENDIX C.....	27
APPENDIX D	28
APPENDIX E.....	30
APPENDIX F	32
APPENDIX G	34



ACRONYMS

AIC	Akaike's information criterion
CDOM	Colored Dissolved Organic Matter
Chla	Chlorophyll-a
GNATS	Gulf of Maine North Atlantic Time Series
GoSLM	Gulf of St. Lawrence and Maine
ESA	European Space Agency
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
OCR	Ocean Color Radiometry
PFT	Phytoplankton Functional Types
PMZA	Programme de Monitoring de la Zone Atlantique
R_{rs}	Remote sensing reflectance
SDM	Species Distribution Modeling
SEABASS	SeaWiFS Bio-optical Archive and Storage System
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
VIIRS	Visible Infrared Imaging Radiometer Suite

LIST OF FIGURES

Figure 1. Performance of the Chla OC5 algorithm (Gohin et al., 2002, 2011) in the North West Atlantic (left) and the GoSL (right).	6
Figure 2. Frequency distributions of Chla concentration. The green line indicates the curve of frequency distribution. The red line shows 25%, 50%, 75% quantiles.	8
Figure 3. Matchup datasets used to estimate chlorophyll-a concentration over the GoSLM.	9
Figure 4. Number of matchup datasets from all 4 sources.	9
Figure 5. Raw R_{rs} (top) and log linearized R_{rs} (bottom) of OLCI sensor used in the PCA-based chlorophyll retrieval. R_{rs} obtained from ACRI are fully normalized at their respective wavelengths. The solid line represents the mean and the two dashed lines above and below it represent mean R_{rs} - standard deviation and mean of R_{rs} + standard deviation, respectively.	11
Figure 6. Spectral shape of the first 4 principal components of OLCI sensor and their eigenvalues on y axis.	12
Figure 7. Comparison of the Chla retrievals from GSM, OCx, OC5 and the PCA-based algorithms. Different sensors are represented by shapes and different models are represented by colours.	17
Figure 8. Taylor diagram displaying a statistical comparison of estimated Chla from six sensors. The x and y axis show the SD of each sensor relative to Chla SD. The purple arc and dot on the x axis represent the SD and reference point, respectively. The green lines indicate the CRMSD relative to the Chla SD.	19
Figure 9. Chla concentration ($\text{mg}\cdot\text{m}^{-3}$) retrieved from PCA, OCx, GSM and OC5 algorithms from a MERIS data acquisition on 2003-08-20.	20



LIST OF TABLES

Table 1. Correlation matrix used for PCA analysis (number of observations x number of bands)	12
Table 2. Model selection for MERIS sensor by Stepwise AIC model.....	15
Table 3. Statistical metrics showing the performance of all algorithms.....	18
Table 4. Summary statistics for estimated Chla from the PCA-algorithm and measured Chla.....	18

1. INTRODUCTION

1.1. Aim of the document

This document is produced as part of the project entitled “SIMBA - Système Intégré de Modélisation de la Baleine noire de l’Atlantique (Integrated modelling system for the North Atlantic Right Whale)” which is within the portfolio of projects of the *smartWhales* initiative, supported by the Canadian space agency (CSA), Fisheries and Oceans Canada (DFO) and Transport Canada (TC). It describes the implementation of a regional algorithm to estimate chlorophyll-a concentration in the Gulf of St. Lawrence and Maine (GoSLM) using remote sensing reflectance (R_{rs} , sr^{-1}) products compiled within the framework of the GlobColour project (www.globcolour.info).

In SIMBA, Ocean Color Radiometry (OCR) is a major source of information for the species distribution model (SDM). The work package 2 (WP2) deals with OCR data processing including both archive and NRT. Briefly, we propose:

- 1) to develop a regional chlorophyll-a (Chla) algorithm for the GoSL, Scotian shelf and GoM (GoSLM);
- 2) to extract phytoplankton bloom metrics from Chla time series,
- 3) to compute phytoplankton primary production,
- 4) to identify Phytoplankton Functional Types (PFT; Xi *et al.*, 2020),
- 5) to explore novel presence/absence algorithm for direct detection of surface *Calanus* swarms.

This ATBD is about the very first tasks of the WP2, i.e. **Task 2.1 Regional Chlorophyll-a Algorithm Evaluation and Development.**

1.2. Scope of the document

In this document, we first provide the rationale for the development of a regional algorithm to retrieve chlorophyll-a (Chla) concentration from normalized R_{rs} (see section 2.1), for the GoSLM. Next, we describe the application of the principal component analyses (PCA) to the matchup data set (*i.e.* satellite-to-in situ data matching). The empirical algorithm proposed is based on the PCA and is used for estimating Chla concentration in the GoSLM. The approach is based on Fichot et al. (2008) and Laliberté et al. (2018). The R_{rs} products used in the processing framework include those derived from satellite imagery of the sensors MERIS, MODIS, OLCI (OLCIA and B), SeaWiFS, VIIRSJ and VIIRSN. The algorithm is uniquely designed for the GoSLM and is implemented in the Globcolor processor. The implementation steps and the equations are also presented.

1.3. Rationale

The presence of a high concentration of colored dissolved organic matter (CDOM) in the Estuary and Gulf of St. Lawrence (EGSL) has been shown to negatively affect the retrieval of the Chla concentration distributed by the Space Agencies (Bélanger *et al.*, 2017). Recently, Clay *et al.* (2019) evaluated the performance of two standard empirical Chla algorithms based on band-ratios, and the Garver-Siegel Maritorena semi-analytical algorithm (GSM, Maritorena *et al.*, 2002) against a large *in situ* data set from the Northwest Atlantic (NWA; Scotian shelf and Labrador Sea). Three ocean colour sensors were evaluated, *Sea-viewing Wide Field-of-view Sensor* (SeaWiFS), *Moderate Resolution Imaging Spectroradiometer* (MODIS) and *Visible Infrared Imaging Radiometer Suite* (VIIRS), and all presented large uncertainties with root mean square errors (RMSE) greater than 0.29 (\log_{10} mg chl·m⁻³) when compared to *in situ* Chla concentration data. Using only SeaWiFS, Laliberté *et al.* (2018) focused on the waters of the EGSL and also evaluated the blue-to-green ratio with the Garver-Siegel-Maritorena (GSM) algorithm, as well as the *Generalized Inherent Optical Property* semi-analytical algorithm (GIOP, Werdell *et al.* 2013). Consistently, they also found RMSE all greater than 0.29 (\log_{10} mg chl·m⁻³). Both Clay *et al.* (2019) and Laliberté *et al.* (2018) agreed that global algorithms performed poorly in the peculiar waters of the NWA and the EGSL.

A preliminary assessment of an operational GlobColour Chla product is shown in Figure 1. The performance of the OC5 Chla product (Gohin *et al.*, 2002, 2011), an empirical model designed for

coastal waters, is relatively good for the North Atlantic Region, but an overestimation of Chla is obvious in the Gulf of St. Lawrence.

As an alternative solution, Laliberté *et al.* (2018) proposed an algorithm for the SeaWiFS sensor specific to these prevailing environmental/optical conditions, using a PCA (also referred to as empirical orthogonal functions, or EOF) to capture the signal variations that can be attributed to phytoplankton. Although the published algorithm yielded satisfying results with the data originating from the EGSL regions, it is flexible and can be adapted to any particular set of *in situ* data. Building upon these previous efforts, we propose to merge all *in situ* Chla data currently available in the EGSL and the Scotian shelf available from the Department of Fisheries and Oceans Canada, and combine it with *in situ* data from the Gulf of Maine collected from the Bigelow research institute. After match-up extraction, we applied the PCA and derived an empirical relationship to retrieved Chla for each sensor currently available in Globcolour. The method is presented in the next section.

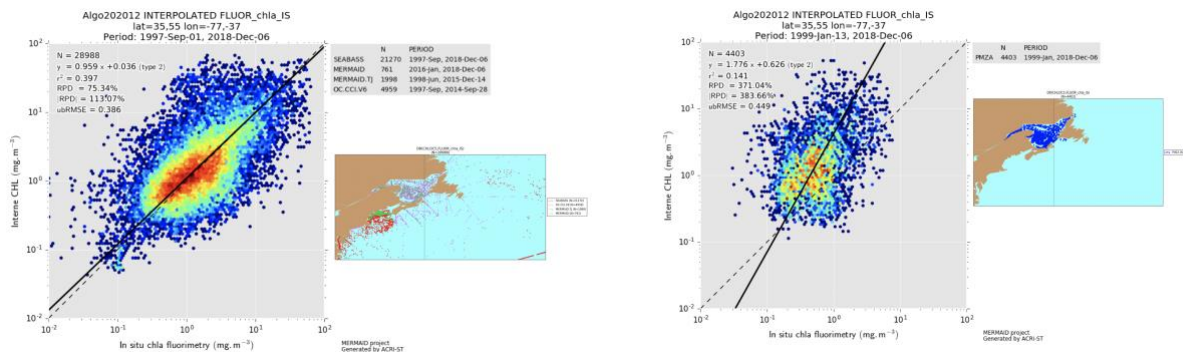


Figure 1. Performance of the Chla OC5 algorithm (Gohin et al., 2002, 2011) in the North West Atlantic (left) and the GoSL (right).

2. METHODOLOGY

2.1. Description of data

Remote sensing reflectance (R_{rs}) data were acquired from the Copernicus GlobColour processor (GC), an operational ocean colour radiometry (OCR) data processing chain that has been developed by ACRI-st (Sophia Antipolis, France). GC processor has been the main ocean colour data provider of the Copernicus Marine Service since 2015. The GC datasets are a continuous global dataset of L3 (binned and mapped) ocean colour products from different ocean colour sensors from 1997 to the present at 4 km spatial resolution. These datasets include normalized remote sensing reflectance products which were used in this work. The normalized remote sensing reflectance products were re-processed from Level 2 R_{rs} products of NASA and ESA.

In situ Chla data for the 1997 - 2021 period for the area bounded by 35°N to 55°N; 77°W to 37°W covering the entire GoSLM were obtained from Fisheries and Ocean Canada (DFO) and the Bigelow Laboratory. Figure 2 shows the frequency distribution of all in situ data for each sensor. All in situ Chla data were matched up with R_{rs} dataset from the GlobColour sensors used in this analysis. The final matchup datasets were obtained by excluding negative and zero R_{rs} values (including missing values) from the initial R_{rs} datasets. These negative values result from atmospheric correction failures and flagging by the atmospheric correction algorithm used by NASA and ESA. Figure 3 shows the matchup datasets (MERIS, MODIS, OLCI, SeaWiFS, VIIRSJ and VIIRSN) for each sensor used in this analysis. All datasets were obtained from 4 different sources/databases (Figure 4).

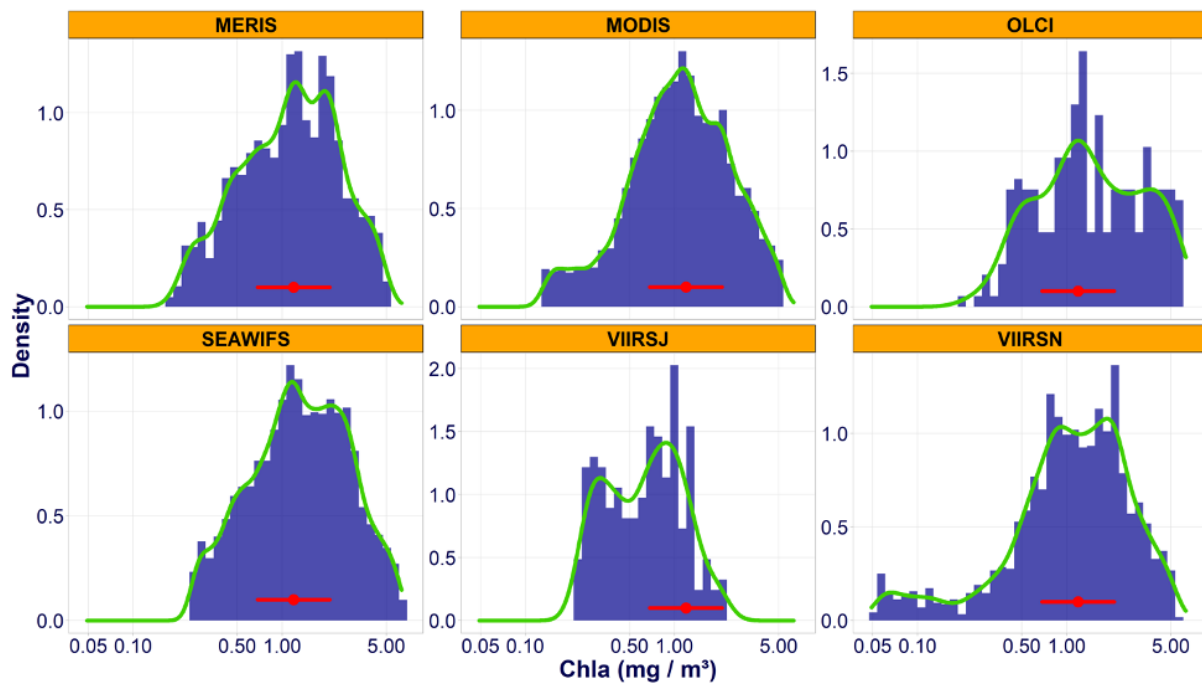


Figure 2. Frequency distributions of Chla concentration. The green line indicates the curve of frequency distribution. The red line shows 25%, 50%, 75% quantiles.

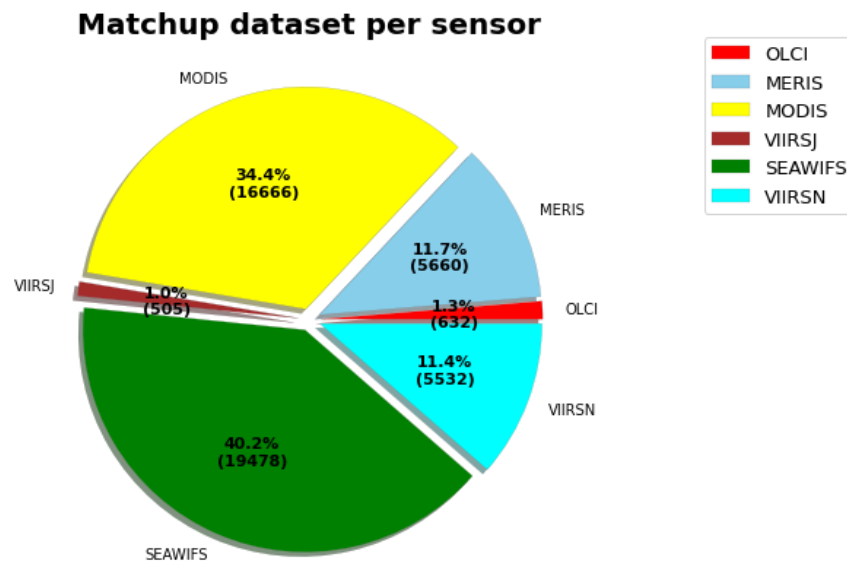


Figure 3. Matchup datasets used to estimate chlorophyll-a concentration over the GoSLM.

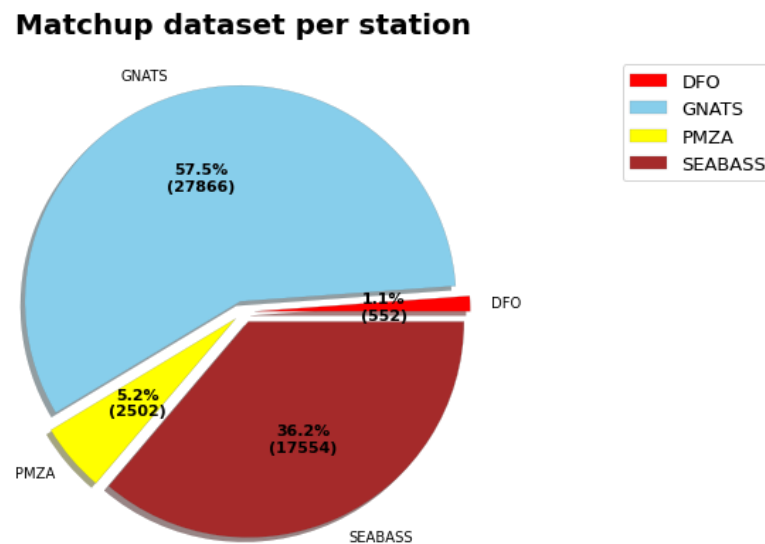


Figure 4. Number of matchup datasets from all 4 sources.

2.2. Principal Component Analysis

Principal Component Analysis (PCA), also known as empirical orthogonal functions (EOF), is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that contains most of the information in the original set. It uses orthogonal transformation to convert a set of correlated data into a set of values of linearly uncorrelated variables such that redundant information is removed from. PCA has been used in various remote sensing applications (e.g. Sathyendranath et al., 1994). PCA was applied on all the R_{rs} datasets used for training in this analysis to obtain a set of uncorrelated R_{rs} predictors called principal components (PCs). The PCs generated provide simpler data that describes the covariance structure of the original R_{rs} datasets.

2.3. Applying the PCA algorithm to the matchup dataset

Table 1 shows the matrix used for each sensor for the analysis described in this document. Basically, three major steps are needed to compute chlorophyll-a concentration from satellite imagery using the PCA algorithm: (1) standardization and log-transformation of R_{rs} datasets, (2) computation of the principal components using the eigenvectors, and (3) estimation of chlorophyll concentration from the computed principal components. Before running the PCA, all R_{rs} values in each wavelength were log-linearized to reduce skewness in the distribution of the R_{rs} datasets. This transformation is needed as the original R_{rs} datasets are not normally distributed, a requirement for the PCA algorithm. Figure 5 shows both the raw R_{rs} and log-linearized transformation of the R_{rs} spectra for OLCI sensor (we used natural log-transformation). The PCA is subsequently applied to the log linearized R_{rs} dataset. We used the *princomp* function in R to compute the eigenvalues and eigenvectors to transform the R_{rs} into PCs. Note that calculations are applied to the correlation matrix rather than the covariance matrix (Fichot et al., 2008) to avoid spectral dependency among R_{rs} spectra.

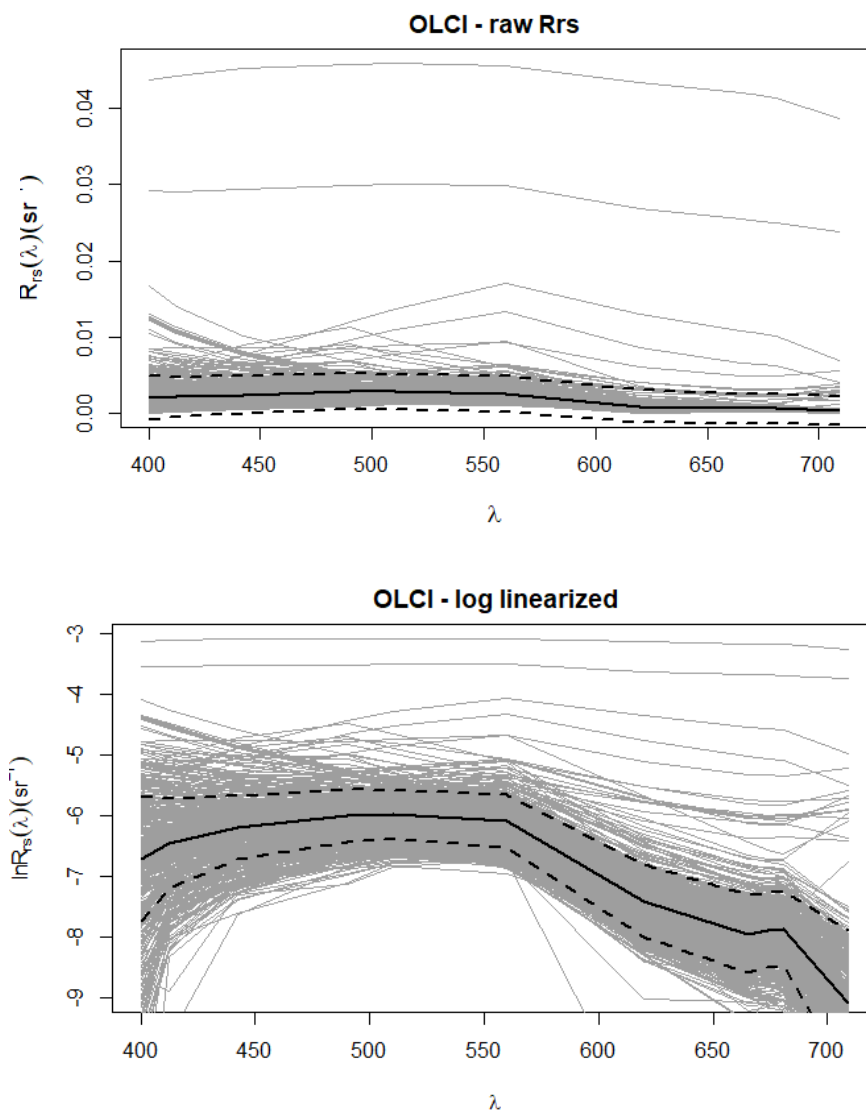


Figure 5. Raw R_{rs} (top) and log linearized R_{rs} (bottom) of OLCI sensor used in the PCA-based chlorophyll retrieval. R_{rs} obtained from ACRI are fully normalized at their respective wavelengths. The solid line represents the mean and the two dashed lines above and below it represent mean R_{rs} - standard deviation and mean of R_{rs} + standard deviation, respectively.

Table 1. Correlation matrix used for PCA analysis (number of observations x number of bands)

Sensor	Matrix
MERIS	5660 x 10
MODIS	16666 x 10
OLCI	632 x 11
SEAWIFS	19478 x 6
VIIRSJ	505 x 5
VIIRSN	5532 x 5

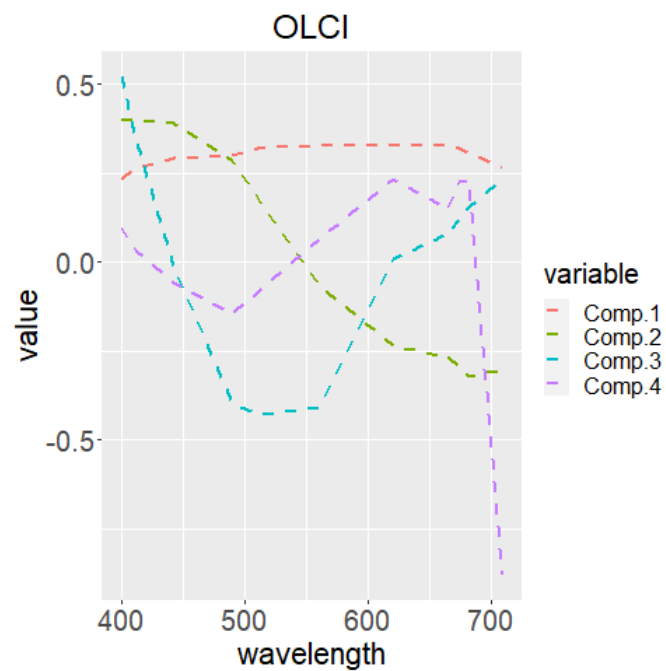


Figure 6. Spectral shape of the first 4 principal components of OLCI sensor and their eigenvalues on y axis.

2.4. Implementation of the model on satellite images

This section describes the implementation of the PCA algorithm on satellite images. The three main steps are described below. They are required to obtain the parameters needed for implementing the algorithm in the Globcolour processor. These parameters are generated in R programming and are saved as csv files. The CSV files were provided to ACRI for the implementation in the GlobColour processor.

The algorithm was implemented in the GlobColour processor as a standalone tool written in Python language. The tool's inputs are 1) a list of input GlobColour R_{rs} mapped products ("L3m") and 2) CSV tables root directory. The tools generate outputs following the GlobColour file format. The processing includes:

- auto-detection of the input R_{rs} products sensor
- selection and reading of the corresponding CSV tables
- filtering and reading of the input R_{rs} products depending on the CSV tables wavelengths list
- implementing the core algorithm
- writing the output product following the GlobColour file format, using R_{rs} input products as template.

The following Python packages are used:

- pandas to read and manipulate CSV tables
- NumPy for fast multi-dimensional arrays computing
- xarray for NetCDF reading/writing

Step 1: Standardizing the R_{rs}

The measured R_{rs} (R_{rsj} ; where j stands for each observation, or each pixel) must be first log-transformed and standardized. The natural log transformed R_{rs} is standardized using Equation 1:

$$X_j(\lambda) = \frac{\ln(R_{rsj})(\lambda) - \overline{\ln(R_{rs})(\lambda)}}{\sigma \ln(R_{rs})(\lambda)} \quad (\text{Equation 1})$$

where $\overline{\ln(R_{rs})}(\lambda)$ and $\sigma \ln(R_{rs})(\lambda)$ are the mean and standard deviation of the log linearized R_{rs} at each of the R_{rs} wavelength (λ), for each sensor in the matchup datasets, respectively. Mean and standard deviations for each sensor computed in this transformation are provided in Appendix E. These values, needed for implementation in the Globcolour processor, are saved in a csv file as *mean-sd_XXX.csv* (where 'XXX' is the sensor name).

Step 2: Computation of the principal components from the standardized R_{rs}

In the second step, the PCs scores are computed from $Xj(\lambda)$ using the so-called eigenvectors. The values of the eigenvectors are denoted as $\Delta_{i,k}$ with indices i , and k are the i th principal component number and k th wavelength, respectively. For example, in the case of OLCI sensor, 11 PC can be calculated from the 11 spectral bandwidths. To compute the first PC (PC1) for a given sensor, we used the first eigenvector, noted as $\Delta_{1,k}$

$$PC1 = \Delta_{1,1}X(400) + \Delta_{1,2}X(412) + \Delta_{1,3}X(442) + \dots + \Delta_{1,11}X(709) \quad (\text{Equation 2})$$

$$PC_i = \Delta_{i,1}X(\lambda_1) + \Delta_{i,2}X(\lambda_2) + \dots + \Delta_{i,k}X(\lambda_k) \quad (\text{Equation 3})$$

The eigenvectors values for each sensor are presented in Appendix B, where i , k are represented in columns and rows, respectively. Values of the eigenvector for each sensor are saved in csv files as *eigenvector_XXX.csv* (where 'XXX' is the sensor name).

Step 3: Predicting chlorophyll-a concentration from the principal components

We used Equation 4 to estimate chlorophyll concentration. Equation 4 is of the form:

$$\log_{10}[Chla] = a_0 + a_1[PC1] + a_2[PC2] + a_3[PC3] + \dots + a_m[PCm] \quad (\text{Equation 4})$$

where a_0 , a_1 , a_2 , a_3 and a_m are fitting parameters of the multiple linear regression and **PC1**, **PC2**, **PC3** and **PC m** are the **PC** numbers selected using a stepwise model selection procedure. By adding and removing modes, the stepwise model tries to find the optimal selection of variables. Table 2 shows a sample file of the parameters needed to retrieve $\log_{10}[\text{Chla}]$ from the **PC** scores for the MERIS sensor (values for other sensors are presented in Appendix F). These coefficients will be needed in the Globcolour processor and are saved in a csv file with file name `coef_XXX.csv` (where 'XXX' is the sensor name).

Table 2. Model selection for MERIS sensor by Stepwise AIC model.

PC coefficients	coefficients
0.04561588312	a_0
0.009524024588	a_1
-0.1392631551	a_2
0.04366867915	a_3
0.2367146972	a_4
-0.07521136779	a_5
0.2916770508	a_6
-0.2782882344	a_7

3. PERFORMANCE EVALUATION

To evaluate the performance of the PCA-based algorithm, we compared the estimated Chla from the PCA-based algorithm with those from selected state-of-the-art algorithms. A random stratification procedure was used on the datasets (using the *createDataPartition* function in R) to produce 50% Chla datasets to evaluate the performance of these algorithms. These algorithms include two traditional band-ratio algorithms (i.e., OCx and OC5) and the Garver-Siegel-Maritorena (GSM) method. Derived Chla products from these algorithms were available and downloaded from the GlobColor database during the R_{rs} matchup extraction step. The algorithm performance was compared using four metrics including the mean absolute error (MAE), the root mean square error (RMSE), Bias represented in log-transformed residuals, and the coefficient of determination (R^2). RMSE and MAE metrics are computed in log-transformed space to provide a better assessment of the algorithms owing to the log-normal distribution of Chla data. Figure 7 shows the graphical illustration of the summary statistics used for evaluation as presented in Table 3. Appendix G shows the residual plots of estimated Chla from all algorithms and the corresponding measured Chla values. Our analyses indicated that the PCA-based algorithm outperformed other existing algorithms. For all sensors it has the lowest RMSE, Bias and MAE compared to the other three algorithms (Figure 7, Table 3). As observed in Appendix G, the PCA-based algorithm has the smallest RMSE across all sensors when compared to in-situ measurements. Based on the mean RMSE of $0.25 \log_{10} \text{ mg chl} \cdot \text{m}^{-3}$ from all sensors, it can be concluded that the PCA-based algorithm was able to generate reasonable Chla products comparable with in situ measurements for the GoSLM as this value is similar to what was reported by Laliberte et al. (2018) for SeaWiFS (i.e., $0.22 \log_{10} \text{ mg chl} \cdot \text{m}^{-3}$). In fact, average RMSE for all sensors (except VIIRSN) is $0.224 \log_{10} \text{ mg chl} \cdot \text{m}^{-3}$, with the exception of VIIRSN where RMSE is $0.38 \log_{10} \text{ mg chl} \cdot \text{m}^{-3}$. It should be noted that this is the sensor with the highest RMSE for all sensors as well. To further show the accuracy of the PCA-based algorithm, we computed the summary statistics of the estimated and measured Chla and presented their result in Table 4. Results show that the percentage difference between the estimated and measured mean values of all sensors is within 14%, except for VIIRSN with a value of 28%. While the PCA algorithm performs well to retrieve the mean Chla, it tends to reduce the range of Chla compare to in situ observations (Appendix G). The algorithm systematically overestimates Chla in the lower range and underestimates Chla in the higher range.

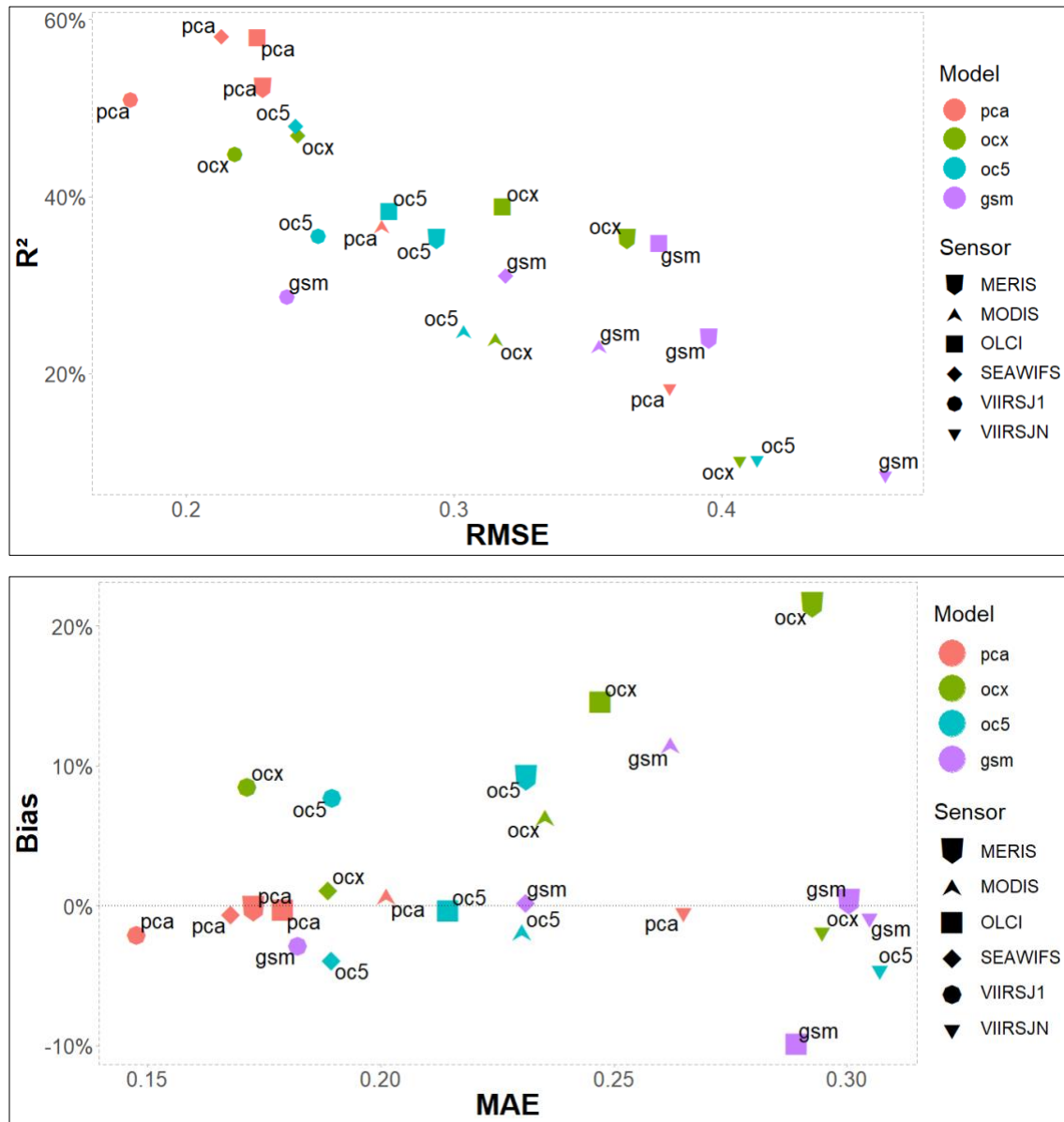


Figure 7. Comparison of the Chla retrievals from GSM, OCx, OC5 and the PCA-based algorithms. Different sensors are represented by shapes and different models are represented by colours.

Table 3. Statistical metrics showing the performance of all algorithms.

	RMSE (mg·m ⁻³)	BIAS	MAE	R2	N
MERIS					
PCA	0.23	-0.0015	0.17	0.34	2287
GSM	0.4	0.0034	0.3	0.28	
OCx	0.36	0.2160	0.29	0.30	
OC5	0.29	0.0924	0.23	0.34	
MODIS					
PCA	0.27	0.0073	0.2	0.40	7158
GSM	0.35	0.1155	0.26	0.33	
OCx	0.32	0.0638	0.24	0.22	
OC5	0.3	-0.0182	0.232	0.22	
OLCI					
PCA	0.23	-0.0020	0.18	0.57	270
GSM	0.38	-0.098	0.29	0.37	
OCx	0.32	0.147	0.25	0.26	
OC5	0.28	-0.0027	0.21	0.13	
SEAWIFS					
PCA	0.21	-0.0047	0.17	0.54	8262
GSM	0.32	0.0036	0.23	0.35	
OCx	0.24	0.0124	0.19	0.32	
OC5	0.24	-0.0375	0.19	0.37	
VIIRSJ					
PCA	0.18	-0.0193	0.15	0.42	228
GSM	0.24	-0.0269	0.18	0.41	
OCx	0.22	0.0865	0.17	0.33	
OC5	0.25	0.0787	0.19	0.16	
VIIRSN					
PCA	0.38	-0.0029	0.26	0.29	2139
GSM	0.46	-0.0068	0.30	0.23	
OCx	0.41	-0.0172	0.29	0.19	
OC5	0.11	-0.442	0.31	0.13	

Table 4. Summary statistics for estimated Chla from the PCA-algorithm and measured Chla.

	Measured Chla concentration				Estimated Chla concentration			
	Min	Mean	SD	Max	Min	Mean	SD	Max
OLCI	0.210	1.944	1.498	5.719	0.242	1.780	1.279	8.388
MERIS	0.181	1.1452	1.046	4.953	0.109	1.317	0.967	20.8177
MODIS	0.133	1.118	1.063	5.353	0.256	1.216	0.616	8.78
SEAWIFS	0.242	1.710	1.264	6.30	0.191	1.50	0.895	10.934
VIIRSJ	0.215	0.741	0.433	2.118	0.305	0.654	0.293	1.717
VIIRSN	0.049	1.453	1.109	5.388	0.181	1.100	0.477	4.156

To show the sensor with predictions that most closely match in-situ observations, we used Taylor Diagram (Taylor, 2001) (Figure 8) to depict three metrics: correlation coefficients (R), standard deviation (SD) and central root mean square difference ($CRMSD$). The Taylor diagram (Taylor, 2001) provides concise and graphical statistics of the amplitude of variations (given by the SD s), the correlation (R) and the $CRMSD$. The position of each point on the plot shows how close they are to the observation. In our case, if a sensor's point is close to the observed point (purple point on the x-axis), it means that it is similar to in situ Chla and has a very high correlation and $CRMSD$ close to zero. The relative closeness of each sensor can be inferred from Figure 8. All sensors differed by their amplitude of variations (SD), with value ranges between 0.29 and 0.97, while the R values were between 0.55 and 0.76. VIIRSJ has the highest variation, while MERIS and SeaWiFS have the lowest SD of 0.97 and 0.93, respectively, compared with SD of 1.05 for in situ Chla. It can thus be concluded that MERIS and SeaWiFS generally agree best with Chla observations, each with points closer to the purple arc.

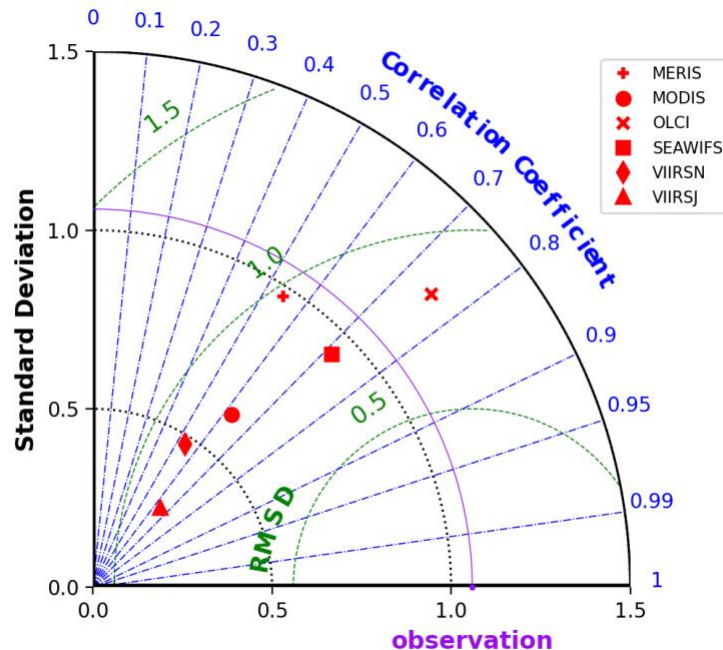


Figure 8. Taylor diagram displaying a statistical comparison of estimated Chla from six sensors. The x and y axis show the SD of each sensor relative to Chla SD . The purple arc and dot on the x axis represent the SD and reference point, respectively. The green lines indicate the $CRMSD$ relative to the Chla SD .

Regarding spatial distribution, the PCA algorithm presents similar patterns to those from OCx, GSM and OC5 algorithms (Figure 9). PCA unsurprisingly retrieves lower Chla concentrations than OCx, GSM and OC5, which were known to overestimate those values in the Gulf of St. Lawrence (Laliberté et al. 2018 and references therein). The number of flagged data is higher for PCA and GSM, especially in coastal waters around Prince Edward Island, which is also the location of highest Chla concentration retrieved by OCx and OC5.

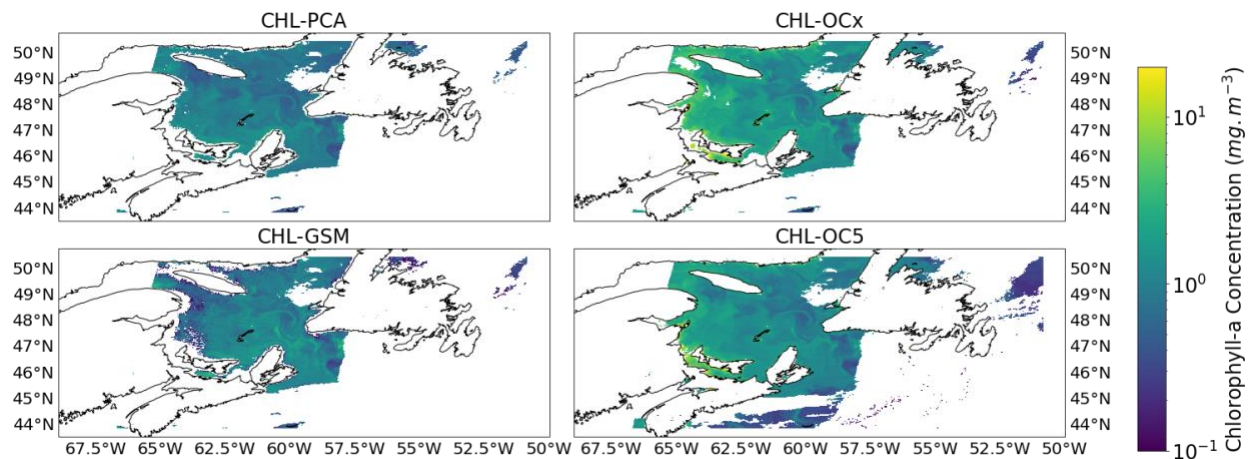


Figure 9. Chla concentration (mg·m⁻³) retrieved from PCA, OCx, GSM and OC5 algorithms from a MERIS data acquisition on 2003-08-20.

4. CONCLUSIONS

The main outcome of our work is a new regional algorithm for estimating Chla in surface waters of the GoSLM from a suite of sensors from the GlobColour database. The algorithm was developed on the basis of in situ data and satellite measurements in the study area. The in situ measurement included spectral values of the remote sensing reflectance R_{rs} and Chla concentrations. Testing this new algorithm with field data showed that it can reasonably estimate Chla concentration and produce better estimates than global algorithms. Developing regional algorithms using ocean colour products from the GlobColour database has improved our ability to represent the concentration of surface Chla in the GoSLM. As revealed in this work, existing algorithms such as GSM, OCx and OC5 produce less accurate estimates of Chla products over the GoSLM. To further improve estimates of Chla in this region, efforts could be tailored towards optimizing remote sensing reflectance. Atmospheric correction is a major source of uncertainty in remotely sensed products. The algorithm developed here has used ready-made remote sensing reflectance products as we did not set out to address the issues associated with atmospheric correction. The algorithm described here effectively scales remote sensed reflectance and uses the most distant bands (i.e., decomposing the remote sensing reflectance into Principal Components (PC), and discarding PCs with less variance associated). Through this procedure, variance due to inherent error in atmospheric correction algorithms used to process the remote sensing reflectance we used may have been accounted for, at least partially. The main limitation of the new algorithm is the reduction of the range of Chla. PCA algorithm is best to retrieve the mean Chla but it will reduce the spatial pattern by overestimating low Chla and underestimating high Chla. Machine learning based algorithm may be an alternative to improve the Chla retrievals, but preliminary tests made using neural network resulted in similar performance than the PCA. In the context of SIMBA, we expect, however, that seasonal variability in Chla will be more important than the absolute retrieval of Chla to predict the arrival of the NARW in the Gulf of St Lawrence.

5. REFERENCES

- Bélanger, S., Carrascal-Leal, C., Jaegler, T., Larouche, P., & Galbraith, P. (2017). Assessment of radiometric data from a buoy in the St. Lawrence estuary. *Journal of Atmospheric and Oceanic Technology*, 34(4), 877–896. <https://doi.org/10.1175/JTECH-D-16-0176.1>
- Clay, S., Peña, A., DeTracey, B., & Devred, E. (2019). Evaluation of satellite-based algorithms to retrieve chlorophyll-a concentration in the Canadian Atlantic and Pacific Oceans. *Remote Sensing*, 11(22), 1–29. <https://doi.org/10.3390/rs11222609>
- Gohin, F., Druon, J. N., & Lampert, L. (2002). A five channel chlorophyll concentration algorithm applied to Sea WiFS data processed by SeaDAS in coastal waters. *International Journal of Remote Sensing*, 23(8), 1639–1661. <https://doi.org/10.1080/01431160110071879>
- 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>
- Fichot C. G., Sathyendranath, and W. L. Miller. (2008) SeaUV and SeaUVC: Algorithms for the retrieval of UV/Visible diffuse attenuation coefficients from ocean color. *Remote Sensing of Environment*, 112:1584-1602.
- Laliberté, J., Larouche, P., Devred, E., & Craig, S. (2018). Chlorophyll-a concentration retrieval in the optically complex waters of the St. Lawrence Estuary and Gulf using principal component analysis. *Remote Sensing*, 10(2). <https://doi.org/10.3390/rs10020265>.
- Maritorena, S., Siegel, D.A., and Peterson, R. (2002), Optimal tuning of a semi-analytical model for global applications, *Applied Opt.*, 41, 2705–2714.
- Sathyendranath, S., Hoge, F.E., Platt, T. and Swift, R.N (1994). Detection of phytoplankton pigments from ocean color: improved algorithms. *Appl. Opt.* 33, 1081-1089.
- Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 2001, 106, 7182–7192
- Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., Liu, Y., d’Andon, O. H. F., & Bracher, A. (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. *Remote Sensing of Environment*, 240(February), 111704. <https://doi.org/10.1016/j.rse.2020.111704>
- Werdell, J.; Franz, B.; Bailey, S.; Feldman, G.; Boss, E.; Brando, V.; Dowell, M.; Hirata, T.; Lavender, S.; Lee, Z.; et al. Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl. Opt.* 2013, 52, 2019–2037

APPENDIX A

EIGENVALUES FOR ALL SENSORS

OLCI	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	7.574236922	68.85669929	68.85669929
Dim.2	2.18256158	19.84146891	88.6981682
Dim.3	0.76012165	6.910196818	95.60836502
Dim.4	0.284363266	2.585120597	98.19348562
Dim.5	0.088713979	0.806490722	98.99997634
Dim.6	0.055428316	0.503893782	99.50387012
Dim.7	0.032028144	0.291164944	99.79503506
Dim.8	0.011340352	0.103094109	99.89812917
Dim.9	0.006527122	0.05933747	99.95746664
Dim.10	0.002910956	0.026463232	99.98392988
Dim.11	0.001767714	0.016070124	100

MERIS	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.961680202	66.24089114	66.24089114
Dim.2	1.888474052	20.98304502	87.22393616
Dim.3	0.630772658	7.008585086	94.23252125
Dim.4	0.29874148	3.319349782	97.55187103
Dim.5	0.096968671	1.077429677	98.62930071
Dim.6	0.077115358	0.856837315	99.48613802
Dim.7	0.028019077	0.311323073	99.79746109
Dim.8	0.015605347	0.173392747	99.97085384
Dim.9	0.002623154	0.02914616	100

MODIS	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	6.268406959	62.68406959	62.68406959
Dim.2	2.593164713	25.93164713	88.61571673
Dim.3	0.754152575	7.541525747	96.15724247
Dim.4	0.164101003	1.64101003	97.7982525
Dim.5	0.147028929	1.470289294	99.2685418



Dim.6	0.029476872	0.294768718	99.56331052
Dim.7	0.026017739	0.260177389	99.8234879
Dim.8	0.014348978	0.143489782	99.96697769
Dim.9	0.003210118	0.032101183	99.99907887
Dim.10	9.21E-05	0.000921131	100

SEAWIFS	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.286177872	71.43629786	71.43629786
Dim.2	1.081016212	18.01693686	89.45323472
Dim.3	0.474137603	7.902293387	97.3555281
Dim.4	0.129382466	2.156374436	99.51190254
Dim.5	0.020131455	0.335524249	99.84742679
Dim.6	0.009154393	0.152573211	100

VIIRSJ	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.29970567	65.99411341	65.99411341
Dim.2	1.274008378	25.48016755	91.47428096
Dim.3	0.324801063	6.496021262	97.97030222
Dim.4	0.090135515	1.802710309	99.77301253
Dim.5	0.011349374	0.226987471	100

VIIRSN	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.247478443	64.94956886	64.94956886
Dim.2	1.136050449	22.72100898	87.67057784
Dim.3	0.499721034	9.994420686	97.66499853
Dim.4	0.092787431	1.855748617	99.52074715
Dim.5	0.023962643	0.479252854	100



APPENDIX B

EIGENVECTOR VALUES FOR PC CALCULATION FOR ALL SENSORS. ONLY SELECTED PCs NEEDED FOR THE CHL-A
ARE PROVIDED

OLCI	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
R_{rs} 400	0.2335199	0.399596	0.5205853	0.4920605	0.1146558	0.4800431	0.0252605	0.0019715
R_{rs} 412	0.2659855	0.4043262	0.3485679	-0.125768	0.0507184	-0.584168	-0.089269	-0.023415
R_{rs} 442	0.2922386	0.3896897	-0.008333	-0.404053	-0.175779	-0.17237	0.1194653	-0.020325
R_{rs} 490	0.3009116	0.287082	-0.397654	-0.191834	-0.036395	0.3249032	0.0322379	0.6038555
R_{rs} 510	0.3217358	0.1816161	-0.4276	-0.003462	0.0564055	0.211269	-0.133111	-0.758299
R_{rs} 560	0.3295269	-0.064154	-0.407196	0.5791066	0.3359998	-0.378623	0.0768905	0.1914186
R_{rs} 620	0.3300813	-0.234025	0.011055	0.2583615	-0.761035	-0.139063	-0.116251	0.0011463
R_{rs} 665	0.3296814	-0.264867	0.0802565	-0.156227	-0.179475	0.2592911	0.3895206	-0.003273
R_{rs} 674	0.3195498	-0.294579	0.1215071	-0.246558	0.2665205	0.1205842	-0.74873	0.122323
R_{rs} 681	0.3090842	-0.32217	0.1542736	-0.223186	0.3917424	-0.052002	0.4753223	-0.088081
R_{rs} 709	0.2671675	-0.299961	0.2356188	0.0697318	-0.010059	-0.035036	-0.026133	-0.001006

MERIS	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
R_{rs} 413	0.2558	0.4554	0.4513	0.5141	0.1614	0.2305	0.3312
R_{rs} 443	0.3076	0.4566	0.2083	-0.0194	-0.0429	-0.1442	-0.5163
R_{rs} 490	0.3391	0.3542	-0.2080	-0.3817	-0.1533	-0.1585	0.0189
R_{rs} 510	0.3705	0.2190	-0.3224	-0.2387	-0.2338	0.0504	0.1590
R_{rs} 560	0.3554	-0.1371	-0.5167	0.2386	0.0716	0.5387	0.1530
R_{rs} 620	0.3682	-0.2320	-0.2022	0.2436	0.5885	-0.2898	-0.4103
R_{rs} 665	0.3661	-0.2810	0.1580	-0.0212	0.0234	-0.5721	0.5702
R_{rs} 681	0.3131	-0.4061	0.2012	0.3138	-0.6925	0.0305	-0.2805
R_{rs} 709	0.3063	-0.3069	0.4820	-0.5635	0.2499	0.4426	-0.0505

MODIS	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
R_{rs} 412	0.2828637	0.3007179	0.6692531	0.3204058	0.4175691	0.2500977	0.2043264	0.0115102
R_{rs} 443	0.3461659	0.3353787	0.2382689	-0.075404	-0.218602	-0.409135	-0.674683	-0.183315
R_{rs} 469	0.3702968	0.2944086	-0.05715	-0.189315	-0.338065	-0.102574	0.3155419	0.7158031
R_{rs} 488	0.3809497	0.2474652	-0.208902	-0.163238	-0.278447	0.2033423	0.3861905	-0.624671
R_{rs} 531	0.4042252	-0.012494	-0.375019	-0.001116	0.3414833	0.0451885	-0.007004	-0.083031
R_{rs} 547	0.3897528	-0.118486	-0.396999	0.1225652	0.4676012	-0.032693	-0.224464	0.1785431



R_{rs} 645	0.2808979	-0.426424	0.009759	0.7323141	-0.441311	-0.06167	0.0399775	-0.011674
R_{rs} 667	0.2746644	-0.448076	0.2381297	-0.398833	-0.163891	0.617103	-0.300736	0.0991094
R_{rs} 678	0.2196939	-0.502364	0.3055711	-0.344663	0.170736	-0.574856	0.33447	-0.122246

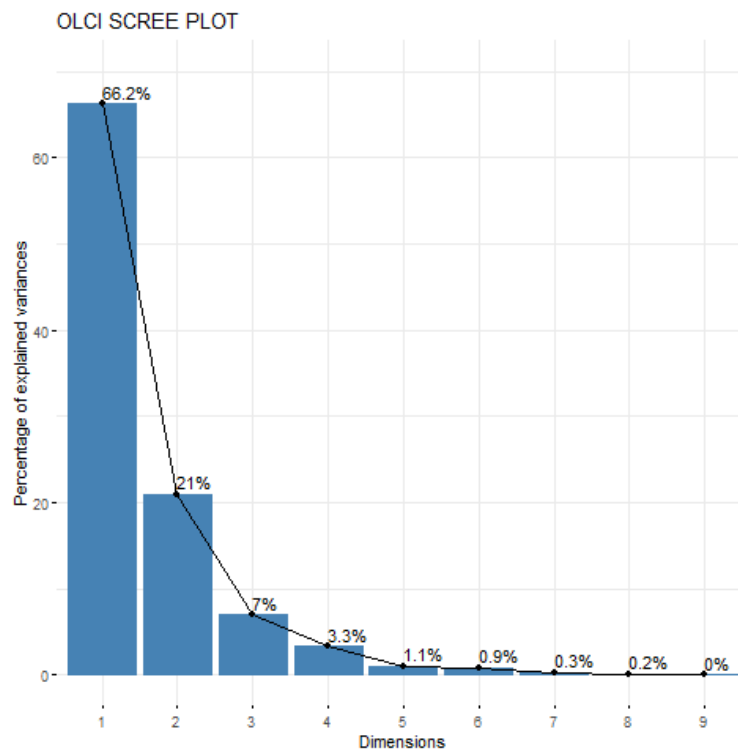
SEAWIFS	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
R_{rs} 412	0.3818	0.4107	0.5845	0.4669	0.3349	0.1164
R_{rs} 443	0.4420	0.3580	0.1086	-0.2205	-0.76647	-0.1688
R_{rs} 490	0.4547	0.1829	-0.3361	-0.3540	0.5137	-0.5074
R_{rs} 510	0.4656	-0.0452	-0.3562	-0.1373	0.0612	0.7947
R_{rs} 555	0.3906	-0.4785	-0.2945	0.6587	-0.1720	-0.2610
R_{rs} 670	0.2875	-0.6622	0.5655	-0.3940	0.0554	-0.0248

VIIRSJ	Comp.1	Comp.2	Comp.3	Comp.4
R_{rs} 411	0.5005	0.2170	0.6164	0.2002
R_{rs} 445	0.5418	0.1500	-0.3973	-0.7083
R_{rs} 489	0.5464	0.0277	-0.4848	0.6498
R_{rs} 556	0.3966	-0.5133	0.4335	-0.1805
R_{rs} 667	0.0017	-0.8162	-0.1982	0.0587

VIIRSN	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
R_{rs} 410	0.4492	0.2990	0.6627	0.5000	0.1402
R_{rs} 443	0.5178	0.2885	0.0407	-0.5195	-0.6141
R_{rs} 486	0.5210	0.1575	-0.3812	-0.2540	0.7028
R_{rs} 551	0.4544	-0.4020	-0.4800	0.5545	-0.3067
R_{rs} 671	0.2287	-0.8006	0.4283	-0.3289	0.1233

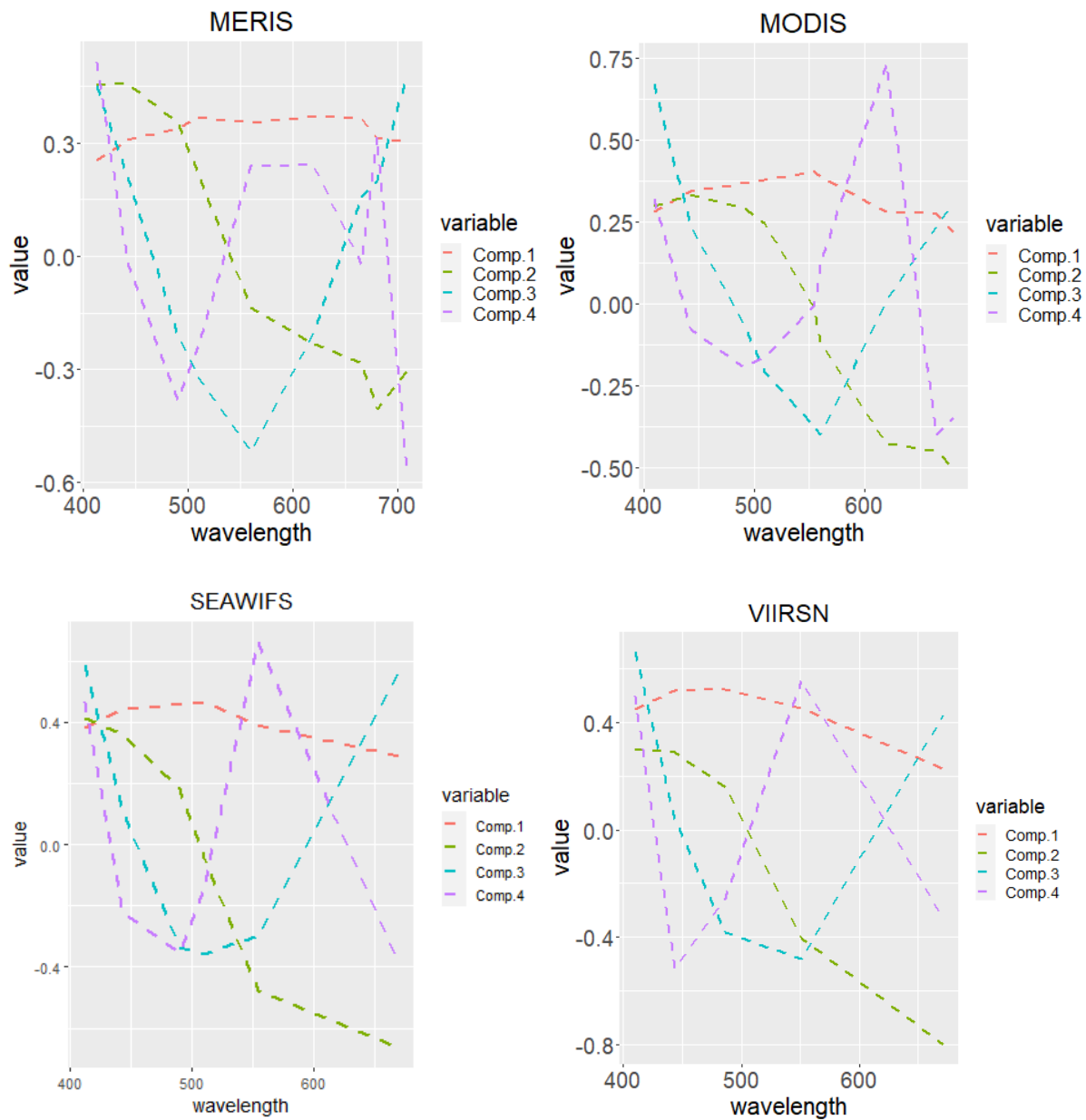
APPENDIX C

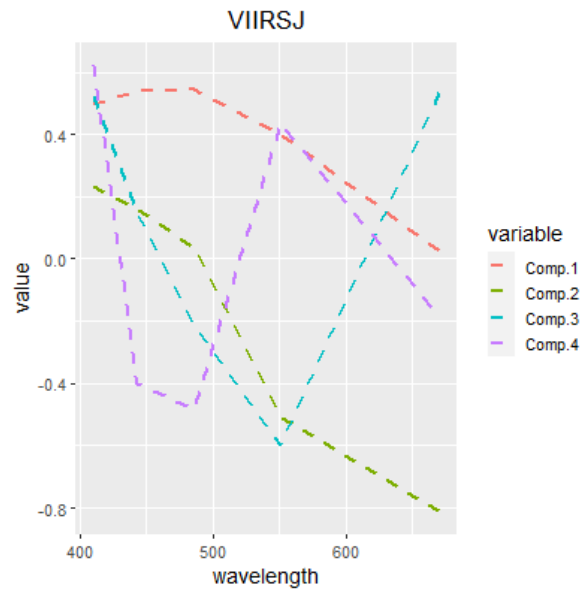
SCREE PLOTS SHOWING THE EIGENVALUES FOR EACH INDIVIDUAL PC FOR OLCI SENSOR.



APPENDIX D

SPECTRAL SHAPE OF THE FIRST 4 PRINCIPAL COMPONENTS OF ALL SENSORS





APPENDIX E

LOG LINEARIZED MEAN (MEAN LN) AND LOG LINEARIZED STANDARD DEVIATION (SD LN) FOR ALL SENSORS

OLCI	mean ln (R_{rs})	sd ln (R_{rs})
R_{rs} 400	-6.713117567	1.02996503
R_{rs} 412	-6.461058089	0.73793148
R_{rs} 442	-6.200315809	0.52612808
R_{rs} 490	-6.001424893	0.43776111
R_{rs} 510	-5.990909844	0.40363044
R_{rs} 560	-6.09662596	0.43262793
R_{rs} 620	-7.410506265	0.592356
R_{rs} 665	-7.957359006	0.64019099
R_{rs} 674	-7.891845052	0.62056109
R_{rs} 681	-7.866880001	0.6225399
R_{rs} 709	-9.094725947	1.19100511

MERIS	mean ln (R_{rs})	sd ln (R_{rs})
R_{rs} 413	-6.5371	0.7977
R_{rs} 443	-6.3558	0.5254
R_{rs} 490	-5.9848	0.3474
R_{rs} 510	-5.9887	0.3189
R_{rs} 560	-6.1217	0.3658
R_{rs} 620	-7.6191	0.5773
R_{rs} 665	-8.0147	0.5779
R_{rs} 681	-7.8862	0.5739
R_{rs} 709	-8.3839	0.6659

MODIS	mean ln (R_{rs})	sd ln (R_{rs})
R_{rs} 412	-6.10584688	0.5633145
R_{rs} 443	-6.031735026	0.3923848
R_{rs} 469	-5.916174137	0.3434811
R_{rs} 488	-5.905221228	0.3152241
R_{rs} 531	-5.944024487	0.2783798
R_{rs} 547	-6.025990874	0.2874204
R_{rs} 645	-7.769633754	0.5121937

R_{rs} 667	-7.997007536	0.5278671
R_{rs} 678	-7.867491638	0.5121048

SEAWIFS	mean $\ln(R_{rs})$	sd $\ln(R_{rs})$
R_{rs} 412	-6.29936624	0.6529859
R_{rs} 443	-6.103821431	0.4124012
R_{rs} 490	-5.931187044	0.3221132
R_{rs} 510	-5.949216925	0.2941422
R_{rs} 555	-6.079642209	0.3178203
R_{rs} 670	-8.119046823	0.7758429

VIIRSJ	mean $\ln(R_{rs})$	sd $\ln(R_{rs})$
R_{rs} 411	-6.585023184	0.9232096
R_{rs} 445	-5.937386898	0.3992793
R_{rs} 489	-5.892880661	0.3613312
R_{rs} 556	-6.183340326	0.2787839
R_{rs} 667	-8.122627304	0.4957144

VIIRSN	mean $\ln R_{rs}$	sd $\ln R_{rs}$
R_{rs} 410	-6.371104055	0.6649559
R_{rs} 443	-6.104049134	0.3802373
R_{rs} 486	-5.995614302	0.327601
R_{rs} 551	-6.207168638	0.3055888
R_{rs} 671	-8.090188225	0.5376861

APPENDIX F

MODEL SELECTION BY STEPWISE AIC MODEL FOR OLCI, MODISA, SEAWIFS-MLCA AND VIIRS SENSORS

OLCI	
PCs coefficients	coefficients
0.067152837	a0
0.027206823	a1
-0.091204758	a2
0.085788324	a3
0.080937936	a4
0.190810854	a5
0.258951567	a6
-0.889019893	a7
1.332396987	a8

MERIS	
PC coefficients	coefficients
0.045615883	a0
0.009524025	a1
-0.139263155	a2
0.043668679	a3
0.236714697	a4
-0.075211368	a5
0.291677051	a6
-0.278288234	a7

MODISA	
PC coefficients	coefficients
0.031358631	a0
-0.012502955	a1
-0.098227547	a2
0.066997509	a3
0.087739685	a4
0.27458948	a5
-0.321039374	a6
-0.083978429	a7
0.286970493	a8

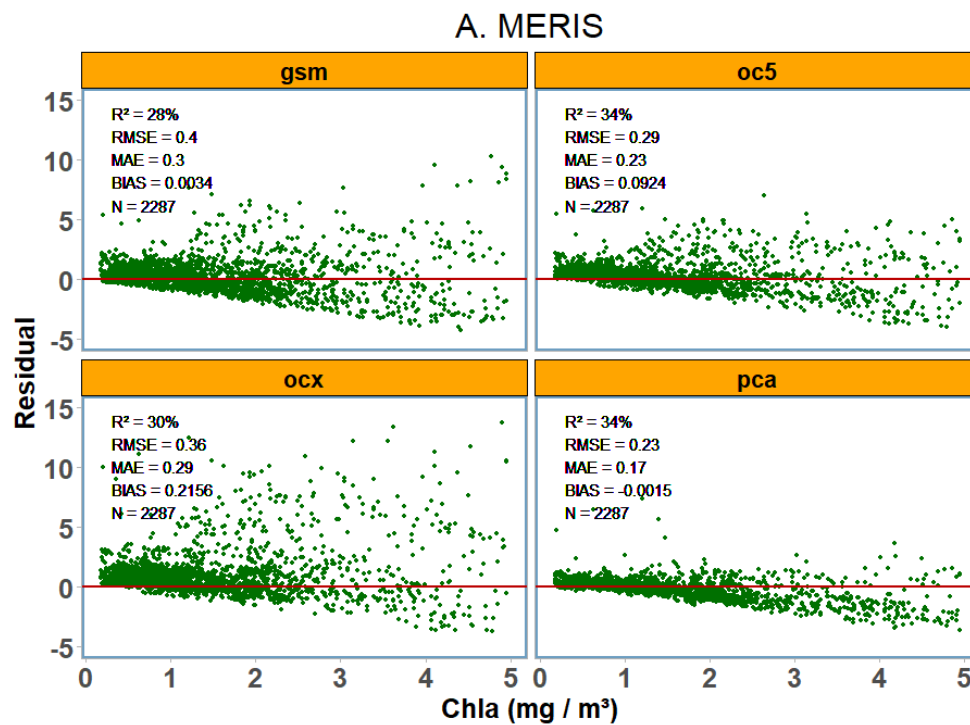
SEAWIFS-MLAC	
PC coefficients	coefficients
0.11205048	a0
-0.029094338	a1
-0.187142105	a2
0.134103165	a3
0.271246128	a4
-0.069616354	a5
0.110488366	

VIIRSJ	
PC coefficients	coefficients
-0.149029954	a0
-0.067588789	a1
-0.130909371	a2
0.428014414	a3
-0.456792333	a4

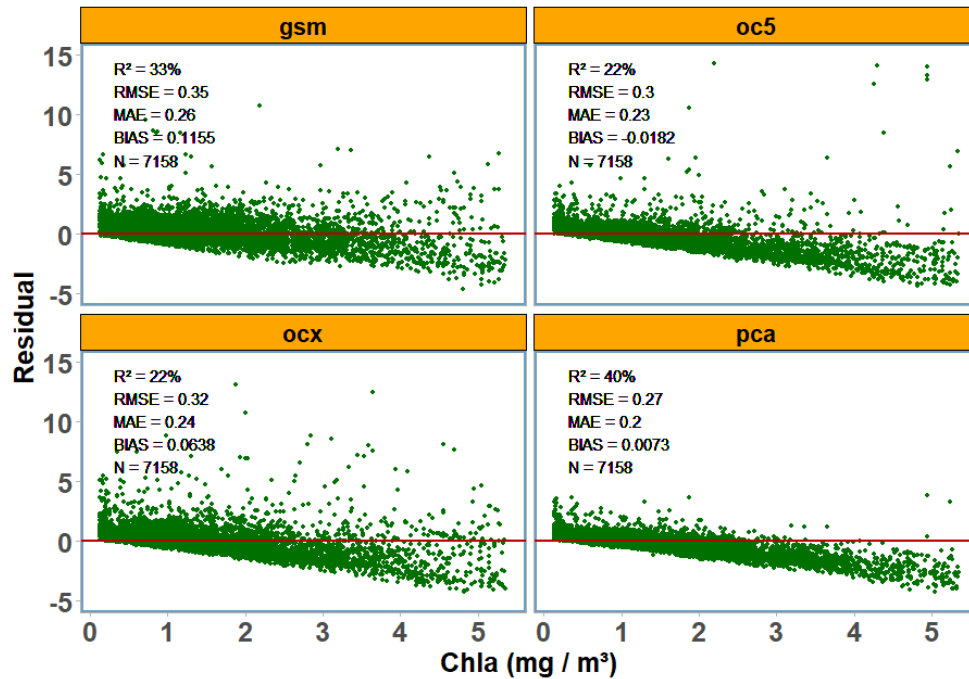
VIIRSN	
PC coefficients	coefficients
-0.007073403	a0
-0.036172424	a1
-0.099372378	a2
0.124724393	a3
0.227766387	a4
-0.460941066	a5

APPENDIX G

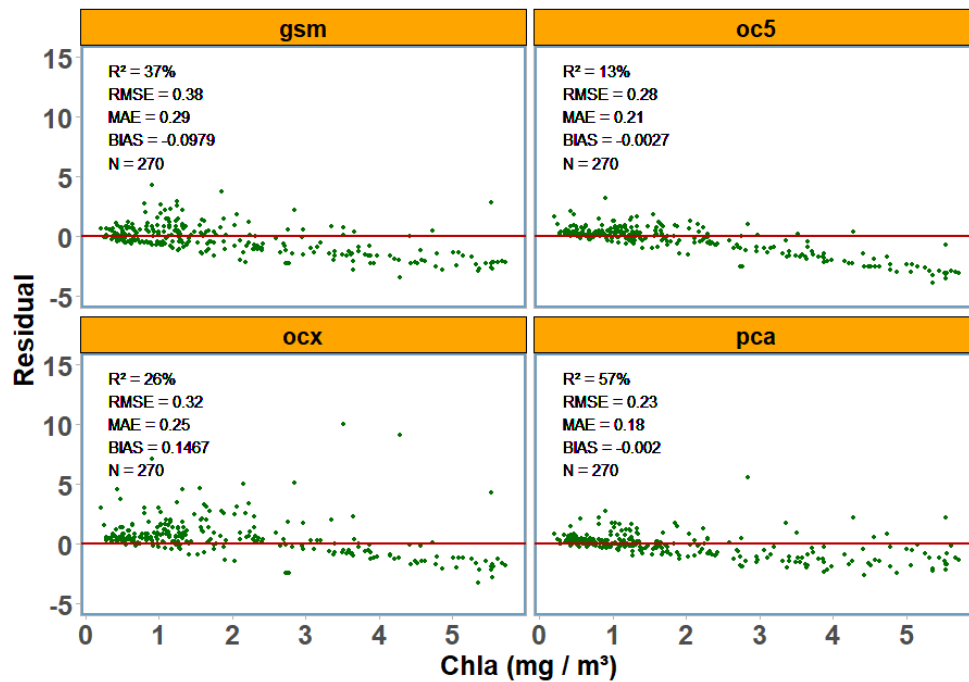
RESIDUAL PLOTS OF ALL ESTIMATED CHLA FROM ALL ALGORITHMS AND THEIR CORRESPONDING IN SITU CHLA VALUES POSITIVE VALUES (ON THE Y-AXIS) MEAN THE PREDICTION WAS TOO LOW, AND NEGATIVE VALUES MEAN THE PREDICTION WAS TOO HIGH; 0 MEANS THE ESTIMATED VALUES ARE EXACTLY CORRECT.



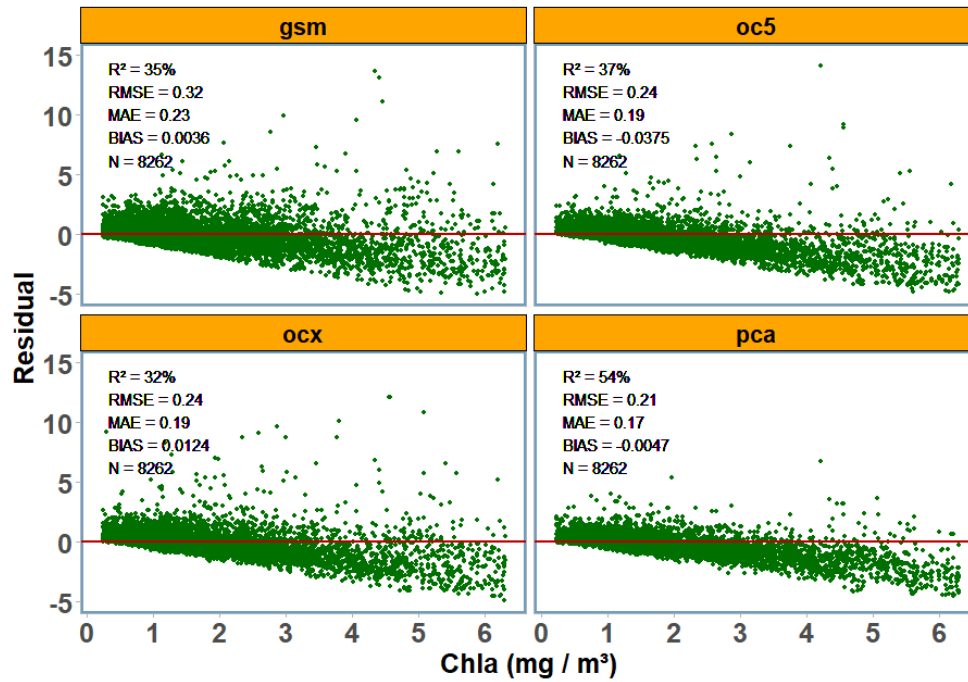
B. MODIS



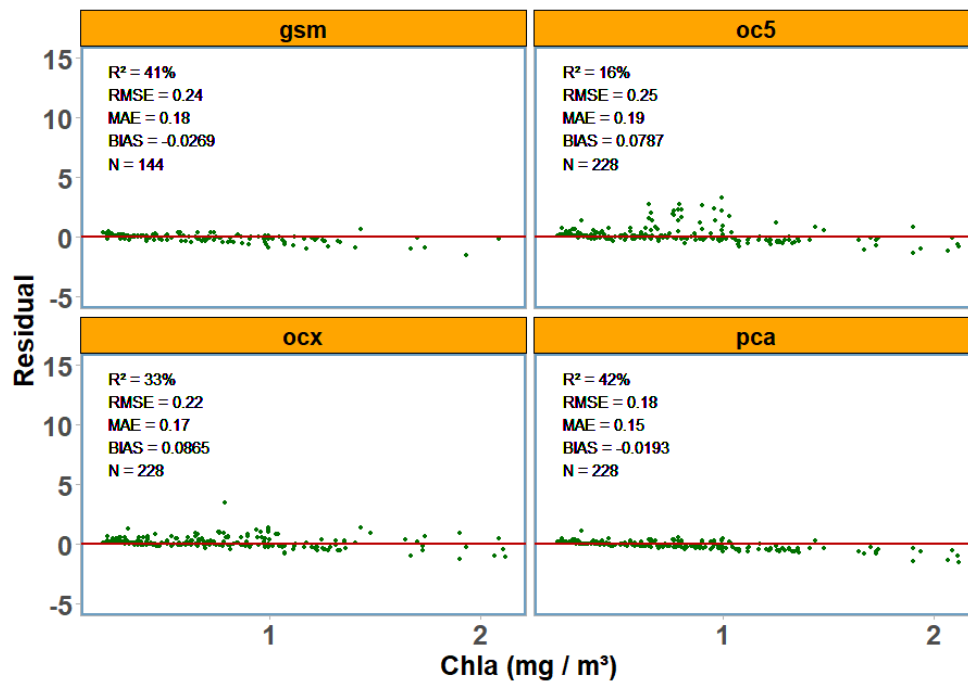
C. OLCI



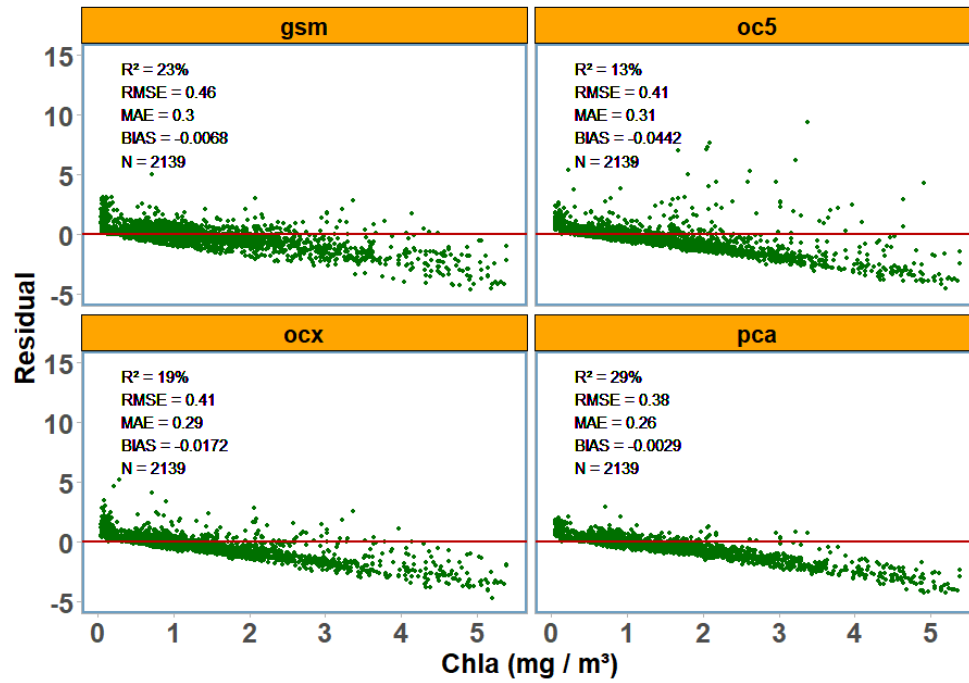
D. SEAWIFS



E. VIIRSJ



F. VIIRS





SIMBA

Chlorophyll-a Algorithm Theoretical Based Document

smartWhales initiative
Ref.: 9F040-190633/004
ATBD-Chla

End of the document