SIMBA

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

(Integrated modelling system for the North Atlantic Right Whale)

Ocean Colour Radiometry processing report

Milestone #3 Progress report

Date: March 31, 2022

Customer: CSA

Project no.: 9F040-190633/004

VERSION AND SIGNATURE

Prepared by	Version	Date	Modifications
Christiane Dufresne	1.0	01/2022	Initial version
Simon Bélanger, Marine Bretagnon, Philippe Garnesson, Thomas Jaegler, Yanqun Pan	2.0	03/2022	Revised version
Verification by			**Modifications**
Simon Bélanger	2.1	03/2022	
Final Authorization			**Signature**
Simon Bélanger	2.1	03/2022	

APPLICABLE AND REFERENCE DOCUMENTS

ID	Description	Reference
PMTP	Project Management and Technical Plan	M2-T1.3.3
ATBD Chla	Technical document: Chla	M3-T2.1.4
ATBD Pheno	Technical document: phenology metrics	M3-T2.3.5
MS3-PR	Milestone#3 Progress Report	M3-PR

TABLE OF CONTENTS

ACRONYMS

ADG	Absorption coefficient due to color organic matter and non-pigmented particles
API	Application Programming Interface
ATLNW	Atlantic Northwest
AV	Simple averaging
AVW	weighted averaging
BBP	Particulate back-scattering coefficient
CCMEP	Canadian Centre for Meteorological and Environmental Prediction
CHL	Chlorophyll-a
CIOPS-E	Coastal Ice-Ocean Prediction System for the East Coast of Canada
C-TEP	Coastal Thematic Exploitation Platform
CMEMS	Copernicus Marine Environment Monitoring Service
CSA	Canadian Space Agency
DPMC	C-TEP processing system
DFO	Department of Fisheries and Oceans
ECCC	Environment and Climate Change Canada
EO	Earth Observation
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
ESA	European Space Agency
GC	GlobColour
GEOSS	Group on Earth Observations System of Systems
GoMSL	Gulf of Maine and Gulf of St. Lawrence
GSM	Garver, Siegel, Maritorena Model
IUCN	International Union for Conservation of Nature
K_d	Diffuse attenuation coefficient
L2 /L3/L4	Level 2, Level 3, Level 4
L4-STD	Standard Level 4 Globcolour product
L4-REG	Regional Level 4 Globcolour product
MERIS	Medium Resolution Imaging Spectrometer
MODIS	*Moderate Resolution Imaging Spectroradiometer*
NASA	National Aeronautics and Space Administration
NARW	North Atlantic Right Whale
NetCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
OCR	Ocean Colour Radiometry
PAR	Photosynthetically Available Radiation
PCA	Principal component analyses
PFT	Phytoplankton Functional Types
PM	Processing Manager
PP	Primary production
RRS	Remote sensing reflectance
SDM	Species Distribution Model
SeaWiFS	*Sea-viewing Wide Field-of-view Sensor*
SIMBA	Système Intégré de Modélisation de la Baleine noire de l'Atlantique (*Integrated modelling system for the North Atlantic Right Whale*)
SPM	Suspended particle matter
SST	Sea surface temperature
TC	Transport Canada
VIIRS	*Visible Infrared Imaging Radiometer Suite*
WP	Work Package
WPS	Web Processing Service
ZSD	Secchi disk depth

LIST OF FIGURES

LIST OF TABLES

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). The *Ocean Colour Radiometry (OCR) Processing Report* is part of the Milestone 3 deliverables, along with two technical documents introducing new processing algorithm (see Applicable and Reference Documents). It aims to summarize the satellite OCR data products and operational processing chains being implemented for the project. However, since several modeling tasks of the project are complementary and strongly linked, multiple iterations are expected. The developed products will be tested as inputs for the modelling system and will likely improve as modelers provide feedback. In contrast, some OCR product may be discarded if not relevant. Consequently, the list of produced datasets and the processing chains presented here are not final and are likely to change during the course of the project.

1.2. Project overview

In July 2020, the International Union for Conservation of Nature (IUCN) announced that it had changed the status of North Atlantic right whales (NARW) from *endangered* to *critically endangered*. Among the many threats they face, NARW often sustain severe injuries due to collisions with ships or entanglement in fishing gear that occurs when in pursuit of its main food source, zooplankton. The distribution and abundance of the lipid-rich copepod *Calanus* spp., one of NARW's preferred prey, changed dramatically at the turn of the 2010s and NARW responded rapidly by looking for yet unexploited feeding grounds, especially further north in Canadian waters. The arrival of NARW in Canadian waters in spring or summer is driven by the presence of abundant *Calanus* biomass that accumulates near the sea surface. Our integrated modelling system for the NARW (referred to as SIMBA, for ***S****ystème* ***I****ntégré de* ***M****odélisation de la* ***B****aleine noire de l'****A****tlantique*) will develop user-friendly predictive tools and critical data that will help identify the presence of NARW's preferred prey in specific areas at a given time, in order to support decision-makers and mariners make more accurate recommendations when determining restriction zones for anthropogenic activities in areas likely to encounter NARW. This project aims at developing a modelling platform for producing short term (< 10 days) predictions of the NARW habitat suitability at the scale of the Northwest Atlantic shelf, including the Gulf of Maine and Gulf of St. Lawrence (GoMSL). The forecasts will be based on an ensemble of Species Distribution Models (SDMs) informed by satellite-derived bio-optical variables and 3D oceanographic outputs from high-resolution operational circulation models (CIOPS-E model, developed by Environment and Climate Change Canada - ECCC), which are also used to simulate the distribution of Calanus.

Satellite remote sensing of the surface ocean properties are a major source of information for this prediction system. It includes improved and high-level operational Ocean Colour Radiometry (OCR) products such as phytoplankton **biomass**, **primary production**, **phenology**, **functional types** and, if possible, direct detection of **surface *Calanus* swarms**. The modelling system will be implemented into two Earth observation (EO) big-data processing and analytics platforms: the Coastal Thematic Exploitation Platform (C-TEP) and the GEO Analytics Canada (*GEOAnalytics.ca*). The C-TEP is a dedicated cloud computing platform developed under ESA funding for the management and analysis of large satellite ocean-related data (mainly OCR). Only specific OCR processing tasks will be achieved in the C-TEP, while other data products generated in the Globcolour processor (see below) will be pushed directly to the GEOAnalytics.ca platform. The later will be used to operationalize the near real time (NRT) generation of outputs from enhanced NARW models. This satellite EO exploitation platform, which provides a data science and development environment along with integrated tools for EO data processing, was built on modern Open Geospatial Consortium protocols by *Hatfield C*. Integration of *GEOAnalytics.ca* with the Globcolour and C-TEP will be done through Web Processing Service (WPS) API calls, which will result in the C-TEP generating the requested EO products on-demand asynchronously. Once all OCR products and 3D ocean model datasets are available on *GEOAnalytics.ca* object storage system, a variety of containerized operations encapsulating model processes will be orchestrated to generate the novel products.

To make the final output products accessible to end-users, we will create a public web portal providing information on the project and methodologies used to generate these output products, a user's guide explaining the products and their uncertainty margins, information on how to access archived products and how to subscribe to new product notifications, electronic communication. The portal will also feature a web map interface allowing output products to be interactively navigated and displayed in a user-friendly manner, with mechanisms allowing end-users to provide feedback on the products.

Figure 1 presents a schematic of the project depicting the satellite and ocean model data flow within the cloud computing platforms in which the modelling system will be developed. For further details regarding Work Packages (WP) and related tasks, refer to the *Project Management Plan* and the *Milestone#3 Progress Report* (see Applicable and Reference Documents).

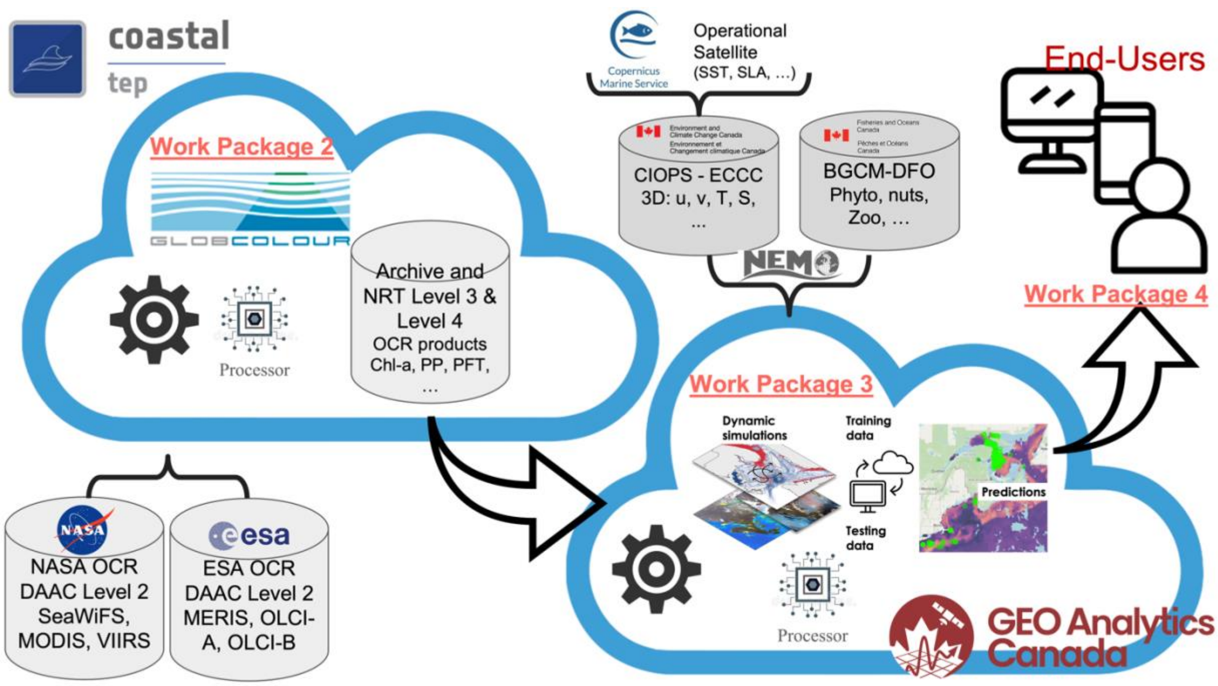

Figure 1. Schematic of the project workflow.

2. OCR PROCESSING REPORT

2.1. Overview

As mentioned above, OCR data products will be delivered to the *GEOAnalytics.ca* platform, which will ingest all needed data layers to perform the NARW modeling. The OCR data products will be generated mainly using the GlobColour (GC) processor, but some additional processing, such as the phenology metric and regional primary production calculations, will be performed in the C-TEP before entering *GEOAnalytics.ca.* Figure 2 presents an updated overview of the OCR data flow from their production sources (i.e. NASA and ESA) to *GEOAnalytics.ca.* Section 2.2 briefly describes the GC processor, and the main products generated that are considered potentially relevant to the SIMBA modeling team. The GC processor includes the regional algorithm for chlorophyll-a developed as part of SIMBA (C-ATBD). We focus more on the product description and the ocean variables being produced. More information on Globcolour could be find in the Globcolour Product User Guide. Section 2.3 describes the C-TEP platform and the specific processing algorithms implemented (or to be implemented). These include, to date, bloom phenology metrics, regional primary production (PP), and ocean front detection. These high-Level OCR products are specific to SIMBA and could not be implemented in the GC processor, which, however, will provide the input for the algorithms.

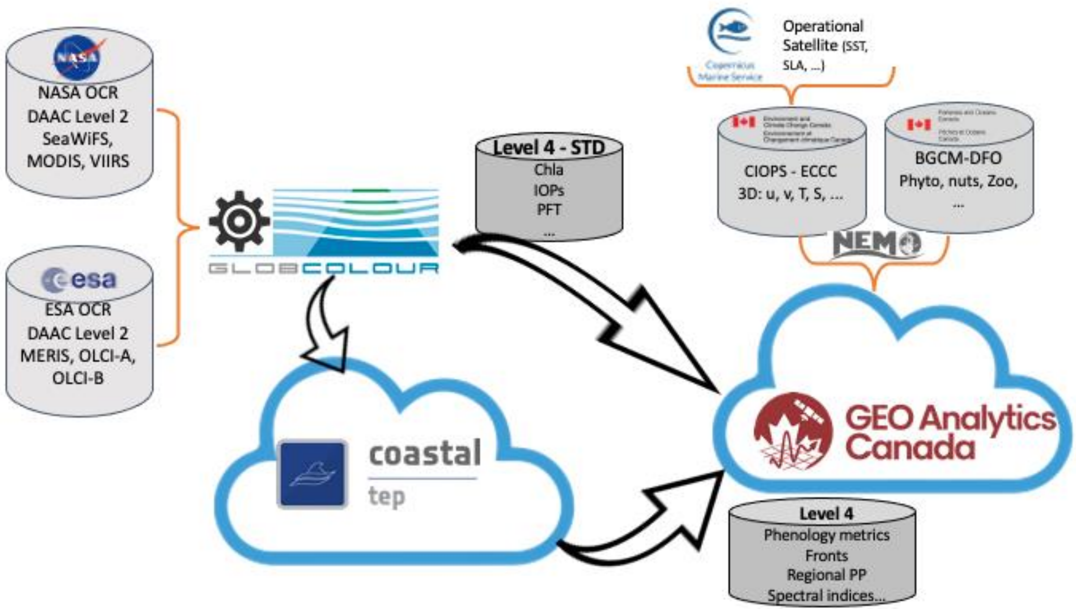

Figure 2: Schematic of the OCR processing illustrating the data flow through the Globcolour processor and the C-TEP.

2.2.GlobColour processor

The GlobColour processor development started in 2005 as an ESA Data User Element project to provide a continuous data set of merged Level 3 Ocean Colour products generated using Level 2 NASA and ESA upstream. In May 2015, GC processor was adapted to provide selected OCR products to the Copernicus Marine Environment Monitoring Service (CMEMS), including NRT products designed for operational modeling purposes. The primary data sets ingested in GC (Section 2.2.1 Upstream Data sets) are part of the core data set of the Group on Earth Observations System of Systems (GEOSS). Here we detail how the OCR products identify as L4-STD and L4-REG products in the SIMBA proposal and MS2 report are being produced and their main characteristics. Figure 3 illustrates schematically the processing from Level 2 (L2 geophysical in sensor swath projection) to the Level 4 (L4) interpolated products. The processing starts with the production of the Level 3 (L3) products (section 2.2.2 Level 3 processing and products) for each satellite sensor individually. Next, Level 4 data products are generated after merging sensor-specific L3 products (merged L3) and interpolating the data (section 2.2.3 Level 4 processing and products). The L4 interpolated products will be produced daily on Chlorophyll-a (CHL) variable only.

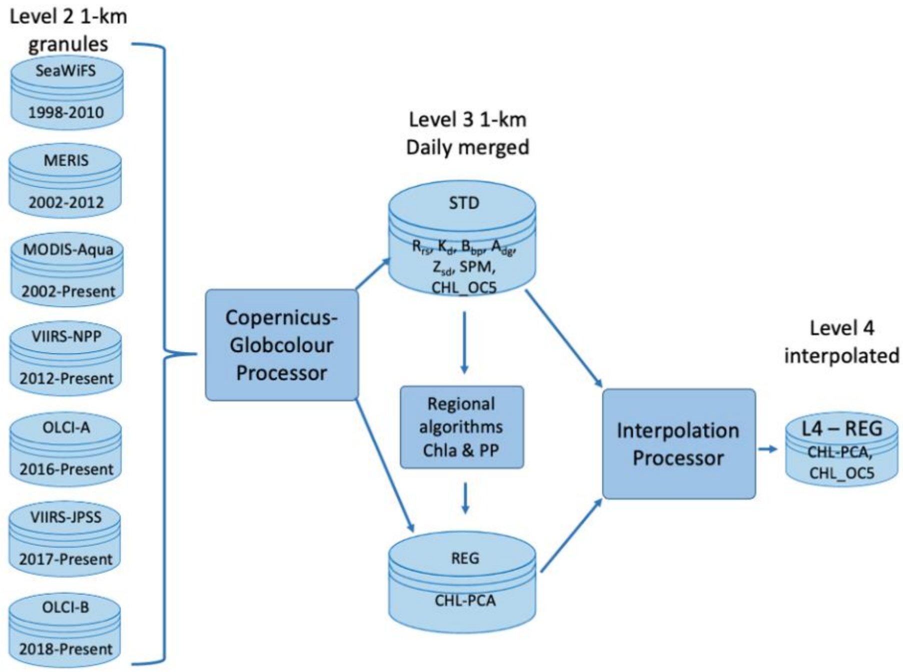

Figure 3: Schematic of the processing steps within the GC processor.

2.2.1 Upstream Data sets

The OCR products rely upon upstream data provided mainly by the NASA, NOAA, ESA and EUMETSAT. They are summarized in Table 1. Due to the degradation of the instruments (MODIS and VIIRS-NPP) the corresponding NASA processing has been updated several times. At the date of this report, the current NASA release used for SeaWiFS, VIIRS-NPP, VIIRS-JPPS1 and MODIS is called R2018. This major reprocessing has been completed by NASA beginning of 2018. The level 2 data set includes the remote sensing reflectance obtained after atmospheric correction. GC also used standard bio-optical variables available in the Level 2 data. Those are not detailed here as they are not being part of SIMBA higher level products.

Table 1: Upstream Level 2 data ingested by GC and the current reprocessing.

Sensor/Provider	Resolution	Reprocessing	Start date	End date
SeaWiFS/NASA	4-km	NASA R2018.0	1997-09-04	2010-12-11
MERIS/ESA	1-km	ESA 3rd reprocessing	2002-04-28	2012-04-08
MODIS-Aqua/NASA	1-km	NASA R2018.1	2002-07-03	Present
VIIRS NPP/NOAA	1-km	NASA R2018.0	2012-01-02	Present
OLCI A/EUMETSAT	1-km & 300-m	EUMETSAT PB.2.73	2016-04-25	Present
VIIRS JPSS/NOAA	1-km	NASA R2018.0	2017-11-29	Present
OLCI B/EUMETSAT	1-km & 300-m	EUMETSAT PB.2.73	2018-05-14	Present

2.2.2 Level 3 processing and products

The level-2 products are transformed from the sensor swath projection to a global sinusoidal grid with a pixel resolution around 1-km. This binning is separately applied to each level-2 input product for each instrument and the outputs are intermediate spatially binned level 3 products for each instrument, also called level 3 at track level. As multiple satellite swaths can overlap in a given day, the next step converts the L3 track to L3 daily for each single instrument.

In the frame of this project, L2 products were used to generate a regional mapped (plate-carré projection) L3 products (L3m) covering the **Atlantic Northwest** domain at full resolution. The spatial extension of the new L3 products (35°N to 55°N ; 77°W to 37°W) cover the same area as the operational East Coastal Ice Ocean Prediction System (CIOPS-East) in operation at the Canadian Centre for Meteorological and Environmental Prediction (CCMEP) of Environment and Climate Change Canada. One of the motivations was to generate OCR data products ready to assimilate in CIOPS-East in the future (beyond the scope of SIMBA). The regional daily L3m dataset for each sensor has already been generated by ACRI-ST and are identified with the acronym ATLNW. Figure 4 shows the extent of the area of interest common to CIOPS-East.

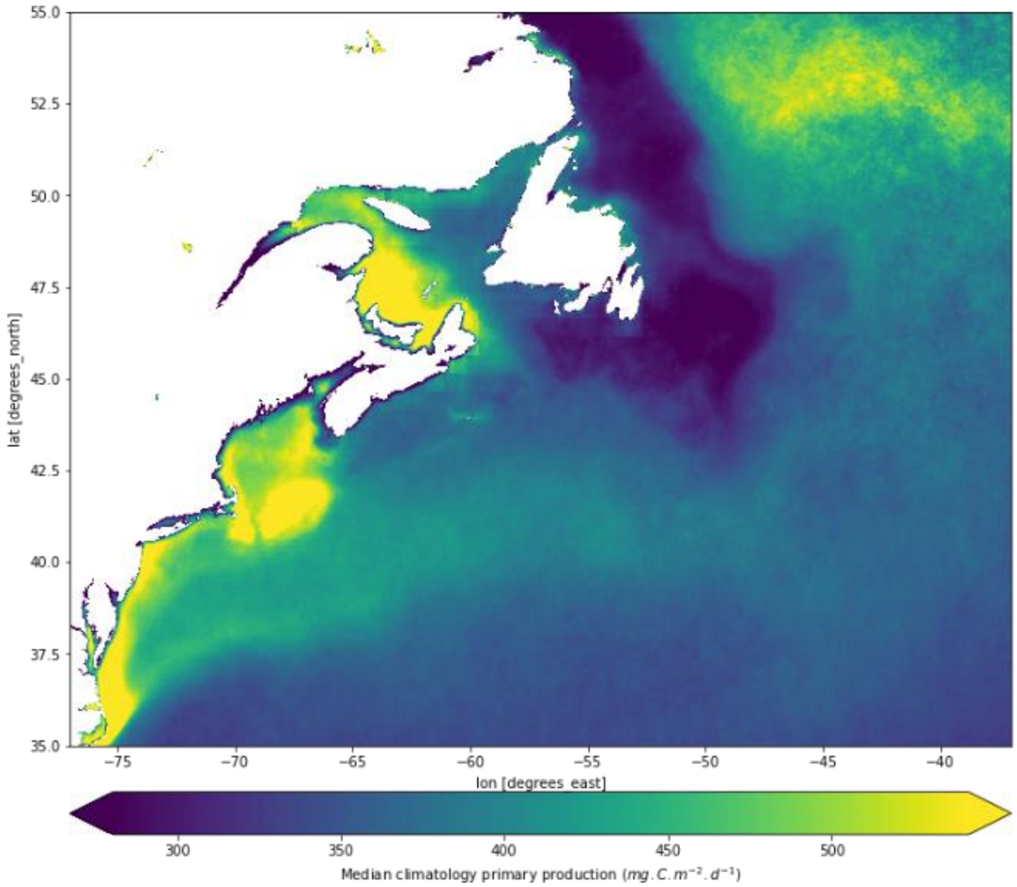

Figure 4: Extent of the area of interest for the SIMBA project. The climatology median of the primary production estimated using MODIS data is shown as an example (see section 2.3.2).

Table 2 presents the single sensor nomenclature for the products already generated and store in the ARCTUS server for the moment. As an example, MODIS daily chlorophyll-a product file for the 4th July 2002 will be named L3m_20020704__ATLNW_1_AV-MOD_CHL-PCA_DAY_00.nc. All GC products are in NetCDF format.

Table 2: Level 3 products nomenclature generated for the ATLNW using GC processor including both single sensor and merged products.

Single sensor product nomenclature	Temporal resolution	Spatial resolution	Available sensors	Merging method
L3m_*YYYYmmdd*__ATLNW_1_*mergingmethod-sensor_products_DAY_00.nc*	Daily	1 km	Merged, SeaWiFS, MODIS-Aqua, MERIS, VIIRS-SNPP, OLCI-S3A & S3B	AV, GSM
L3m_*YYYYmmdd-YYYYmmdd*__ATLNW_1_*mergingmethod-merged_products_8D_00.nc*	8 days	1 km	Merged	AV, GSM
L3m_ *YYYYmmdd-YYYYmmdd* __ATLNW_1_*mergingmethod-merged_products_MO_00.nc*	Monthly	1 km	Merged	AV, GSM

The next step in the GC processor merges the L3 daily for each single instrument to merged L3 daily product (e.g., merging MODIS-Aqua and MERIS). The merged products are generated by different simple averaging techniques (see IOCCG reports N°4 and 5) or using a semi-analytical algorithm, the GSM model (see Maritorena and Siegel, 2005; Garnesson et al., 2019), that considered the full spectral remote sensing reflectance information of each sensor (Maritorena et al., 2010). From the merged product, temporally binned products are generated for 8-day and Monthly, as illustrated in Figure 5.

The averaging method depends upon the variable (parameter) and the sensor available. Standard GC L3 products includes numerous optical and bio-optical variables and only the selected variables were retained. Following the SIMBA proposal, we distinguish two categories of parameters grouped into STD and REG databases. These selected parameters are found in Level 3 products and are very similar to those found in the CMEMS (Table 3). The merging method for each parameter listed in Table 3 is provided in Appendix A.

Table 3: List of products and associated algorithms for the Level 3 ATLNW products generated in GC. Note that regional PP variable is generated in a stand-alone processor and added to the GC data set for the archived products. The variable marked with a * will be processed to L4.

Products delivered by ACRI-ST			
Database name	Parameter	Description	Algorithm
REG	PCA-CHL*	PCA-based Chlorophyll-*a* concentration ($mg \cdot m^{-3}$)	Region-specific Principal component algorithm (ATBD-Chla)
	PP	Primary Production ($mgC \cdot m^{-2} \cdot d^{-1}$)	Regionally adapted Antoine and Morel (1996) (see section 2.3)
STD	RRS_xxx	Fully Normalized reflectance at available wavelength (sr^{-1})	Weighted average merging for multi sensors
	K_d*	Diffuse attenuation coefficient at 490 nm (m^{-1}) of the downwelling irradiance at 490nm.	Morel 1997
	CHL_OC5*	OC5-based Chlorophyll-*a* concentration ($mg \cdot m^{-3}$)	Gohin 2011
	PFT	Phytoplankton functional types (mg/m3). The PFT concentration is provided for: Diatoms, dinophytes, haptophytes, green algae, prokaryotes and Prochlorococcus.	Xi et al., 2020
	BBP	Particulate backscattering coef. at 443 nm (m^{-1})	GSM merging for multi sensors
	ADG	Absorption coef. due to colored dissolved organic matter and non-pigmented particles at 443 nm (m^{-1}).	GSM merging for multi sensors
	ZSD*	Secchi disk depth Transparency (m)	Morel 1997
	SPM	Suspended Particle Matter ($g \cdot m^{-3}$)	Gohin 2011

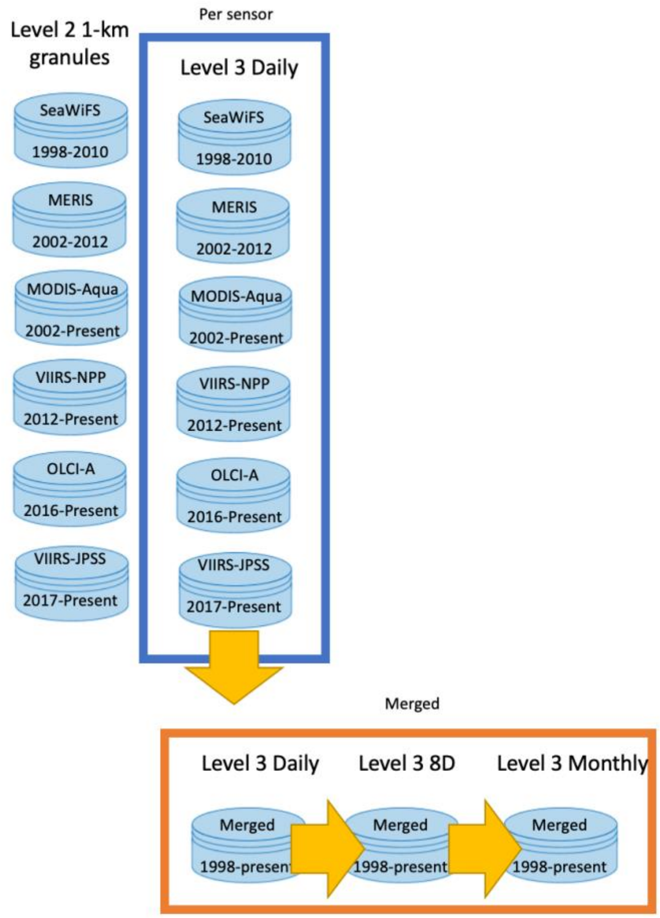

Figure 5: Flow chart of the data merging in the GC processor.

2.2.3 Level 4 processing and products

Following the L3 single sensor merged product computation, as the final step a space-time interpolation scheme is applied to produce “cloud free” images (Saulquin et al., 2019). Interpolated L4 products only include CHL variable and its derived products, i.e. the K_d and ZSD (empirical formula). The actual processor was developed for OC5-based CHL variable, but it is being adapted for PCA-based CHL. These “gap-free” images provide useful information for regions with high cloud coverage but can increase uncertainty in the parameters. Heavy cloud cover in the ATLNW zone may introduce noise in the OCR products, resulting in large uncertainties in the daily interpolation. To mitigate this risk, we will investigate larger time windows for the interpolation (currently +/- 4 days), or considered the interpolation on 8-day (8D) binned

L3 products. Table 4 summarize the lists of L4 products generated in the frame of the SIMBA project for the Atlantic Northwest zone.

Table 4: Summary of the main characteristics of the merged L4 interpolated products.

Globcolor product ref	Temporal resolution	Availability	Used sensors
L4_ YYYYmmdd-YYYYmmdd __ATLNW_1 _L4-STD.nc L4_ YYYYmmdd-YYYYmmdd __ATLNW_1 _L4-REG.nc	Daily or 8-Day gap-free	J0+1 (NRT), updated at J0+31	MODIS-Aqua, VIIRS-SNPP, VIIRS-JPPS1, OLCI-S3A & S3B

2.3. Description of the C-TEP

The aim of this section is to describe the Coastal Thematic exploitation platform (C-TEP) funded by the European Space Agency. C-TEP provides a dedicated service for the observation and monitoring of our coastal environment through the provision of access to large volumes of EO, computing resources and the fundamental processing software required to extract temporal and spatial information from Big Data.

For the SIMBA project, the processing chain for Earth Observation derived products will be implemented in the C-TEP. For now, the current processing system (DPMC) is able to schedule and run a generic batch on cluster nodes. Adding a processing chain inside the DPMC is currently not possible without making it by hand. So, the Processing Manager (PM) must be a module interacting with the DPMC. The DPMC will only be used as a resources allocator. The scheduling will be managed by the Processing Manager. This tool allows having different processing modes: on-demand processing, time-driven processing, and event-driven processing.

On-demand mode

The on-demand processing is similar to the classic earth observation processing integrated in the DPMC. It means we are able to define a processing chain and execute it on a sample of data for a specific time range.

Time-driven mode

This mode allows a user to run a processing chain periodically (e.g., every day) and make the starting date programmable (e.g., first planned run on 08/02/2021 at 4pm).

Event-driven mode

This mode describes an event to be the trigger of the processing chain. An event can be different things like a data ingestion, a notification from another system or a custom trigger.

For each mode, the parameters for the configuration will be available through the unique PM's application programming interface (API). A User Guide is available to describe how to define a workflow and its use.

Functional scheme

The Processing Manager is a standalone module. We can interact with it by API. The most common interaction is:
Users send a workflow command to the PM by API, then the PM transforms the command in Prefect Workflow. Next, the PM sends the tasks to the DPMC as a job. Figure 6 shows a simplified scheme of the workflow of the C-TEP processing.

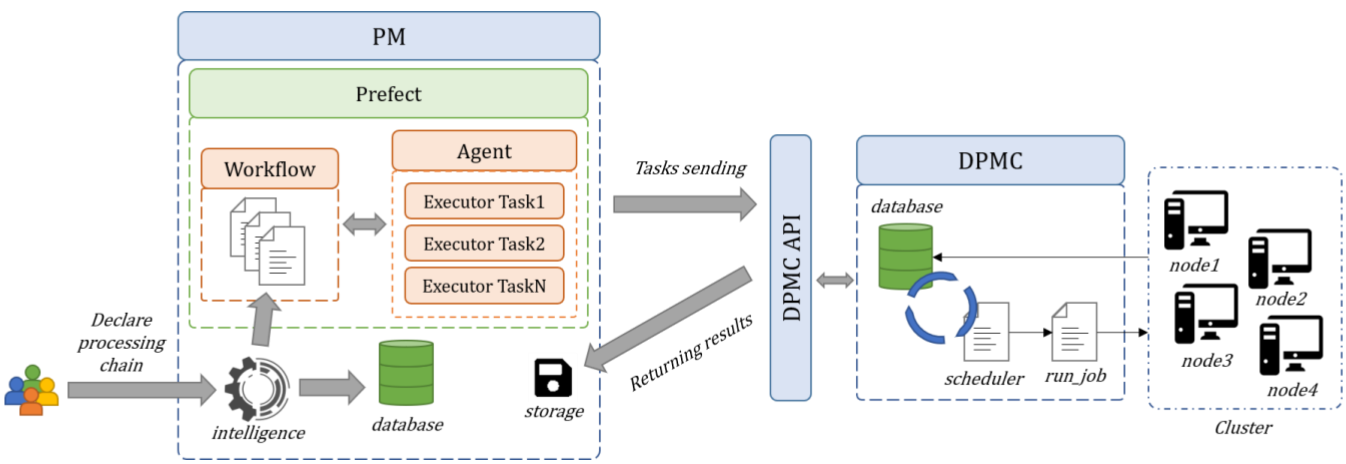

Figure 6: Workflow of the processing scheme in the C-TEP.

For the SIMBA project, the regional algorithms for phenology metrics, primary production and front detection will be implemented in the C-TEP to generate the dedicated products in NRT for the species distribution model. Figure 7 shows the input and output of the selected algorithms implemented in the C-TEP. Input of the algorithms and the output product generated with the C-TEP are described in the following sections. We briefly detailed these three components in the next sections.

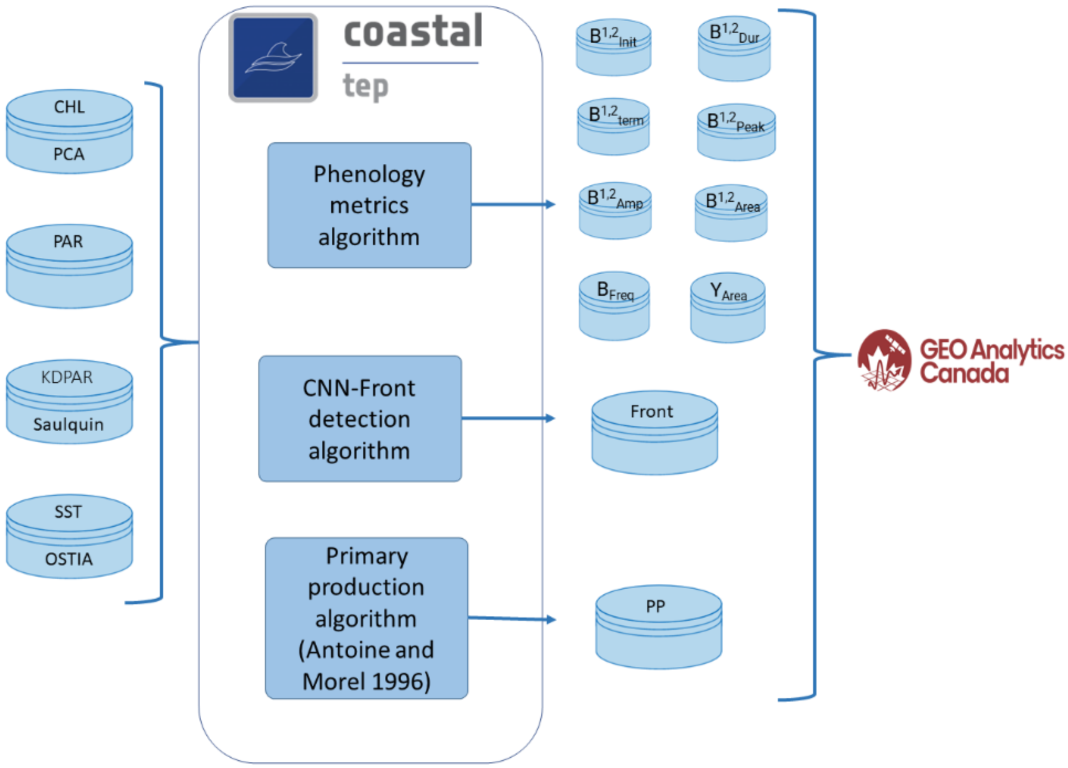

Figure 7: Workflow of the C-TEP regional algorithms implementation.

2.3.1 Phenology metric

In the frame of the SIMBA project, metrics of the phenology of the phytoplankton are generated using the C-TEP. For the complete description of the algorithms and the output products (see Applicable and Reference Documents).

Input of the phenology algorithm is the L4 interpolated Chlorophyll-a retrieved using the PCA regional algorithm as produced by the GC processor (section 2.2.3). The processing of the algorithm for the metrics calculation will run on the C-TEP and compute eight bloom metrics. Six metrics (output) will be generated in quasi NRT, bloom initiation (B_{init}), bloom termination (B_{Term}), bloom duration (B_{dur}), bloom amplitude (B_{Amp}), day of year of CHL maximum of the bloom (B_{peak}), and the bloom area (B_{Area}), which is the cumulative biomass during the bloom. As explained in the ATBD, a bloom is detected if the Chla concentration remains above a threshold for at least 15 days. Therefore, B_{init} metric could be obtained at NRT+16 days or so. To assess B_{dur}, B_{Amp}, B_{peak} and B_{Area}, the bloom termination must be reached, which could take several weeks. On the GoMSL, B_{dur} of the first bloom is generally <45 days, but can reach up to 60 days in some years. As soon as a bloom terminates, its associated metrics will be calculated and provided to the SDM. In addition to these metrics, two additional metrics could be computed after the completion of the whole year, i.e. bloom frequency during the seasonal cycle (B_{freq}) and the Biomass cumulated over the year (Y_{area}). We expect that the algorithm will output the bloom metrics for each bloom occurring between February and November to avoid the winter season.

2.3.2 Regional adaptation of the NPP algorithm

In the frame of the SIMBA project, the Antoine and Morel (1991) (AM96) net primary production (NPP) algorithm has been adapted to answer to the specificities of this region (for details description of the AM96 see Appendix B). Indeed, studies have previously demonstrated that usual ocean colour product do not perform well in the arctic or sub-arctic region (Lee et al., 2015). With sea ice presence, the GoSL could be considered as sub-arctic region. It is also a Case 2 water system influence by significant terrestrial input of colored dissolved organic matter. To address this issues, specific algorithms were used. The NPP was then estimated using the PCA-chlorophyll concentration product generated by the GlobColour processor (archived L3-CHL-PCA; L3m_YYYYmmdd__ATLNW_1_AV-SENSOR_CHL_DAY_00.nc), the Photosynthetically Available Radiation from Frouin et al. (2003) available on the GlobColour archive (L3-PAR) (https://hermes.acri.fr), the coefficient of diffuse attenuation developed by Saulquin et al., (2013) also available on the GlobColour archive (L3-KdPAR-SAULQUIN), and the OSTIA product (Ocean Surface Temperature and Sea Ice Analysis, from CMEMS: https://doi.org/10.48670/moi-00167) for the sea surface temperature. All these products were projected on a horizontal grid with a spatial resolution of 1km to retrieve daily products of primary production. Description of these satellite products can be found in the GC documentation elsewhere.

To dispose of the largest temporal archive possible, and benefit of the best spatial coverage, the primary production estimated from different ocean colour sensor should be merged. Thanks to that, NPP observations will be available from 1997 up to now. At present, the timeseries is available for MODIS (Figure 8) and will be soon available for the other sensors. The NPP will be calculated on the archived data using a stand-alone processor and the data will be included in the GC products as described in section 2.2. It includes daily PP, 8D and monthly averaged as described in section 2.2.2 Level 3 processing and products.

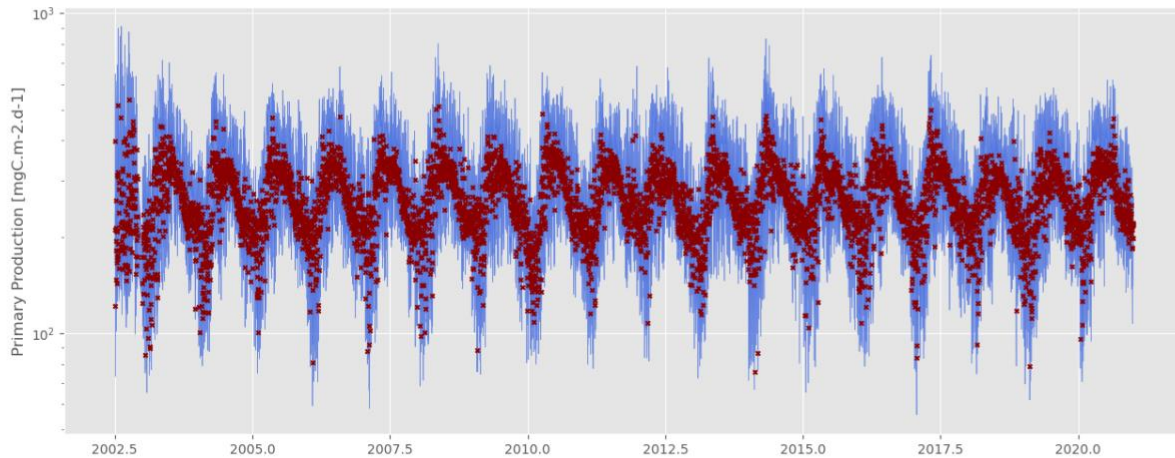

Figure 8: Timeseries of the primary production for GoMSL region estimated from MODIS sensor.

Additionally, the NPP algorithm will be implemented in the C-TEP for NRT purposes. Daily NPP will be calculated in NRT using L4 interpolated Chla. PAR and SST are also gap-free. Kd product, in contrast, will be from L3 with gaps. A solution to fill the gaps in the Kd product will be to use a Kd climatology, as done in Bélanger et al. (2013). In addition to daily NPP, additional phenology metric will be computed. These NPP-based metrics could add crucial information on the seasonal productivity of our area of interest. Cumulative NPP starting from January will be computed in NRT daily (PP_{cumul}). In addition, accumulated (sum) biomass for the last 45 days and 90 days will be also calculated daily. These products (PP_{cumul_45D}, PP_{cumul_90D}) will be added layers to the SDM for testing. We are also considering the same phenology metrics as for the Chla, but the method will likely need adaptation (to be confirmed).

2.3.3 Front detection

Oceanic fronts, defined as areas of sharp gradients between adjacent water masses, are often sites of increased physical and biological activity affecting oceanic ecosystems. Relationships between fronts and fish abundance have been established, and maps of front frequency based on observation of time series were incorporated as a proxy for pelagic biodiversity. In the SIMBA project, the Chlorophyll-a/Sea Surface Temperature (SST) front is one of the key inputs for the species distribution modeling. Algorithms of automatically extracting gradient magnitudes and front contours from satellite images have been proposed, the most widely cited algorithms are the Belkin & O'Reilly (2009) method and the Single Image Edge Detection (SIED) algorithm (Cayula & Cornillon 1992). Besides, some common-used edge detection algorithms, e.g., Canny (1986) in computer vision also have the potential to be used for the oceanic front detection purpose. Performance of these available algorithms have been evaluated using the level-3 daily CHL-PCA products (see section

2.2.2 Level 3 processing and products) and the results indicate that the Belkin & O'Reilly method and SIED implemented in Marine Geo-spatial Ecology Tools (MGET) (https://mgel.env.duke.edu/ ; hereafter MGET-SIED) outperform the others in terms of gradient calculation and front contours extraction, respectively, and they are thus selected in this project.

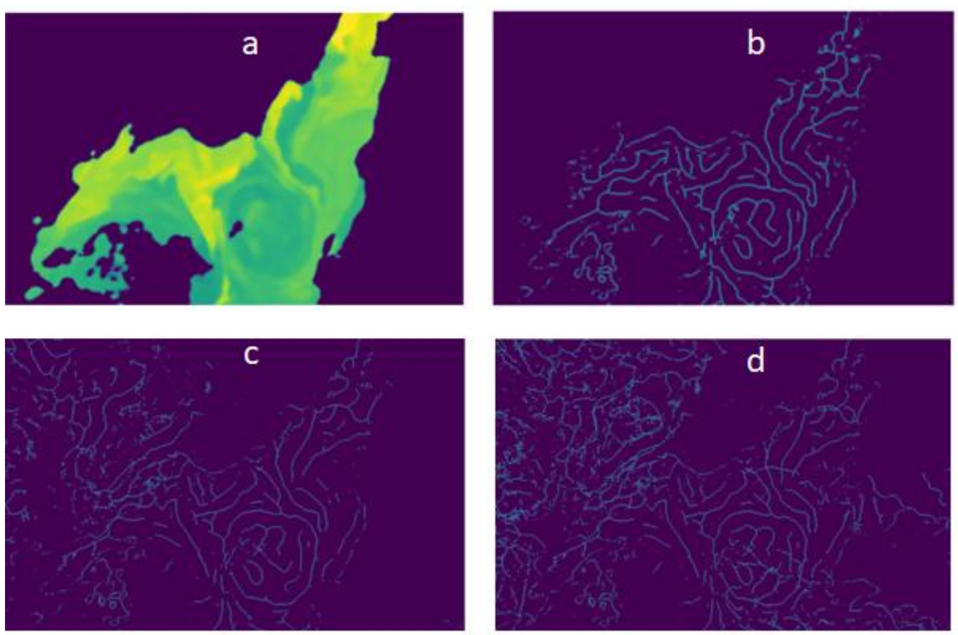

Figure 9. An example of Chla front contours. (a) map of Chla concentration from MERIS (2005/06/02), (b) front contours produced by CNN-FRONT, (c) 7-day composite map of front contours (2005/06/01-2005/06/07), (d) 14-day (2005/06/01-2005/06/14) composite map of front contours

As designed, the front detection function will finally be integrated into the cloud-based C-TEP system for the NRT satellite image processing. Unfortunately, on one hand, MGET-SIED can only be accessed through a Windows operating system but no other system including C-TEP which is Linux. On the other hand, its high complexity makes it almost impossible to be 100% accurately implemented in a short time. To solve this problem, an AI-based model that adapts a classic convolutional neural network architecture for front detection, so called CNN-FRONT, was developed. CNN-FRONT was trained using the inputs and outputs of MGET-SIED as the ground truth. The trained ready model was then applied to image processing. After a few steps of automatic post processing, it could detect front contours with a close accuracy to MGET-SIED. Figure 9(b) shows the front contours produced by CNN-FRONT from Chla shown as Figure 9(a).

Applying SIED to a typical satellite image reveals only fragments of fronts observed through the gaps in cloud cover. To produce a synoptic view of dynamic ocean features, some specific compositing technique that combines a sequence of satellite images is needed. In this project, the algorithm proposed by Miller (2009) was selected. Figure 9(c)-(d) show the composite front map using 7-day and 14-day singe gradient magnitude and front contours products, respectively.

Up to now, the implementation of the front detection function including CNN-FRONT, compositing technique and batch processing is completed. In the next step, it could be easily integrated to the C-TEP system.

3. STATUS OF THE PROCESSING AND THE IMPLEMENTATION

The general framework of the OCR products processing has been developed in collaboration with our partner ACRI-ST. The following section give an overview of the status of the implementation of the OCR processing.

Several steps are necessary to derive crucial information from the satellites data tailored for the NARW species distribution models. Our method largely builds upon the GC, an operational OCR data processing chain that has been developed by ACRI-ST (Sophia Antipolis, France) since 2005. The status of the OCR product

- Regional algorithm Chlorophyl-a development; For this task, the implementation of the new improved regional algorithm for the GOMSL to the GC processor has been done. This algorithm has been applied to Level 2 remote sensing reflectance from available sensor in the GC database.
- Level 3 daily per sensor processing; Following the implementation of the new Chlorophyl-a algorithm, 1km daily data per sensors has been generated for available sensors (see Table 1). A complete set of satellite based remote sensing data is now available for analyses.
- Level 3 daily merged generation; The next step of the OCR processing, will be the creation of daily merged sensors Level 3 data. Using the GC processor merging algorithm, data from all sensors will be merged into one file to improve the spatial coverage (if no clouds) and create a continuous time series from 1997 to present days.
- Level 4 investigation; To mitigate the lack of data introduce by the cloud coverage, following the generation of the Level 3 merged products, a space-time interpolation scheme is applied to produce “cloud free” images. Thus, due to the high variability of the water constituent in the GOMSL this interpolation could introduce uncertainty in the generated product. An investigation of a prior binning method (8 days, month) will be done to verify the applicability of the interpolation.
- Primary production product generation; The GC primary production algorithms (Antoine & Morel, 1996) has been adapted to use the regional CHL-PCA along with a specific coefficient of diffuse attenuation developed by Saulquin et al., (2013). The processing for all sensors has started. Following the computation of the primary production for L3 daily per sensor, data will be merged to obtain a merged L3 daily PP 1 km product. To fill the gaps of the cloud coverage we will investigate the feasibility of applying the same interpolation scheme than the CHL to the primary production.
- C-TEP implementation; More work needs to be done for the implementation of the phenology metrics and the CNN-Front algorithms to the C-TEP computing environment. The product

generation has been developed locally into Arctus facilities. The code needs to now be transfer to the C-TEP to benefit from the computation power and the direct access to the Globcolour database produced in the frame of the project.

4. REFERENCES

Antoine, D., & Morel, A. (1996). Oceanic primary production: 1. Adaptation of a spectral light-photosynthesis model in view of application to satellite chlorophyll observations. Global Biogeochemical Cycles, 10(1), 43–55. https://doi.org/10.1029/95GB02831

Bélanger, S., Babin, M., & Tremblay, J.-É. (2013). Increasing cloudiness in Arctic damps the increase in phytoplankton primary production due to sea ice receding. Biogeosciences, 10(6), 4087–4101. https://doi.org/10.5194/bg-10-4087-2013

Belkin, I. M., & O'Reilly, J. E. (2009). An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. Journal of Marine Systems, 78(3), 319–326. https://doi.org/10.1016/j.jmarsys.2008.11.018

Canny, J., A Computational Approach To Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.

Cayula, J. F., & Cornillon, P. (1992). Edge detection algorithm for SST images. Journal of atmospheric and oceanic technology, 9(1), 67-80.

Frouin, R., B. A. Franz, and P. J. Werdell, 2003: The SeaWiFS PAR product. In Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, S. B. Hooker and E. R. Firestone, Editors, CC NASA/TM-2003-206892, Vol. 22, 46-50.

Garnesson, P., Mangin, A., Fanton d'Andon, O., Demaria, J., and Bretagnon, M.: The CMEMS GlobColour chlorophyll a product based on satellite observation: multi-sensor merging and flagging strategies, Ocean Sci., 15, 819–830, https://doi.org/10.5194/os-15-819-2019, 2019.

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

Lee, Z., Marra, J., Perry, M. J., & Kahru, M. (2015). Estimating oceanic primary productivity from ocean color remote sensing: A strategic assessment. Journal of Marine Systems, 149, 50-59.

Maritorena, S. and Siegel, D.A. 2005. Consistent Merging of Satellite Ocean Colour Data Sets Using a Bio-Optical Model. Remote Sensing of Environment, 94, 4, 429-440.

Maritorena S., O. Hembise Fanton d Andon, A. Mangin, and D.A. Siegel. 2010. Merged Satellite Ocean Color Data Products Using a Bio-Optical Model: Characteristics, Benefits and Issues. Remote Sensing of Environment, 114, 8: 1791-1804.

Miller, P. (2009). Composite front maps for improved visibility of dynamic sea-surface features on cloudy SeaWiFS and AVHRR data. Journal of Marine Systems, 78(3), 327–336. https://doi.org/10.1016/j.jmarsys.2008.11.019

Morel, A., & Berthon, J. F. (1989). Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. Limnology and oceanography, 34(8), 1545-1562.

Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research: Oceans, 106(C4), 7163-7180.

Regaudie-de-Gioux, A., Huete-Ortega, M., Sobrino, C., López-Sandoval, D. C., González, N., Fernández-Carrera, A., ... & Duarte, C. M. (2019). Multi-model remote sensing assessment of primary production in the subtropical gyres. Journal of Marine Systems, 196, 97-106.

Saulquin, B., Hamdi, A., Gohin, F., Populus, J., Mangin, A., & Fanton, O. (2013). Remote Sensing of Environment Estimation of the diffuse attenuation coefficient K dPAR using MERIS and application to seabed habitat mapping. Remote Sensing of Environment, 128, 224–233. https://doi.org/10.1016/j.rse.2012.10.002

Saulquin B., Gohin, F. & Fanton d'Andon, O. (2019) Interpolated fields of satellite-derived multi-algorithm chlorophyll-a estimates at global and European scales in the frame of the European Copernicus-Marine Environment Monitoring Service, Journal of Operational Oceanography, 12:1, 47-57, DOI: 10.1080/1755876X.2018.1552358

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

APPENDIX A – MERGING METHODS

- AV: simple averaging
- AVW: weighted averaging
- GSM: GSM model

Table A1 – Merging methods and sensor available for parameters found in the CHL data products

Parameters	L3 Merging methods	Sensor availability						
		SWF	MER	MOD	VIR	VJ1	OLA	OLB
CHL-PCA	AV	X	X	X	X	X	X	X
PP		X	X	X	X	X	X	X

Table A2: Merging methods and sensor available for the Normalized Remote Sensing Reflectance found in the STD data products

Parameters	L3 Merging method	Sensor availability						
		SWF	MER	MOD	VIR	VJ1	OLA	OLB
NRRS400	AV						X	X
NRRS412	AVW	X	X	X	X	X	X	X
NRRS443	AVW	X	X	X	X	X	X	X
NRRS469	AV		X					
NRRS490	AVW	X	X	X	X	X	X	X
NRRS510	AVW	X	X				X	X
NRRS531	AV			X				
NRRS547	AV			X				
NRRS551	AV				X	X		
NRRS555	AVW (1)	X	X	X	X	X	X	X
NRRS560	AV		X				X	X
NRRS620	AV		X				X	X
NRRS645	AV			X				

NRRS670	AVW	X	X	X	X	X	X	X
NRRS674	AV						X	X
NRRS678	AV			X				
NRRS681	AV						X	X
NRRS709	AV						X	X

(1): spectral inter-calibration is applied prior to the merging.

Table A3: Merging methods and sensor available for bio-optical variables found in the STD data products

Parameters	L3 Merging methods	Sensor availability						
		SWF	MER	MOD	VIR	VJ1	OLA	OLB
BBP	GSM	X	X	X	X	X	X	X
CDM	AV		X				X	X
	GSM	X	X	X	X	X	X	X
SPM	AN	X	X	X	X	X	X (1)	X(1)
KD	AN	X	X	X	X	X	X	X
ZSD	AN	X	X	X	X	X	X	X

APPENDIX B – PRIMARY PRODUCTION ALGORITHM

Primary production algorithm description

Primary production (PP) drives the production of zooplankton and fish communities since phytoplankton is at the basis of the marine food-web. With the aim to modelled upper trophic level as predict the presence of whales, get an estimate of the primary production appears to be a prerequisite. While in situ measurement of the primary production is temporally and spatially punctual, the algorithms developed to estimate this parameter from ocean colour observations reveals to be a precious source of data. Indeed, satellites offer a relatively high frequency of sampling, and this for more than two decades, over the global ocean.

The estimation of the primary production yield is based on a very basic principle: primary production efficiency is function of the light and the nutrient availability. Starting from this point, and in reason of the high interest of the oceanographic community to monitor the primary production, several algorithms based on satellite observations were developed during the last decades. All the algorithms take, at least, the remotely sensed chlorophyll concentration, which is considered as an index of biomass, and the quantity of light available for the photosynthesis (PAR). To estimate the efficiency of the primary production the sea surface temperature is also an important variable. Thanks to these variables, algorithms can estimate the quantity of carbon which is uptake, and its conversion into energy.

The algorithm used in the frame of SIMBA project is the one developed by Antoine and Morel 1996 (AM96) and adapted for the Arctic Ocean. In the following, one will find a description of the algorithm, the inputs required, an analysis of the validation, results for the region of interest.

Basically, the estimation of the primary production from satellite ocean colour observation is based on the following principle. First, it is recall that the primary production corresponds to the amount of inorganic carbon which is fixed by phytoplankton in one day. To derive an estimate of the primary production, one should therefore take into account physiological properties of phytoplankton. For instance, the linked between seawater temperature, and the

Basically, one can considered that the primary production algorithm is based on the following equation:

$$PP = \overline{PAR} * Chla_{tot} * \psi \qquad (1)$$

Where PP is the primary production expressed in gC.m-2.d-1, PAR represents the solar radiation available for the photosynthesis (integrated between 400 and 700 nm). Chlatot represents the chlorophyll concentration integrated over the productive layer.

And ψ, correspond of the amount of solar energy which is stored by photosynthesis relatively to the chlorophyll concentration in the sunlit layer. In other word, it represents the efficiency of the photosynthesis.

The AM96 algorithm estimates the primary production over the productive layer, i.e. 1.5 times the euphotic layer, or equivalently, the depth where the light is equal to 1% of its surface value. This allow to considered cases where a deep chlorophyll maximum is established, notably at the basis of the euphotic layer.

Considering that primary production is vertically, temporally and spectrally dependant, the previous equation becomes:

$$PP = 12 \times \iiint_{0\,0\,\lambda 1}^{D\,L\,\lambda 2} Chl(z) \times PAR(\lambda, z, t) \times a^*(\lambda) \times \phi_\mu(\lambda, Z, t) \times d\lambda \times dz \times dt \quad (2)$$

Parameters are then integrated: vertically between the surface and the bottom of the productive layer, temporally between sunrise and sunset, and spectrally between 400 and 700 nm.

The vertical profile of the chlorophyll concentration is based on the estimation from Morel and Berthon 1989, which allow to model subsurface chlorophyll maximum in function of the chlorophyll concentration estimated from satellite observation.

$$c(\zeta) = C_b - s \times \zeta + C_{max} \times \exp\left\{-\left[\frac{\zeta - \zeta_{max}}{\Delta\zeta}\right]^2\right\} \quad (3)$$

Where Cmax represents the maximal chlorophyll concentration, ζmax the depth where this maximal concentration occurs, Δζ represents the width of this peak in chlorophyll concentration, s represents the slope with which the chlorophyll concentration decrease, starting from the surface, where the concentration is noticed Cb.

And ζ represents the depth normalized by the euphotic depth (z/Zeu).

The light vertical profile over the productive layer is modelled from Morel 1990 according:

$$PAR(z) = PAR(0) \times e^{-Kd(PAR) * z} \quad (4)$$

Therefore, it then possible to retrieve a profile of light and biomass in function of satellite observation. All these parameters were integrated following the equation 2 with trapezoidal integration.

Validation of the algorithms

Prior to applying the AM96 algorithm over the whole ocean colour archive, a validation of this algorithm was performed thanks to a comparison between in situ measurements of primary production and satellite estimations. However, due to the paucity of publicly available datasets concomitant with satellite observations, the validation was performed with in situ measurements sampled from mid to law latitudes. The in situ dataset is extracted from the MALASPINA circumnavigation (Regaudie-de-Gioux et al., 2019), which present the advantage to sampled different ecosystem (from coastal to open ocean) with a standardized protocol.

Satellite estimates of the primary production are extracted from the Copernicus Marine Service (OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082). This product is given with a temporal resolution of one month, and a spatial resolution of 4 km.

Both datasets were compared in function of the longitude. This comparison points out that AM96 appears to reproduce the spatial distribution of the primary production relatively well. In addition to the spatial validation of the product, usual validation metrics were analysed. As primary production presents a lognormal distribution, statistics are given for data in log-scale.

Both datasets appear to be in agreement (Figure 10) with relatively low error and bias (RMSD=0.22, bias=-0.05), even if some discrepancies appear. These discrepancies can be linked with the difference in spatial and temporal scales, between in situ data and satellite observations.

Nevertheless, results from this validation give confidence in the use of this algorithm.

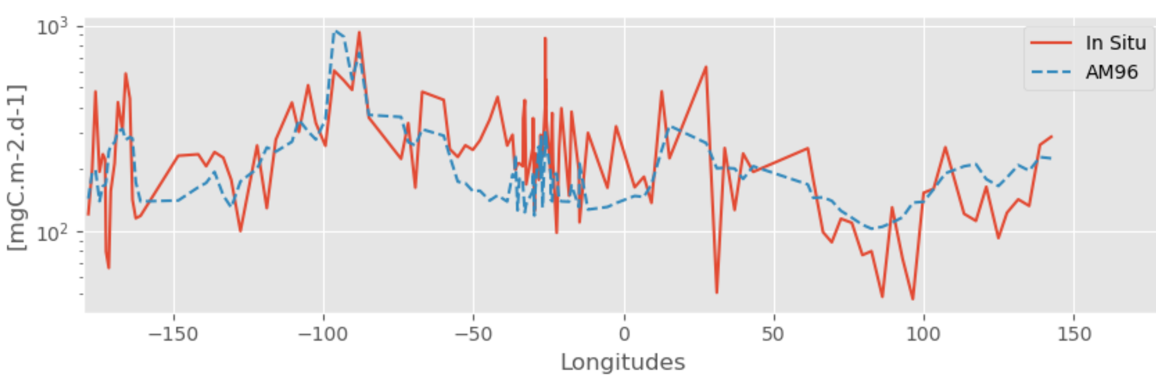

Figure 10. Primary production during the MALASPINA cruise. Red line corresponds to the in situ measurements and the blue dashed line correspond to satellite observations.

End of the document