



Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories - JERICO-NEXT	
<b>Deliverable title</b>	Novel methods for automated <i>in situ</i> observations of phytoplankton diversity and productivity: synthesis of exploration, inter comparisons and improvements
<b>Work Package Title</b>	WP 3
<b>Deliverable number</b>	D3.2
<b>Description</b>	Report on the technical and analytical improvements of innovative techniques and recommendations on their use
<b>Lead beneficiary</b>	CNRS
<b>Lead Authors</b>	Felipe Artigas, Véronique Créach, Emilie Houliez, Bengt Karlson, Fabrice Lizon, Jukka Seppälä, Guillaume Wacquet
<b>Contributors</b>	Hedy Aardema, Michael Brosnahan, Reinhoud de Blok, Pascal Claquin, Gérald Grégori, Florent Colas, Elisabeth Debusschere, Klaas Deneudt, Jacco Kromkamp, Soumaya Lahbib, Alain Lefebvre, Arnaud Louchart, Klas Möller, Emilie Poisson-Caillault, Thomas Rutten, Machteld Rijkeboer, Suvi Rytövuori, Lars Stemmann, Melilotus Thyssen, Lennert Tyberghein, Jochen Wollschläger and Pasi Ylöstalo.
<b>Submitted by</b>	Felipe Artigas
<b>Revision number</b>	5
<b>Revision Date</b>	August 12, 2019
<b>Security</b>	Public





History			
Revision	Date	Modification	Author
1	14/05/2019	Compilation and revision	G. Wacquet
2	14/05/2019	Revision	F. Artigas
3	01/07/2019	First complete version submitted	F. Artigas
4	31/07/2019	Last version revised with last addings	F. Artigas
5	12/08/2019	Last revision	F. Artigas
6			
7			
8			
9			

Approvals				
	Name	Organisation	Date	Visa
Coordinator	Patrick Farcy	Ifremer	14/08/2019	PF
WP Leaders	George Petihakis Laurent Delauney	HCMR Ifremer	13/08/2019	

### **PROPRIETARY RIGHTS STATEMENT**

THIS DOCUMENT CONTAINS INFORMATION, WHICH IS PROPRIETARY TO THE **JERICO-NEXT** CONSORTIUM. NEITHER THIS DOCUMENT NOR THE INFORMATION CONTAINED HEREIN SHALL BE USED, DUPLICATED OR COMMUNICATED EXCEPT WITH THE PRIOR WRITTEN CONSENT OF THE **JERICO-NEXT** COORDINATOR.





## Table of contents

1. Executive Summary .....	4
2. Introduction .....	6
3. Main report.....	8
3.1. Imaging in flow and <i>in situ</i> imaging of plankton.....	8
3.1.1. Overview of methodology and instruments commercially available.....	8
3.1.2. Results – developments, evaluations and experiences .....	16
3.1.3. Conclusions and recommendations for the use of automated image acquisition/analysis .....	25
3.1.4. Analytical improvements.....	25
3.2. Single-cell optical characterization.....	30
3.2.1. Short overview of methodology and instruments commercially available .....	30
3.2.2. Results and discussion - developments, evaluations and experiences .....	32
3.2.3. Conclusions and first recommendations for the use of automated flow cytometry .....	43
3.2.4. Technical improvements.....	43
3.2.5. Analytical improvements.....	46
3.3. Bio-optical Instrumentation.....	50
3.3.1. Short overview of methodology and instruments commercially available .....	50
3.3.2. Results and discussion - developments, evaluations and experiences .....	51
3.3.3. Technical and analytical improvements .....	63
3.3.4. Conclusions and recommendations.....	82
4. Conclusions and future work.....	83
5. References .....	84





## 1. Executive Summary

### Novel methods for automated *in situ* observations of phytoplankton diversity and productivity

This is a summary of the activities and results of JERICO-NEXT Work Package 3 Innovations in Technology and Methodology, Task 3.1 *Automated platform for the observation of phytoplankton diversity in relation to ecosystem services*. The aim is to provide an advanced report on the last developments dedicated to the observation of the phytoplankton diversity by applying novel techniques on automated platforms. The work was carried out in close connection with task 4.1 Biodiversity of plankton, harmful algal blooms and eutrophication. The partners involved in these developments are CNRS, SYKE, SMHI, HZG, RWS, VLIZ CEFAS and Ifremer. Subcontractors in WP4, task 4.1 are WHOI, Scanfjord AB, Tomas Rutten b.v., CytoBuoy b.v. and UGent - PAE.

The work was carried out mainly in the field with activities in the Baltic Sea, the Kattegat-Skagerrak, the Celtic seas-English Channel-North Sea Area, the Western Mediterranean, as well as in shared studies with other WP3.4 and WP4.4 in the Bay of Biscay and, out of Europe, in the Benguela Current. Instrument platforms included continuous recording (Ferrybox or assimilated) systems on research vessels, on ships of opportunity, instrumented oceanographic buoys/fixed platforms and land-based systems.

Common implementations were carried out in some observatories, allowing inter comparison of sensors and some techniques. In addition, work on developing and testing new algorithms have been carried out in offices and laboratories. Three international workshops have been successfully arranged, one in Wimereux, France (June 2016 – organised by CNRS-LOG), one in Gothenburg, Sweden (September 2016, organised by SMHI) and one in Marseille (March 2019 – organised by CNRS MIO) connected to a EuroMarine workshop. Partners presented, discussed, were able to inter compare the sensors and techniques to be implemented in the field (Goetborg) and were trained to different automated analytical procedures and tools (for automated flow cytometry, Marseille).

The work was divided into three sections but there is substantial overlap and cooperation. One example is that reference samples analysed in the microscope were used for completing and/or evaluating the quality of some of the automated methods.

#### **Imaging in flow and *in situ* imaging of plankton (led by SMHI)**

The work included evaluating instruments and developing algorithms for automated identification of phytoplankton automated image acquisition (in flow or *in situ*). Three different commercial instruments and one instrument prototype were used. On the Swedish west coast (Skagerrak coast) a study of harmful algae and other phytoplankton was carried out near a mussel farm. The Imaging Flow Cytobot was deployed *in situ* and collected samples at six different depths for approximately two months. In the English Channel the old generation of FlowCAM and a prototype system, the FastCam, were used to analyse samples on research vessels or in the laboratory. A colour version of FlowCAM was used both during a cruise in the Baltic Sea-Kattegat-Skattegat (July 2017) as well as in coastal monitoring in the Baltic and the Channel). In addition, the CytoSense and CytoSub were used to collect images in flow.

The *in situ* video system UVP5 was implemented during a cruise in the Baltic Sea-Skagerrak-Kattegat area, together with a new generation of FlowCAM of faster acquisition and providing colour images and CytoSense.

A major task was to develop and evaluate plankton identification algorithms. This included the use of a subset of images of organisms for training the software tools. Existing software were improved (as the PhytoZoolmage) and an image data system/platform named EcoTaxa was described and is currently available for storing and cooperative analysis/discrimination of plankton images.

#### **Single-cell optical characterization (led by CEFAS)**

Automated flow cytometers (FCM, CytoSense/CytoSub, Cytobuoy b.v.) were implemented on a Ferry line, on research vessels and a fixed platform to investigate functional groups of phytoplankton. In the Western Mediterranean the main targets were the picoplankton and the nanoplankton while in the other areas pico-, nano- and microplankton were in focus. Several cruises were carried out in the Channel and North Sea to follow combined diatoms and *Phaeocystis* bloom development. A cruise covering the Baltic Sea and Skagerrak-Kattegat area had a main focus on cyanobacteria and dinoflagellates. Moreover, inter comparisons of machines and on clustering analysis methods were performed. Finally, a combination of FCM and multi-spectral fluorometer continuous recording was coupled with physical and hydrological continuous measurements in the southern Bay of Biscay.





### **Bio-optical Instrumentation (led by SYKE)**

Novel multi-wavelength fluorometers for detecting phycoerythrin indicative e.g. of certain cyanobacteria and of cryptophytes were evaluated in the Baltic Sea. Multi wavelength fluorometers were also used in the Benguela current, during the Gothenburg workshop, as well as on a variety of field cruises from the southern Bay of Biscay to the E. Channel and North Sea, in order to discriminate amongst main phytoplankton pigmentary groups. The manufacturers' algorithms were found to be partly inaccurate for detecting algal groups based on photosynthetic pigment composition. New dedicated fingerprints were used in field work to improve discrimination amongst phytoplankton groups. A principle component analyses approach was also evaluated.

Single wavelength fluorometers were evaluated in several sea areas. Sun induced photoquenching had a strong effect on fluorescence yield. In the North Sea and the Norwegian Sea multi spectral absorption was used to detect chlorophyll and phytoplankton groups based on pigment content.

Variable fluorometers were implemented on both samples, continuous recording and profiles in the E. Channel and North Sea, as well as in the Baltic and Skagerrak-Kattegat, for studying photosynthetic parameters and potential primary productivity. Recommendations are made on the strategy and type of measurements to carry out.

### **Recent work (since mid 2017) in task 3.1**

Some field work was continued, especially at the Utö observatory in the Baltic Sea. The new data collected was processed together with old data and used for improving the discrimination of phytoplankton taxa or functional groups by inter-comparison of techniques and continued algorithm development which are described in the present deliverable 3.2. Scientific publication of results is in progress. A special issue in the open access journal Diversity (MPI) is being discussed. Some results and strategy were presented during a symposium in Hannover, in October 2017 and during the FerryBox workshop on board the ship Colour Fantasy later in October 2017 and in FerryBox meetings in 2018 and 2019. Results were also presented during the third JERICO-NEXT workshop on automated plankton observation in Marseille in March 2018 and some analytical tools for flow cytometry were presented and further directions as well as connections with modelling and remote sensing were discussed during the EuroMarine meeting that followed. A special session was held during the International Conference on Harmful Algae (ICHA) in Nantes in October 2018, and more presentations were carried out in 2019 meetings (IMBER, OceanObs, etc.).

### **Main conclusions**

1. The methods used are reliable for automated observation of phytoplankton biodiversity (functional groups, size classes, taxa when possible) and biomass, complementing manual methods for sampling and microscope analyses.
2. Operating the equipment and interpreting the results still need a lot of knowledge and time. Even though some operational procedures can be established, the standardization of analytical and data processing as well as data management need more development. The degree of automation varies depending on the method considered.
3. Imaging in flow and *in situ* imaging provide means for identifying and counting phytoplankton at the genus or species level. Also, biomass based on cell volume of individual cells can be estimated. Development of classifiers for automated identification of organisms is time consuming and requires specific skills on signal analysis and on taxonomy.
4. Flow cytometry has proven to be a useful tool for counting phytoplankton and for describing the phytoplankton community as size-based classes and functional groups. There was an agreement to report the phytoplankton count in four groups for inter comparison purposes: *Synechococcus* (pico-cyanobacteria), pico-eukaryotic organisms, nanoplankton and microplankton.
5. Single and multi-wavelength fluorometry makes it possible to estimate phytoplankton biomass (at a chlorophyll-a basis) and to differentiate phytoplankton based on photosynthetic pigments. Sunlight induced photo-quenching is a problem for estimating chlorophyll *a* from fluorescence. For instruments mounted buoys or vessels, night time data can be used to minimize the problem.
6. Multi-wavelength absorption is a useful tool for estimating chlorophyll *a* and is also useful for discriminating between phytoplankton groups based on pigment content.
7. Variable (active) fluorescence is available for addressing phytoplankton physiology, photosynthetic parameters and we could estimate primary productivity on both continuous sub-surface recording and water column profiles, mediating careful coupling with other optical and also biogeochemical analysis.



## 2. Introduction

Phytoplankton forms the base of the marine food web. The number of phytoplankton taxa in the sea have been estimated to be over 10 000. All of them are primary producers but the ecological function of the different taxa varies. Many species can not only utilize light as an energy source but also feed on other organisms. Some of the species are harmful, e.g. producing phycotoxins that may accumulate in sea food and pose a threat to human health. Phytoplankton vary in size and shape; the size range is approximately 0.8  $\mu\text{m}$  to 0.5 mm. Colonies of cells may be a few mm in size. Traditionally phytoplankton is monitored by collecting water samples and analysing them manually using microscopy. This is a good but labour-intensive method. The last few decades novel methodologies have been developed to be able to process a much larger number of samples compared to microscopy and to do it automated and autonomously. The novel methodologies include optical methods and also molecular biological methods described in JERICO-NEXT deliverable 3.7 Progress report after development of microbial and molecular sensors. An overview of current methods is presented in table 1. Remote sensing is outside the scope of JERICO-NEXT.

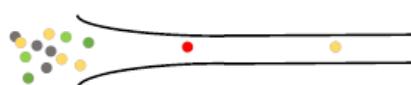
The aim of this report is to describe results from JERICO-NEXT on the development and evaluation of novel methodology for observing phytoplankton *in situ*. There are three main approaches used:

1. Imaging in Flow systems (Imaging Flow Cytometry) - Describing the phytoplankton composition based on morphology by imaging individual cells. Describing the plankton community imaging organisms and colonies of cyanobacteria in the free water mass *in situ*.
2. Single-cell optical analysis (Pulse shape-recording Flow cytometry) - Describing the phytoplankton composition based on the fluorescence properties (pigment content) and scattering properties of individual cells.
3. Bulk optical approaches (multi-spectral Fluorescence or absorption; variable/active fluorescence) – describing the phytoplankton community based on bulk properties: fluorescence or absorption of a large number of cells. Multi wavelength approaches makes it possible to differentiate pigment groups of microalgae, whereas variable fluorimetry addresses photosynthetic parameters and potential productivity.

**Imaging flow cytometry**  
Single cells – morphology of organisms



**Flow cytometry**  
Single cells – fluorescence and scattering



**Fluorescence and absorption**  
Pigment based methods – bulk properties

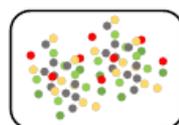


Figure 2.1. Overview of three methods for observing phytoplankton.



Table 2.1. Overview of methods for observing phytoplankton biomass, abundance and biodiversity. Not all methods are being used in JERICO-NEXT

Method	Biodiversity	Biomass estimates	Functional groups	Sample throughput	Level of automation	Horizontal coverage
Light microscopy	Good	Good	Good	Low	Low (automated water sampling is available)	Low
Fluorescence microscopy	Medium	Medium	Medium-Good	Low	Low	Low
Electron microscopy	Very good	Medium	Low	Very low	Low	Low
Flow cytometry	Low-Medium	Medium	Good	High	Semi-automated on research vessels and/or in fixed stations	Medium (Ferrybox)
Imaging flow cytometry	Medium-Good	Medium-Good	Good	Medium	Semi-automated on research vessels and/or in fixed stations	Medium (FerryBox)
Gene probes	Medium (only a limited number of species)	Low-Medium	Medium	Medium	Semi-automated (ESP)	Low-medium
Barcoding	Good	Low	Medium	Medium	Automated sampling and preservation in development	Low-medium
Chlorophyll a analyses (water sampling)	-	Medium	-	Low-Medium	Low (automated water sampling is available)	Low-Medium
HPLC-analysis of photosynthetic pigments	Low-medium	Medium	Medium (pigment groups)	Low	Low (automated water sampling is available)	Low
In vivo fluorescence methods based on the fluorescence of photosynthetic pigments	Low-Medium	Medium	Low-Medium	High	High	High
Methods based on the absorbance of photosynthetic pigments	Low-Medium	Medium	Low-Medium	High	High	Medium-high
Satellite remote sensing (ocean colour-reflectance of photosynthetic pigments)	Very low	Medium	Low	High	High	High (during cloud free conditions)



### 3. Main report

#### 3.1. Imaging in flow and *in situ* imaging of plankton

Guillaume Wacquet & Felipe Artigas (CNRS-LOG), Bengt Karlson (SMHI), Florent Colas & Alain Lefebvre (IFREMER), Machteld Rijkeboer (RWS), Lars Stemmann (CNRS-LOV)

##### 3.1.1. Overview of methodology and instruments commercially available

###### 3.1.1.1. Imaging FlowCytobot (McLane Research Laboratories) - SMHI



**Instrument.** The Imaging FlowCytobot (IFCB) is an in-situ automated submersible imaging flow cytometer that generates high resolution (1380x1034 pixels) images of suspended particles in-flow, in the size range <math><10</math> to 150  $\mu\text{m}$  (such as diatoms and dinoflagellates).

The instrument continuously samples at a rate of 15 ml of sea water per hour, and, depending on the target population, it can generate on the order of 30,000 high resolution images per hour.

IFCB uses a combination of flow cytometric and video technology. Laser-induced fluorescence and light scattering from individual particles are measured and used to trigger targeted image acquisition; the optical and image data are then transmitted to the computer in real time, through an Ethernet communication.

Figure 3.1.1 The imaging FlowCytobot with the pressure proof housing removed.

**Software.** Collected images during continuous monitoring can be processed externally thanks to a specific automated image classification software. For each imaged particle,  $\sim 200$  different parameters are extracted. Images can then be automatically classified to the genus or even species level with demonstrated accuracy comparable to that of human experts (Olson and Sosik, 2007).

This software is open source and developed in Matlab (license needed). However, a Python version is in development.

###### 3.1.1.2. FlowCAM (Fluid Imaging Technologies)



**Instrument.** As for Imaging FlowCytobot and instruments developed by CytoBuoy, the FlowCAM combines selective capabilities of different technologies: flow cytometry, optical microscopy and fluorescence detection. It can generate high resolution (1280x960 pixels) images of particles in-flow, in the size range 2 $\mu\text{m}$  to 2000 $\mu\text{m}$  (depending on the combination « magnification/flow cell » used for the optical system, ie. 2X/600 $\mu\text{m}$ , 4X/300 $\mu\text{m}$ , 10X/100 $\mu\text{m}$  or 20X/100 $\mu\text{m}$ ).

Figure 3.1.2. A benchtop version of the FlowCam. There are other versions available.

The sample introduced in the device is attracted by a peristaltic or a syringe pump into a flow cell (or flow chamber) with known dimensions, located in front of a microscope objective which is connected to a camera video.

Two mode of detection can be used: « Trigger » or « AutoImage ». For the first one, when a particle passes through the laser, the scattering of laser light is measured and a value of fluorescence is calculated and compared with a



fluorescence threshold value. If the obtained value is higher, the camera is triggered to take an image. For the « Autolmage » mode, the particles in the field of view are imaged and captured at a regular user-defined interval. Principle of FlowCAM is illustrated on Figure 3.1.3.

**Software.** « *Visual SpreadSheet* » is the software provided with FlowCAM. It is essential for all the major aspects of sample analysis: setup for data acquisition through the context settings, for controlling the device, managing files and setting preferences; data acquisition and post-processing of collected data. For each particle, the software provides a set of 26 image parameters.

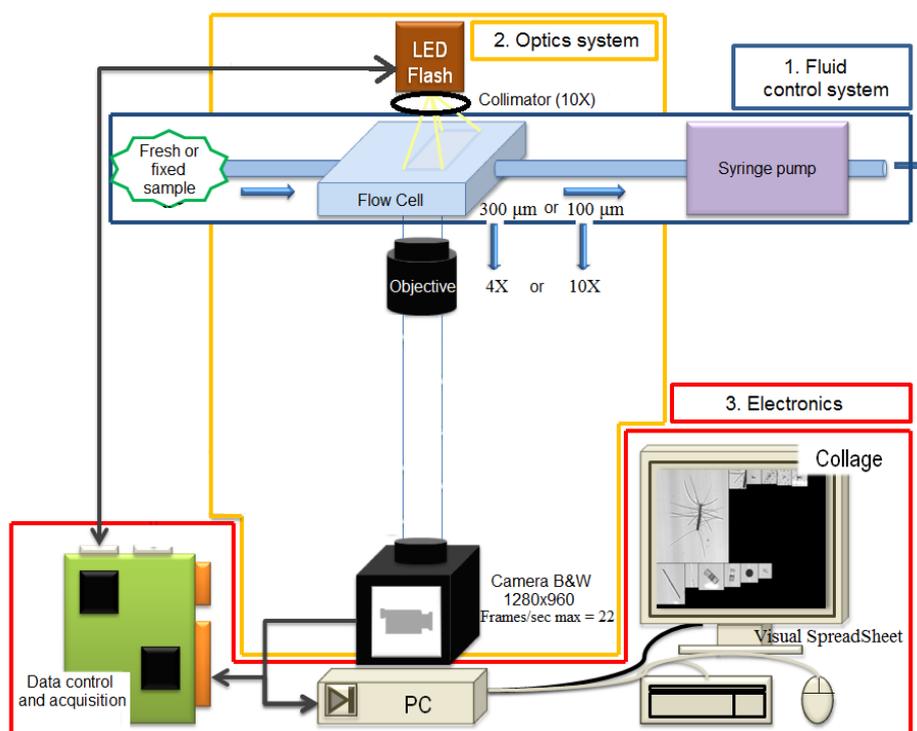
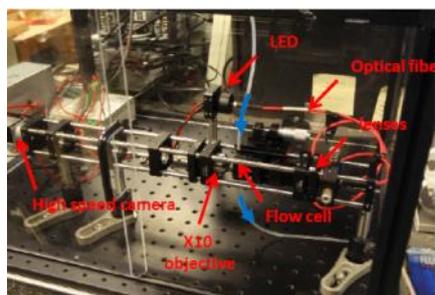


Figure 3.1.3. Overview of the FlowCAM principle.

### 3.1.1.3. FastCAM prototype (IFREMER - LDCM)



**Instrument.** The FastCAM system is based on a high resolution (2 Megapixels) and high speed camera allowing the acquisition to 340 frames per second. It digitizes 10 mL of sample with a X10 magnification within only 15 min (which is not possible with the first generation of FlowCAM devices). Comparison of greyscale images with those obtained with the first generation of FlowCAM showed that this new system analyses samples much faster with high image quality. (Figure 3.1.4).

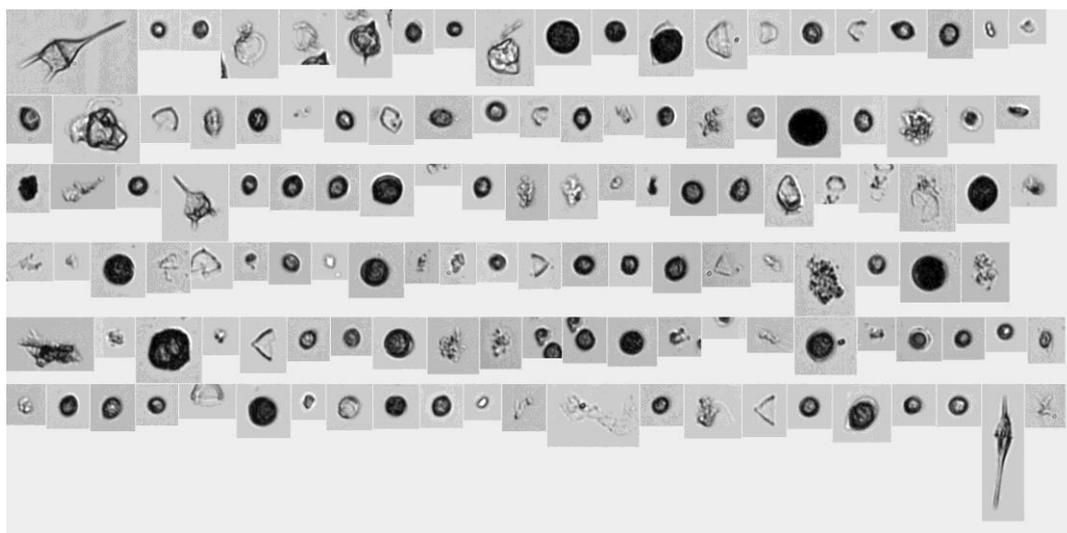


Figure 3.1.4. Thumbnails generated by the FastCAM during the JERICO-NEXT Phytoplankton workshop in Gothenborg, 27-30 September 2016.

An LED driven by a control box emits light pulses of 5  $\mu$ s duration. Light is injected into a large core diameter (1 mm) optical fibre to homogenize the beam. Light out of the optical fiber illuminates the flow cell. A 10X magnification microscope objective associated with a tube lens images the organisms that circulate in the flow cell on a camera. The frame grabbing is synchronized with the LED emission. A pixel of the image corresponds to 0.5  $\mu$ m. The images are saved on the PC in real time thanks to a fast hard drive. Principle of FastCAM is illustrated on Figure 3.1.5.

**Software.** For the image acquisition, a specific software is developed in Visual Basic 12. A second software developed in C is used for image processing. Thanks to the « Matrox MIL 10 » library, ~50 parameters are computed on each image. These parameters then are used to classify images by applying existing classification tools, like « Plankton Identifier » (Delphi and Tanagra environments) or « Zoolmage » (R environment).

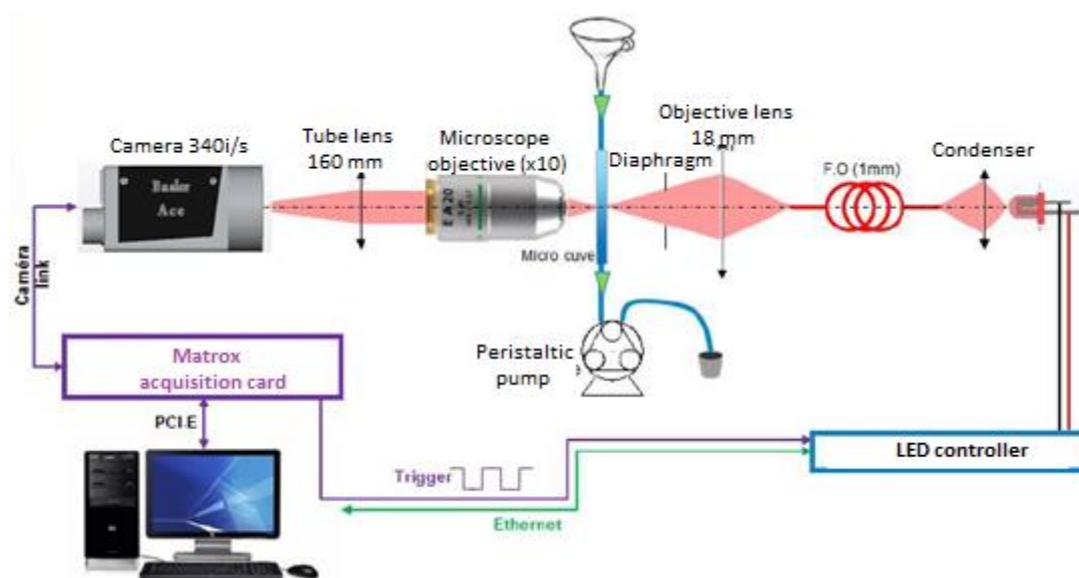


Figure 3.1.5. Overview of the FastCAM principle.

#### 3.1.1.4. Underwater Vision Profiler UVP5 (Hydroptic) - CNRS LOV



**Instrument.** The UVP5 images large plankton (equivalent spherical diameter, ESD >600  $\mu\text{m}$ ; Picheral et al., 2008) usually metazoans but also large unicellular organisms or colonies of those (diatom mats, rhizarians and prokaryotes such as cyanobacterial colonies) (Biard et al., 2016; Guidi et al., 2012). The UVP5 sampling volume varies from 0.5 to 1 L and images are recorded every 5 to 20 cm along vertical profiles, leading to an observed volume of 1  $\text{m}^3$  for a 100 m depth profile. Mounted on a CTD rosette frame, the UVP5 starts recording below a few meters, eventually leading to an underestimation in the quantification of objects just beneath the sea surface.

Figure 3.1.6. The Underwater Vision Profiler UVP5.

**Software.** Images produced by the UVP5 are extracted using the ZooProcess software. Image identification is possible for objects larger than 600  $\mu\text{m}$  (total number ~100000 during the Baltic Sea cruise). A computer-assisted method is used to classify all organisms with a Random Forest classification. All images are checked using the Ecotaxa web application (<http://ecotaxa.obs-vlfr.fr/>) by experts to discriminate plankton (including cyanobacterial colonies) from other plankton and detritus. Differences in shapes and grey level were used to distinguish between categories.

#### 3.1.1.5. CytoSense and CytoSub instruments (CytoBuoy b.v.) - CNRS LOG, MIO, RWS



**Instrument.** Each particle intercepts a laser beam and the generated pulse shape of optical properties (two scatters, up to three fluorescences) induced by the particle are recorded. Pulse shapes recording allow chains forming cells to be recorded. Increase in laser power and optimisation of sheath cleaning enable the resolution of *Prochlorococcus*.

An image-in flow device records pictures of preselected groups of cells, resolving cells at its best above 20  $\mu\text{m}$  but able to collect pictures of 2  $\mu\text{m}$  beads (but with low resolution). Particles are recorded above a defined threshold (scatter or fluorescence) and phytoplankton cells are separated from non-photosynthetic particles thanks to their red auto-fluorescence. The CytoSense sensors are adaptable on ships of opportunity and scientific vessels, whereas the submersible version (CytoSub) fits in fixed stations and buoys, running samples from a subsampling dedicated system isolating sea water from a continuous flow of pumped sea water.

Figure 3.1.7. The CytoSense.

**Software.** Recorded samples are analysed using dedicated software for manual and automatic clustering. For manual clustering, each particle is represented on two dimensional cytograms. The different cytograms available makes the manual clustering possible. CytoBuoy company did build its own manual clustering software (CytoClus). However, today, CytoClus does not process images. It is why automated methods using images are in development for size calibration, species recognition in microphytoplankton, and cells counting in colonies. These new functionalities will be integrated to the RclusTool package (R environment) developed by the LISIC laboratory in collaboration with CNRS LOG (ULCO).

### 3.1.1.6. ECOTAXA application (Institut de la Mer de Villefranche and Biological Station of Roscoff)



ECOTAXA is a web based application for collaborating on large plankton image datasets. Once the images collected, they can be uploaded and stored on the ECOTAXA server. Today, it is considered as the most important worldwide dataset of annotated plankton images.

ECOTAXA handles images of individual organisms, proposes identifications using « machine learning » (algorithms developed in Python), and keeps all metadata associated with each image, from the acquisition to the final identification. However, the founding principles of ECOTAXA are that:

- (i) the identification of organisms is collaborative, through the internet,
- (ii) every change is explicit and recorded in a robust relational database (including the simple confirmation of a correct identification),
- (iii) identifications are based on a universal taxonomy that allows to link the morphology of organisms with genomic information,
- (iv) ECOTAXA can easily import image datasets from any instrument (UVP5, ZooSCAN, FlowCAM, ZooCAM, HCS1, ISiS, IFCB, automated microscope, etc.).

The application already hosts over 80 million images of plankton from 450 users in 105 institutes, and about 30% which have been verified by experts. It also uses classification tools to assist taxonomist classifying large datasets: « Random Forest » and « Deep Learning ». Short overview of ECOTAXA is shown on Figure 3.1.8.

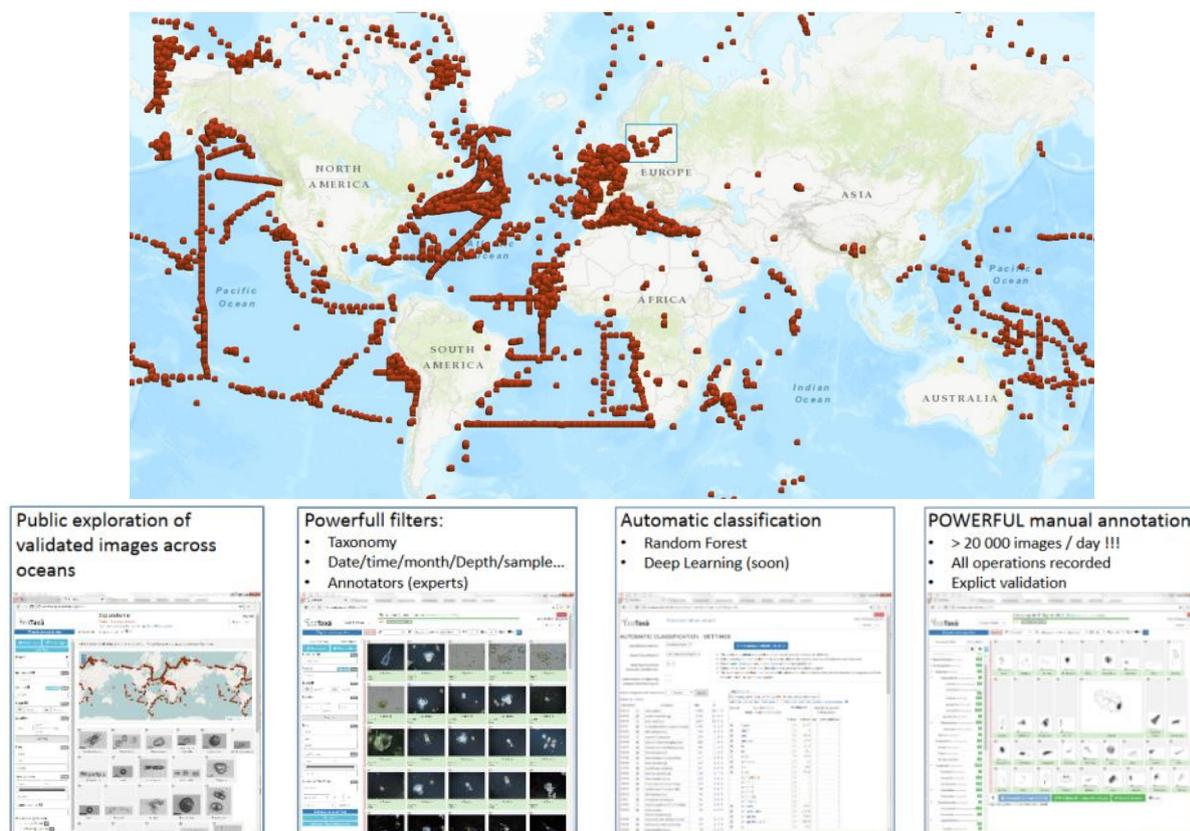


Figure 3.1.8. Above: Map showing the data coverage into ecotaxa. More than 1000 images of plankton are observed at each of these locations. Below: Screenshots of the EcoTaxa application.

### 3.1.1.7. Zoolmage software (University of Mons)



« Zoolmage » is a software developed by the University of Mons (UMONS) in collaboration with IFREMER, for image processing and particle classification. It is a computer-assisted plankton image analysis software for predicting taxonomic identification of plankton samples. Zoolmage is an open source software bundled with Java-based ImageJ and R, statistical software.

It can be modified to meet the user requirements and constraints defined for monitoring networks and/or campaigns, and accommodate many different imaging systems (FlowCAM, FastCAM, CytoSense, etc.). The parameters extracted from each image are used by a classification tool to automatically predict the taxonomic group and, when possible, the species the particles belong to (Figure 3.1.9).

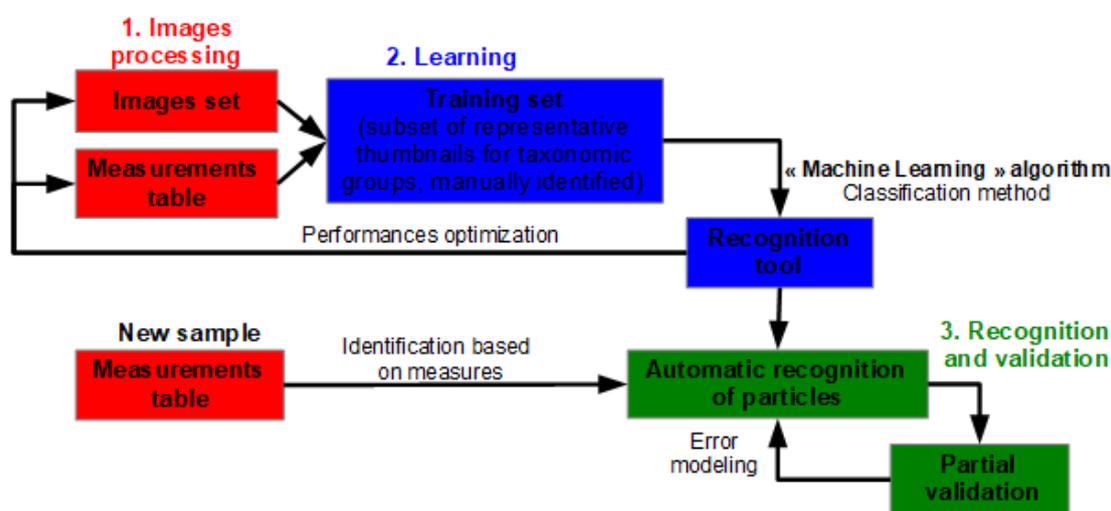


Figure 3.1.9. Overview of the classification process in Zoolmage.

Thanks to an ergonomic Graphical User Interface, it is possible to easily explore data through simple statistical descriptions (number of particles, number of samples, size distribution for each sample, etc.), visualization of particles images in the sample and measurements table associated to each image (Figure 3.1.10).

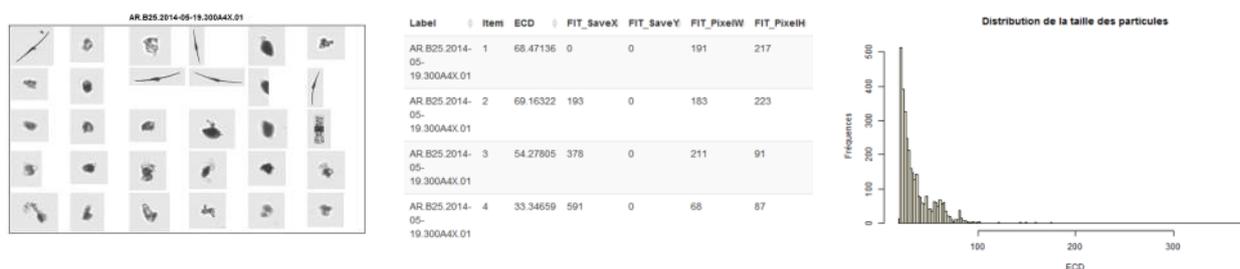


Figure 3.1.10. Examples of sample description tools in Zoolmage.

During the classification process in Zoolmage, it is possible to use « contextual » samples in order to adapt the training set, and consequently, the associated recognition tool, to the current sample to analyse. According to the



set of contextual samples selected, the active learning can lead to a significant reduction of the prediction error, and to a considerable time saving concerning the manual validation process for the error correction. Intuitively, the contextual samples selection (which contain groups of particles already validated) can be based on different criteria: same geographical area, same period (+/- one week, +/- one month, ...), same digitization protocol, etc. Once the selection done, the active learning process is run, and the groups in the initial training set are completed with the new particles from these contextual samples. This adapted training set is then used for the sample analysis. Classification results are then presented in terms of abundances, biovolumes and size spectra for each taxonomic group (Figure 3.1.11).

```
--- AR.B25.2014-05-19.300A4X.01 ---
Échantillon contenant 3329 particules numérisées.
                                Id Abd Cerataulina Abd Ceratium_furca+lineatum
1 AR.B25.2014-05-19.300A4X.01                                7600                                19600
  Abd Ceratium_fusus Abd Chaetoceros_cf_curvisetus Abd Ciliophora
1                                32300                                11700                                19300
  Abd Dinophysis_spp Abd Lepididinium_mucous_shape Abd Prorocentrum_spp
1                                11800                                11600                                21500
  Abd Scrippsiella Abd Thalassiosira_spp Abd [other] Abd [total]
1                                10200                                22100                                165200                                332900
  Bio Cerataulina Bio Ceratium_furca+lineatum Bio Ceratium_fusus
1                                241632                                832570.8                                1979526
  Bio Chaetoceros_cf_curvisetus Bio Ciliophora Bio Dinophysis_spp
1                                287478.1                                556080.4                                422519.9
  Bio Lepididinium_mucous_shape Bio Prorocentrum_spp Bio Scrippsiella
1                                540673.2                                587245                                242234.3
  Bio Thalassiosira_spp Bio [other] Bio [total]
1                                570554.5                                5907500                                12168014
```

Figure 3.1.11. Example of analysis results for a sample in Zoolmage.





Table 3.1.1 summarizes the main technical specifications of each imaging devices, but also the processing and classification software used for each of them.

Device	Size range (um)	Sampling rate (mL/h)	Image size (pix)	Resolution (µm/pixel)	Magnification of lens	Color module	Processing/Classification software	Environment	Training set available	Results validation
<b>IFCB</b>	<10 to 150	15	1380x1034	~0.3	x10	no	IFCB Image Analysis and Classification software	Matlab/Python	North Sea (Skagerrak), the Baltic Sea, US East coast, US West coast	Partial manual validation
<b>FlowCAM</b>	<10 to 2000	108 9	1280x960	1.3462 0.5515	x4 x10	no	Visual SpreadSheet	not defined	no	no
							ZooImage	R/Java	Eastern English Channel (2013-2014)	Partial manual validation
							ZooProcess/Plankton Identifier	Delphi/Java	no	Total manual validation
<b>FastCAM</b>	10 to 1000	54	2048x1088	0.5	x10	no	ZooImage	R/Java	no	Partial manual validation
							ZooProcess/Plankton Identifier	Delphi/Java	no	Total manual validation
<b>UVP5</b>	>600	up to 20 Hz, 20L every meter of water column	2048x2048	n/a	n/a	no	Zooprocess/ECOTAXA	ImageJ/Python/postsql	Baltic sea (in 2018)	Total manual validation (in 2018)
<b>CytoSense/Sub (old IIF system)</b>	>20	3.6 to 72	1280x1024	0.6	X11	no	CytoClus	.NET/VisualBasic	no	no
							EasyClus	Matlab	no	no
							RclusTool	R	Eastern English Channel (2016-2017)	no
<b>CytoSense/Sub (new IIF system) Since 2017</b>	>20	3.6 to 97	1920x1200	0.3	x16	no	CytoClus	.NET/VisualBasic	no	no
							EasyClus	Matlab	no	no
							RclusTool	R	Eastern English Channel (2017)	no

### 3.1.2. *Results – developments, evaluations and experiences*

#### 3.1.2.1. **Development of classifiers for an automated identification of plankton using the Imaging Flow Cytobot - SMHI**

The Imaging FlowCytobot (IFCB, Figure 3.1.12) can be described as an underwater microscope. The instrument was originally developed at Woods Hole Oceanographic Institution (WHOI) in the USA (Sosik and Olsen, 2007 and Olsen and Sosik 2007). It has been further developed since then and is available commercially from Mclane Inc. The main characteristics are described in Table 3.1.1. In JERICO-NEXT an instrument has been provided by WHOI as part of subcontracting to SMHI. The IFCB is a true flow cytometer in the sense that sheath fluid is used to focus a very narrow flow of water sample water containing the plankton organisms. A laser provides excitation light for chlorophyll fluorescence and the emission triggers a camera. Essentially every organism containing chlorophyll, also micro zooplankton with phytoplankton prey inside, is imaged. A sample of 5 mL results in several thousands of images, sampling every 25 minutes for a few months results in millions of images. Thus, automated image analysis is a requirement to work with the data efficiently.



Figure 3.1.12. An automated underwater microscope, the Imaging FlowCytobot was used next to a mussel farm in Sweden August to October 2016. An automated winch was used to move the instrument to different depths. The collage of images to the right shows images collected using the IFCB illustrating some of the phytoplankton observed.

A simplified description on the steps needed to develop and use automated phytoplankton analyses using the IFCB are described in Figure 3.1.13. IFCB images may be classified manually and/or automatically by a computer. Both approaches are facilitated by a suite of open source, MATLAB-based software (Sosik and Olson 2007; <https://github.com/hsosik/ifcb-analysis>) and have been applied to images collected during a deployment of an IFCB at a mussel aquaculture facility in Tångesund, Sweden. Manual annotation involves review and sorting of images into a set of user-defined categories, typically representing genus- or species-level taxonomic groups. Computer-based classification applies a random forest based machine that is trained from sets of training images. Each image within these training sets is processed by computer to determine cell boundaries and to numerically evaluate a set of 237 features (length, width, orientation, etc.). The distribution of these feature attributes is assessed by class during construction of the random forest based classification machine. Once created, the random forest classification machine assigns posterior probabilities of class affiliation to features extracted from new images in an IFCB data set. Classification decisions may be made by assigning images to the class having the highest posterior probability or through application of probability thresholds that vary by class. The two approaches have trade-offs with respect to error sources. The former approach tends to improve the probability that examples of

a given class will be detected but also increases rates of misclassification. Use of optimized thresholds tends to reduce misclassification rates but lowers the probability of cell detection. In practice, the performance of the computer based classifier varies with the changes in the attributes of a given population. Changes in cell size or co-occurrence of highly similar species may lead to higher rates of misclassification or lower rates of detection. In most cases, best practice is to combine the manual and computer based approaches so that changes in computer-based classifier performance can be evaluated through time.

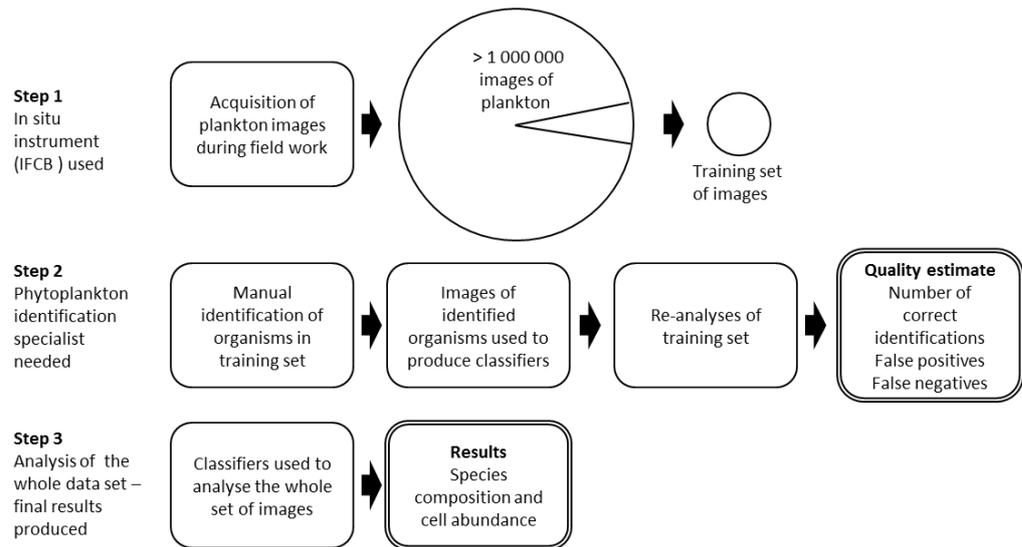


Figure 3.1.13. An overview of the steps needed for automated phytoplankton analyses using the Imaging Flow Cytobot.

During the Tångesund study next to a mussel farm on the West coast of Sweden the IFCB was used to observe phytoplankton with a focus on harmful algae. A summary of the results related to the development of automated classifiers for automated detection of phytoplankton is presented here. A selection of the large data set of images was used for manual identification of the organism in the images. The organisms chosen (see Figure 3.1.14) were either very abundant or harmful algae. The classification works best if at least 500 images of an organism is used to produce the classifiers. This was reached for some of the chosen organisms but not for all. It should be noted that a specialist in identifying the local phytoplankton is needed to correctly identify the organisms. The process of assessing many hundreds of images is time consuming but when complete automated analysis is possible. The optimal score threshold for the evaluated organisms is shown in Figure 3.1.14. One of the most abundant harmful algal bloom (HAB) organisms present was the dinoflagellate *Lingulodinium polyedrum* (syn. *Gonyaulax polyedra*). An evaluation of the quality of the data is presented in Figure 3.1.15 together with the time series of the depth distribution of *L. polyedrum* from mid-August to mid-October.

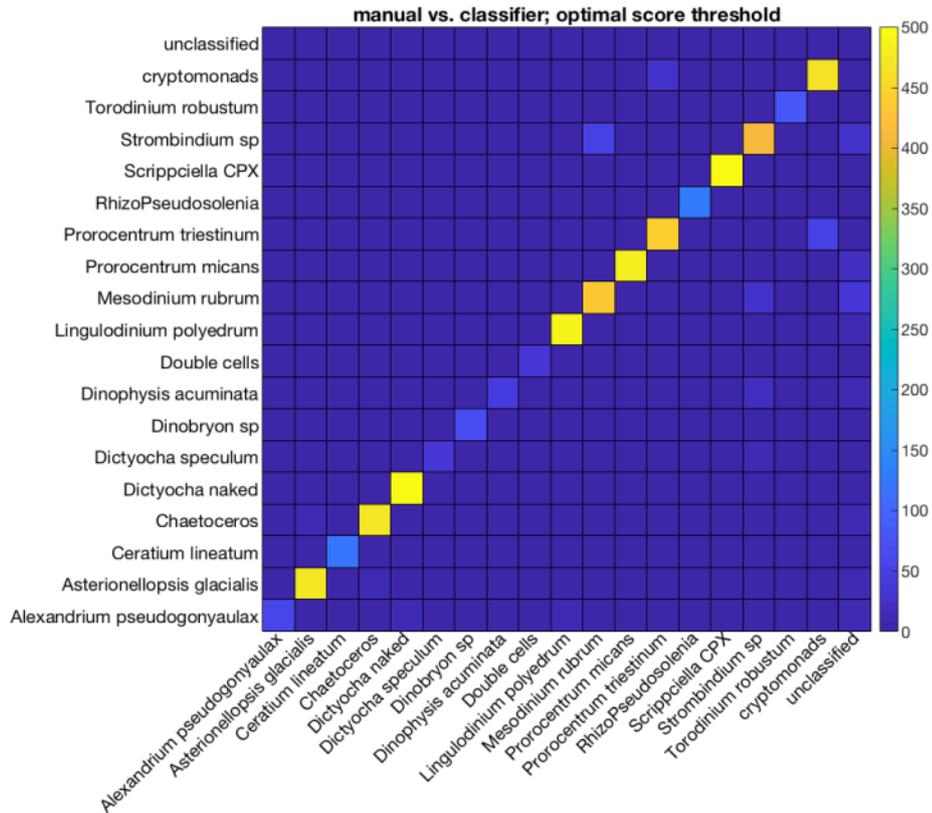


Figure 3.1.14. The graph illustrates the accuracy of automated identification of selected phytoplankton species using classifiers developed using the Tångesund data set. A higher value of optimal score threshold, i.e. closer to 500, indicates a more accurate detection.

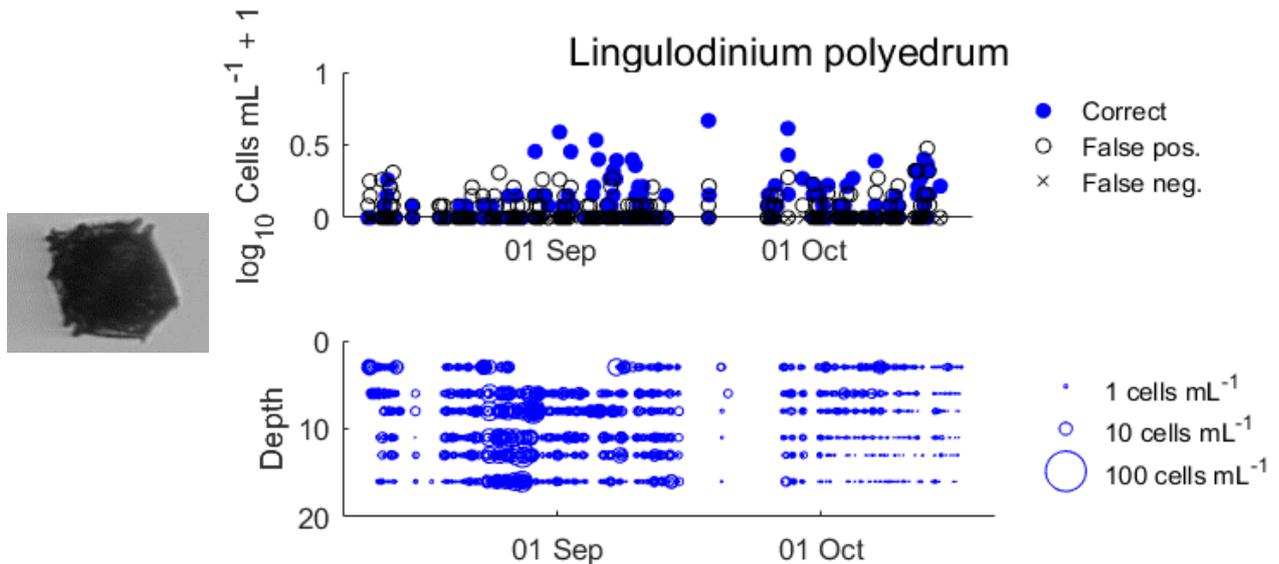


Figure 3.1.15. An example of results using automated identification and quantification of a harmful algae during the Tångesund study in 2016. *Lingulodinium polyedrum* is a producer of yessotoxins that may accumulate in shellfish. Upper panel: The cell abundance of *L. polyedrum* as estimated using the IFCB. Lower panel: A time series of the cell abundance of *L. polyedrum* at six different depths.



### 3.1.2.2. FlowCAM-Zoolmage training set - CNRS-LOG, IFREMER

**Material and methods.** A living training set representative of each plankton community met in the Eastern English Channel was built using samples taken throughout 2013 (in the frame of the IFREMER Regional Nutrients and REHY Monitoring networks) and 2014 (in the frame of the CNRS LOG DYPHYRAD transect). Samples were digitized using an 8-bit greyscale benchtop FlowCAM® VS, for which the pump speed was set at 1.8ml.min<sup>-1</sup>. A 4X objective (40X overall magnification) coupled with a 300µm-depth flow-cell was used and samples were run in “Autolmage” operation mode. The analysis process, from raw images processing to statistical analysis, was carried out using the Zoolmage 3.0-5 R package (Grosjean & Denis 2014).

A total of 3585 images were manually classified into 29 plankton groups. Moreover, instead of manually removing detritic particles and artefacts as it is commonly done (Zarauz *et al.* 2007), we added twelve groups for floating dark and light dead particles, bubbles, fibers, etc., to the thirty-seven plankton groups. Finally, 5154 images were sorted into 40 groups (Figure 3.1.16). From this training set, a classifier was trained using the “Random Forest” algorithm (Breiman 2001, Fernandez-Delgado *et al.* 2014). Global error measured by using 10 folds cross-validation is thus equal to 25.89% for these groups.

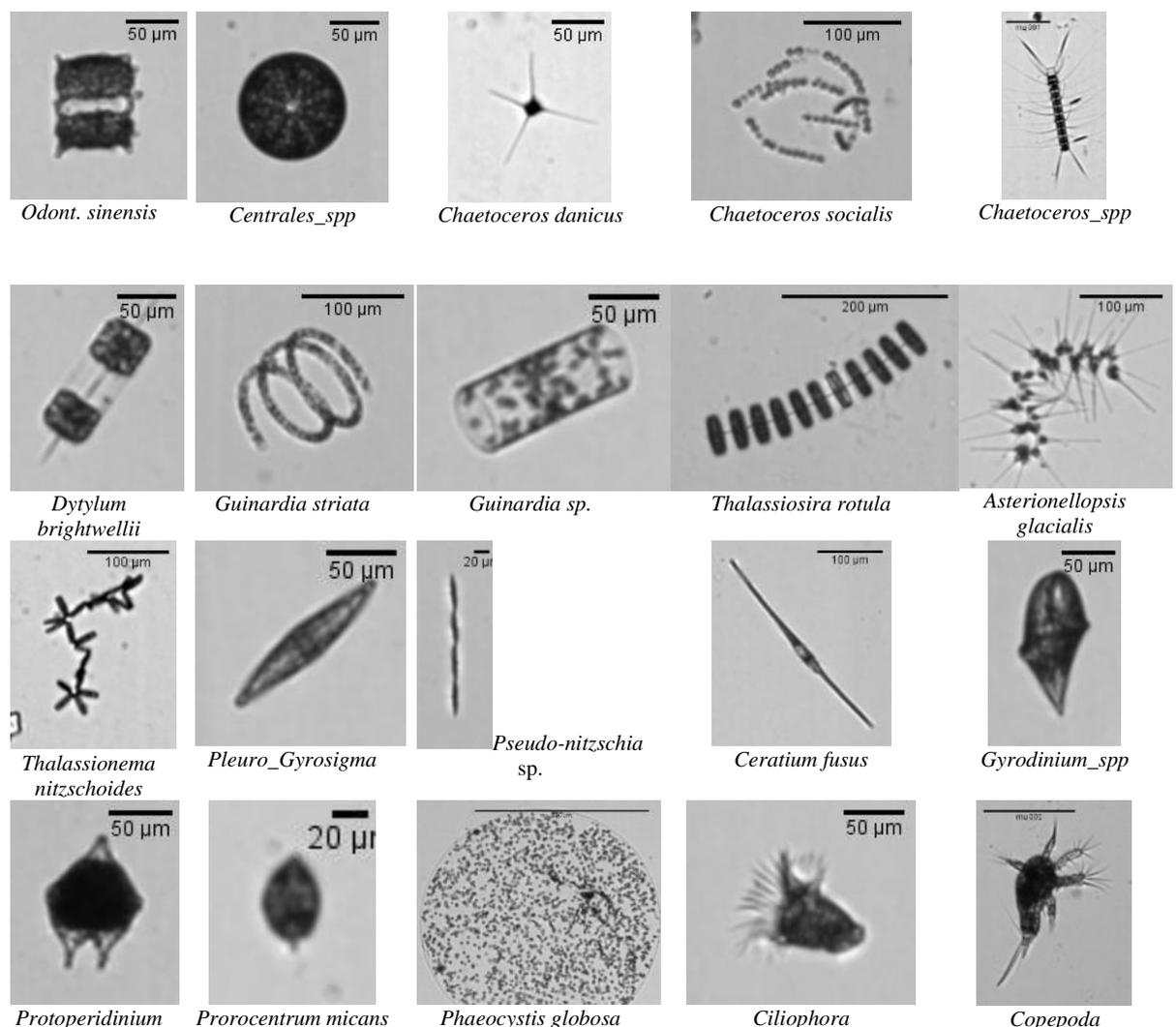


Figure 3.1.16. Examples of images in some taxonomic groups in the FlowCAM training set.



**Results.** Abundance and error rates per group are presented in Table 3.1.2. Here, the most important errors concern the phytoplankton groups *Paralia* spp., *Ceratium fusus* and *Dinophysis acuminata* (upper than 70%). However, these groups are represented in only a few images (32, 6 and 3 images respectively). In order to highlight the groups for which particles are misclassified, a visual representation of the confusions between groups is proposed.

A confusion matrix is a square contingency table between the manual identification of particles sorted in the training set, and the automatic identification of these particles. All particles counted on the diagonal are correctly identified by the recognition tool, while errors are located off diagonal.

Table 3.1.2. Abundance and error rates per group for the FlowCAM recognition tool built with Random Forest. The red font indicates the species or genus for which the percentage of misclassification goes over 50%.

Taxa name	Nb images	Error %	Taxa name	Nb images	Error %
<b><u>PHYTOPLANKTON</u></b>					
<b>Diatomophyceae</b>			<b>Dinophyceae</b>		
<i>Asterionellopsis glacialis</i>	200	22	<i>Ceratium fusus</i>	6	100
<i>Biddulphia sinensis</i>	256	11.6	<i>Ceratium</i> spp.	30	50
<i>Chaetoceros curvisetum</i>	71	47.9	<i>Dinophysis acuminata</i>	3	100
<i>Chaetoceros danicus</i>	38	7.9	<i>Gyrodinium</i> spp.	195	5.1
<i>Chaetoceros</i> spp.	200	44	<i>Prorocentrum micans</i>	103	6.8
<i>Centrales</i> spp.	67	19.4	<i>Protoperdinium</i> spp.	44	38.6
<i>Dactyliosolen fragilissim.</i>	158	13.9	<b>Total nb of images</b>	<b>381</b>	
<i>Ditylum brightwellii</i>	200	41.5	<b><u>ZOOPLANKTON</u></b>		
<i>Guinardia delicatula</i>	54	62.9	Ciliates	119	30.3
<i>Guinardia flaccida</i>	200	16.5	<b>Total nb of images</b>	<b>119</b>	
<i>Guinardia striata</i>	45	31.1	<b><u>DETRITIC</u></b>		
<i>Lauderia Schroederella</i>	19	42.1	aggregates	200	62
<i>Leptocylindrus danicus</i>	200	13.5	black opaque particles	200	18.5
<i>Nitzschia longissima</i>	22	18.2	air bubbles	200	4
<i>Paralia</i> spp.	32	81.2	clear particles	200	12.5
<i>Pleuro-Gyrosigma</i> spp.	385	8.3	dark particles	200	55
<i>Pseudo_Nitzschia</i> spp.	200	9.5	drop particles	17	76.5
<i>Rhizosolenia imbricata</i>	200	13	fecal pellets	84	52.4
<i>Thalassionema nitzschoid.</i>	200	15	fibers	132	34.1
<i>Thalassiosira rotula</i>	135	36.3	granular particles	38	42.1
<b>Total nb of images</b>	<b>2882</b>		long thin particles	200	50
<b>Prymnesiophyceae</b>			membranous particles	31	74.2
<i>Phaeocystis globosa</i>	200	13.5	short thin particles	70	71.4
<b>Total nb of images</b>	<b>200</b>		<b>Total nb of images</b>	<b>1572</b>	



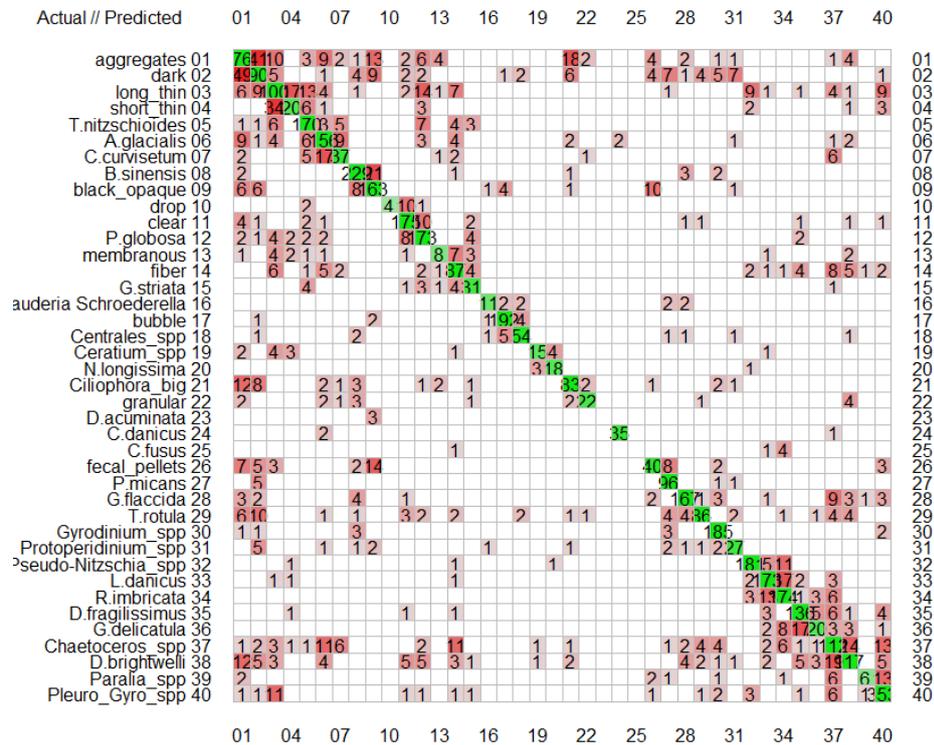


Figure 3.1.17. Confusion matrix associated to the FlowCAM recognition tool built with Random Forest.

Figure 3.1.17 shows that particles of the different detritic groups (aggregates, dark, long\_thin, short\_thin) have the most important confusions. For phytoplankton groups, the highest confusions concern particles of *Thalassionema nitzschooides*, *Asterionellopsis glacialis* and *Chaetoceros curvisetum*. These confusions can be explained by the morphological similarity of the colonies of these species (whereas cells are different in shape).

### 3.1.2.3. CytoSense training set - CNRS LOG

**Material and methods.** To build a training set based on images, we focused on micro-phytoplankton particles. Indeed, with the CytoSense device, and in order to have high resolution and high recognizable particles images, only large particles were imaged.

A living training set representative of each micro-phytoplankton community met in the Eastern English Channel was built using samples taken throughout 2016 and 2017 in the frame of the IFREMER Regional Nutrients and REPHY Monitoring networks (SRN-REPHY), the CNRS LOG DYPHYRAD transect and the SOMLIT CNRS INSU observatory. Samples were digitized using a CytoSense device, for which the trigger channel was set on the Red Fluorescence with a threshold equal to 15mV. The training set building was carried out using the RclusTool package.

A total of 482 images were manually classified into 15 plankton groups (Figure 3.1.18). From this training set, a classifier was trained using the “Random Forest” algorithm (Breiman 2001, Fernandez-Delgado *et al.* 2014). Global error measured by using 10 folds cross-validation is thus equal to 26.76% for these groups.

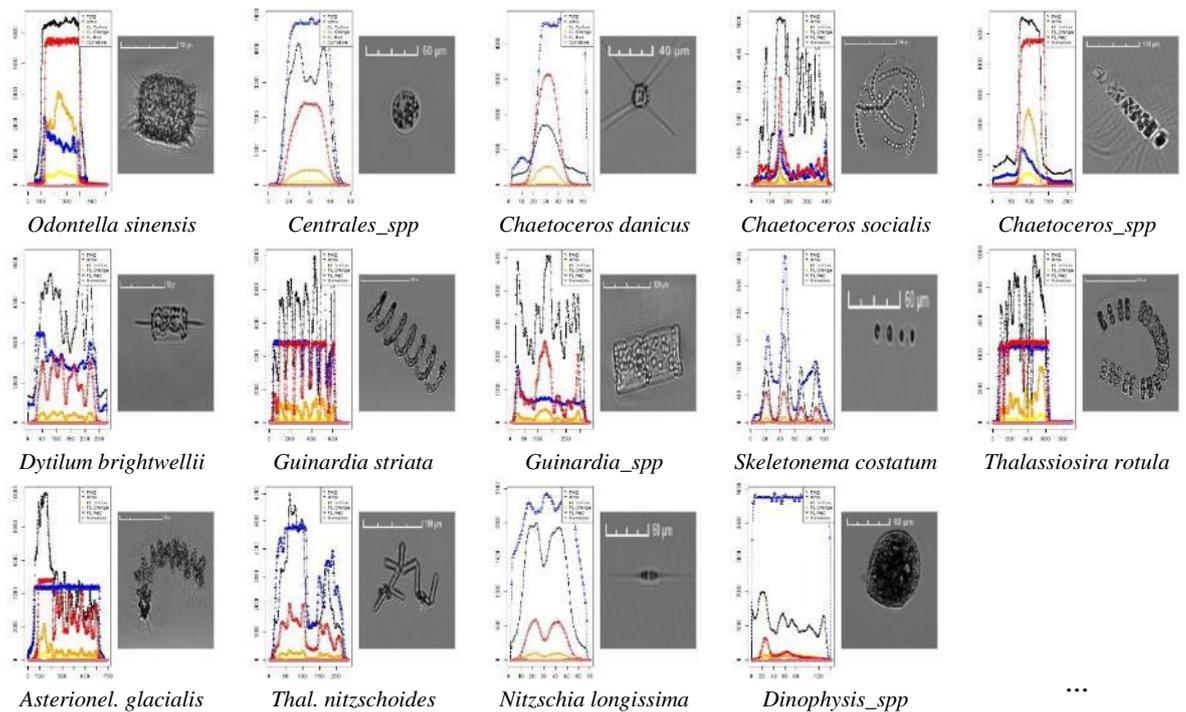


Figure 3.1.18. Examples of images in some taxonomic groups in the CytoSense training set.

**Results.** Abundance and error rates per group are presented in Table 3.1.3. The phytoplankton groups *Asterionellopsis glacialis*, *Ditylum brightwellii*, *Guinardia striata* and *Dinophysis acuminata* have the most important errors (equal to 100%). However, these groups are represented in only a few images (3, 4, 9 and 2 images respectively). In order to highlight the groups for which particles are misclassified, a visual representation of the confusions between groups is proposed.

Table 3.1.3. Abundances and error rates per group for the CytoSense recognition tool built with Random Forest. The red font indicates the species or genus for which the percentage of misclassification goes over 50%.

Taxa name	Nb images	Error %	Taxa name	Nb images	Error %
<b>PHYTOPLANKTON</b>			<b>Dinophyceae</b>		
<i>Asterionellopsis glacialis</i>	11	72.2	<i>Dinophysis acuminata</i>	2	100
<i>Biddulphia sinensis</i>	3	100	<b>Total nb of images</b>	2	5.1
<i>Chaetoceros danicus</i>	22	18.2	<b>ZOOPLANKTON</b>		
<i>Chaetoceros socialis</i>	82	32.9	<b>Ciliates</b>	7	85.7
<i>Chaetoceros spp.</i>	114	15.8	<b>Total nb of images</b>	7	
<i>Centrales spp.</i>	5	60			
<i>Ditylum brightwellii</i>	4	100			
<i>Guinardia striata</i>	9	100			
<i>Guinardia_spp</i>	29	93.1			
<i>Nitzschia longissima</i>	72	1.4			
<i>Skeletonema costatum</i>	21	0			
<i>Thalassionema nitzschoid.</i>	80	16.2			
<i>Thalassiosira rotula</i>	21	19			
<b>Total nb of images</b>	<b>473</b>				

Figure 3.1.19 shows that the highest confusions concern particles of *Guinardia\_spp*, *Chaetoceros socialis* and *Chaetoceros\_spp*. These confusions can be explained by the proximity of optical properties for these taxonomic groups (whereas amplitudes and lengths can differ, as shown in Figure 3.1.20).

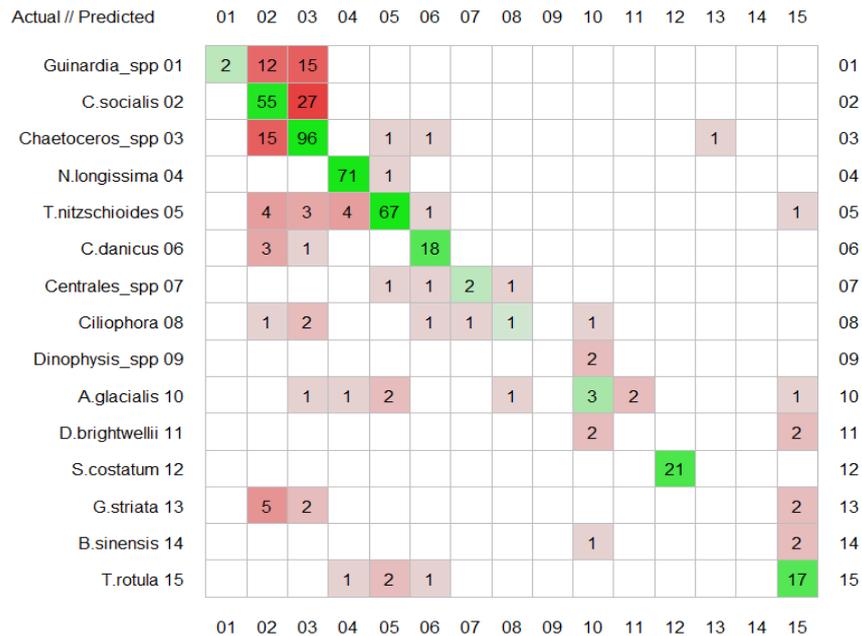


Figure 3.1.19. Confusion matrix associated to the CytoSense recognition tool built with Random Forest.

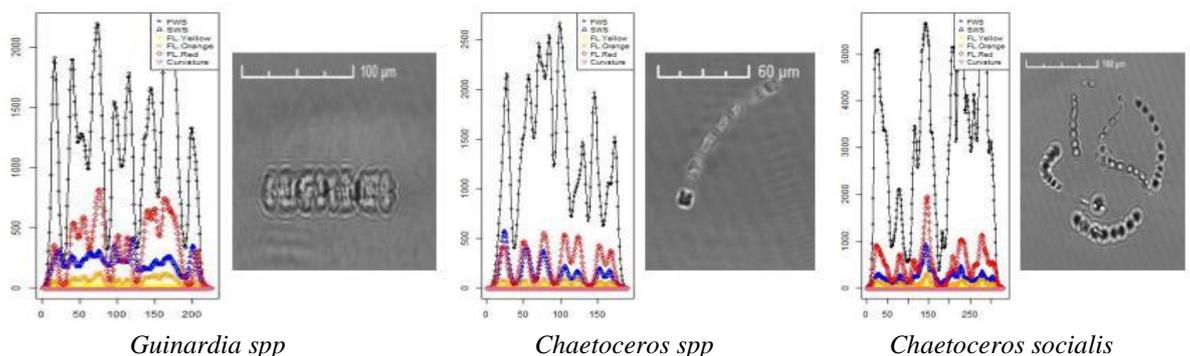


Figure 3.1.20. Cytometric profiles of taxonomic groups presenting high confusions.

Perspectives for FlowCAM and CytoSense training sets

The recognition performances of the FlowCAM and CytoSense classifiers are often associated to the number of images in each taxonomic groups of their respective training sets. So, in order to improve the recognition scores, several aspects must be considered:

- Even if some groups have a good recognition score with a low number of images, in all cases, a significant improvement of recognition performances can be observed when the number of images per group is important (upper than 100). However, this threshold can be higher when the morphological variability of particles is important (particularly for colonial diatoms).
- Ideally, the number of images between groups must be equivalent. Indeed, the groups with the largest number of images are also the groups with the better recognition rates. It seems that the number of images give proportionally more weight to a group, according to the classification by the machine learning algorithm.
- The detritic particles are often a source of important confusion with the small-sized plankton particles. It is thus essential to carefully define these groups.

### 3.1.2.4. FastCAM deployment - IFREMER

The FastCAM was deployed during a technological campaign in Brest bay and Iroise sea (IPARO) in 2015. It participated also to the JERICO-NEXT Phytoplankton workshop in Gothenborg, in 2016. An automated sampling unit was planned to be designed, the aim being of making fully automated analysis from a seawater inlet. The module was supposed to run the process from the sample injection to the taxonomic classification, without operator. However, as FlowCAM has in parallel developed new machines allowing a much faster image acquisition (and accurate focus of particles), then the continuity of the development of the FastCAM is being reconsidered.

### 3.1.2.5. UVP5 deployment - CNRS LOV

The UVP5 was used during the Swedish monitoring cruise (SMHI) on the R/V Aranda from Helsinki to the Oslo Fjord (and return) in July 2017 (Figure 3.1.21). 28 profiles from surface to 335 m depth were performed (one failed), we extracted 70000 vignettes as the UVP was lowered. First results show that bacterial filaments can be detected in the upper 20m depth at stations only in the Baltic sea. The vignettes were uploaded on ECOTAXA and are now available for further sorting to build a training set for computer assisted recognition (based on a Random Forest algorithm).

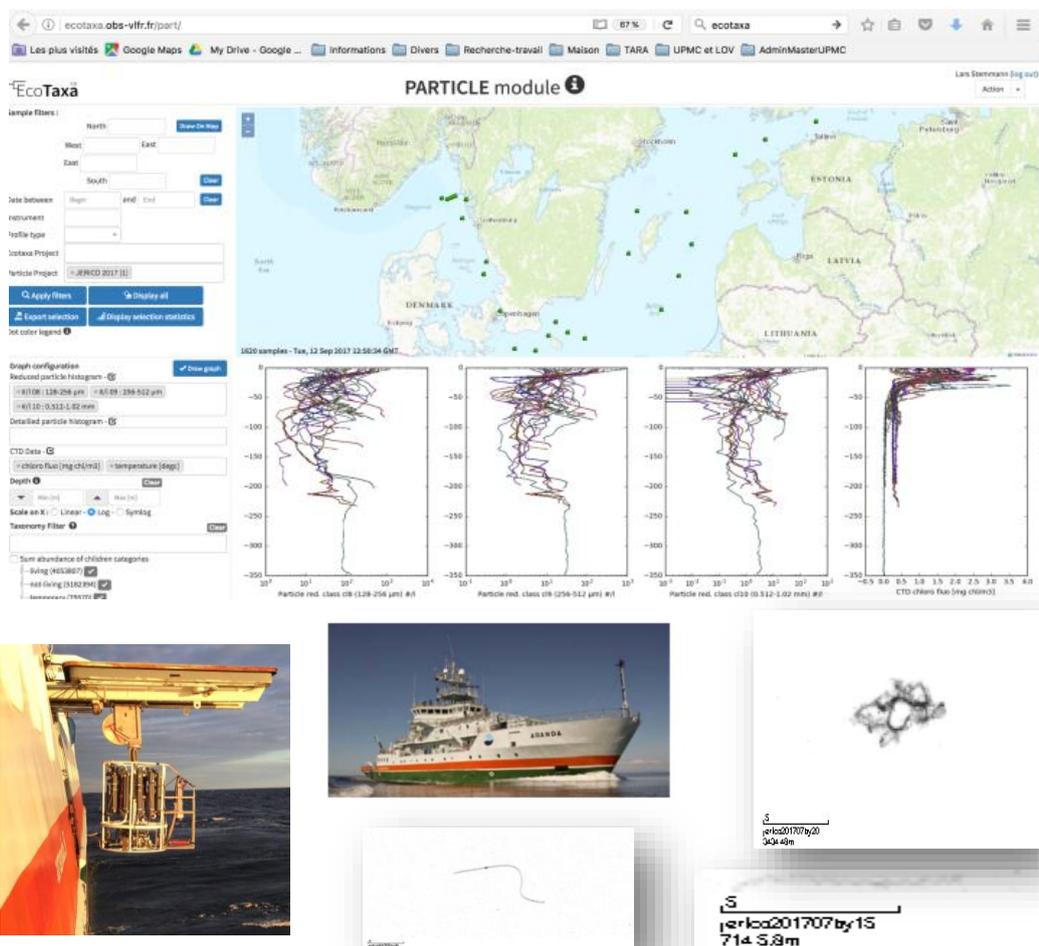


Figure 3.1.21. (A) Aranda ship during the JERICO-NEXT cruise in the Baltic Sea (July 2017) and UVP5 mounted on the CTD rosette. (B) Three vignettes from the UVP5 of three type of cyanobacterial colonies (probably *Aphanizomenon*, *Dolichospermum*, *Nodularia*). All validated images are on ECOTAXA <http://ecotaxa.obs-vlfr.fr> (direct link to JERICO-NEXT, <http://ecotaxa.obs-vlfr.fr/explore/?taxo=&taxochild=1&ipp=1000&zoom=80&magenabled=0&MapN=60.9025&MapW=8.7032&MapE=27.9292&MapS=53.0727&samples=&instrum=&projid=> ).

### 3.1.3. Conclusions and recommendations for the use of automated image acquisition/analysis

The in-flow or *in situ* imaging instruments applied in JERICO-NEXT have different optical and fluidic characteristics, which lead to various results in terms of image resolution (magnification/size of particles analysed) and measurements (features). That's why specific training sets are built for each instrument. However, some inter comparison and inter calibration exercises started to be operated in the field and laboratory and will need to be continued in order to better define the acquisition capacity of each device and the interoperability of available training sets.

### 3.1.4. Analytical improvements

#### 3.1.4.1. Adaptive learning (FlowCAM data) - CNRS LOG

During the classification process with Zoolmage, it is possible to use "contextual" samples in order to adapt the training set, and consequently the associated recognition tool, to the current studied sample. Depending on the set of contextual samples selected, adaptive learning can lead to a significant reduction of prediction errors, and considerable time savings during the manual validation process for error correction.

Intuitively, the selection of contextual samples (which contain groups of particles already validated) can be based on different criteria: same geographical area, same period (+/- one week, +/- one month, ...), same protocol digitization, etc. Once the selection is done, the active learning process is started, and the groups in the initial training set are completed with the new particles from the contextual samples, in order to give a more important weight to the target groups. This adapted training set is then used to build an adapted recognition tool for samples classification.

In the frame of the analysis of the CAMANOC cruise (multidisciplinary campaign from 16th September to 12th October 2014, in the Channel, aboard the R/V « Thalassa II » - IFREMER), one sample was selected every 2 days, then manually and fully validated, and finally added to the training set for the analysis of the samples of the two next days.

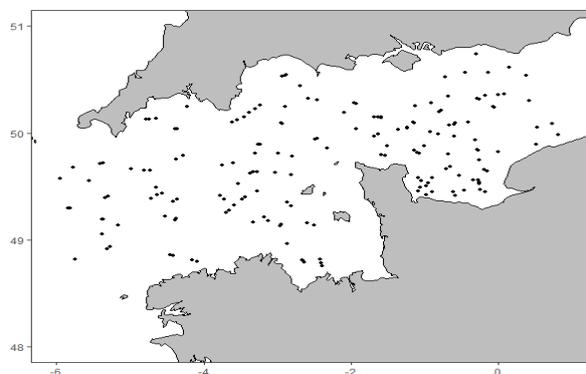


Figure 3.1.22. Sampling stations during CAMANOC (Sept-Oct 2014) cruise.

The temporal distribution of the different groups during the campaign was computed and is presented in the Figure 3.1.23. The Western English Channel is mainly characterized by high abundance of *Pseudo-nitzschia* (with a peak at 7400 particles.L<sup>-1</sup> on September 21th), *Chaetoceros* (with a peak at 5900 particles L<sup>-1</sup> on September 22nd) and *Guinardia* (with a peak at 4800 particles L<sup>-1</sup> on September 22nd) genera. For the Central English Channel, the dominant observed groups are *Prorocentrum* (with a peak at 6700 particles L<sup>-1</sup> on September 27th) and *Phaeocystis* (with a peak at 5900 particles L<sup>-1</sup> on September 28th) genera. However, the highest abundance were recorded in the Bay of Seine: *Leptocylindrus* (with a peak at 24300 particles L<sup>-1</sup> on October 09th), *Pseudo-nitzschia* (with a peak at 17000 particles L<sup>-1</sup> on October 09th), *Chaetoceros* (with a peak at 13100 particles L<sup>-1</sup> on October 09th) and *Prorocentrum* (with a peak at 10500 particles L<sup>-1</sup> on October 12th) genera mainly make up the observed phytoplankton community.

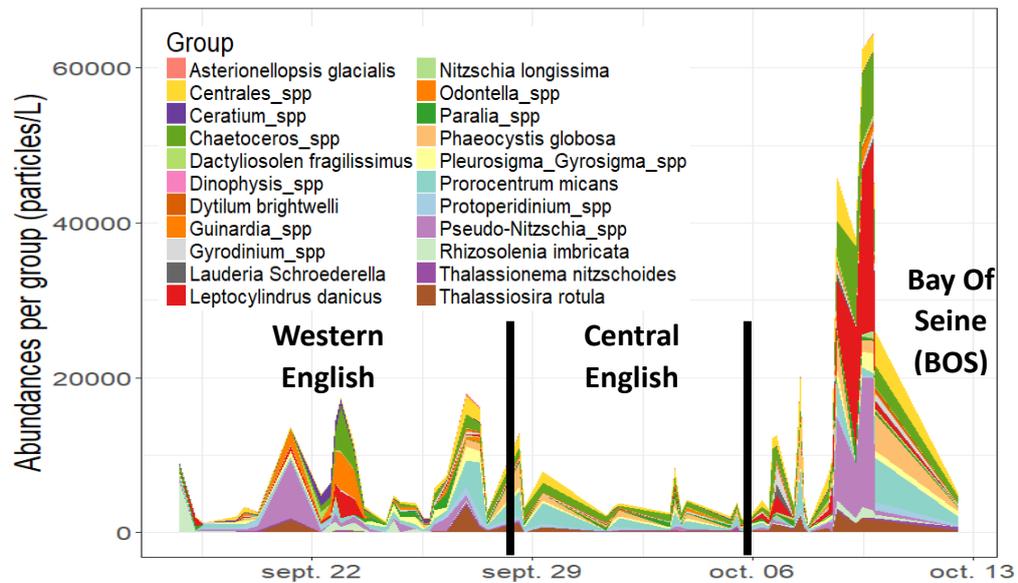


Figure 3.1.23. Temporal distribution of phytoplankton groups during the CAMANOC cruise.

#### 3.1.4.2. Automated estimation of the number of cells in colonies - CNRS LOG

Imaging-in-flow systems do not distinguish a cell from a colony. However, although colonies largely contribute to total biomass and annual productivity, all abundance and biomass estimators are mainly calibrated on the abundance in terms of cells per unit volume.

The proposed method consists in building and calibrating predictive models for the automated estimation of the number of cells per colony, based on the manual counts made by an expert on the particles of the training set. To help the expert in its counting task, a visual and interactive tool has been developed. The method involves marking each identified cell with a simple mouse click.

The main objective of this work is to define and calibrate (from features extracted from images for the FlowCam, or from optical profiles for the CytoSense/Sub) a statistical model for the estimation of the number of cells per colony, specific to each taxonomic group with ecological interest. For this, different statistical tools were used and tested: linear and nonlinear regressions, estimation of the prediction quality using cross-validation, etc.

##### Interactive tool for manual counting

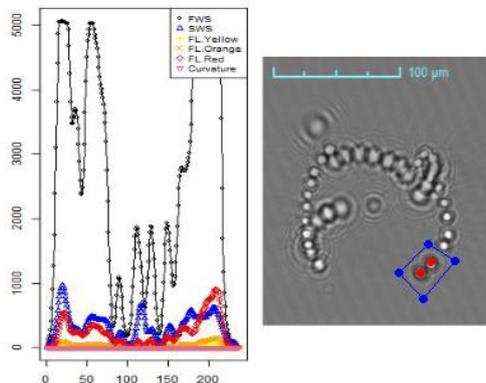
This tool was developed with R. It consists in:

- Displaying image of the particle (and optical profile for flow cytometry), taxonomic group in the training set, and some parameters measured (for instance, length, width, etc.),
- Counting manually (with simple mouse click) the number of cells in the colony. This enumeration can be made by marking ALL cells in the colony, or on a user-defined region of the colony followed by an extrapolation of the result.

Once these two steps done, a new input is automatically created in the training set for each particle.

**Class: C.socialis**

Particle: 663\_img\_Mer2\_2016-07-19\_fir26\_medium 17u53



Length FWS: 102.865 µm  
 Nb cells SWS: 6.2413 µm  
 Nb cells FLR: 4.19332 µm

1. Click 4 polygon vertices (subregion).
  2. Click items and right-click when done.
- Close windows to end.



**Methodology:**

**Total enumeration.**

- Count all cells in the image by simple mouse click (here, about 36 cells).

**OR**

**1. Partial enumeration.**

- Define a sub-region in the image.
- Count all cells in this sub-region by simple mouse click (here, 2 cells).

**2. Binarize the image and apply morphological transformations.**

**3. Compute the number of pixels of interest in the sub-region (here, equal to 270 pixels), and in the whole image (here, equal to 5157 pixels).**

**4. Extrapole**

$$\frac{2 \text{ cellules}}{270 \text{ pixels}} * 5157 \text{ pixels} = 38.2 \text{ cellules}$$

**Calibration of predictive models**

Four colonial species were selected in the training set of the CytoSense, and one specie in the training set of the FlowCam (Table 3.1.4). For each species, the cells on each image were counted thanks to the tool presented in the previous section.

**Table 3.1.4.** Number of images and number of cells manually counted for each studied species.

Instrument	Species	Nb images	Total number of cells counted
CytoSense	<i>Chaetoceros socialis</i>	82	9980
CytoSense	<i>Chaetoceros_spp</i>	114	851
CytoSense	<i>Thalassionema nitzschooides</i>	71	428
CytoSense	<i>Thalassiosira rotula</i>	37	199
FlowCam	<i>Phaeocystis globosa</i>	39	4895

The performances of the predictive methods are compared by cross-validation (10-folds) on linear and no-linear regression models:

- LM : Linear Model,
- LDA : Linear Discriminant Analysis,



- MDA : Mixture Discriminant Analysis. The aim of this technique is to model each class by a mixture of two or more Gaussian components with different centroids, but with each Gaussian component (intra- and inter-classes) sharing the same covariance matrix. This allows to define more complex frontiers.

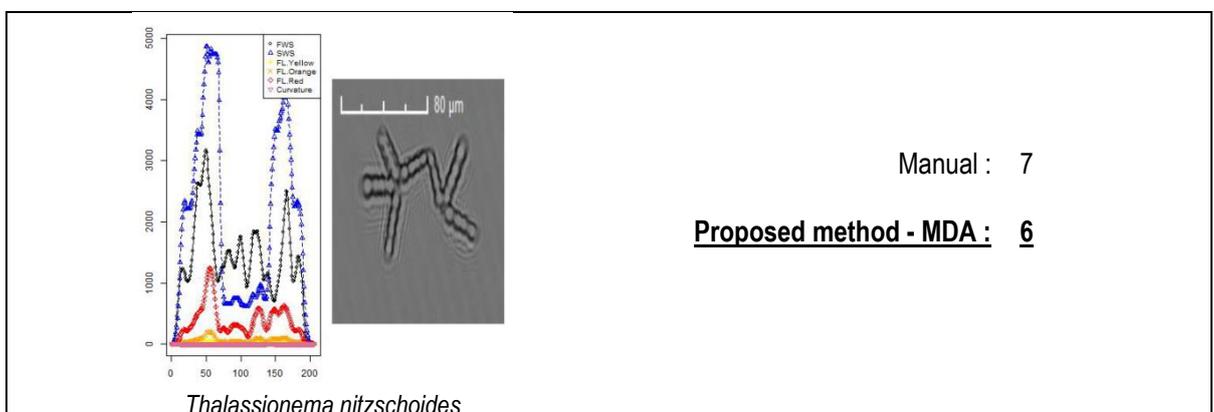
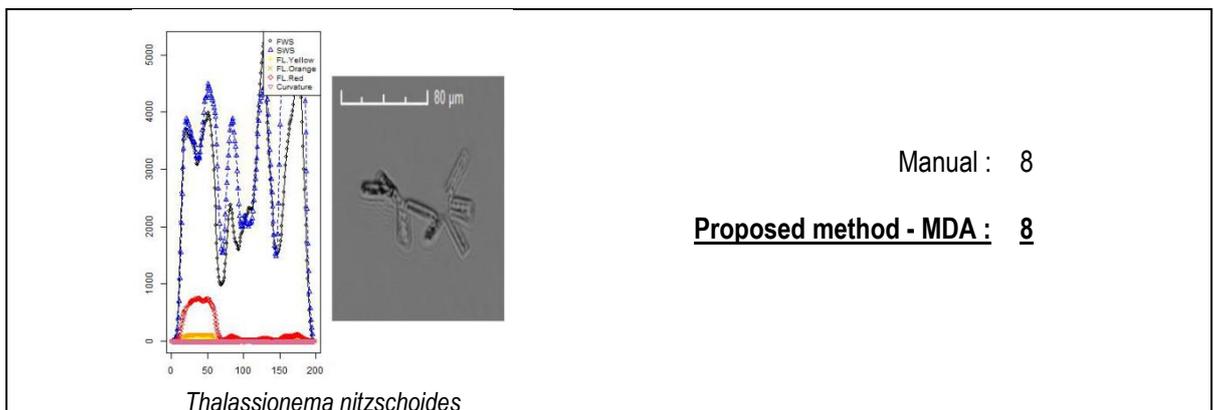
Thanks to the cross-validation process, it is possible to evaluate the recognition rates of the different predictive methods, and in particular the percentage of the total number of cells predicted over the total number of cells counted manually.

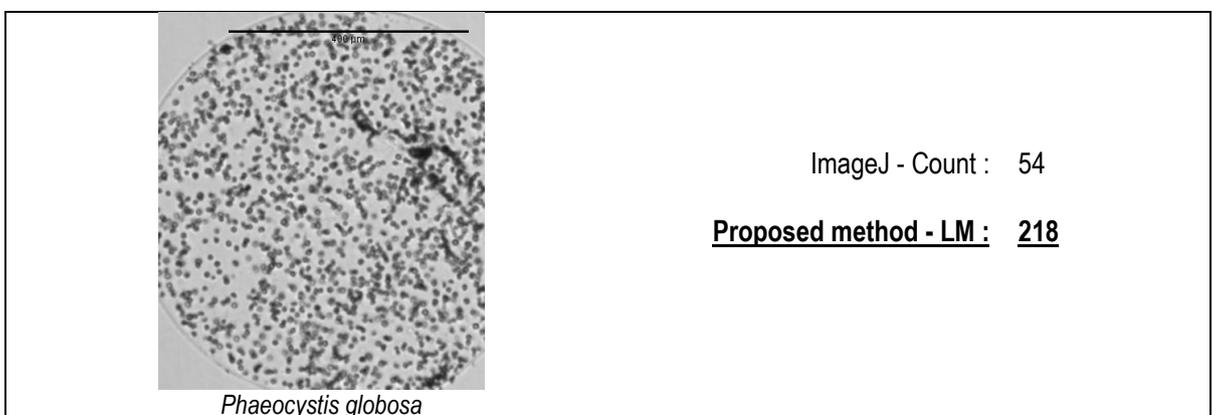
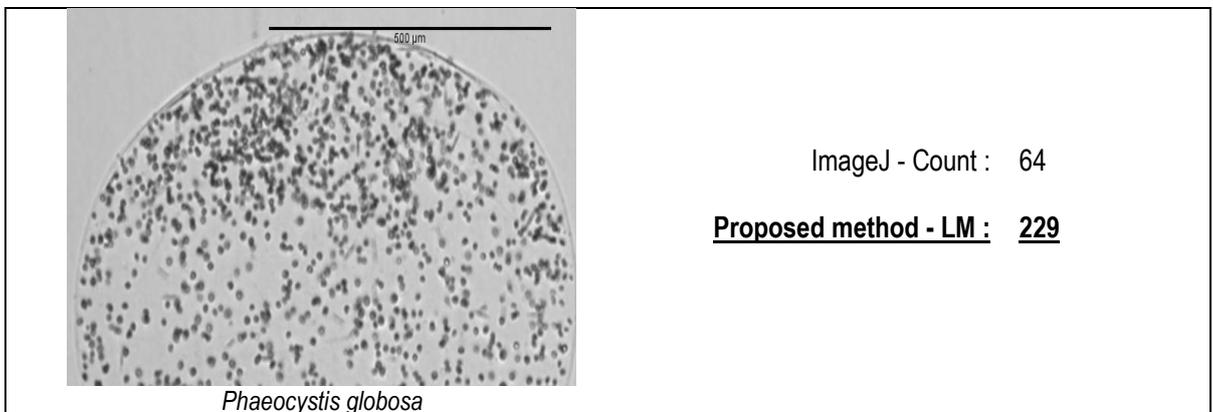
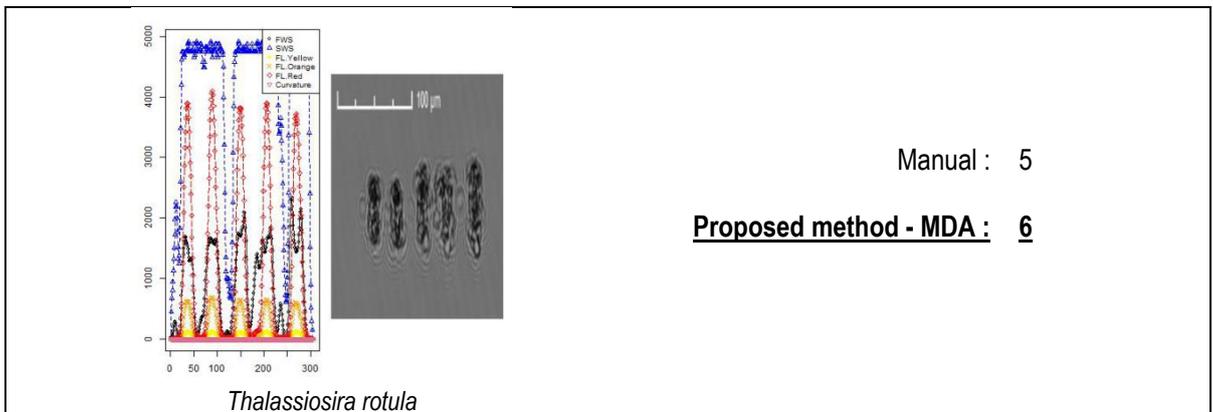
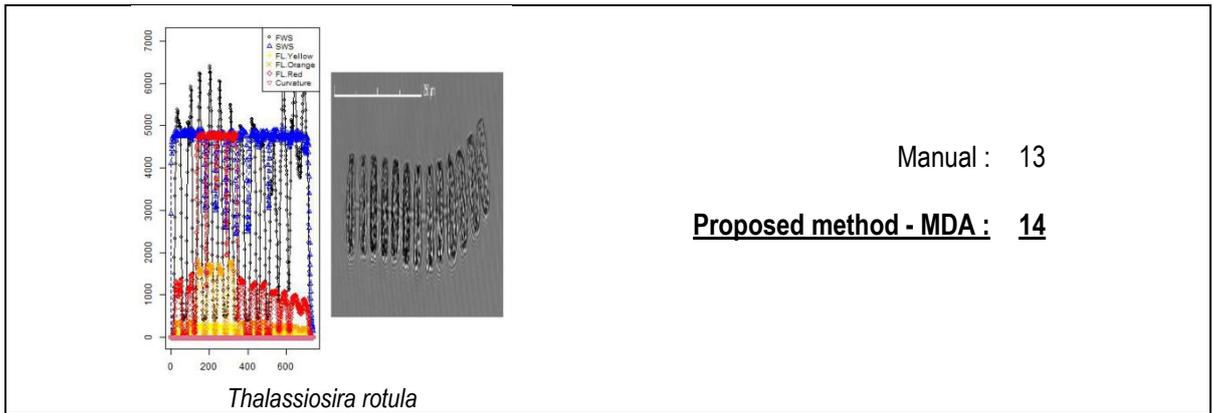
**Table 3.1.5.** Percentages of the total number of cells predicted over the total number of cells manually counted, for each studied species and for each predictive method.

Instrument	Species	R <sup>2</sup>	LM	LDA	MDA
CytoSense	<i>Chaetoceros socialis</i>	0.72	90.49	104.72	<b>101.24</b>
CytoSense	<i>Chaetoceros_spp</i>	0.64	96.94	98.24	<b>99.29</b>
CytoSense	<i>Thalassionema nitzschoides</i>	0.88	98.13	99.77	<b>99.77</b>
CytoSense	<i>Thalassiosira rotula</i>	1.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
FlowCAM	<i>Phaeocystis globosa</i>	0.93	93.81	X	X

The results presented in Table 3.1.5 show that the MDA method provides the best recognition rates for all species acquired with the CytoSense instrument. For the FlowCam, the building of the predictive model for *Phaeocystis globosa* was possible only with the LM method, because for the two other methods (LDA and MDA), the enumeration of additional images is required.

**Examples - Application on new images**







### 3.2. Single-cell optical characterization

Véronique Créach (CEFAS), Guillaume Wacquet, Arnaud Louchart & Felipe Artigas (CNRS-LOG), Melilotus Thyssen & Gérald Grégori (CNRS-MIO), Machteld Rijkeboer & Arnold Veen (RWS), De Blok (VLIZ), Thomas Rutten (TRP)

#### 3.2.1. Short overview of methodology and instruments commercially available

Flow cytometry was first applied to oceanographic data collection in the 1980s (Yentsch & Yentsch 1979, Olson et al. 1983) and was quickly adopted as a valid technique for the analysis of marine microbes (Olson et al. 1985). It is now considered amongst the methods of choice for reproducible measurements of phytoplankton abundance and community structure (Collier 2000). As research activity at this time centred on exploration of the picophytoplankton, this new technology soon yielded advances in this field (Li & Wood 1988, Olson et al. 1990). In 1994, flow cytometry was then responsible for discovering the smallest eukaryote identified to date, *Ostreococcus tauri* 59 (Courties et al. 1994). Since then, the use of flow cytometry to study the phytoplankton diversity and biomass increases, not only focusing on small cells but all the Phytoplankton Functional Groups (PFG). By providing large volumes of comparative data on the abundance and distribution of all PFG, the technique has enhanced understanding of the seasonal cycles of nano- and microphytoplankton. It has also provided a tool with which to begin assembling the same depth of knowledge about the picophytoplankton. However, this technique is not without limitations. Problems arise when multiple species possess similar optical characteristics, or when a single species displays a wide range, e.g. cells which are liable to clump or form chains or colonies (Jonker et al. 2000). The vast diversity within and between phytoplankton groups can create issues in inferring taxonomic meaning to flow cytometric output alone (Veldhuis and Kraay 2004). It should also be noted that flow cytometry only provides a snapshot of community diversity. The information gained by analysis of a single sample is enough to give an indication of phytoplankton community composition, but cannot be extrapolated to a population census (Li 2009). The results provided by flow cytometry will also be weighted significantly in favour of the pico- and nanophytoplankton. This bias is unavoidable, as these small cells are more numerically dominant within the phytoplankton. To attain a more statistically equivalent balance between PFG, greater volumes of water samples need to be analysed in order to counteract lower numbers of large cells in natural assemblages (Li 2009). Problems are also encountered when natural populations occasionally produce parameters which cannot be accurately measured by flow cytometry. Some cells cause light scatter beyond the range measurable by the instrument, e.g. extremely large or highly fluorescent cells cause saturation of the light sensors. Others suffer the converse; electronic detectors may not be sensitive enough to capture very small quantities of light scatter and fluorescence (Li, 2009). In efforts to counteract these issues, flow cytometry has undergone many improvements and refinements since its inception for marine use. Instruments have become increasingly sophisticated, with broader detection ranges and greater sensitivity to morphological features (Dubelaar and Gerritzen 2000; Veldhuis and Kraay 2000; Dubelaar et al., 2004). With certain types of flow cytometer, it is now possible to acquire an image of each particle analysed (Campbell et al. 2010), or sort them into groups based on size or optical properties (Zubkov et al. 2004). Some models are automated, and capable of continuous analysis whilst submerged under water for weeks at a time (Thyssen et al. 2009); whereas others are equipped with multiple lasers for in depth investigation of multiple pigment fluorescence (Katano & Nakano 2006). Machines are now available with a range of specifications, dependent on the target population: some focus on a narrow size range (e.g. the Apogee A50-Micro, the Accuri C6 or the BD FACSCalibur), whilst more generalist machines process particles across a large size range (CytoBuoy CytoSense).

The CytoSense and CytoSub (CytoBuoy) is a flow cytometer dedicated to the detection of phytoplankton cells and colonies (from 1µm to 800µm width). Its development has been funded originally in 1992 by a European grant (MAS20001:1992-1995). Twenty years later, the Cytosense/CytoSub fulfils all the characteristics described in the aims of the project. It is a user-friendly flow cytometer for phytoplankton analysis with a large dynamic range, with an increased discrimination level by incorporating pulse shape analysis, light diffraction measurements, with a video imaging-in-flow of a selected group of individual particles, and with the possibility of on-line identification of PFG with the aid of sophisticated data-analysis tools of cells for routine shipboard analysis and a modular set-up of the instrument with a high degree of ease of deployment and operation. The CytoSense (benchtop) instrument



records the entire pulse shape of the particles flowing in front of a laser beam (blue, green and red lasers are available, single and two lasers configurations). The sample is collected and sent to the optical unit thanks to a calibrated peristaltic pump, with a speed controlled by the operator. The sample is surrounded by a sheath fluid generating a hydro-dynamical focusing in order to separate each particle at a flow rate decided by the user. Each particle intercepts a laser beam and the generated pulse shape of optical properties (two scatters, up to three fluorescences) induced by the particle are recorded. Pulse shapes recording allow chains forming cells to be recorded. An image in flow device records pictures of preselected groups of cells, resolving cells at its best above 20  $\mu\text{m}$  but is able to collect pictures of 2  $\mu\text{m}$  beads (at low resolution). Particles are recorded above a defined threshold (scatter or fluorescence) and phytoplankton cells are separated from non-photosynthetic particles by their red auto-fluorescence. The CytoSense sensors can be used on ships of opportunity and scientific vessels and connected to other instruments such as a FerryBox, whereas the submersible version (CytoSub) fits in fixed stations and buoys, running samples from a subsampling dedicated system isolating sea water from a continuous flow of pumped sea water. The Cytosense/CytoSub runs automatically and can be remotely controlled, to analyse as frequently as the operator wants according to the protocol reaching kilometer scales resolution of phytoplankton distribution by its high frequency. Principle of automated flow cytometry is illustrated on Figure 3.2.1.

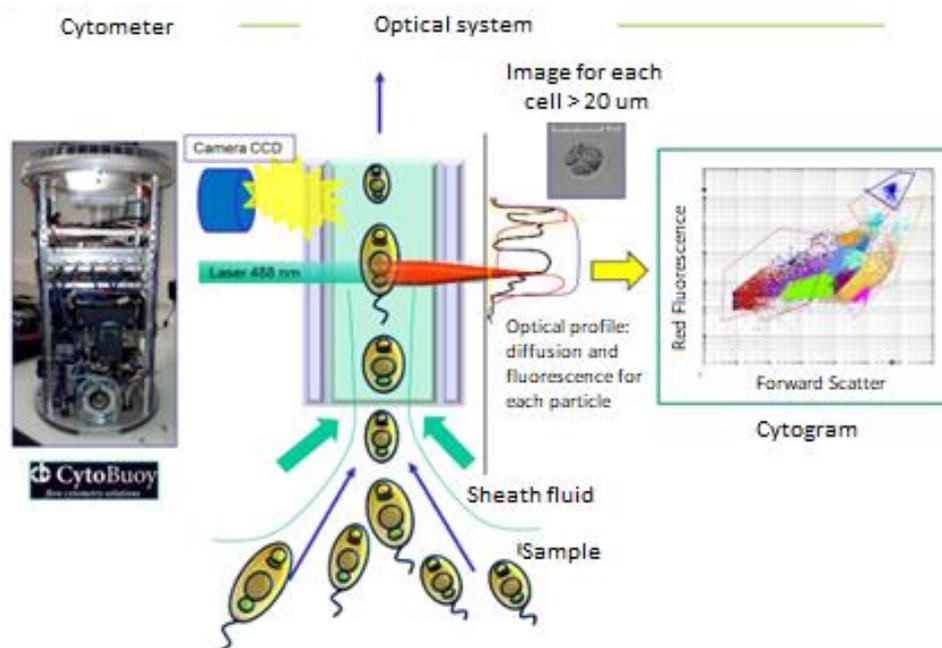


Figure 3.2.1. Overview of the flow cytometry principle.

Several studies revealed flow cytometer as a powerful tool to count and discriminate the species. Sometimes, the optical profile which is composed by scattering and fluorescence signals are integrated to clustering software to increase the capability of discrimination. Two optical properties concern the scattering: the forward scatter and the sideward scatter. Moreover, three fluorescence signals can be used: red, orange and yellow fluorescence. The arrangement of these five signals and their features can be mapped on two-dimensional plots (i.e. cytograms).

In the North Sea and the English Channel, many publications focused on the spring season especially due to *Phaeocystis globosa* and diatoms blooms. In different studies (Bonato et al., 2015; 2016; Guiselin, 2010; Houliez et al., 2012; Rutten et al., 2005), the CytoSense was able to discriminate *Phaeocystis globosa* according to different single cell states as well as colonies (Figure 3.2.2).

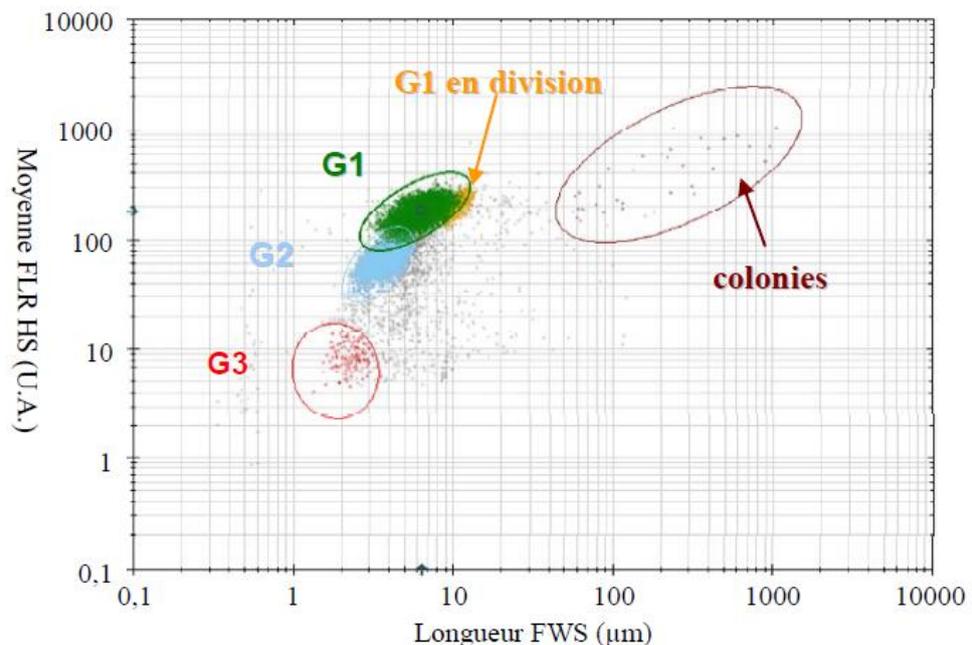


Figure 3.2.2. Discrimination of *Phaeocystis globosa* groups from a culture. G1, G2 and G3 groups were single-cells and colonies clusters were composed by G1 cells bulk (Guiselin, 2010).

Cells forming colonies were an association of single cells and thus showed high red fluorescence signal and the biggest size. Moreover, three single cells were discriminated. The G1 group corresponded to the biggest single cells (i.e. between 3.5 and 9  $\mu\text{m}$ ) with the highest fluorescence content per cell. Finally, the G2 and G3 groups showed approximately the same size (G2: 3 to 6  $\mu\text{m}$  and G3: 1.5 to 4.5  $\mu\text{m}$ ) but G2 was characterized by a higher fluorescence content than G3.

### 3.2.2. Results and discussion - developments, evaluations and experiences

In recent years, improvements in data acquisition procedures have been carried out to reflect the high variability of data composition. However, these acquisition and digitization techniques, including those concerning « pulse shape recording » flow cytometry, still generate an important quantity of data which cannot be easily processed manually. If within these data, we are sure to have the most complete and useful information, this information may be lost due to the large amount of data. In this way, automated classification is usually carried out for data exploration and structuring in order to extract useful information.

For this purpose, analytical tools were and are being built to allow greater automation in data analysis. However, classification context is directly associated to the available prior knowledge. Indeed, depending on the kind of prior information, it is possible to distinguish three contexts:

- **Unsupervised classification**, which can be defined as an exploratory analysis of the data structure. No additional information other than the data themselves, are provided to the classification algorithms. The aim is to organize the data into clusters, such as particles in a cluster are similar and particles belonging to different clusters are different.
- **Supervised classification**, for which prior knowledge are available as sets of labelled particles, allowing to model the relationship between particles and classes. This training set is used to create decision rules. These rules will be then used to determinate the class of a new particle.
- **Semi-supervised classification**, which is an intermediary solution between unsupervised and supervised context, and for which the data are partially labelled.



### 3.2.2.1. Description of clustering tools

#### Manual clustering with CytoClus© (CytoBuoy, The Netherlands)

Each particle passing through the laser is analysed by the photomultiplier of the flow cytometer. Five parameters are obtained (forward scatter, sideward scatter, red fluorescence, orange fluorescence and yellow fluorescence). They define the optical profile of the particle and they are recorded and stored by the flow cytometer. A dedicated software designed by the CytoBuoy company is required to perform the manual clustering: CytoClus©. For each parameter, CytoClus© calculate nine traits (+ Time of flight):

*Length:* The raw length is determined from the time of flight between the crossings of the 50% of maximum threshold. This raw length is subsequently corrected by applying a correction curve which has been determined from physical considerations combined with measurements

*Total:* The total is simply the detector value at each data point summed over the length of the particle.

*Maximum:* This is the maximum detector value occurring along the length of the particle as it passes through the detector unit.

*Average:* The average is equal to the total divided by the number of data points, thus removing the length-dependence of the total fluorescence.

*Inertia:* The inertia is defined as the second moment of the pulse form.

*Centre of gravity:* The centre of gravity is found by dividing the first moment of the pulse shape by the total%.

*Fill factor:* Gives an indication of the solidity of the pulse shape.

*Asymmetry:* Gives an indication of the distribution of the signal over the particle length.

*Number of cells:* Gives an indication of the number of cells in particle. The algorithm works by moving across the pulse and counting the sign changes. If the difference in value between the points of two sign changes is greater than 75% of the signal extreme value, this is seen as one signal peak.

*Time Of Flight:* Signal length based on where the signal is above the set trigger level.

#### General case for clustering:

A two-dimensional plot (i.e. cytogram) allows the operator discriminating particles with the same characteristics (traits and parameters). Each axis is characterized by one parameter and one trait or two of each in the case of ratio calculation. The maximum combination is  $[(5 \text{ parameters} * 9 \text{ traits}) + \text{TOF}] * 2 = 92$  combinations (i.e. cytograms). A particle viewer window helps recognizing similar particles. Particles with similar characteristics often form bulks well-defined from others. When a bulk is identified, the operator has to draw a gate around and creates a set. Several choices are proposed for this:

*Rectangle:* Draw a rectangle by clicking one corner and then clicking at the opposite corner

*Polygon:* Draw a polygon by clicking consecutive points and close by clicking near to the first one.

*Ellipse:* an ellipse is defined by its two axes; first click the centre of the ellipse and then click to define its first axis and finally click to define the second axis and complete the ellipse.

This step ends the manual clustering for one file. Then, the operator can export the results in CSV format or Matlab® file.



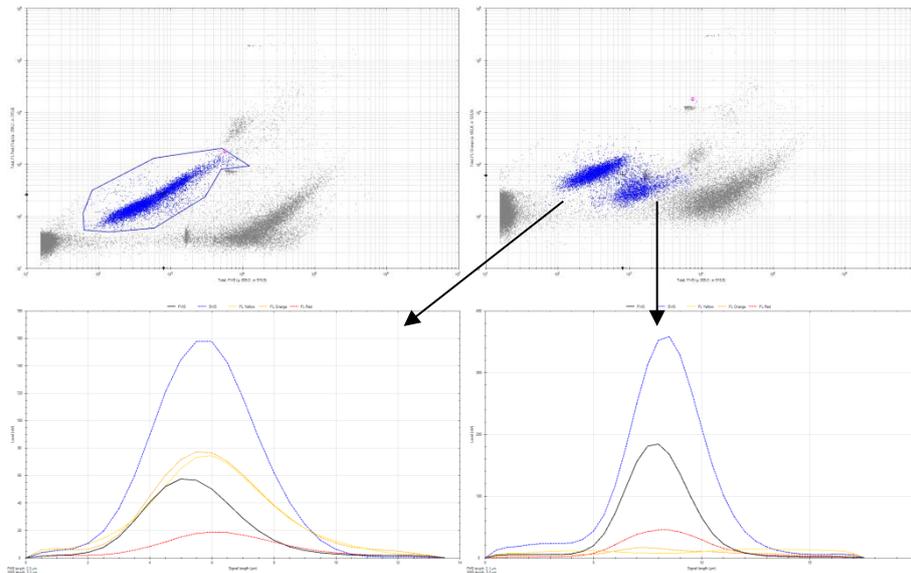


Figure 3.2.3. The use of multiple combinations is very useful and sometimes the operator can discriminate several groups in one plot that could not be seen in one other. In the above figure, the first combination (Total forward scatter vs. Total FLR) highlights only one blue group. A second combination (Total forward scatter vs. Total FLO) highlights two groups in the blue one. The two optical profiles, below the cytograms, reveal the different particles. Left: with high yellow and orange content, it is the *Synechococcus* sp. group; right: with a small size but higher red fluorescence content, it is a picoeukaryote group.

#### Batch-process mode:

For large raw-data sets (particularly data from the same area or with very similar particles) a batch-process mode can be used. It processes the files by using the same gates. If there are changes between files, the operator can resume the analysis and redefine the gates.

Several outputs can be getting from Cytoclus®:

- Quantities and concentrations
- Sums of total fluorescence per set
- Average properties of the active set
- Listmode particle data
- Raw particle data

#### Additional features:

Some machines have an Image-in-flow system which has the capability of taking targeted pictures. The target area is defined according to a rectangle selection on a cytogram. The association of pictures and optical profile for the targeted selection can reveal sub-groups in one cluster.

#### **RclusTool® package (LISIC-ULCO/LOG-CNRS, France)**

The RclusTool® is a toolbox to classify data in unsupervised, supervised or semi-supervised way, through a complete Graphical User Interface (GUI). This package is written in R (fully open source), and was developed within the DYMAPHY project (Hébert *et al.*, 2014). It is currently improved in the frame of JERICO-NEXT by Guillaume Wacquet (CNRS-LOG/ULCO) in collaboration with the Laboratoire d'Informatique du Signal et de l'Image de la Côte d'Opale (LISIC/ULCO, France), via funding by the French Ministry for an Ecological and Solidarity Transition (MTES) for the implementation of the MSFD monitoring program, at the laboratory of Oceanology and Geosciences of the Centre National de la Recherche Scientifique (CNRS).

This tool is designed both to automatically cluster the phytoplankton functional groups, and to proof and eventually correct the results. This interaction is first covered by many visualization tools, then by queries allowing adding some prior knowledge such as classes' models, labelled observations, pair wise constraints, etc. An important effort was brought to make easier the data exploration, as well as the insertion of data knowledge/constraints, through convenient graphics outputs and user interaction. The combination of operator analysis and automatic clustering method improves the performance scores for classification. Most of the analysis and classification methods used in this interface are components of existing packages on CRAN website (<https://cran.r-project.org/>). GUI was designed and added to reduce computing time of large datasets, and some pre-processing. The objective of the interface is to bring together and simplify the use of these various methods into an ergonomic and interactive classification system.

The main features of the toolbox are available through a user-friendly interface:

- A. **Importation**: loading of input data files. The (required) main file must contain the numeric features of all observations, and some optional files may add metadata, signals and/or images of some observations. Features are used in the automatic processing, whereas optional information is intended to help the operator in its supervision task.
- B. **Pre-processing**: features selection, transformation, creation or filtering. Several projection spaces and visual statistical tools are proposed to help in the feature selection.
- C. **Unsupervised**: clustering process settings (estimation of the number of clusters) and application of the selected unsupervised classification algorithm.
- D. **Semi-supervised**: clustering constraints by manual selection of pairwise relationships in a reduced space, and application of the selected semi-supervised algorithm.
- E. **Supervised**: prototypes loading and selection (a dataset called "training set"), and application of the selected supervised classification algorithm.
- F. **Batch process**: multiples datasets processing (only supervised and unsupervised classification).

**Data viewing.** The package proposes the following representations:

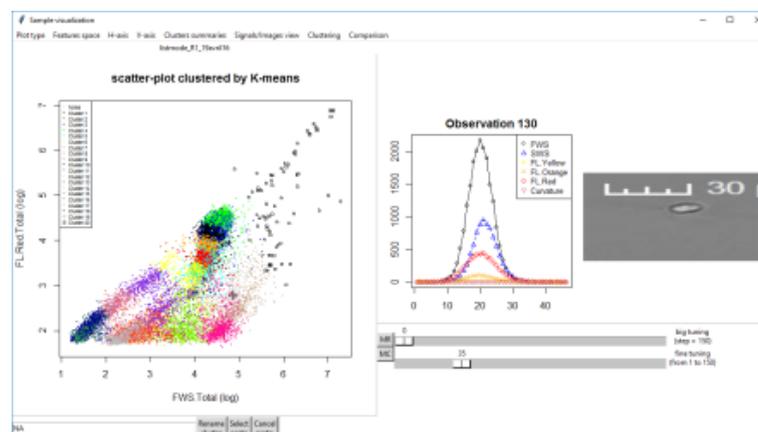


Figure 3.2.4. Scatter plot (by default): a global 2D representation of all observations contained in the dataset, each one being plotted as a point, whose coordinates match its measurements according to the 2 selected features. The colour used for each point matches its corresponding cluster.

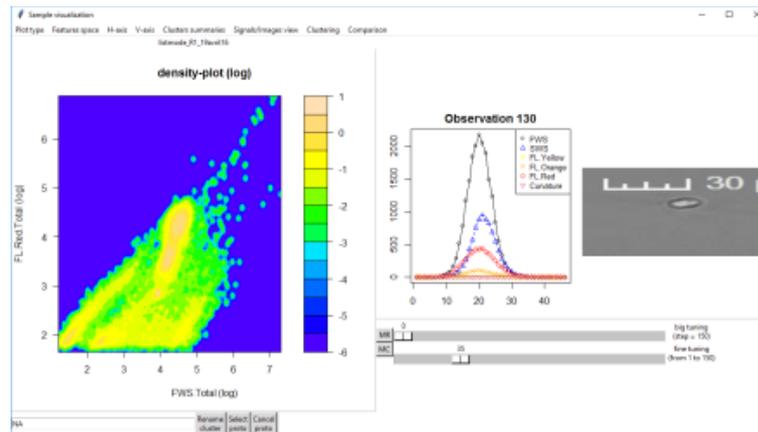


Figure 3.2.5. Density plot: a density estimation of the dataset in the selected features space, allowing a better visualisation of the concentrations than the 2D scatter plots. The colours correspond to the densities of particles with the warmest colour for the highest concentration of particles.

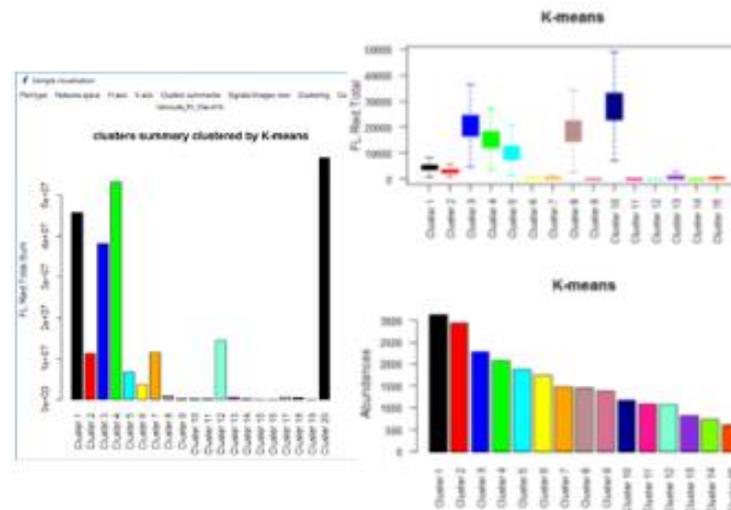


Figure 3.2.6. Cluster summary: in the menu “Clusters summaries”, by-cluster synthetic graphics proposed (histograms of abundance, features min, max, sum, average or std) or observation features boxplots.

**Data pre-processing (B).**

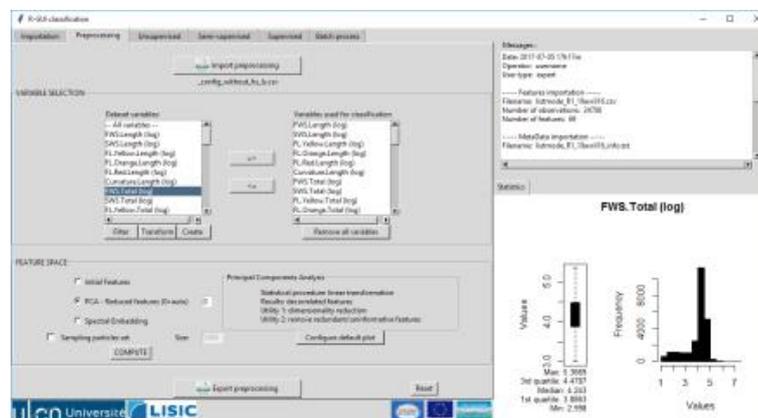


Figure 3.2.7. Processing page in RClusTool®.



In the proposed GUI, several functions of pre-processing are available, such as: selection of relevant features, feature thresholds to remove outliers/errors, feature normalisation, integration of new features (by addition, subtraction, division or multiplication), and reduction of the number of features while preserving the analytical results. In the pre-processing, the feature reduction by Principal Components Analysis aims at finding a projection subspace of uncorrelated features, while maximizing the global dispersion of the observation points. This reduction of variables not only simplifies the clustering computation, but also provides the GUI with a more synthetic subspace and some interpretation and visualisation tools (such as correlation circle, variance explained). For example, the operator may use PCA to identify the most informative features and remove the redundant ones.

While PCA provides linear combinations of initial features, Spectral Embedding technique has the advantage to transform data non-linearly. It can then deal with complex data structures, both “non-globular” and non-linearly-separable. RclusTool GUI proposes views of these data subspaces, which can be used to verify how convenient the boundaries revealed by the clustering methods are.

**Data classification (C,D,E).** The classification methods implemented in the GUI can be separated into three categories, according to the quantity of prior knowledge used in their process:

- **unsupervised classification** (K-Means, PAM, EM, HCA, Spectral Clustering): observations are clustered without any previous knowledge on their identity. Observations sharing the same feature values are separated from others, by different algorithms. The groups defined are unlabelled: “Cluster 1”, “Cluster 2”, ...
- **supervised classification** (K-NN, MLP, SVM, Random Forest): a set of labelled observations is used as a training set to deduce the class of new observations. It becomes possible to obtain class-names or functional groups, but the process generally requires validation, because of the high complexity of data (intra-classes heterogeneity and/or inter-classes similarity).

Anybody can build its own training set thanks to the simple design: signals and/or images are sorted in different groups (represented by directories), and associated measurements tables (containing all features for each observation, in CSV format) are added to the root directory of the training set.

- **semi-supervised classification** (Constrained K-Means, Constrained Spectral Clustering): the methods implemented compute a “constrained” clustering, in such a way to put together or conversely to separate, two observations (or several pairs of observations). On the contrary to the supervised mode, the only added information is a set of “constrained” pairs of observations: no observation needs to be a priori identified.

But this information is for indication only: if some proposed constraints violate the dataset spatial structure, they will not succeed in modifying the initial clustering. This method is proposed through an iterative process: after each semi-supervised clustering, the operator is invited to link or to separate some new pairs of observations. Those pairs should be selected by studying the clusters processed, and particularly the signals and/or images of the observations located near the boundaries.

For those three kinds of clustering processes, the operator has the opportunity (i) to set the output summaries (Minimum, Maximum, Sum, Average, Standard deviation) to be computed on each cluster; (ii) to sort by-cluster images and corresponding signals (stored in distinct folders on disk); (iii) to automatically identify and label some observations, as particularly “representative” of their cluster: ie. the prototypes. Those identified observations are used to compose some training sets, that may later be used in supervised classification.

#### **EasyClus® software (Thomas Rutten Projects, Middelburg, The Netherlands)**

The EasyClus® software proposes many tools to organize, cluster and handle flow cytometric data (of many types of instruments) and uses the Matlab® environment.

EasyClus LIVE® is especially made for CytoSense or-Sub instruments (CytoBuoy bv, Woerden, The Netherlands) for the online monitoring, data processing and upload to the internet ([www.phytoplanktonlive.com](http://www.phytoplanktonlive.com)) of several types of analysis results and the analysis performance.

Processing of flow cytometric data by EasyClus has many options, as:

1. Total concentration of particles and/or phytoplankton as a function of time
2. Size distribution of particles and/or phytoplankton as a function of time





3. Unsupervised clustering combined with database matching
  - C1. GO-method
  - C2. DESIGN 1 method
  - C2. DESIGN 2 method
4. Supervised clustering combined with database matching
  - D1. LASSO method
  - D2. FIXED cell method
  - D3. RULES method
5. Several bio-indicators calculations based on clustering results
6. Trend analyses methods
7. Specific species (LIVE) monitoring
8. Other features: database handling options, beads analysis, control & check analysis and instrument performance, normalization towards beads, automatic peak shaping of flattened (saturated) peaks, add & date-time synchronize data tools, plot results in Google maps plots.

Short explanation:

- A. Total concentration of particles and/or phytoplankton as a function of time

The total number of counts of particles or phytoplankton per volume, the biovolume expressed as the total Sideward or Forward Light Scatter per volume, the biomass expressed as the total of Red Fluorescence (chlorophyll a -a of phytoplankton) per volume, the Orange Fluorescence (phycoerythrin-phycoerythrin) per volume can be calculated and plotted as a function of time of space.
- B. Size distribution of particles and/or phytoplankton as a function of time

As above, but the contribution is represented as cumulated bars according to self-chosen particle length fractions.
- C. Unsupervised clustering combined with database matching.

Unsupervised means that data is clustered without pre-trained data. Database matching is done as a second step after clustering using a predefined database, which matches clusters with species within this database.

  - C1. GO-method: Data is clustered on basis of principal component analysis. The two main principal components are used to process scatterplots in the background. Data clusters are identified in this scatter plot. Each identified cluster is clustered again by new calculated principal components until this cluster is not separated into two new clusters.
  - C2. DESIGN 1 method: Data is clustered on basis of a combination of given scatter plots, the density distribution of data within the scatterplots and the similarity of data events in all dimensions (ie. using all parameters available from the analysis) with each other. Particles with high mutual similarity are merged together as a cluster. The degree of 'cluster resolution' and 'high similarity' can be set by the user in the software.
  - C2. DESIGN 2 method: Almost similar to DESIGN 1, but different what happens with single events, which falls just outside clusters. In DESIGN 1 these single events form single event clusters. In DESIGN 2 single events are assigned to the best matching cluster afterwards, resulting in a lower number of total clusters.
- D. Supervised clustering combined with database matching

Supervised means that data is clustered with pre-trained data.

  - D1. LASSO method: Clustering of data takes place on basis of the cluster data borders of many automatic (smart) chosen scatterplot combinations, which are stored as lasso's or selections sets in EasyClus. These lasso's of interest are defined and stored by the user.
  - D2. FIXED cell method: This rather confusing name (fixed cell) has nothing to do with fixing phytoplankton cells. The name FIXED cells stands for the distribution of data in the multidimensional space. Each particle (cell) in the multidimensional space can be predefined by the user, so each data event will be assigned to one of the multidimensional cells. All data is clustered according these cells and each cell forms a unique cluster.
  - D3. RULES method: Clustering is based on rules, which are produced on basis of stored events of species in a database. These rules are defined to obtain the best discrimination of the species in the database.





- E. Several bio-indicators calculations based on clustering results  
The Shannon-Wiener, the Simpson, as well as a clusters dominance index (Rutten index) are calculated on basis of clusters found by above described cluster methods. An amoeba indicator (polygon of several variables) is calculated as an overall status description variable.
- F. Trend analysis methods  
Trend analysis is useful for LIVE monitoring as an indicator to what degree a water system changes between subsequent samples. Several methods are available based on the (dis)similarity between scatterplots or recognition of outliers for self-chosen variables.
- G. Specific species (LIVE) monitoring  
The monitoring of specific (e.g. harmful) species groups can be monitored in the LIVE modus, through the display of several graphs, such as the contribution of several size algae fractions in water (count/ml, chlorophyll/ml), scatter plots for automatic unsupervised classification of species groups, etc.
- H. Other features: Beads are used to check the stability of the instrument and analysis. EasyClus has a tool to automatically detect beads in the data and put the beads results in calibration charts. Database handling options is used to add species to, rename species in or remove species in an existing database. Diverse checks are carried out to control and check the analysis and instrumental performance (e.g. particle rate in relation to coincidence). The analysis and instrumental variation can be used to normalize data against beads using the normalization towards beads tool. An automatic peak shaping of flattened (saturated) peaks tool can be used to recalculate the flattened peak shape as well as the shape variables. A tool is provided to synchronize (date and time) and merge EasyClus data results with other data (e.g. Ferrybox) using (linear) interpolation or the closest time criterion. A Google maps plot tool is available, which enables the plotting of results (colour change as a function of value) as locations projected in a map.

#### “FLOW C.A.R.S.®” R Shiny App

FLOW C.A.R.S.® is an R based (R Core Team, 2019) clustering and classification software application (App) developed specifically for Cytosense Data with the use of R Shiny (Chang *et al*, 2019) into a fully Graphical User Interface (GUI). The app is developed in such a way that the interface builds itself dynamically around the data the user loads into it. It is currently still undergoing development and is as of yet not yet available as a package or open source code, but will become so when the development has reached the level of functionality required to do so. Work on the first versions of the GUI was carried out under various fundings within Italy at Stazione Zoologica Anton Dohrn (SZN, Napoli, Italy) and Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS, Trieste, Italy), and is currently being improved in the frame of JERICO-NEXT project via a funding by the French Ministry for an Ecological and Solidarity Transition (MTES) for the implementation of the MSFD monitoring program, at the laboratory of Oceanology and Geosciences (LOG) of the Centre National de la Recherche Scientifique (CNRS).

The application is designed to automatically cluster the data into phytoplankton functional groups in an unsupervised to semi-supervised approach in combination with extensive visualization tools to aid the user from start to finish. In its core it uses the automatic flowpeaks clustering algorithm (Ge and Sealfon, 2012), which uses an overall density function generated by K-means to find peaks in multidimensional data. This approach was also tested by Mridul *et al.* (2018) to cluster their Cytosense data, and FLOW C.A.R.S.® is partially based on the results found in this study.

The FLOW C.A.R.S.® app applies this algorithm on training samples, created by random subsampling of the raw data files in a batch analysis, and combines it with machine learning (Breiman, 2001) to build a classification tree with the use of a random forest algorithm (Liaw & Wiener, 2002). In summary, the interface allows to go through the entire process of file selection, quality control, data preparation, training sample creation, parameter selection, clustering, cleaning the data, machine learning and classification of individual files. This first round of clustering is intended to remove clusters that are not of interest such as artefacts, instrument noise and non-living particles in order to better cluster the living phytoplankton into functional groups. The classified, and cleaned individual files



are then re-sampled into a new trainings sample followed by a second round of, clustering, machine learning, classifying and quantifying of the clusters of interest and ends at various tools to visualize the results.

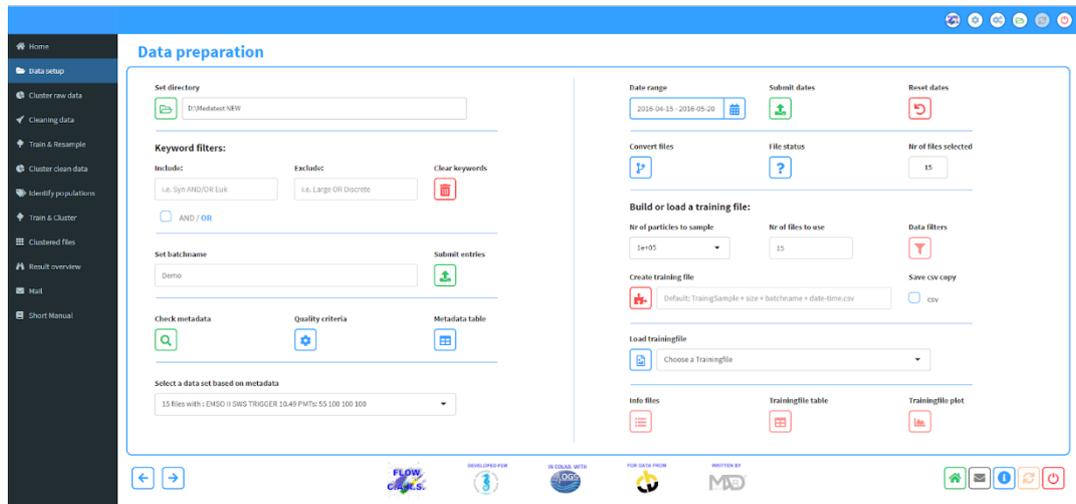


Figure 3.2.8. Overview of the data preparation page of FLOW C.A.R.S.® showing the pipeline approach going from 'page 1' top left to bottom right. Here the user can i.e. select files, filter them by keywords, date range and instrument settings before creating a training sample, and inspect the data.

### Stage 1: Clustering raw data

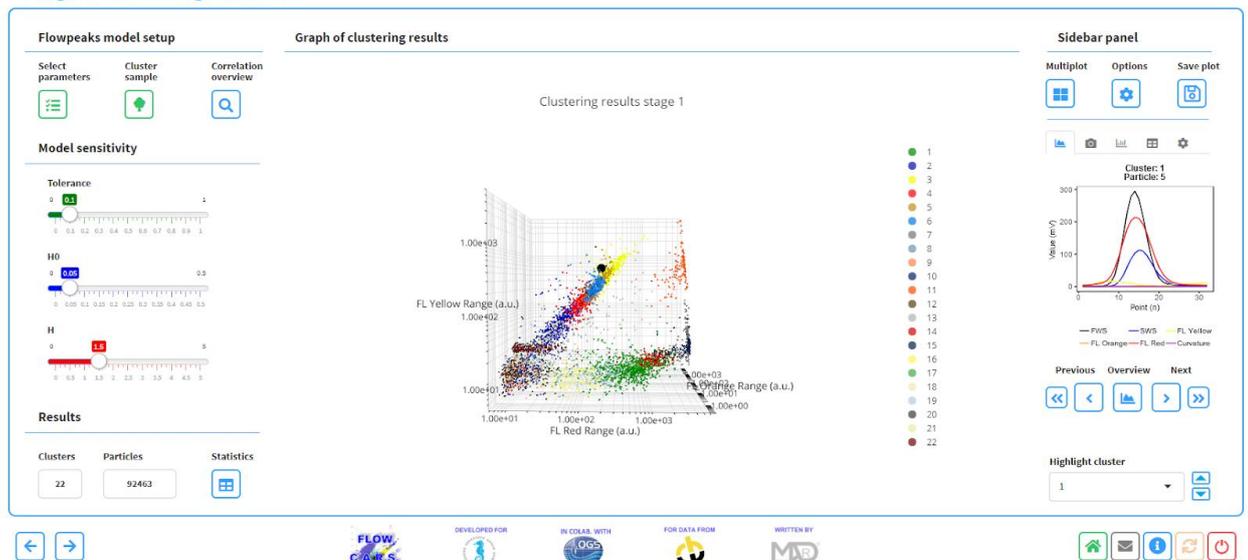


Figure 3.2.9. Screenshot of the first clustering stage where a central 3D scatter plot is surrounded by various tools to setup the model on the left, and further visual tools on the right.

Most of the analysis and classification functionality applied within FLOW C.A.R.S.® are components of existing open-source R packages and main objective of the app was to tie all the needed functionality together in a user-friendly graphical interface that is as dummy proof as one can make it without sacrificing analytical performance. It follows a pipeline structure to go from raw data to final results and guides the user along with simple colour schematic as well as optimized information messages during the various processes. Current development involves further improving and increasing the interactive functionality of the visualization tools, further automation of the optimal parameter selection to be used in the clustering algorithm, using normality and unimodality tests, tools to validate the clustering results, and additional quality control functionality to detect problems with the data or CytoSense instrument.



### 3.2.2.2. Standardisation of flow cytometry measurements and data

Besides the heterogeneity of the data collection in terms of acquisition mode and equipment and the differences in subsequent techniques of processing the data (manual or automated), quality-control processes and standardised labels need to be applied before transferring any flow cytometry information to a data portal such as SeadataCloud or EMODnet-Biology. With the new marine observing technologies, the flow of measurements has increased by orders of magnitude and in a same way has increased the challenge of delivering of data in a format which can be used to better understand marine ecosystems for healthy oceans under global / climate change by integrating new information on phytoplankton dynamics and diversity into biogeochemical models, to build scenarios for marine socio-ecological systems under changing oceans by adding in situ data for improving forecasting harmful algal bloom and better understanding the size-dependence in the food web. Great progress has been made in physical oceanography to make available the data to users in a standardised format through portals such as EMODnet-PHYSICS. In terms of biological variables, it is very different. In the case of water quality assessment, the phytoplankton diversity in the different marine areas has mainly been studied from discrete samples collected manually, with the species determined later by microscopy. The development of new technologies and particularly flow cytometry (since the 1980s) has revealed new functional groups such as picophytoplankton, and has greatly increased the sampling density. The integration of high frequency data from flow cytometry into already established data infrastructure is a challenge.

During the INTERREG IVA “2 Seas” (2010-2014) project called DYMAPHY ([www.dymaphy.eu](http://www.dymaphy.eu)), standardisation and data format have been discussed, and inter calibration exercises using the CytoSense were organised between the partners to define the best procedures and to increase the confidence for using the data. The main conclusions were:

1. some parameters related to instrument and the analysis are crucial to report such as the specifications and configuration of the instrument (lasers used, adjustment of optical parts, detector settings, amplifier settings) and the settings of the instrument during analysis (warming time lasers, minimal and maximal sample are throughput, triggering parameter and levels, number of particles and or minimum volume that should be analysed).
2. Each CytoSense needs to be calibrated each analysis day. Beads results should be stored in calibration charts to validate the well-functioning and good alignment of a CytoSense during a day and during long periods (e.g. years) to enable the comparison of yearly gathered data (drift check). Another reason to use calibration beads is to intercalibrate between CytoSense instruments. A standard bead should meet several requirements: visible in most of the detectors by each CytoSense although different instrumental configurations are used, good (=small) short and long-term variability, size not too small, not too large and acceptable concentration, not expensive, not sticky, well discriminating and therefore not overlapping with species fingerprints.
3. Performance indicators should be listed in an instrument dependent specification table by each user and they are related to the repeatability, the reproducibility, the calibration with respect to the size and the concentration of particles, the accuracy, the correctness, the sensitivity, the selectivity, specificity and the robustness of the instrument.

An operational Common protocol for Pulse-shape recording flow cytometry (Rutten et al. 2013) has been also established as well as a reporting format for data but never implemented to date.

### 3.2.2.3. Towards a standardised labelling and inter calibration for high frequency flow cytometry measurement: activity within JERICO-NEXT during the last 24 months

The aim of the consortium was to implement and improve in JERICO-NEXT some of the procedures already established in other projects to make visible and available the flow cytometry data collected at high frequency through a data portal. Taking advantage of workshops and cruises organised by WP4 in JERICO-NEXT as well as discussion with WP5 and collaborators also involved in SeaDataCloud, intercalibration procedures and standardised labelling for the high frequency data from flow cytometry have been established. Even if the





phytoplankton community is very diverse in a sample, it has been decided to combine and report the phytoplankton into 4 main categories: *Synechococcus*, *Picoeukaryotes*, *Nanophytoplankton*, and *Microphytoplankton* with mandatory and optional information for each sample (see below).

Mandatory information:

- Total number of phytoplankton particles per ml
- Contribution of the phytoplankton particles to the total particles (%)
- Total number of particles by functional groups: picoeukaryotes and *Synechococcus*, nano- and microphytoplankton per ml per sample
- recognized microalgae (pictures)
- Contribution relative of the main category to total red fluorescence (%)

Optional information :

- Total red fluorescence standardised to total chlorophyll *a* for each sample
- Median size of the phytoplanktonic community
- Number of sub-groups in each main 4 categories and number of phytoplankton particles in the sub-groups)

#### **Comparison between the different modes of clusterisation for PFGs**

16 Files from different JERICO-NEXT cruises (2016-2017) have been provided by the partners (Cefas, VLIZ, CNRS-LOG/ULCO, CNRS-MIO) to a FTP site established by VLIZ. They covered the North Sea (Cefas-RWS, VLIZ, CNRS-LOG/ULCO), the Channel (CNRS-LOG/ULCO, VLIZ), and the Mediterranean Sea (CNRS-MIO). The task of each partner was to use CytoClus software to cluster the PFGs and the results will be compared with the RclusTool outputs to estimate the discrepancy between the different clustering modes. Some results were processed already (see section 3.2.5) and other are still being processed. A scientific publication is being written.

#### **Inter calibration between Cytosenses/Cytosub**

##### Analyses of natural community during cruises for inter comparison

The same natural samples have been analysed at high frequency by on-line flow cytometry during:

- Second workshop on Phytoplankton Automated Observation, Gothenburg, September 2016: the workshop was held at the Oceanographic unit of SMHI (Swedish Meteorological and Hydrological Institute), at SMHI Tångesund observatory in Mollösund, and on the Swedish Skagerrak coast. Partners from RWS, MIO CNRS-LOG/ULCO and VLIZ participated and were able to compare two flow cytometers (Cytosenses from RWS and CNRS-LOG/ULCO) on bench mode in the laboratory in Gothenburg.
- CytoBuoy Workshop, March 2017: participants from Cefas, RWS, VLIZ, CNRS-LOG/ULCO, CNRS-MIO attended a CytoBuoy workshop in Woerden (NL) in order to present their current and past work and to discuss about technical and analytical improvements together with the manufacturers and with scientists and engineers from different European countries and abroad using automated flow cytometry.

A common cruise in the English Channel/ southern North Sea in May 2017: the VLIZ spring cruise (May 2017, 8<sup>th</sup> - 12<sup>th</sup>) onboard the "Simon Stevin" RV investigated mainly the North Sea and Strait of Dover. Four institutes were involved on this cruise: CNRS-LOG/ULCO, VLIZ, RWS and NIOZ. 44 stations were investigated with several measurements (see D4.1), and 3 on-line flow cytometers made analysis of the phytoplankton community in continuous mode. Comparisons of the results are still being processed and a scientific publication is being written.



### 3.2.2.4. Next steps

#### 3.2.3. Conclusions and first recommendations for the use of automated flow cytometry

Results obtained from automated flow cytometry depend to some extent on the specifications, configuration and settings of the instrument (wavelength and intensity of the laser used, multiple or single-laser, optical parts, detector settings, amplifier settings, triggering parameter and levels applied, etc.), as well as the clustering method applied (manual or automated) by each operator. During DYMAPHY project (INTERREG IV-A “2 Seas”, 2011-2014), intercalibration exercises were performed with one type of flow cytometer (CytoSense, Cytobuoy b.v.) in order to study the specificities and the complementarity of each device/settings. A common operational procedure was also defined for flow cytometry measurement. As the technology changes very quickly and new coastal areas have been added during JERICO-NEXT, analytical procedures need to be reassessed. Guidelines on quality control during the measurement and results analysis are still in process. Because the flow cytometer analysis is user and machine dependant, a single analytical procedure is not recommended today. However, by setting a quality control procedure for sample and result analysis, and intercalibration exercises in regular base, we expect to finally reach the degree of confidence needed for the data to be reported in European data portal.

#### 3.2.4. Technical improvements

##### 3.2.4.1. Coupling of a flow cytometer with a FerryBox system

Méllilotus Thyssen, Gérald Grégori, Michel Denis, Cherif Sammary, Pierre Marrec & Nagib Bhairy (CNRS MIO), Véronique Créach (CEFAS), Machteld Rijkeboer (RWS), Felipe Artigas (CNRS-LOG)

Sampling of phytoplankton functional community structure at the basin scale, hourly to monthly scale, with a sub meso to meso scale resolution and from single cell automated acquisition is a scientific challenge that has now accomplished its first successful trials. Phytoplankton functional community structure resolution is a step forward in the understanding of marine ecosystems, especially of the main biogeochemical fluxes triggered by the phytoplankton, and strongly tuned by the presence of the main functional groups observed and their spatial and temporal dynamics. Dedicated flow cytometers such as Pulse Shape flow cytometers (PSFCM) (Cytobuoy bv, Woerden, The Netherlands) are capable of sampling at the single cell level to ensure absolute counting with size and functional group identifications and running on a remotely controlled automated way. They can be coupled with a FerryBox (FB) with measures in a same time different biogeochemical variables. The first coupling between a flow cytometer and automated temperature and salinity sensors from a continuous flow through system was run in 2007 onboard a sailing vessel (Thyssen et al., 2008) from the Azores to Brittany for 10 days. A second coupling was validated during the DYMAPHY-PROTOCOL cruise with a pocket FB (Thyssen et al., 2015), on the RV CEFAS Endeavour for 5 days. The first known test between a flow cytometer and a Ferry line was run in 2013 onboard “L’Armorique” (Brittany Ferries) between Plymouth (UK) and Roscoff (F), in the frame of a LEFE-CYBER project. During JERICO-NEXT project (2015-2019), several coupling between the PSFCM and FerryBox (or Pocket Ferry Box) have been operated with success by the partners (CEFAS, RWS, CNRS-LOG, CNRS-MIO, VLIZ) in separated or join cruises, on research vessels (CEFAS “Endeavour”, CNRS “Côtes de la Manche”, IRD “Antéa”, VLIZ “Simon Stevin”, RWS “Zirfaea”, Ifremer “Thalassa”) as well as onboard “Le Carthage” (Compagnie Tunisienne de Navigation) between Marseille (F), Tunis (T) and Genova (I) by CNRS-MIO.

##### Installation on platforms (research vessel and Ferries)

FB are systems that can operate with foreign sensors as far as the implemented sensor’s company and FB constructor and provider collaborate to share their competencies for optimising the coupling in terms of water flow, location and time. The PSFCM CytoSense (Cytobuoy bv, Woerden, The Netherlands) has been coupled to a FB (4H-JENA engineering GmbH, Germany) onboard in two ferry lines (“L’Armorique”, Brittany Ferries, LEFE CEL2SAT project 2013, and “Le Carthage”, Compagnie Tunisienne de Navigation, A\*MIDEX CHROME project, 2015, PI M. Thyssen, CNRS-MIO) and RV CEFAS Endeavour (Monitoring programme for Eutrophication, ICES International Bottom Trawl Survey, and integrated surveys, P.I. Véronique Créach). The flow cytometer runs on the FB flow through, from existing spare outlets before the acid wash connection. Flow rate from the FB is



approximately max 1 bar, and 7 to 8 dm<sup>3</sup>/min. The connections and additional fittings on the PSFCM depend on the platform and time of the survey which can be between days and several months.

Each equipment is connected to its own computer and the analysis processing is setting independently. It can be installed on a bench (Figure 3.2.10A) or on a dedicated frame (Figure 3.2.10B). The flow cytometer can be directly connected to the FerryBox system (Figure 3.2.11A, B). A chamber can also be placed in between the two systems to avoid any overpressure in the main flow through during acquisition of the sample (Figure 3.2.11C) in fully automated system in ferries (Thyssen et al., 2008).



Figure 3.2.10. 2 CytoSenses side by side connected onboard the RV CEFAS Endeavour (A) and onboard the "Le Carthage". They are connected to the 4H-JENA FerryBox.

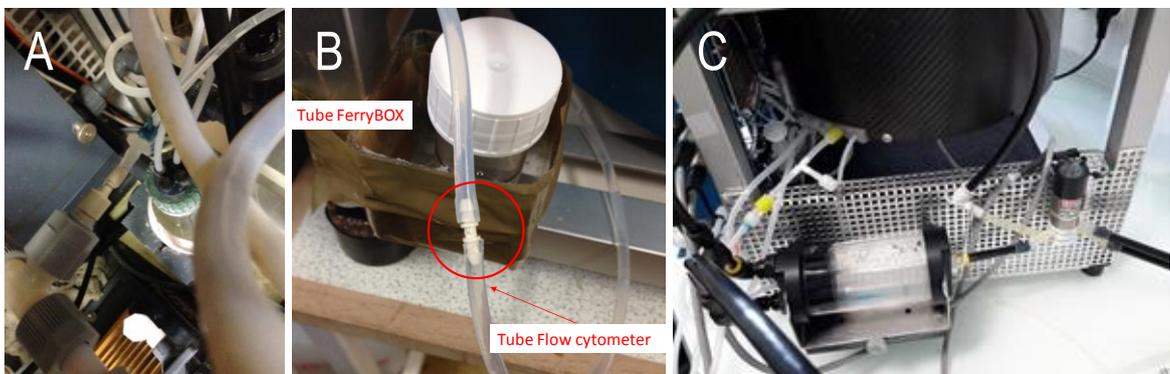


Figure 3.2.11. Connection between the FerryBox and the flow cytometer: 1. direct (A, B), 2. through a chamber (C).

For a fully automated system, an additional device fixed on the frame of the flow cytometer called "Beads-Sheath-Toxins" systems ensures an increase in deployment capacities by cleaning the sheath with coal filters and additional 1  $\mu$ m filters (Figure 3.2.12). The BST system programmes bead injections (Figure 3.2.12B) for quality control and biocide for keeping the sheath non affected by biofouling.

#### **Connection to the FB for GPS, data retrieval, internet and remote access to the PSFCM**

The GPS information can be retrieved by the CytoClus software (Cytobuoy bv, Woerden, The Netherlands). GPS reading from the PSFCM enables scheduling the sampling with the FB, out of the harbor zones. The GPS information is recovered separately from the FB and merged after PSFCM analysis. The schedule of the PSFCM depends on the start of the FB, and should be controlled by the geographical position.



Figure 3.2.12. Beads-Sheath-Toxins (BST) system with coal filter and classical 0.1  $\mu\text{m}$  filter ensure long life clean sheath fluid.

Without accessing GPS data from the FB, any schedule must be manually programmed based on route and timeline of the ferries or research vessel. The success of managing the start to the end of the PSFCM analysis depends on the reliability and the regularity of the timeline for ferries and but also on people for research vessels. The CytoSense embedded computer is directly connected to a SIM card for internet connection and remote control. Quality checks (internal pressure and temperature sensors, number of samples collected, size of the samples) for the instrument and analysis can be done by using remotely CytoClus or EasyClus software (Thomas Rutten projects, Middelburg, The Netherlands, Figure 3.2.13). Data may be real-time analyzed and made accessible via websites like [www.fytoplankton.nl](http://www.fytoplankton.nl) (EasyClus live module). For the operator to follow the performance of the flow cytometer or other devices and for the water manager to keep track on the behavior of the phytoplankton community and other relevant data.

## PhytoplanktonLIVE

Online automatic phytoplankton monitoring at several locations performed by flowcytometry (CytoSense-CytoSub)



Figure 3.2.13. Example of output from near real time data by phytoplankton live (EasyClus software) to be used by operator and water managers.

### 3.2.5. Analytical improvements

#### 3.2.5.1. Comparison between automated and manual classification of a sample

An inter-comparison exercise between manual (by several experts) and automated clustering was performed. The aims of this exercise mainly concern the harmonization and standardisation of raw data, metadata and results for their integration in a database. Some points must be considered:

- Provide possibilities to partners for checking the quality of their analysis methods,
- Compare the different approaches for the data classification (manual and automated),
- Highlight the criteria and needs of each institute for the result files for further statistical analyses.

Data were analyzed by each partner, depending on the classification approach used:

- manually, with CytoClus (CEFAS, CNRS-MIO, VLIZ, CNRS-LOG),
- automatically, with the RclusTool package (CNRS-LOG).

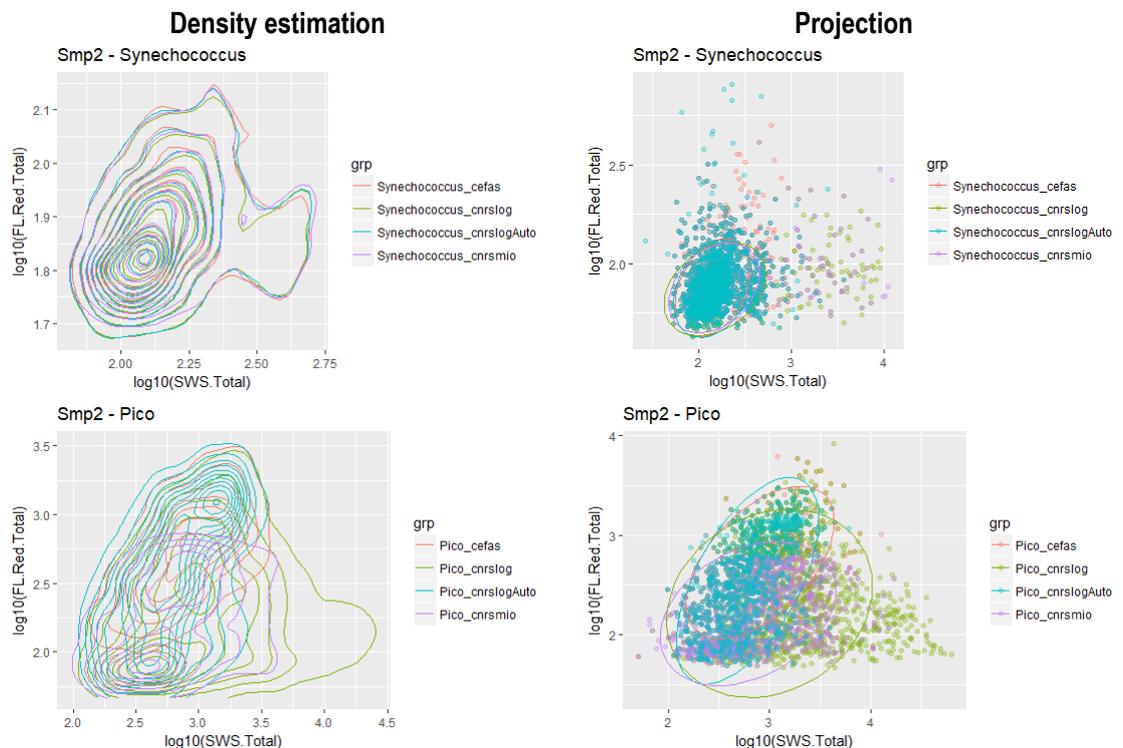
Here, a sample taken in the Eastern English Channel was digitized with the CytoSense instrument and clustered by all partners. In order to easily compare the different kinds of classification, only 4 functional groups were discriminated: *Synechococcus*, Pico-, Nano- and Micro-phytoplankton. The obtained results were presented in the Figure 3.2.14.

**Sample name:** R1\_flr10\_5min\_5µls\_high 2016-05-19 12u30

**Digitized volume:** 1.02 ml

**Trigger channel:** FLR10

**Total number of particles:** 45846



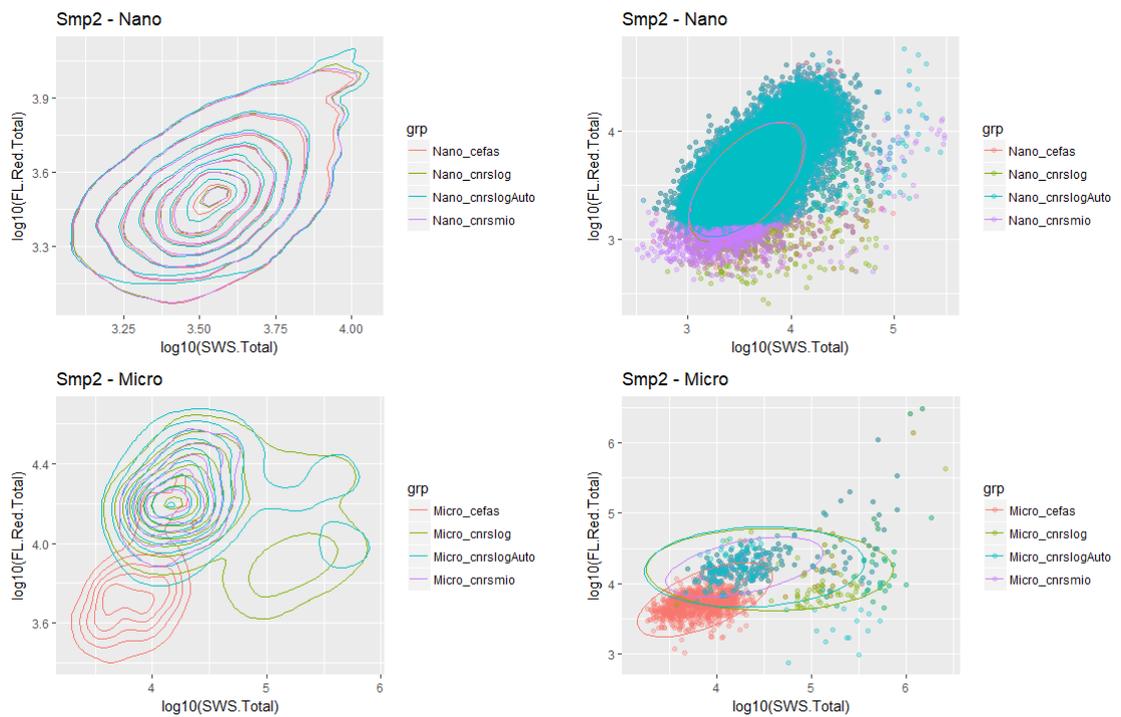


Figure 3.2.14. Overlay of groups obtained by the different classification methods (manual and automated) for each functional group. The first column presents the graphs of density estimation, and the second column presents the projection of particles in the space defined by Total SWS and Total FLR).

The graphs in Fig. 3.2.14 show a high similarity in the definition and the discrimination of the different groups, and more particularly for *Synechococcus* and Nano-phytoplankton. For the Pico-phytoplankton, the manual definition of the group for CNRS-LOG is wider than for other partners. In order to quantify the global similarities between manual and automated clustering, the Rand Index (RI) is computed. The results are presented in Table 3.2.1. This index represents the proportion of pairs of particles (x,y) which are similarly classified in the 2 types of clustering: either in the same group or in different groups.

$$RI = \frac{a + b}{N(N - 1)/2}$$

with:  $a$  = number of times the 2 clusterings assign the particles x and y to the same group.  
 $b$  = number of times the 2 clusterings assign the particles x and y to different groups.  
 $N$  = total number of particles.

The Rand Index has a value between 0 and 1, with 0 indicating that the 2 clusterings are completely different and 1 indicating that the clusterings are exactly identical.

Table 3.2.1. Rand indices between manual and automated clusterings for each partner (CEFAS, CNRS-LOG\_Manual, CNRS-MOI, CNRS-LOG\_Auto).

	CEFAS	CNRS-LOG_Manual	CNRS-MIO	CNRS-LOG_Auto
CEFAS	1	0.95	0.94	0.89
CNRS-LOG_Manual		1	0.97	0.91
CNRS-MIO			1	0.90
CNRS-LOG_Auto				1

The results presented in Table 3.2.1 show high similarity scores between clusterings. Even if the lowest Rand indices concern the automated method, the obtained values are important with an average of 0.90, which offers interesting perspectives for automated classification.



This methodology for inter-comparison was applied on 12 samples from 3 different geographical areas (Eastern English Channel, North Sea and Mediterranean Sea) and two cruises in North Sea (next section). The obtained results showed high similarities between manual and automated clustering.

### 3.2.5.2. Comparison between automated and manual classification during cruises

**LifeWatch-JERICO-NEXT cruise.** Scientific campaign for analysis of plankton diversity (LifeWatch/JERICO-NEXT cruise, from 8<sup>th</sup> May to 12<sup>th</sup> May 2017 in North Sea, aboard R/V « Simon Stevin » VLIZ).

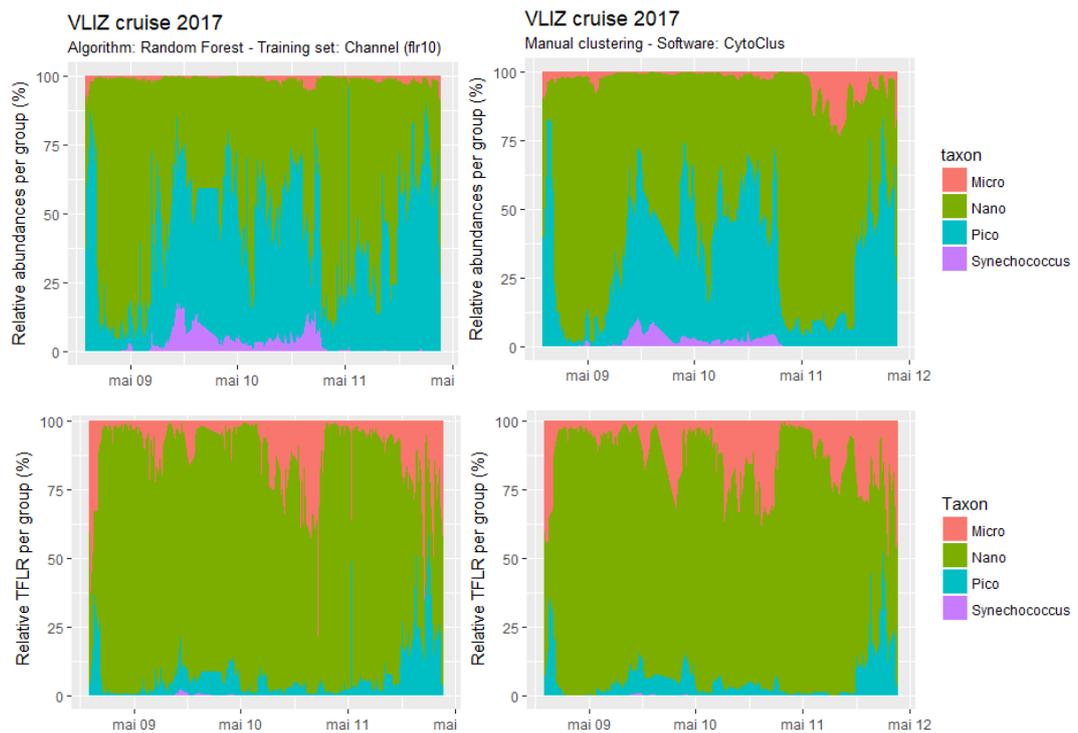


Fig. 3.2.15: Relative abundance (above) and total red fluorescence (below) of phytoplankton groups defined by automated classification (RClusTool, left) vs. manual clustering (right) during the VLIZ cruise (May 8-12, 2019).

The relative contribution of the four main phytoplankton functional groups addressed by flow cytometry to the total phytoplankton abundance and red fluorescence (TFLR as a proxy of total chlorophyll *a*) assessed by manual (CytoClus) and automated (R ClusTool) analysis during the VLIZ cruise showed a similar spatial and temporal pattern (Figure 3.2.15) with a possible over estimation of pico-eukaryotes and underestimation of Microphytoplankton abundance and TFLR by May 11, even though the similarity of both analysis is confirmed by the high  $\rho$  values of Spearman's rho (table 3.2.2).

Table 3.2.2. Spearman's rho between cell abundance and contributions to the total red fluorescence obtained by manual and automated techniques during the LifeWatch-JN cruise.

Groups	$\rho$ abd	pvalue	Groups	$\rho$ tflr	pvalue
Synecho	0.99	<2e-16	Synecho	0.95	<2e-16
Pico	0.84	<2e-16	Pico	0.59	<2e-16
Nano	0.89	<2e-16	Nano	0.95	<2e-16
Micro	0.73	<2e-16	Micro	0.81	<2e-16
Global	0.88	<2e-16	Global	0.95	<2e-16



**RWS cruise.** Scientific campaign for analysis of plankton diversity (RWS/JERICO-NEXT cruise, from 15<sup>th</sup> May to 18<sup>th</sup> May 2017 in North Sea, aboard R/V « Zirfaea » RWS).

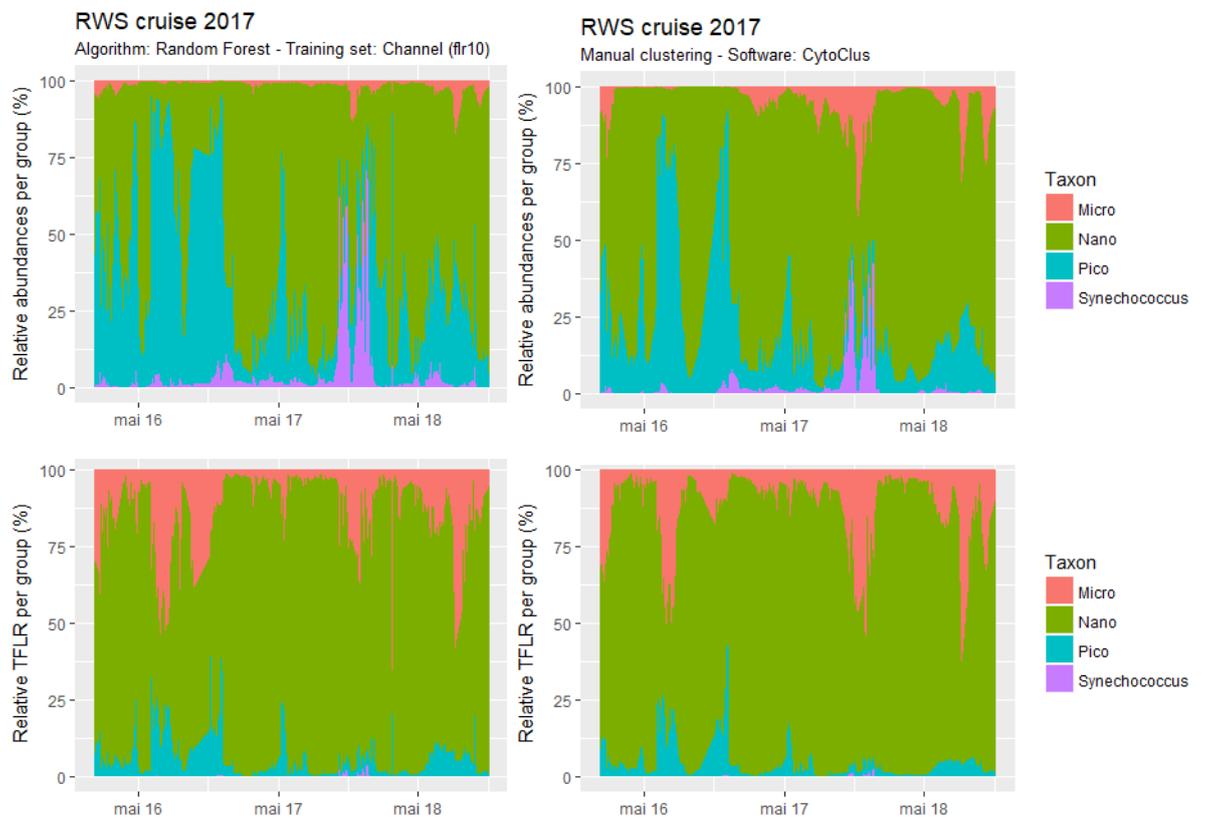


Fig. 3.2.16: Relative abundance (up) and total red fluorescence (bottom) of phytoplankton groups defined by automated classification (left) vs. manual clustering (right) during the RWS cruise (May 15-18, 2019).

The relative contribution of the four main phytoplankton functional groups addressed by flow cytometry to the total phytoplankton abundance and red fluorescence (as a proxy of total chlorophyll *a*) assessed by manual (CytoClus) and automated (R ClusTool) analysis during the RWS cruise showed a similar spatial and temporal pattern (Figure 3.2.16) with a possible over estimation of pico-eukaryotes and *Synechococcus*-like abundance and underestimation of Microphytoplankton by May 15 and 17, even though the similarity of both analysis is confirmed by high  $\rho$  values of Spearman's rho (table 3.2.3).

Table 3.2.3. Spearman's rho between cell abundances and contributions to the total red fluorescence obtained by manual and automated techniques during RWS cruise.

Groups	$\rho$ tflr	pvalue	Groups	$\rho$ abd	pvalue
Synecho	0.94	<2e-16	Synecho	0.96	<2e-16
Pico	0.98	<2e-16	Pico	0.96	<2e-16
Nano	0.95	<2e-16	Nano	0.87	<2e-16
Micro	0.92	<2e-16	Micro	0.80	<2e-16
Global	0.96	<2e-16	Global	0.96	<2e-16



### 3.3. Bio-optical Instrumentation

Jukka Seppälä, Suvi Rytövuori & Pasi Ylöstalo (SYKE), Felipe Artigas·Emilie Houliez, Fabrice Lizon & Arnaud Louchart (CNRS-LOG), Alain Lefebvre (IFREMER), Pascal Claquin (CNRS-BOREA), Bengt Karlson (SMHI), Jacco Kromkamp (NIOZ), Klas Möller & Jochem Wölschlagel (HZG)

#### 3.3.1. *Short overview of methodology and instruments commercially available*

The major technologies developed and tested in this subtask include LED fluorometers, spectral fluorometers, spectral absorption meters and variable fluorescence instruments, as available for continuous online measurements for different coastal observation platforms. The current status of the systems is reported in JERICO-NEXT deliverable D2.2 and the overview of sensors given here is very brief.

Single waveband LED fluorometers are widely available in various models from several manufacturers. Key fluorescing components relevant for this subtask are chlorophyll-a, phycocyanin and phycoerythrin. Chlorophyll-a is used as a proxy chlorophyll-a concentration, while phycocyanin is used as a proxy for filamentous cyanobacteria and phycoerythrin as a proxy for some filamentous cyanobacteria as well as picocyanobacteria, cryptophytes and some other phytoplankton groups. Technically instrumentation is well developed, relatively simple, and major concerns are in the instrument calibration and validation.

Spectral fluorometers measure fluorescence emission of chlorophyll-a after excitation through accessory pigments using LEDs with different wavebands. The major use of the instrumentation is to get information on the abundance of different pigment groups. There are two major manufacturers (JFE Advantech Co, Ltd, Japan, and bbe Moeldanke, Germany), who provide instrumentation with slightly different optical configurations. The main objective of JERICO-NEXT subtask 3.1.3 is to compare and develop algorithms and calculation tools to analyse spectral data and to retrieve as much biological information from spectra as possible.

Spectral absorption meters studied in JERICO-NEXT subtask 3.1.3., HyAbS (HZG, custom made) and the OSCAR-G2 (TriOS, commercially available) are instruments for measuring the absorption coefficients of the water constituents, taking advantage of an integrating cavity for this purpose. In comparison to conventional spectrophotometric measurements, this enhances the sensitivity of the measurements due to an increased optical path length and avoids biases occurring from scattering of light on particles present in the sample (Röttgers et al. 2005). Both sensors are designed to be used in flow-through mode and provide absorption coefficient spectra in high resolution (less than 2 nm) in the visible range of the light. Since this is done in high frequency (distinct spectra every 5-10 s), they deliver optical information in high resolution.

For variable (active) fluorescence and FRRF technique, there are three main types of instruments used in JERICO-NEXT community, 1) the FastOcean (Chelsea Technologies Group Ltd, UK) sensor, which can be used in different modes: as a profiler (ADP), as a bench top model, using the FastActs accessory to measure light curves on discrete samples, as an automated flow-through system on board, using the Act2-accessory. The last two systems are very similar (optically identical), and both can be used as bench top instruments, but the FastOcean/Act2 combination allows to measure fluorescent light curves continuously using a water supply (normally using the water that is also provided to a ferrybox). 2) the FFL-40 (Photon System Instrument, Czech Republic) which the PSI does seem able to deliver anymore. FastOcean sensor has three excitation bands at 450, 530 and 624 nm and PMT detector, while FFL-40 has two bands at 458 and 593 nm and photodiode detector (Houliez et al. 2017). Both instruments allow measurement of rapid light curves, using different levels of actinic light illuminating the sample. The profiling ADP system uses the solar light in the water column. The 530 and 624 nm LEDs used in the FastOcean FRRF are used to add extra excitation light in the case the output is not sufficient to reach the maximum during the fluorescence induction curve, a problem known to occur with cyanobacteria and Cryptophytes. This is based on the fact that the chlorophyll-a in these organisms is associated with photosystem-I, which does not fluoresce at normal temperatures. The green and orange LED specifically target the phycobilisomes of cyanobacteria and Cryptophytes. 3) Pulse Amplitude Modulated (PAM) fluorometers. A PhytoPAM (Heinz Walz GmbH, Germany) is also used in JERICO-NEXT. It has several wavebands (440, 480, 540, 590, and 625 nm) and using the technique similar to spectral fluorometry aims in resolving fluorescence parameters for various phytoplankton pigment groups. The new version of PhytoPAM allows also determination of functional absorption cross section of PSII. Other PAM instruments are also available (e.g. from Photon System Instrument, Czech Republic; Turner Designs, U.S.) and additional techniques measuring phytoplankton productivity with fluorescence



include FIRE - Fluorescence Induction and Relaxation System (Satlantic, U.S.) and Profiling Natural Fluorescence radiometer (Biospherical Instruments Inc., U.S). For variable fluorescence studies, the key research item in JERICO-NEXT subtask 3.1.3. is to determine constrains of the conversion factors from electron transport rate to C-fixation at various spatio-temporal scales.

### 3.3.2. Results and discussion - developments, evaluations and experiences

#### Exploring the effects of photo-quenching on chlorophyll fluorescence

In high irradiance conditions, phytoplankton protect their photosystems from bleaching through nonphotochemical quenching processes (Milligan et al. 2012, Müller et al. 2001). The consequence of this is suppression of fluorescence emission. Daytime fluorescence quenching (i.e. the reduction in the fluorescence quantum yield) is often observed in depth profiles of in situ chlorophyll fluorescence obtained using a CTD with a chlorophyll fluorometer on a research vessel. A decreased signal in chlorophyll fluorescence in the upper five meters is commonly observed in daytime compared to night time.

The effect of photo-quenching was investigated using a data set from the SMHI oceanographic buoy Huvudskär E. located in the NW Baltic Sea at latitude: 58.9333 and longitude: 19.1667 (Figure 3.3.1). A chlorophyll fluorometer (Wetlabs Inc. ECO FLNTU) was mounted at approximately 1 m depth. Excitation wavelength was 470 nm and emission wavelength 695 nm. The instrument was fitted with a copper shutter to minimize biofouling of the optical window. Figure 3.3.2. illustrates the preliminary results for the period 8 March to 12 September 2017. Panel A shows the whole data set. A large variability is obvious. Panel B shows only data around midnight local time (2200, 2300 and 0000 UTC) and panel C data from around noon (1000, 1100 and 1200 UTC). The night time chlorophyll fluorescence is much higher. Unfortunately, there is an offset in the preliminary data. Chlorophyll fluorescence levels are too high.

A similar effect was observed in data from a chlorophyll fluorometer mounted on an oceanographic buoy in the Kattegat (Karlson et al. 2009). A quenching effect resulted in approximately three times higher chlorophyll fluorescence at night compared to daytime chlorophyll fluorescence for the same phytoplankton population.

The main implication for using chlorophyll fluorescence as a proxy for phytoplankton biomass is that irradiance has to be taken into account when interpreting the data. Without correction to photo-quenching, the daytime data need to be treated with care. It is recommended that reference water samples are collected regularly for analysis of chlorophyll in the laboratory.

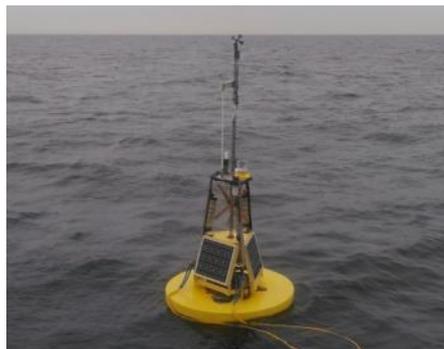


Figure 3.3.1. The SMHI Huvudskär E. oceanographic buoy.

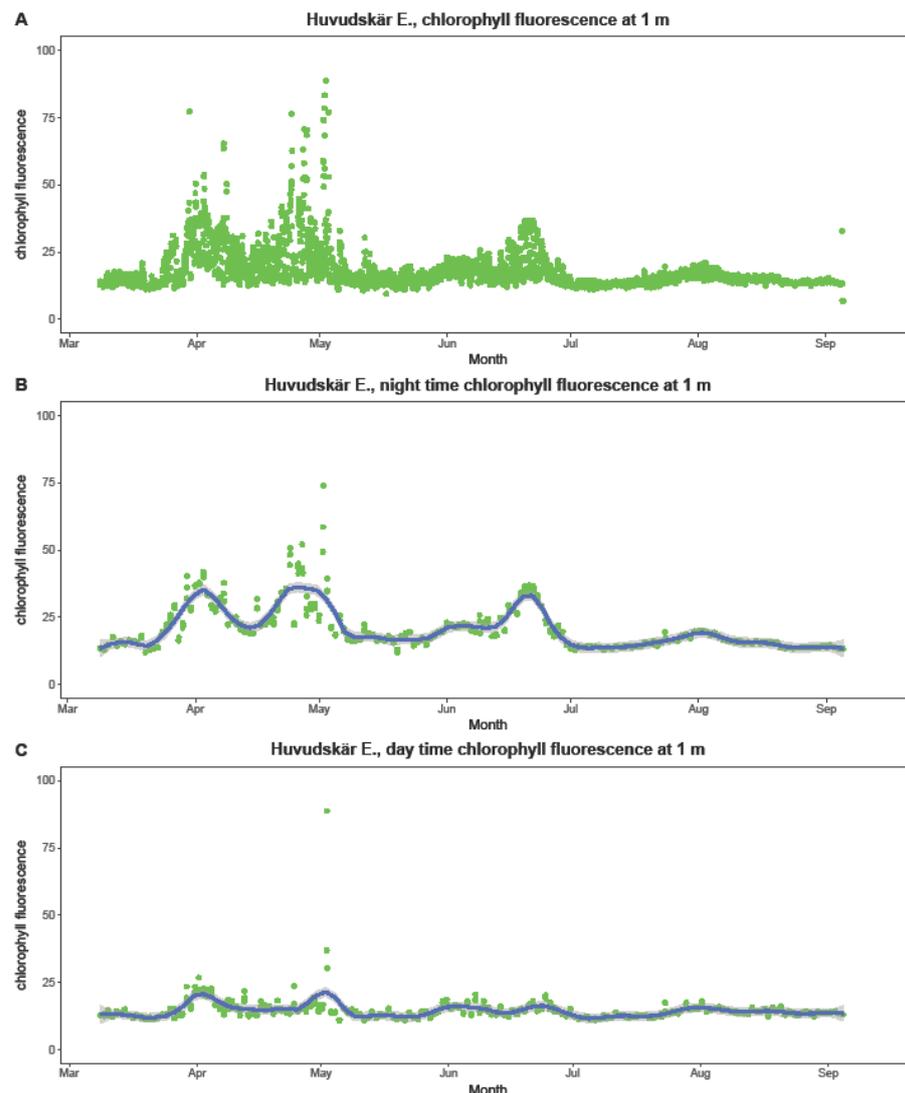


Figure 3.3.2. Chlorophyll fluorescence measured at 1 m depth in the NW part of the Baltic proper in year 2017. The y-axis shows arbitrary units. The graphs illustrate preliminary data. A: data collected every hour, B: data collected at 2200, 2300 and 0000 UTC, i.e. around midnight local time, C: data collected at 1000, 1100 and 1200 UTC, i.e. around noon local time. There is an offset in the data, which means that fluorescence levels are exaggerated.

#### Testing phycoerythrin fluorometers.

Two phycoerythrin (PE) fluorometers were evaluated for Baltic Sea research, microFlu Red (Trios GmbH) and Unilux (Chelsea Technologies Group Ltd, UK). Instruments measure PE fluorescence using green excitation light peaked at 507 nm (microFlu Red) or 523 nm (Unilux) and they detect fluorescence emission  $>590$  nm. We tested their specificity using algae cultures with known pigment content, and signal was noted only for samples with PE containing species, thus there seem to be no influence from chlorophyll-a or phycocyanin fluorescence to PE signal. We also tested that flowthrough cap, used during field measurements, does not affect the signal intensity. We determined the limits of detection (LoD) and limits of quantification (LoQ) for both instruments using ultrapure water and algae-free marine samples. As an example, LoD for microFlu Red changed from 0.0052 V in milliQ water to 0.022V in Baltic Sea water with salinity of 6 PSU, while LoQ varied from 0.0073 to 0.025 V respectively. In Baltic Sea water sample, LoD for Unilux was 0.35 fluorescence units, while LoQ was 0.50. For Unilux the values in MilliQ water were practically zero. For microFlu Red we also noted an increase in the background signal when

salinity of the water decreases, i.e. when there is more humic matter causing the background signal. For example, in the salinity of 6 the background is approximately 0.02 V while in the salinity of 2 the background is 0.04 V.

Instruments were installed on board the Ferry “Finnmaid”, as part of alg@line ferrybox system (Figure 3.3.3.). The natural variability observed for the PE fluorescence varied from 0 to 0.1 V for microFlu Red and from 0 to 3 instrument units for Unilux (Figure 3.3.4.). Linear range for instruments, as tested using algae cultures were at least 0-1.2 V for microFlu red and 0-100 fluorescence units for Unilux. Thus, the natural variability of PE fluorescence in the Baltic Sea is quite close to the LoQ of the instruments and in lower part of the linear range.

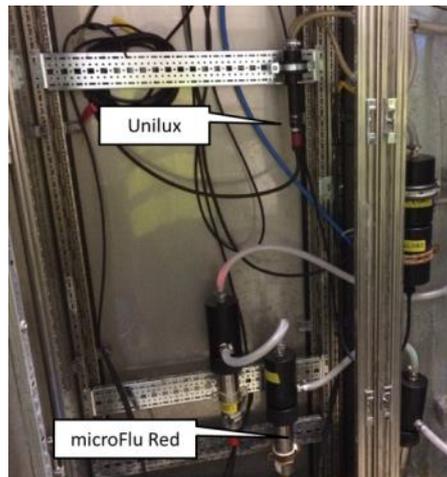


Figure 3.3.3. Phycoerythrin fluorometers installed onboard Finnmaid.

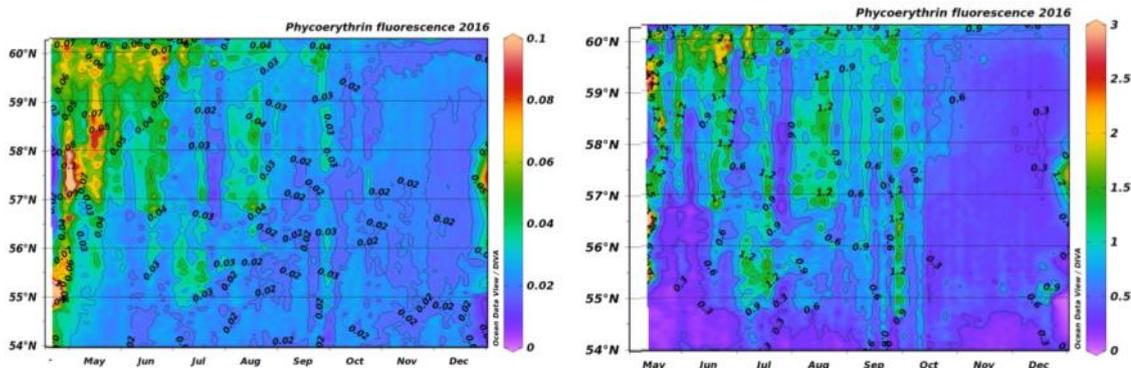


Figure 3.3.4. Variability of phycoerythrin fluorescence measured with microFlu Red (left) and unilux (right) fluorometers). Limits of quantification (LoQ) for instruments were 0.025 V and 0.5 fluorescence units, respectively. The values below these LoQ threshold are to be considered as background noise. Higher values indicate abundance of PE containing species.

During the deployment of PE fluorometers, we collected weekly watersamples from three stations, using automated water sampler onboard and made series of laboratory analyses: Chlorophyll-a, spectral absorption, spectral fluorescence in different size fractions, counts of PE containing picoplankton with epifluorescence microscopy and counts of PE containing nano- and microphytoplankton using flowCAM. Our aim was to resolve the origins of PE fluorescence and to estimate if the PE fluorescence could be used e.g. in tracking the abundance of picocyanobacteria (Seppälä et al. 2005). We did not find a direct correlation between PE fluorescence and abundance or biomass of PE containing species or with surface area of picocyanobacteria. The high PE values in spring are in case indicative for occurrence of ciliate *Mesodinium rubrum* and summer fluctuations are most likely related to variation in picocyanobacteria abundance. Some high PE readings are observed to match upwelling



events. In such events cells from the deeper layers are transported to surface, and as the phycoerythrin is well suited to harvest green light found in the deeper layers high PE readings most likely indicate the uplift of these organisms. An MSc thesis was written based on the results (Rytövuori 2017).

#### Deployments of multi-spectral fluorometers and data analysis

Several submersible spectrofluorometers were deployed during JERICO-NEXT cruises and a plankton workshop in Gothenburg (27 – 30 Sept. 2016): the FluoroProbe (bbe-Moldaenke, Kiel, Germany) by CNRS-LOG, AlgaeOnlineAnalyzer (bbe-Moldaenke, Kiel, Germany) by IFREMER and Multi-Exciter (JFE Advantech Co., Ltd., Japan) by SYKE. These instruments were tested in different ecosystems: the Baltic Sea, North Sea, eastern English Channel, Bay of Biscay and Benguela current (cf. JERICO-NEXT-WP3-D3.1 report).

The FluoroProbe (FP) differentiates 4 algal groups (blue group, green group, brown group and mixed group). It was designed to provide chlorophyll *a* (chl *a*) concentrations for these 4 groups and concentration of chromophoric dissolved organic matter (CDOM or yellow substance). Phytoplankton groups are differentiated thanks to a factory statistical algorithm based on the linear unmixing approach which requires to calibrate the instrument by recording reference fluorescence spectra on CDOM standards and pure algal laboratory cultures representative of the 4 phytoplankton groups to be identified (see Beutler et al. 2002 for more details). However, assessments of FP performance under field conditions highlighted some inaccuracy in chl *a* determination (Gregor and Maršálek 2004) and misclassifications of phytoplankton groups under specific conditions (e.g. Catherine et al. 2012, Kring et al. 2014). A great part of these discrepancies has been attributed to the linear unmixing approach and calibration procedure. This is because the linear unmixing approach relies on the assumption that the group-specific reference fluorescence spectra measured on pure cultures are constant (i.e. independent of the physiological status of the cells and independent of the species within a group) and representative of phytoplankton groups encountered in the field (Beutler et al. 2002). This assumption has, however, been shown to be often violated and this may lead to inaccurate estimations of chl *a* and identification phytoplankton groups because the linear unmixing results are strongly dependent on the similarity between the shape of the fluorescence spectra recorded on samples to be identified and the reference spectra (MacIntyre et al. 2010). To reduce bias due to inaccurate reference spectra, it has been advised to calibrate FP with species representative of the region of interest (Richardson et al. 2010).

This methodology has been shown to be efficient and to greatly improve FP results when reference spectra have been properly selected (Escoffier et al. 2015, Houliez et al. 2012, Leboulanger et al. 2002). The key problem is that only one set of reference spectra can be stored in the FP to run the linear unmixing. As a consequence, if the device has been calibrated for a specific area, using this calibration in other places with different taxonomic composition may lead to misclassification of phytoplankton groups. Ideally, a new calibration with site specific reference spectra should be made before to use FP in a new area. However, this makes the use of the same device for measurements in different systems (as it was the case during the JERICO-Next project) extremely complicated and time consuming. This also raises the question about the right calibration for measurements crossing areas with very different phytoplankton compositions. Moreover, it is not possible to share reference spectra with other FP users because reference spectra are instrument specific.

Another difficulty is related to the calibration which is usually made with laboratory cultures (i.e. phytoplankton cells grown under specific conditions) while environmental conditions experienced by phytoplankton in nature are dynamic and affect the pigmentation and fluorescence properties of the cells (including the chl *a*:fluorescence ratio). Reference spectra recorded on cultures are thus likely to differ from the fluorescence properties of taxa encountered in situ and this may result in misclassification of phytoplankton groups.

In order to face these difficulties, alternative methods to analyse spectral fluorescence data have been experimented using FP data collected during the JERICO-Next campaigns.





### Total chlorophyll estimation

To explore alternative ways of analysing the data, we started with total chlorophyll estimation. By software default Chl-a concentration [Chl-a] is estimated as

$$SFS = C_g * S_g + C_{bg} * S_{bg} + C_d * S_d + C_c * S_c + C_y * S_y + \text{offset} + E$$
$$[Chl-a] = C_g + C_{bg} + C_d + C_c$$

where

- SFS* = sample spectra at 6 wavelengths
- [Chl-a]* = total Chl-a concentration of sample
- C<sub>x</sub>* = concentration of algal class *x* ( $\mu\text{g Chl-a L}^{-1}$ )
- S* = spectral signal at 6 wavelengths
- g* = green algae
- bg* = blue green algae
- d* = diatoms
- c* = cryptomonads
- y* = yellow substances
- offset* = offset due to distilled water
- E* = error spectra at 6 wavelengths to be minimized

With the factory calibration FluoroProbe was overestimating chlorophyll concentrations (Figure 3.3.5), when compared against extraction method. Of course, such a mismatch may be corrected simply by adjusting fingerprint spectra, but there is no rationale how to do this in unbiased way. We analysed some alternative ways to estimate chlorophyll concentration from FluoroProbe data. First, we tested a simple regression between chlorophyll concentration and FluoroProbe fluorescence using excitation at 470 nm (analog to single waveband fluorometers). This yielded slightly lower coefficient of determination than FluoroProbe default method (Figure 3.3.5).

Secondly, we tested multiple regression, where fluorescence from all wavebands was used as independent variables, without constrains due to setting fixed fingerprints. This analysis yielded considerably higher coefficient of determination, but still the power of prediction seemed not quite good (Figure 3.3.5). While multiple regression, using all available wavebands, actually takes into account the variability in the group specific differences in chlorophyll-a specific fluorescence, it still does not take into account the major problem in such predictions, non-photochemical quenching of fluorescence which makes the chlorophyll fluorescence to concentration ratio to vary in depth, especially during sunny days, low values in surface in the midday, higher values in deeper layers and during night. Unfortunately, we did not have any light measurements during the cruise and the light level cannot be taken as additional parameter in the regression equation. However, we noted clear spatiotemporal shifts in the ratio, and including depth, time of day and distance from shore in the analyses yielded actually pretty good prediction of chlorophyll-a concentration when using FluoroProbe raw spectral fluorescence data (Figure 3.3.5). We still need to perform in depth analyses of nonlinearities of the system and residuals, as well as study the generality of the solution using other datasets.



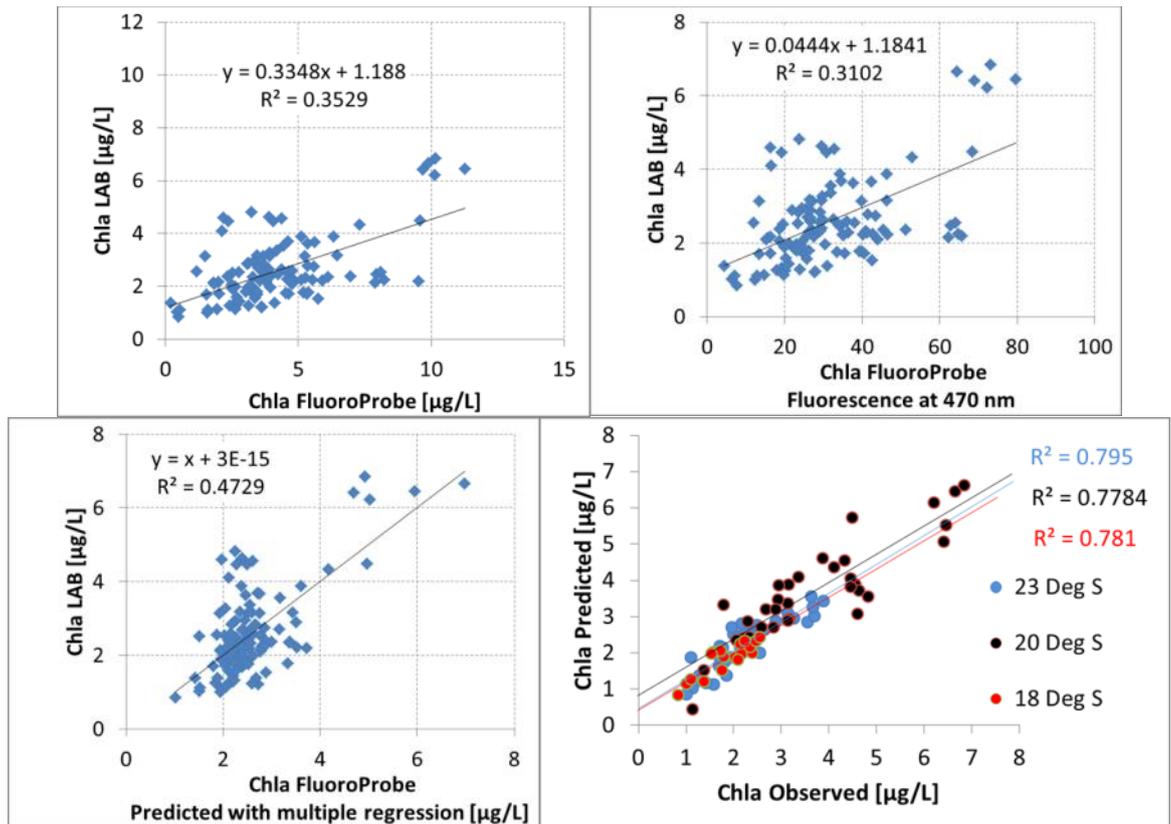


Figure 3.3.5. Estimation of chlorophyll-a concentration with FluoroProbe data. Upper-left: Default analysis using FluoroProbe software versus laboratory analyses, Upper-right: relationship between fluorescence at 470 nm waveband and chlorophyll concentration, Lower-left: relationship between chlorophyll analysis using multiple regression with all FluoroProbe wavebands and chlorophyll concentration, Lower-right: relationship between chlorophyll analysis using multiple regression with all FluoroProbe wavebands and accessory variables (depth, time of day, distance from shore) and chlorophyll concentration.

Major spectral components: applications in the Benguela system and Baltic Sea

FP measures fluorescence emission at ~ 680 nm after chl a excitation by 5 LEDs centered at different wavelengths (370, 470, 525, 570, 590 and 610 nm). Primary output of the instrument is thus fluorescence intensity measured at the 5 excitation wavelengths (data called raw data). Instead of using the calculated chl a concentration by algal groups obtained with the factory statistical algorithm, Alexander et al. (2012) and Harrison et al. (2016) proposed to analyze these raw data. They used ordination analyses, such as principal components analysis (PCA) or non-metric multidimensional scaling (nMDS), as alternative methods to derive spectral signatures from FP raw data and to reveal patterns in CDOM and phytoplankton community composition without reliance on predefined reference spectra for each phytoplankton group. As this alternative method has been applied only in two freshwater systems, it was thus desirable to explore its applicability in marine systems.

This alternative method was tested on FP raw data collected in two systems with very different phytoplankton composition: the Benguela system and Baltic Sea.

In the Benguela system, vertical fluorescence profiles were collected at 34 stations along the Namibian coast. During the cruise, we collected altogether 3848 spectra. The major spectral components in the dataset were analyzed using principal component analysis (PCA). Three first components of PCA (PC1, PC2 and PC3) contributed 99.9 % of the total spectral variability (Figure 3.3.6). Visual inspection of the principal components indicated that the first component (PC1) has an identical shape to the “diatom” spectra of the factory calibration fingerprints. PC2 was obviously similar to CDOM component and the PC3 showed a peak at 530 nm, not found in



any fingerprint. We speculate, however, that the PC3 represents the “true” phycoerythrin rich cyanobacterial group spectra in the Benguela system, unlike the fingerprint spectra showing phycocyanin.

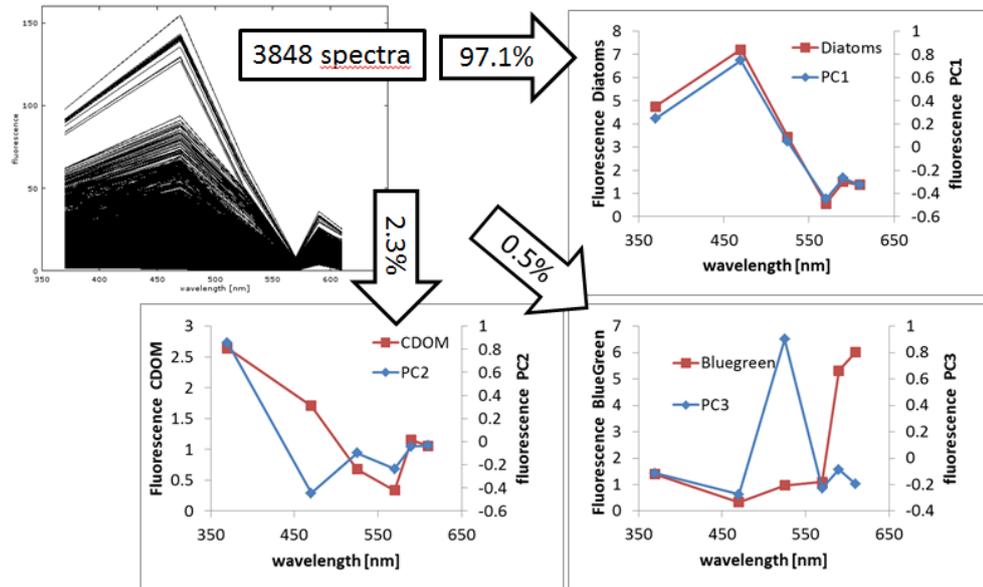


Figure 3.3.6. Principal component analyses of FluoroProbe spectral data acquired in the Benguela system. Upper-left: All spectra measured with FluoroProbe, Upper-right: Principal Component 1, showing similarity with fingerprint spectra of diatoms, Lower-left: Principal Component 2, showing similarity with fingerprint spectra of CDOM, Lower-right: Principal Component 3, showing peak at 530 nm, typical for phycoerythrin, while the cyanobacteria fingerprint spectra shown for comparison shows peak related to phycocyanin.

Using the PCA analyses we could explore how the different end-members (PC1, PC2, PC3) are distributed in the study area. As an example, colour coding the P1 vs. PC2 shows different locations with biomass variations (PC1) and CDOM variations (PC2) (Figure 3.3.7). Using such PCA method and endmember detection may be used as a powerful alternative for factory calibrated fingerprints, or determination of fingerprints with cultures of local species (Harrison et al. 2016). PCA identifies the major components of the spectra and these may be used further in analyzing the abundance of components. The dataset will be analysed further with pigment data (HPLC) and microscopy data, to indicate if PCA components show correlation with these more traditional approaches.

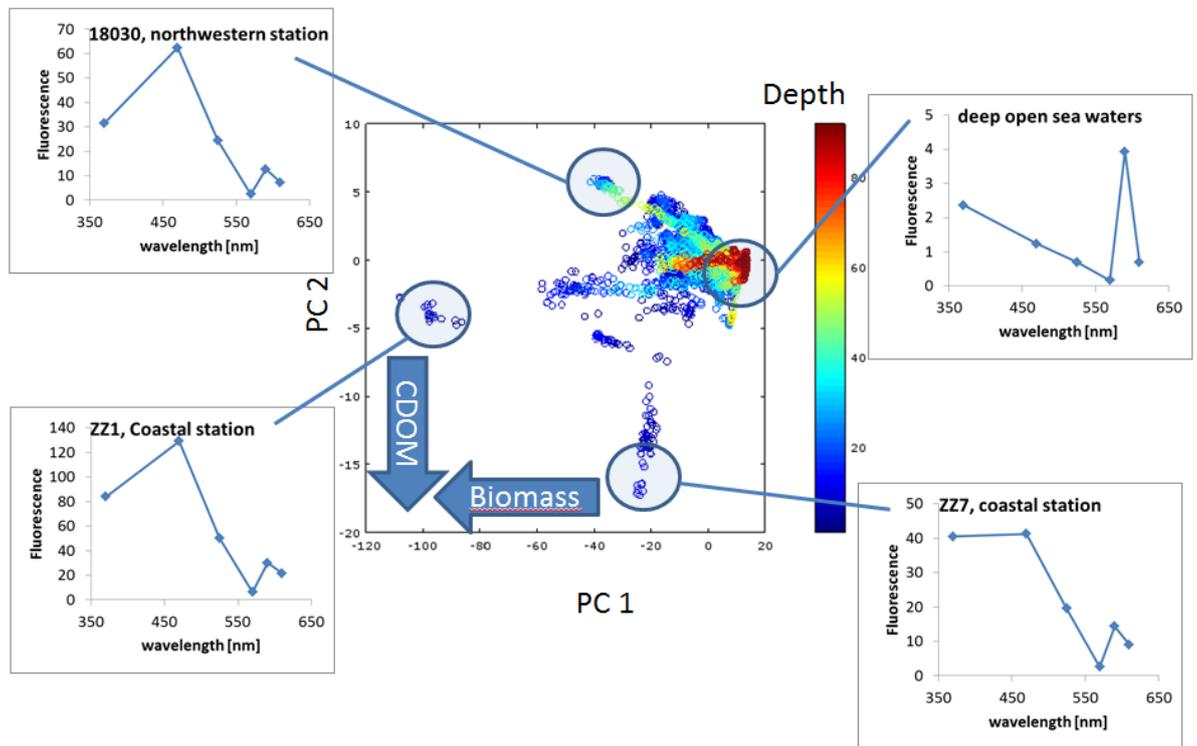


Figure 3.3.7. Relationship between principal components 1 and 2 of the principal component analysis of FluoroProbe spectral data in the Benguela system. Different endmembers may be identified regarding the CDOM and algae biomass.

In the Baltic sea, vertical fluorescence profiles were measured at 11 stations distributed from the Kattegat to the Gulf of Finland along the main environmental gradients (Figure 3.3.8). Raw data showed that phytoplankton distribution in water column was not identical at the different sampling stations and depth of maximum fluorescence differed (Figure 3.3.9). At stations LL7 (Figure 3.3.9A) and LL12 (Figure 3.3.9B), over the whole water column, the highest fluorescence intensity was recorded at 370 nm (i.e. wavelength used to measure CDOM) and was followed by fluorescence measured at 470 nm (i.e. wavelength where chl *a* and *c* present a maximum). At all other stations, the highest fluorescence intensity was measured at 470 nm on the first meters of the water column while at deeper depths, the highest fluorescence intensity was recorded at 370 nm (Figures 3.3.8 C, D, E, F, G, H, I, J, K). These raw data were analyzed using a PCA (Figure 3.3.10). The 2 first principal components explained respectively 49.9% and 30.1% of total inertia. The first principal component (PC1) was mainly built by fluorescence intensity measured at 370 (wUV) and 525 nm (w525) while the second principal component (PC2) was mainly built by fluorescence intensity measured at 570 (w570) and 470 (w470) nm (Fig. 3.3.10A). PC1 had a shape similar to the difference between the brown group and CDOM reference spectra provided in the factory calibration (Figure 3.3.10C). PC2 had a shape similar to the difference between the reference spectra of mixed and green groups (Figure 3.3.10D). PCA individuals distribution (Figure 3.3.10B) reflected fluorescence distribution within the water column (as shown in Figure 3.3.9): fluorescence spectra recorded within the first meters of the water column were dominated by phytoplankton and showed high fluorescence signals at 470 and 525 nm while fluorescence spectra recorded deeper were dominated by CDOM and showed a high fluorescence signal at 370 nm.

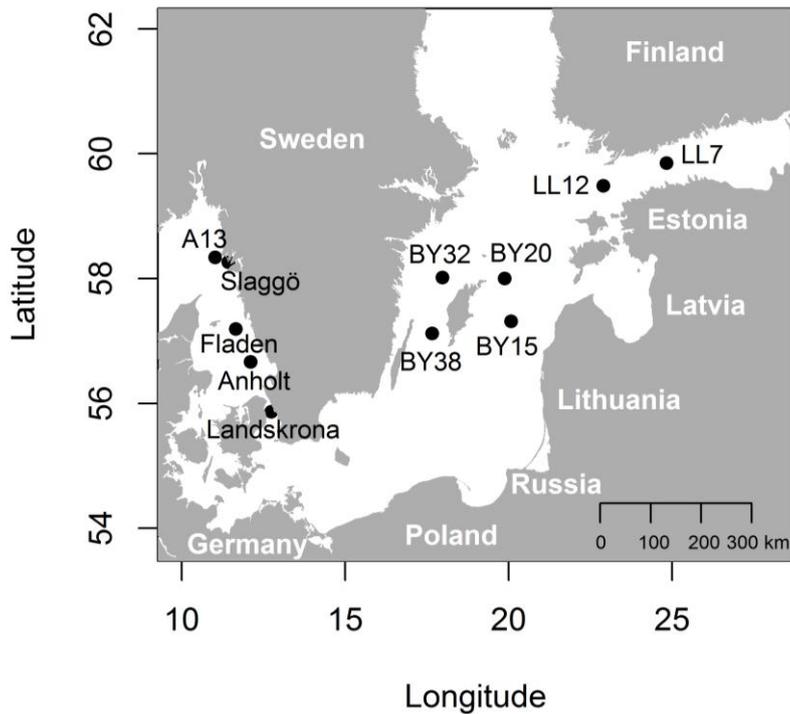


Figure 3.3.8. Map of the Kattegat and Baltic Sea representing the sampling stations.

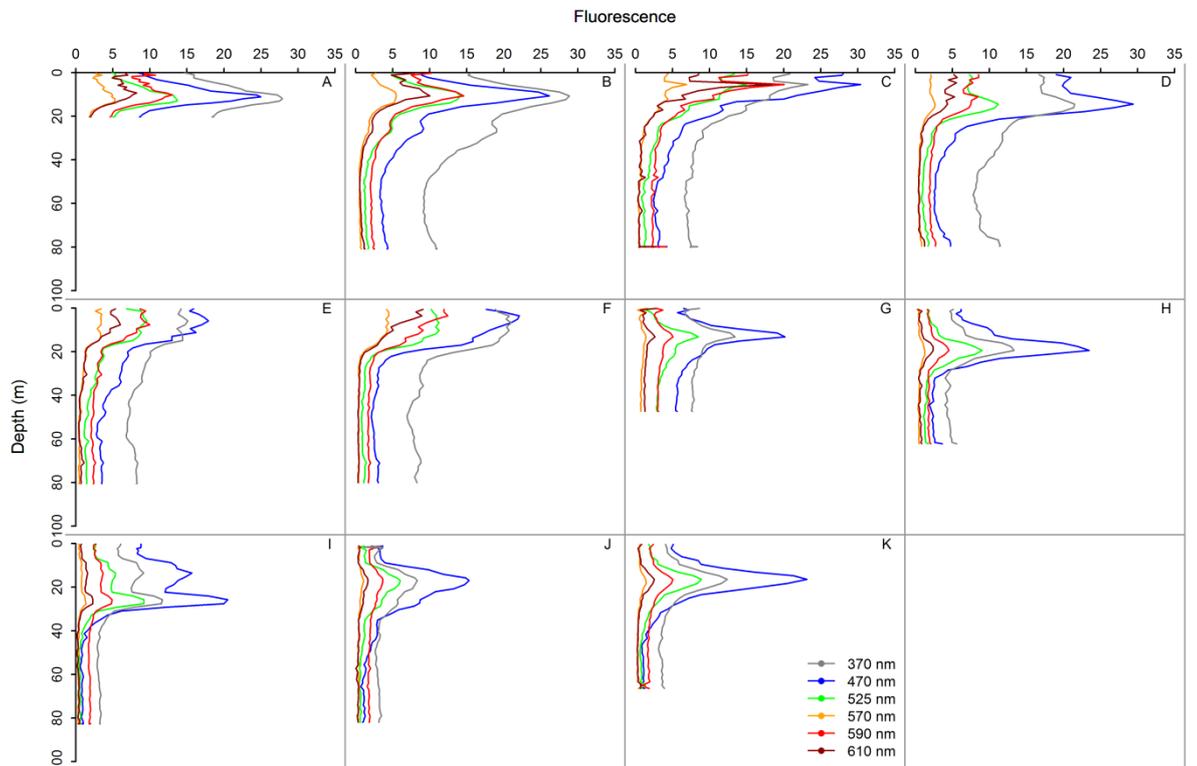


Figure 3.3.9. Vertical profiles of FluoroProbe fluorescence intensity measured after excitation at 370, 470, 525, 570, 590 and 610 nm at stations: LL7 (A), LL12 (B), BY32 (C), BY20 (D), BY38 (E), BY15 (F), Landskrona (G), Anholt (H), Fladen (I), A13 (J) and Slaggö (K).

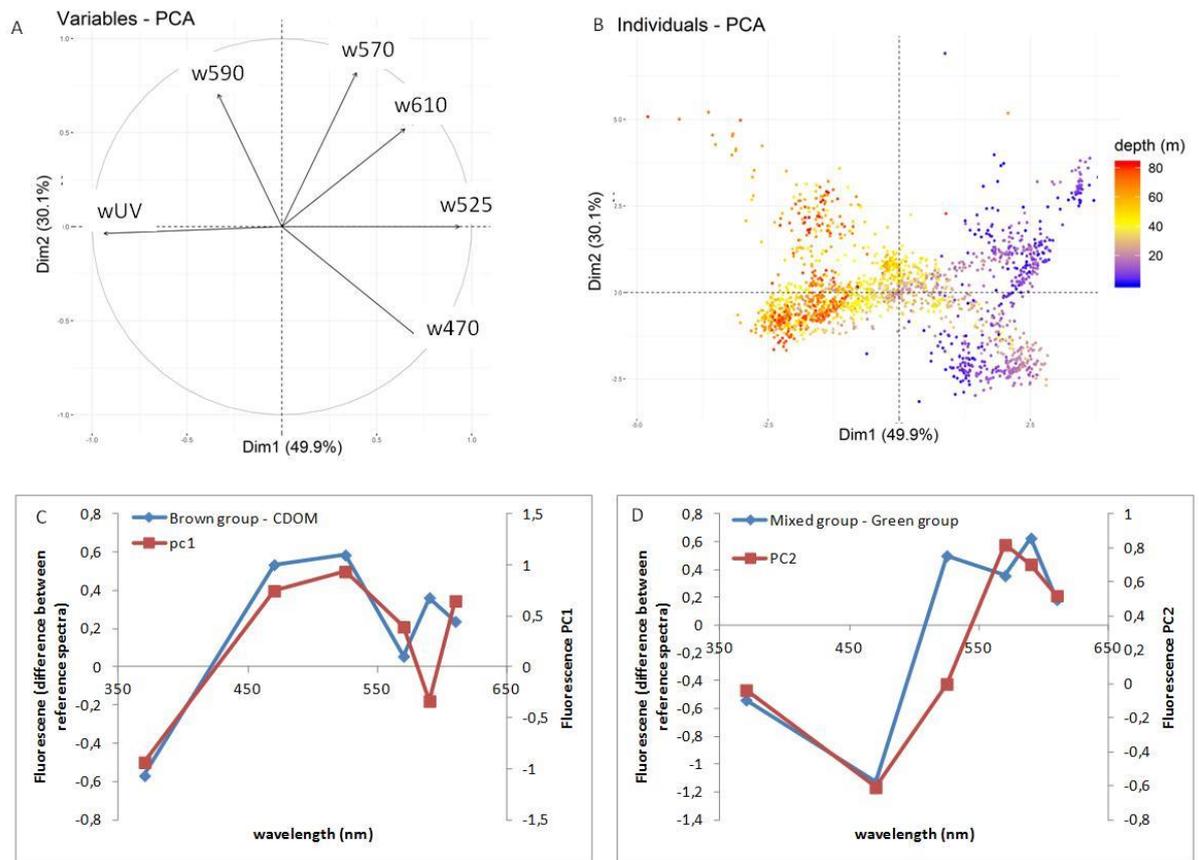


Figure 3.3.10. Principal component analysis of FluoroProbe raw data recorded over the whole water column. (A) PCA variables. (B) PCA individuals. (C) Shapes of PC1 and difference between brown group and CDOM reference spectra. (D) Shapes of PC2 and differences between mixed and green groups reference spectra.

To focus on phytoplankton communities, fluorescence spectra recorded within the first meters of the water column were selected to run another PCA analysis (Figure 3.3.11). The 2 first principal components explained respectively 45.9% and 36.6% of total inertia. PC1 was mainly built by fluorescence intensity measured at 470 (w470) while PC2 was mainly built by fluorescence intensity measured at 370 (wUV). PC1 had a shape similar to the opposite of the brown group reference spectra provided in the factory calibration (Figure 3.3.11B). PC2 had a shape similar to the opposite of CDOM reference spectra (Figure 3.3.11C).

PCA individual's distribution revealed three main termini (Figure 3.3.12A). The fluorescence spectra at terminus 1 presented a maximum at 470 nm and corresponded to the factory brown group reference spectra (Figure 3.3.12B). The fluorescence spectra at terminus 2 presented a maximum at 370 nm and were similar to the factory CDOM reference spectra (Figure 3.3.12C). The fluorescence spectra at terminus 3 presented a maximum at 590 nm and looked like a mixture between reference spectra of mixed and blue groups (Figure 3.3.12D).



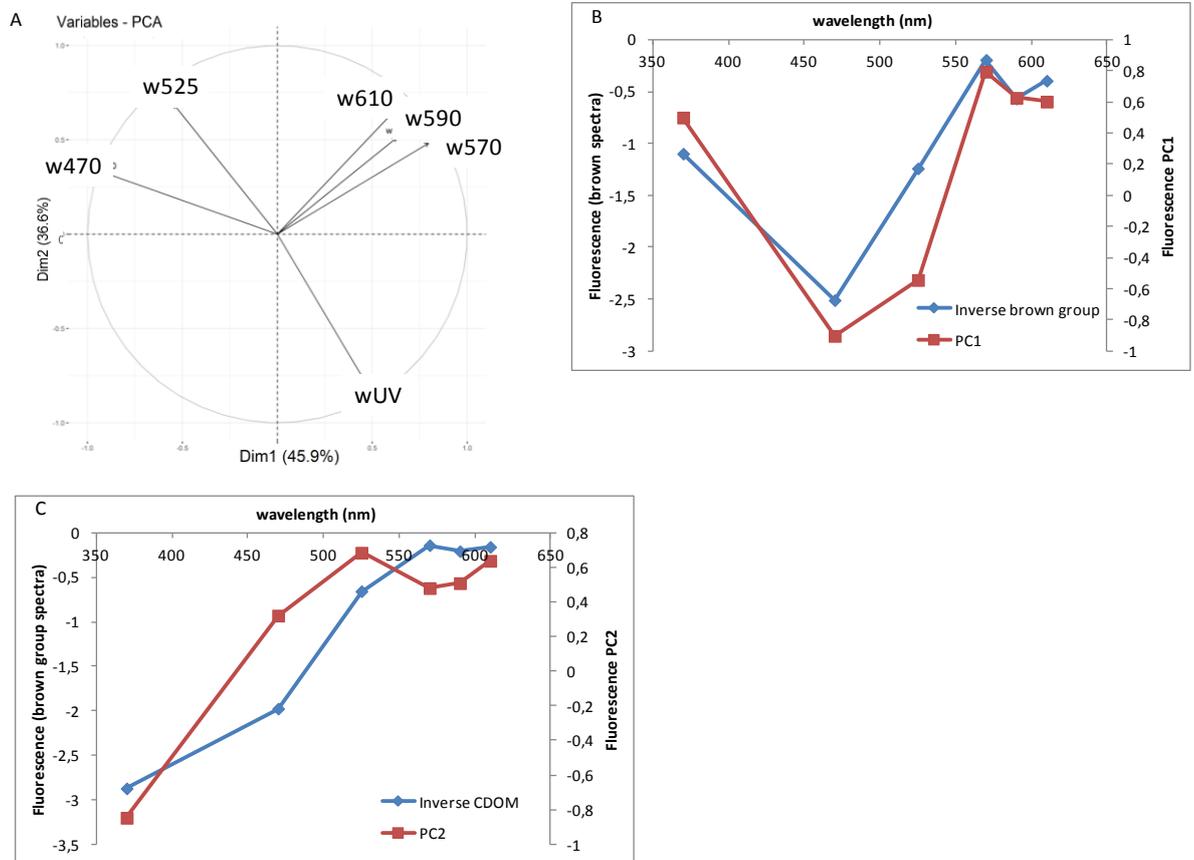


Figure 3.3.11. Principal component analysis of FluoroProbe raw data recorded within the first part of the water column. (A) PCA variables. (B) Shapes of PC1 and opposite of brown group reference spectra. (C) Shapes of PC2 and opposite of CDOM reference spectra.

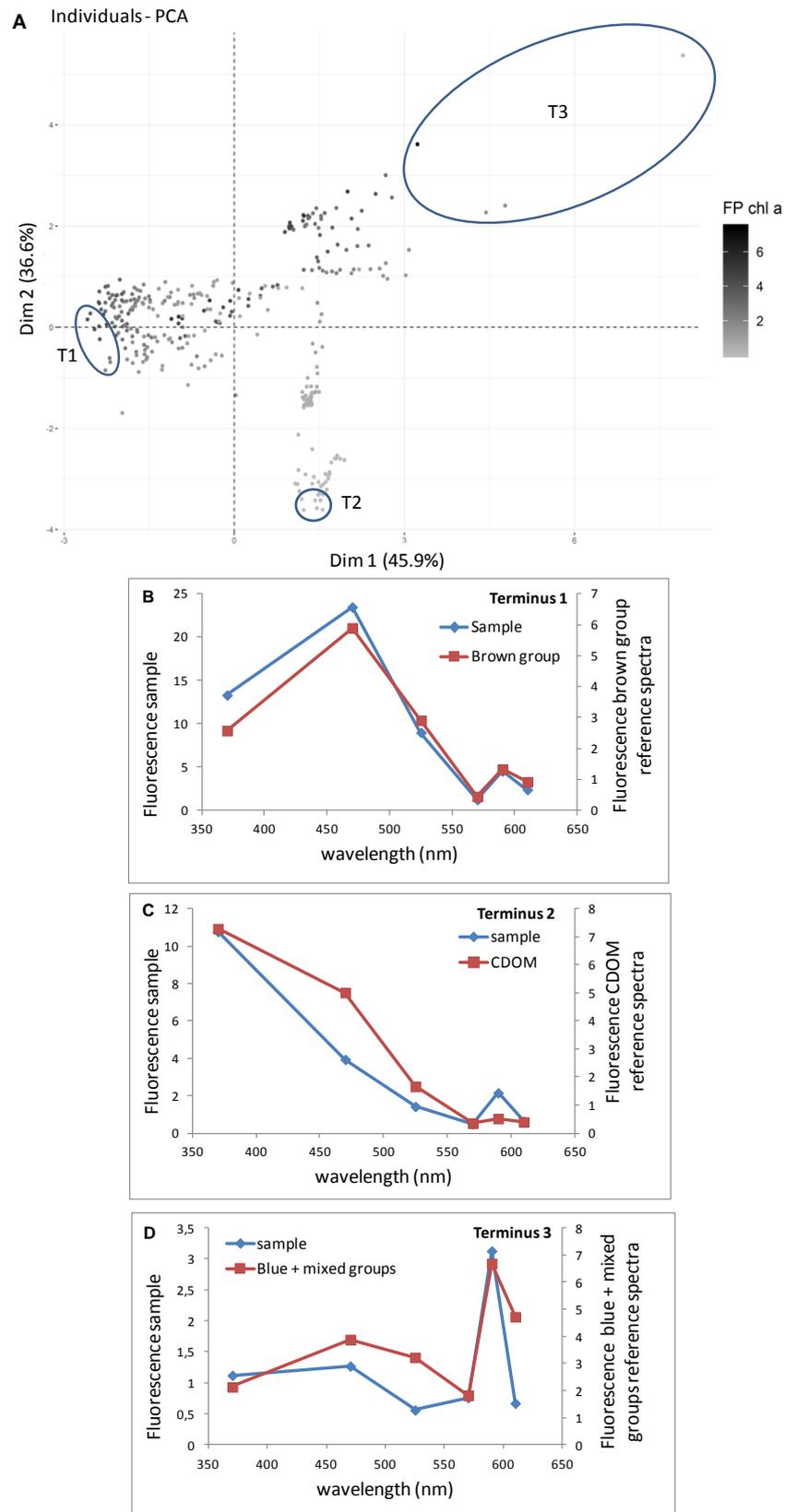


Figure 3.3.12. (A) PCA variables. (B) Fluorescence spectra of samples at terminus 1 and brown group factory reference spectra. (D) Fluorescence spectra of samples at terminus 3 and factory reference spectra of blue + mixed groups. T1 = terminus 1. T2 = terminus 2. T3 = Terminus 3.





This method seems to be efficient to separate the different phytoplankton assemblages by analyzing FP raw data. It would be desirable to test it in additional marine systems with very different phytoplankton structure and to further validate its utility through comparisons of FP measurements with other techniques to analyse phytoplankton such as flow cytometry and HPLC.

#### *Automated data analysis for spectral fluorescence*

Work was initiated during the DYMAPHY project on automated analysis of multi-spectral fluorometry data and a Fluorometry R-toolbox was proposed (Poisson-Caillault et al., 2015), which needs further improvements that are in progress. Moreover, the implementation of high-frequency measuring systems generates large multivariate series with missing values. To take advantage of information at infra-day scale, optimized numeric tools are required.

IFREMER and LISIC-ULCO teams work on methodological developments, originally planned for the instrumented station Marel Carnot database then the Pocket FerryBox series. We hope to be able to extend our approach to other multidimensional high resolution data set. The unsupervised Hidden Markov Model approach (uHMM R-package, <https://cran.r-project.org/>) allows to define environmental states characteristic of a combination of physico-chemical and biological parameters and their dynamics. To improve state characterization and state prediction, semi-supervised machine learning techniques are investigated. To deal with the drawback of missing values or intervals due to periods of sensor maintenance, failure, we work on two automated imputation methods, one for monovariate series, the second for multivariate series using Dynamic Time Warping algorithms.

### 3.3.3. Technical and analytical improvements

#### **3.3.3.1. Evaluation of hyperspectral absorption sensors**

From the spectrally resolved total absorption spectra, information about the optically active water constituents can be derived (Wollschläger et al. 2014). This is specifically the concentration of chlorophyll-*a* as a proxy value for phytoplankton biomass and the concentration of suspended matter. Furthermore, as the shape of the spectra is influenced by the presence of certain, often group-specific, phytoplankton pigments, the evaluation of the shape allows an estimation of the phytoplankton group dominating the investigated water.

The difference between the HyAbS and the OSCAR-G2 as the commercially available sensor is the degree of automation and compactness, respectively. The HyAbS is designed for running autonomously over a longer time period (days to weeks), given a supply for purified water. The whole measurement cycle including calibration and reference measurements can be customized in the LabView-based control software. A solid standard introduced by a motor is replacing the nigrosine dye solution that is normally used for calibration of integrating cavities. However, as the instrument is a custom-made design study and includes besides the components necessary for the actual measurements (light source, spectrometer, integrating cavity, data logger) also the components for automation (valves, pumps, tubing, control unit, PC), its dimensions are that of a box of approx. 60 x 40 x 40 cm, designed as a benchtop instrument.

In contrast, OSCAR-G2 includes the integrating cavity, the LED-based light source, the data logger, and the spectrometer in a compact, pressure resistant single housing. The dimensions of the device are approx. 40 x 16 x 16 cm. OSCAR-G2 can be used both as submersible instrument connected to e.g. a CTD for profile measurements or as a benchtop instrument. The instrument is configured by a web browser interface via PC and starts internal logging when disconnected while powered. However, water and reference or calibration solutions have to be introduced manually or by an external pump. The data stored by the instrument are light intensity spectra that have to be processed into absorption coefficient spectra subsequently to the measurements.

These sensors were employed and tested during two cruises on board the research vessel "Heincke" in 2017 (see Deliverable 3.1). One year later, another cruise (cruise 3) took place in the seas surrounding the United Kingdom and Ireland (18.07.2018 to 14.08.2018), and in 2019 the last cruise (cruise 4) was carried out in the German Bight area (19.03.2019 to 30.03.2019).



Cruise 3: In the course of preparation of this cruise, more detailed experiments were performed in order to characterize the performance of the solid standard calibration approach (results published in Wollschläger et al. 2019). Also, during the cruise, a variety of different (coastal) regimes were crossed (Figure 3.3.13), providing ideal conditions for further test of the HyAbS.

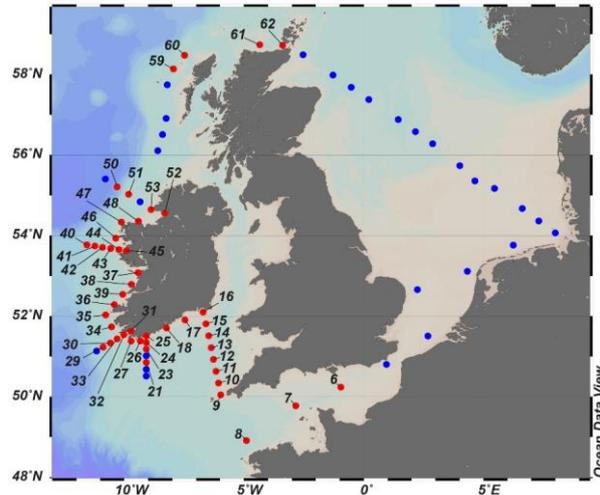


Figure 3.3.13. Area of the third HyAbS testing cruise.

The OSCAR was not used on this field cruise, but discrete measurement with a point-source integrating cavity absorption meter (PSICAM) were made for comparison with the HyAbS. The absorption-based chl-a estimates between the two instruments were in very good agreement (Figure 3.3.14), even if there were still some deviations visible when comparing the whole spectra (Figure 3.3.15). To some degree, they might be attributable to minor errors in reference or calibration measurements, but especially the large deviation in the red region are most likely due to the occasional occurrence of air bubbles in the cavity.

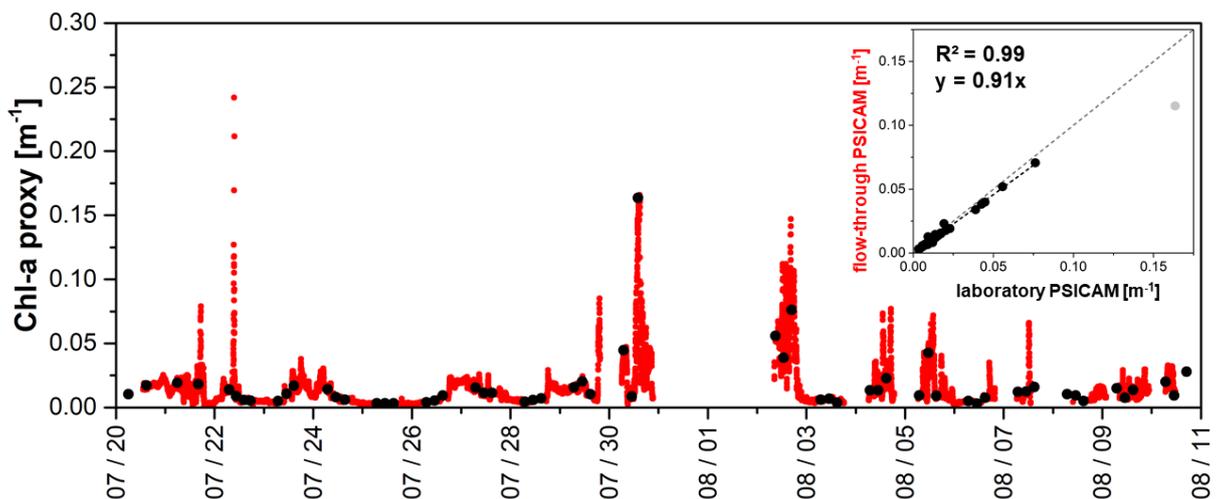


Figure 3.3.14. Comparison of the absorption-based chl estimates of the HyAbS (shown in red, here named as flow-through PSICAM) and the discrete PSICAM measurements (black). The inlay figure in the upper right corner shows the linear regression of the data at the sampling points.

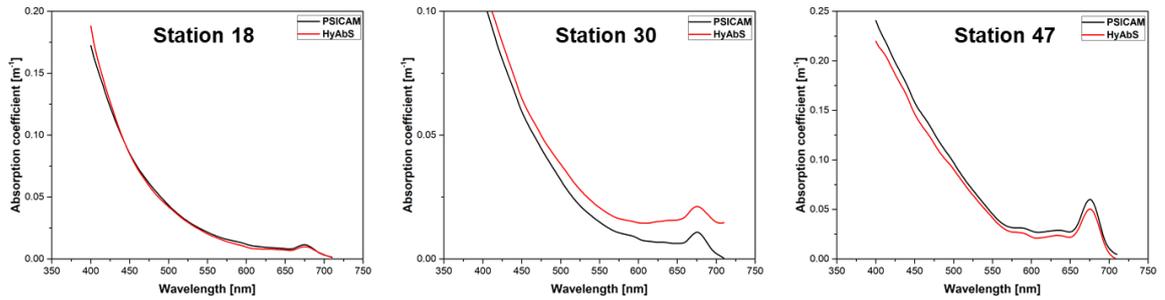


Figure 3.3.15. Examples of absorption coefficient spectra obtained at different station with the HyAbs (red) and the PSICAM (black).

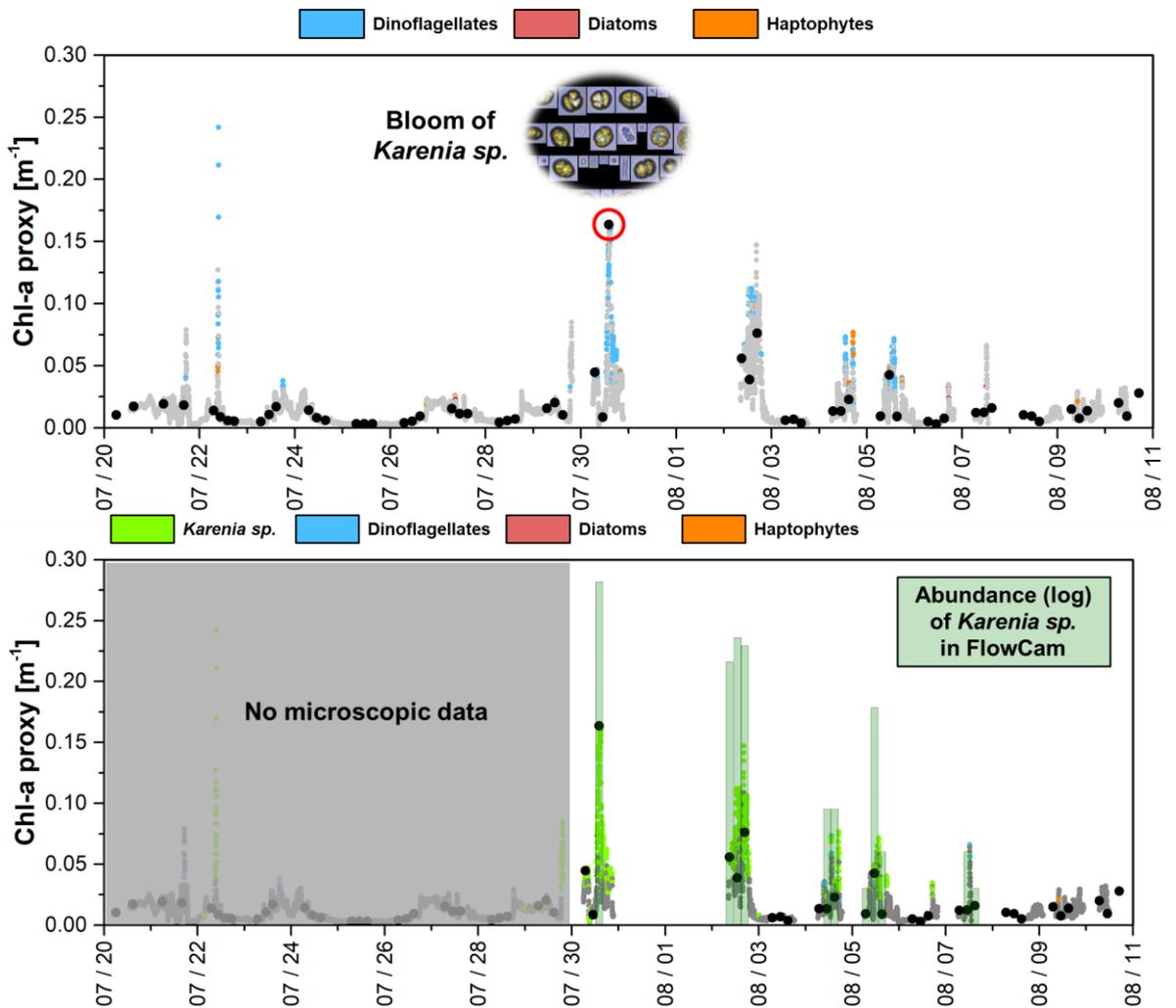


Figure 3.3.16. Results of the phytoplankton identification algorithm. Upper panel: Using laboratory grown spectra in the database. Lower panel: After supplementing the database with one field spectrum containing a high percentage of *Karenia sp.* The green bars show the log-abundance of *Karenia sp.* in the corresponding samples taken.

In the data evaluation subsequently to the cruise, also the performance of the phytoplankton identification algorithm was tested. For this purpose, the continuously measured spectra were compared with the verified lab spectra from our database. This led to the identification of the major chl-a peaks as belonging to dinoflagellate



spectra (Figure 3.3.16, upper panel). Nevertheless, most of the spectra could not be attributed to a certain group with the desired degree of confidence, probably because there is a certain difference in the spectra from phytoplankton grown under artificial light in the lab and phytoplankton grown under natural conditions. In order to test the effect of field spectra on the identification performance, one field spectrum was included in the database from a station where a bloom of the dinoflagellate *Karenia sp.* occurred (red circle in the upper panel of Figure 3.3.16). After reprocessing the data, a much higher percentage of the spectra was identified, mostly as being similar to the *Karenia sp.* field spectrum (Figure 3.3.16, lower panel). The distribution of these identified spectra was quite similar to the distribution of *Karenia sp.* as assessed by the FlowCam data. However, although these results are promising, it could also be that not exactly *Karenia* was detected, but in general dinoflagellates, as *Karenia* was frequently abundant when other dinoflagellates were present. Nevertheless, the higher percentage of identification alone shows the importance of incorporating validated field spectra in the database to improve the results of the phytoplankton identification algorithm, a task which should be tackled in future work more intensively.

Cruise 4: The aim of this cruise was primarily to explore ways to enhance the long-term deployment capacities of the HyAbS. Due to the use of the solid standard calibration approach, the need for a supply with nigrosine dye (the original liquid calibration standard), which was a severe limitation of the deployment time, could be bypassed. However, in order to obtain accurate data, still reference measurements with purified water have to be performed. Thus, the need for a purified water supply is currently the major obstacle in operation the HyAbS over a longer period. For this reason, we tested on this cruise the optical stability of the tap water provided by the ship. In case of high optical stability, the tap water could also be used as a reference, provided that its absorption coefficient spectrum can be measured at the beginning of the measurements. Subsequently, this spectrum could then be used to correct the HyAbS measurements. This approach would overcome the last major limitation factor to HyAbS deployment time and would bring the system closer to a status where it could be used completely unattended or at least with minimal human intervention.

However, the water source aboard the research vessel was relatively unstable in terms of optics on this cruise (Figure 3.3.17). But it has to be noted that with continued use of the tap water source (and thus flushing of the tube system aboard), the variability of the absorption coefficient spectra became less. Thus, intensive flushing prior to use or the additional use of filter cartridges might overcome this problem of high variability in the end. Nevertheless, further experiments are required in this respect. A further evaluation of the data is currently work-in-progress.

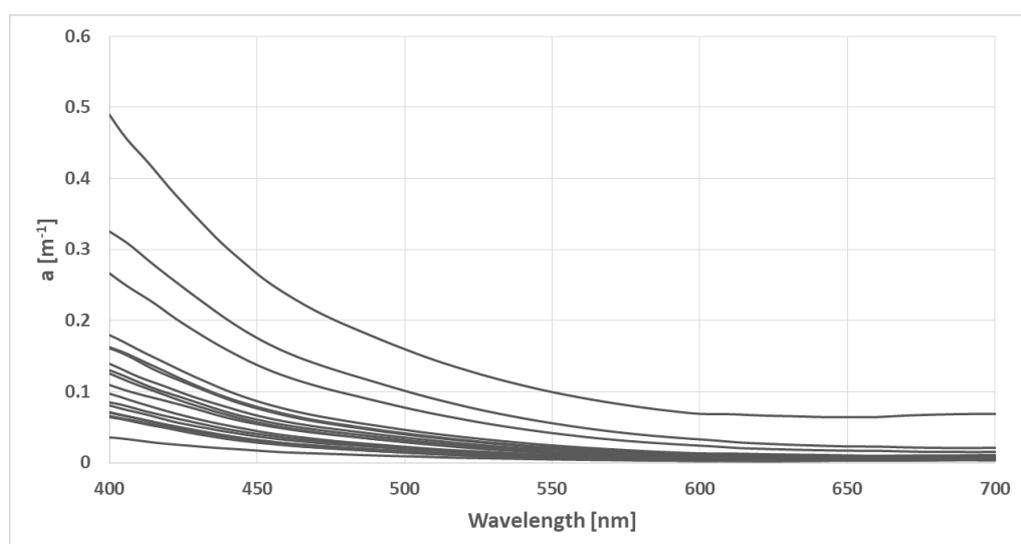


Figure 3.3.17. Absorption coefficient spectra of the tap water aboard the research vessel during the fourth HyAbS testing campaign.

### 3.3.3.2. Evaluation of variable fluorescence methods

#### **Sensor, measurement, parameters and protocols**

For our JERICO-NEXT cruises, continuous and automated online measurements were mainly carried out with the FastOcean Fast Repetition Rate (FRRf) system (Chelsea Technology Group). It is designed for profiling systems (APD system, Ambient Plus Dark sensor) and moorings but when coupled with Act2 system, it can be used in laboratory or in flow through applications, allowing automated measurements of Fluorescence Light Curves (FLCs). The inlet depth varies with ship, but is in general between 3-4 m, so the photosynthetic activity of the phytoplankton in the surface mixed layer is measured.

The FastOcean FRRF sensor system was used in 2016 and 2017 in eleven main cruises involving several laboratories of JERICO-NEXT (Table 3.3.1). On the Skagerrak cruise in 2016 a FastAct accessory was used in combination with a Chelsea technologies Mark-II FRRf: the FastAct is very similar to the Act2, but only discrete samples were measured.

From the eastern English Channel to the Baltic Sea, more than 90 vertical profiles were performed in five of these cruises, to measure photosynthesis in the water column between April and July. (Figures 3.3.18.- 3.3.19.). A PhytoPAM was also deployed on all of our cruises (cf Table 3.3.1) in order to compare another classical technique (but a multiple turnover technique) with the FRRf technique (single turnover).

*Table 3.3.2. Variable and spectral fluorescence measurements during JERICO-NEXT research cruises 2016-2017; C:: for continuous measurement in surface waters ; P : for profile in the water column ; D : discrete measurement in surface and bottom waters (s/b). The number of profiles and discrete measurements (in surface / bottom waters) are also indicated.*

Cruise	Area	Date	FRRf2 (C)	FRRf3 (P)	Phyto-PAM (D)	bbe-FLP (C)	bbe-FLP (P)
<b>CNRS R.V. Côte de la Manche - PHYCO</b>	English Channel 2017	April 20-30	X	37	46/48	X	47X
<b>VLIZ- R.V. S Stevin</b>	North Sea & E. Channel 2017	May 8-12	X	19	25/25		27
<b>RWS- R.V. Zirfea</b>	North Sea 2016	April 11-14	X	-	-	X	-
	North Sea 2017	April 10-14 May 15-18 June 19-22 August 14-17	X	17	18/15	X	
<b>NIOZ-R.V. Pelagia</b>	North Sea 2017	June 15-30	X	-	-	-	-
<b>CNRS R.V. Sepia II - PELRAD</b>	E. Channel, Strait of Dover 2017	July 4	X	6	-	X	17
<b>SMHI- R.V. Skagerrak</b>	Skagerrak-Kattegat 2016	September 5-9	X	-	-	-	-
<b>SMHI- SYKE R.V. Aranda</b>	Baltic Sea, Skagerrak & Kattegat 2017	July 10-17	X	16	19	X	16

Coastal areas see large gradients in phytoplankton biomass and physiological conditions. When automated underway measurements are made this requires continuous adjustment of instrument sensitivity and adjustments of the LED output used to generate the fluorescence induction curves. This is a necessity to optimize the signal to noise ratio and thus minimize the effect of noise in the sample. Fortunately, the FastOcean sensor, in contrast to its predecessors, have an inbuilt routine to so that these settings (autoPMT and autoLED) are optimized during the operation. Another difficulty with automated measurements is that no automated blank correction is possible. Blank fluorescence is fluorescence generated by any substance (often CDOM) that passes the filter (a 0.22 µm





filter) and contributes to the measured fluorescence signal. Samples for blanks were measured at intervals and manually subtracted from the fluorescence parameters ( $F_o$ ,  $F$  and  $F_m$ ,  $F_m'$ ) so that the measured photochemical efficiencies ( $F_v/F_m$  and  $F_q'/F_m'$ ) were not underestimated. The value of  $\sigma_{PSII}$  is not affected by background fluorescence. The calculation of absolute photosynthetic electron transport was performed as described by Oxborough et al. (2012, Silsbe et al. 2015). A detailed description of protocol carried out in most cruises is given below.

Data collection of single turnover (ST) fluorescence measurements were obtained with a classical acquisition protocol (100 saturation flashes, flash duration of 2  $\mu$ s and series of 24 sequences averaged and recorded). For each fluorescence acquisition, all blank correction, fits and calculations of all physiological parameters were made using the FastPro8 software (Oxborough, CTG Ltd, V. 1.2.0, 2017). Fluorescence acquisitions were then corrected for background fluorescence obtained by blank measurements from filtered seawater (on 0.2  $\mu$ m or GF/F membranes according to the cruises) collected in surface waters excepted for the SMHI cruise. For this last sampling in stratified water column, seawater was collected at two or three depths (surface water, stratification layer and sometimes under the stratification). Fluorescence acquisitions were then fitted to the biophysical model of Kolber et al. (1998) to obtain minimum fluorescence ( $F'_o$  or  $F'$  for the dark and light sensor respectively, all measurements made during the daylight period), maximum fluorescence ( $F'_m$ , one by sensor and acquisition) and effective absorption cross section of PSII ( $\sigma_{PSII-\lambda}$  and  $\sigma'_{PSII-\lambda}$ ) for each wavelength protocols used. Indeed, three or four different measurement protocols were implemented by sensor programming, using one (450 nm) and most of the time two different mix (450 + 530 nm and 450 + 624 nm according to the Chelsea recommendations) of the wavelengths available on the last FRRf production model (cf 3.3.1). During the SMHI-Aranda cruise in the Baltic Sea, dominated by cyanobacteria (order Nostocales), a protocol with a mix of the three wavelengths available was also tested (450 + 530 + 624 nm).

Using these corrected fluorescence acquisitions, the standard set of physiological parameters was calculated with FastPro8 incorporating the data of the dark chamber in the ambient light data file.

The light curves are to be analysed with the FRRf-package version 1.0 developed by Soetaert, Kromkamp and Silsbe and based on the paper by Silsbe and Kromkamp (2012), resulting in a set of fluorescence parameters (Table 3.3.2).

Table 3.3.2. The analysis of variable fluorescence data will deliver following important fluorescence parameters (not all parameters are listed here, only those important for data-analysis).

parameter	Meaning	unit
$F_o$ , $F$	Minimal or steady state fluorescence	dimensionless
$F_m$ , $F_m'$	Maximum fluorescence in the dark resp. light	dimensionless
$\sigma_{PSII}$	Functional absorption cross section	$\text{nm}^2$
$F_v/F_m$ , $F_q'/F_m'$ , $F_v'/F_m'$	Maximum (dark acclimated), effective quantum efficiency of PSII under actinic light and after a transient dark period	dimensionless (interpreted as electron photon <sup>-1</sup> )
Chl-a	Concentration of chlorophyll-a, based on calibration factor for $F_o$	$\text{mg m}^{-3}$
$E$	Irradiance used in the light curve	$\mu\text{mol photons m}^{-2} \text{s}^{-1}$
$\tau$	Rate constant for $Q_a$ reoxidation	msec
$p$	Connectivity between antenna complexes	dimensionless
$JV_{PSII}$	PSII flux (electron transport) per unit volume	$\text{Mol electrons m}^{-3} \text{d}^{-1}$
$a_{LHII}$	Volumetric PSII absorption coefficient	$\text{m}^{-1}$
[RCII]	Concentration of reaction centre II	$\text{nmol m}^{-3}$
$NPQ_{NSV}$	Non-photochemical quenching (Normalised Stern-Volmer)	dimensionless

Although the FRRf data files give all these parameters, they should not be used as the data has to be corrected for blank values first. In the D3.1 report a specific section was dedicated to how to calculate other important physiological parameters such as  $a_{LHII}$ , [RCII] and  $n_{PSII}$ . Calculations are here reported in an Appendix (2.)





The current mode of the FastOcean profiler deployment in joint sampling cruises was relatively complex and time consuming. The main difficulty for sampling is to keep the boat facing the sun in the sea current, especially with adverse winds. According to the water column depth and the turbulence level of waters, a typical profile took between 15 and 25 min to be performed. Taking into consideration the data variability due to wave turbulence and the strong vertical gradient of light in the upper part of the water column, a depth-by-depth profiling protocol was chosen in order to optimize data quality but also data number with an adequate  $R_{\sigma_{PSII}}$  factor of measurement fits around 0.045/0.050. To obtain such values of R factors, we must use the auto-gain function of the sensor photomultipliers (PMT of the two sensors) with a start value of 400 volts. In this case, the disadvantage is that the PMT optimisation time for a satisfactory measurement is long: between two and three minutes. So the instrument must stay in the surface-layer during this time before profiling. According to the turbulence level and the solar light variations due to clouds covering, the duration of physiological variable measurements at a given depth range between 25 and 50 sec; the number of depths or measurement levels depending on the water column depth.

On all cruises, the FastOcean profiler was not deployed on a “rosette” or CTD but separately, on an independent winch cable on the side of the ship facing the sun. Most of the time, the used winch allowed to deploy the FRRF away from the ship’s shadow. The main problem with such strategy was to keep the ship facing the sun during the profile. With high current and/or wind speed, the ship can easily turn and the instrument can be on the shadow side of the ship, or under the ship. If the ship used its lateral booster to keep the position facing the sun, that can generate high turbulence level in surface-layer and change the vertical gradient of phytoplankton and light. So, it is necessary to come to a compromise with all these problems of position and timing to realize a satisfactory FRRF profile.

On the VLIZ and RWS/NIOZ North Sea cruises the FastOcean/Act2 recorded approx. 770 fluorescent light curves. In addition, regular discrete measurements were made on the samples taken from a Niskin bottle just below the surface and from just above the bottom. On each of these cruises 3-6  $^{13}\text{C}$ -uptake experiments were performed daily in order to measure the electron requirement for C-fixation. Most of these results are still being analysed.

### ***Results displaying the different approaches implemented***

#### ***Spatial studies.***

**In the eastern English Channel.** Figure 3.3.18. below is an example of horizontal (sub-surface) measurements of variable fluorescence across the different water bodies sampled in spring 2017 during the PhycoManche cruise. Here we report the maximum PSII photochemical efficiency after a transient dark period ( $\pm 1$  mn according to the pipe length and the sample pump speed),  $F'v/F'm$  (see Suggett et al 2011), but the functional absorption cross section  $\sigma_{PSII}$  was also measured in the same time, in a continuous way.

This sampling procedure is not common. Usually photosynthesis-energy relationship (PE or FLCs curves) are directly measured continuously by the FRRF-Act2 technique. However, it is clear that this alternative approach allows to have phytoplankton functional traits at high frequency (10 sec here), in a consistent way with sampled water masses.



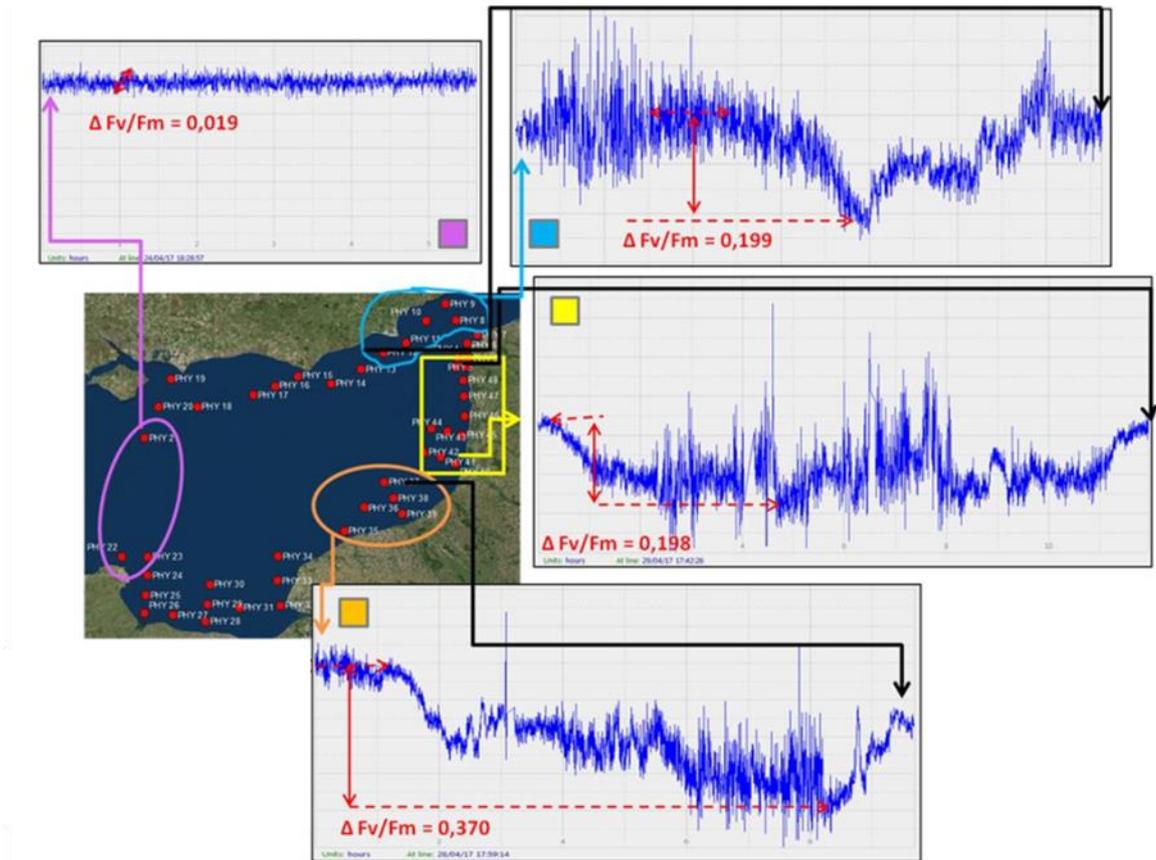


Figure 3.3.18 Spatial variation of effective quantum efficiency of PSII ( $F_v'/F_m'$ ) after a transient dark period during the PhycManche cruise in 2017. Range of variations are calculated for different characteristic water masses.

It is then possible to characterize the maximum range of physiological parameter variations ( $\Delta F_v'/F_m'$  on fig. 3.3.18) without aliasing risk. Maximum range of variations (0.37) are here observed along an inshore-offshore transect in the water body going up to the north under the effect of the residual tidal currents, after the bay of Seine (area under orange ellipse on the Figure 3.3.18).

Figure 3.3.19 below is a typical example of a vertical FRRf profile (APD system added with a TRIOS spectroradiometer) obtained **in the strait of Dover (between English Channel and south bight of the North Sea)** in spring 2017 during **the VLIZ cruise**.

Mean values of each classical **FRRf profiler** parameters are approximated for surface and euphotic depth (blue dashed line on the figure) and reported on each panel of the Figure 3.3.19. Vertical variations of each parameter are here characterized with the Coefficient of Variation (CV) in percent. Our observed variations across the vertical gradient of light in an hydrologically homogeneous water column are very interesting and probably refer first to photoregulation (CV=44% for effective quantum efficiency against 13% for the functional absorption  $\sigma_{PSII}$ ).

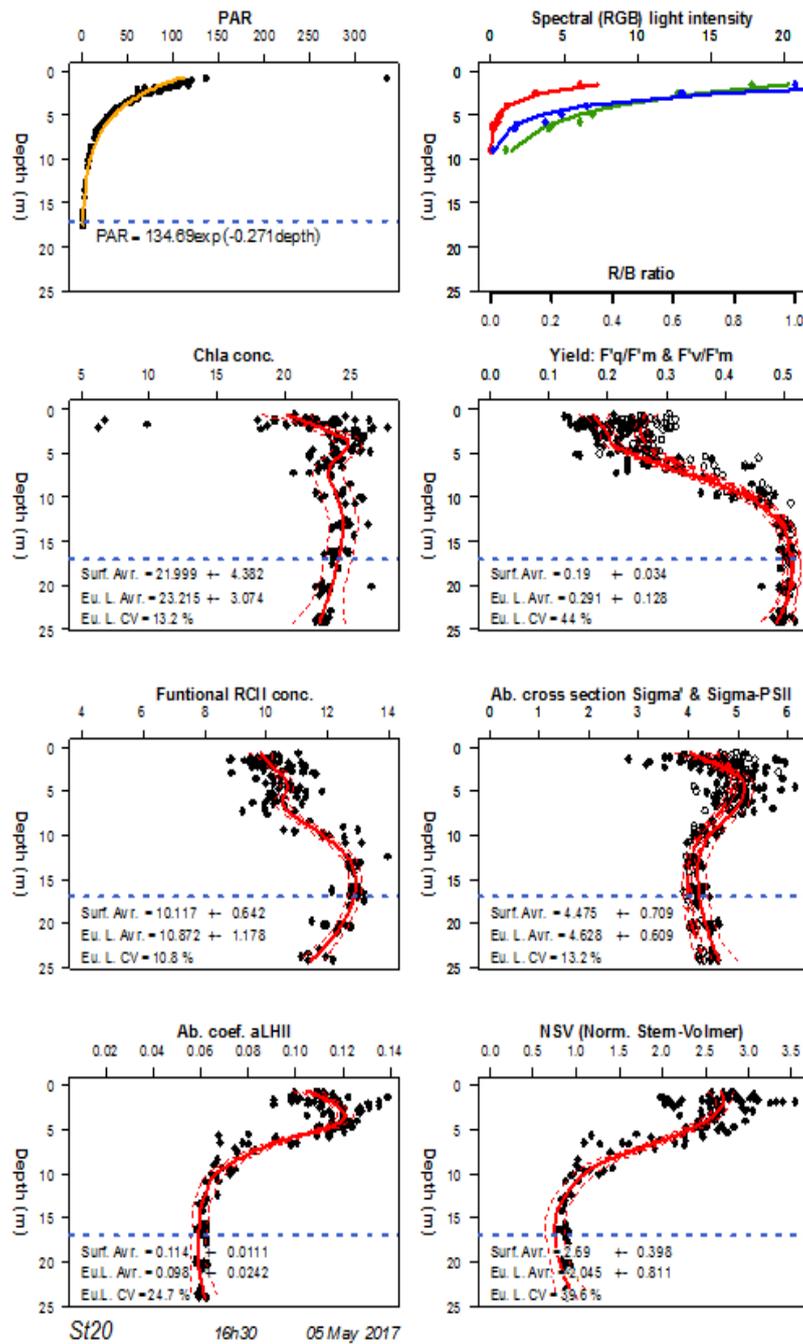


Figure 3.3.19 Physiological parameters obtained from a vertical profile of FRRf APD system added with a spectroradiometer TRIOS (blue, green and red light are integrated on their characteristic spectral ranges); ex. at station n°20 of the 2017 Vliz cruise. Parameter units are reported in table 3.3.2. All data are fitted with a Loess model (red line) ; NB : the ambient light sensor data are the white circles for  $\sigma'_{PSII}$  and effective quantum efficiency of PSII ( $F_q/F_m$ ). The blue dashed line is the depth of the euphotic layer (Eu.L.). Mean values for surface and euphotic depth are reported on each panel, and coefficient of variation (CV) for the euphotic zone.

For this measuring station, FRRf physiological parameters vary mainly in the first ten meters of the water column. The layer thickness of physiological parameter variations is thus lower than the euphotic zone thickness, since the euphotic depth calculated here is 17 m. This is not a general case for our all sampling stations but it is very interesting because 10m depth is here (with a  $K_{dPAR} = 0.252 \text{ m}^{-1}$ ) the so called “mid-point” of the euphotic zone



(Kirk, 2011). In the present case, the partitioning of the physiological processes in the water column could be explained by a strong photoregulation (and perhaps more see below) in the upper water column, and in the lower part by the vertical mixing generated by tidal currents occurring from the bottom of the water column (28 m depth). It should be noted that the range of  $F_v'/F_m'$  (as for  $F_q'/F_m'$ ) through the water column are here in the same order of magnitude than horizontal variations (see Figure 3.3.18.) of this parameter, the maximum quantum efficiency of PSII measured during the daylight period after a transient dark period (2 sec with the Chelsea APD system equipped of a dark chamber).

On the bottom-right side of the Figure 3.3.19, we can see that non-photochemical quenching with the NSV parameter (non-photochemical quenching Normalised Stern-Volmer index) is well developed for phytoplankton in the surface water, even though light intensity is not very high in the present case. This is an additional argument that supports our previous paragraph on the occurrence of a strong photoregulation in the upper water column. It is important to notice that under high sunlight, it is often difficult to generate a fluorescence rise and a good iterative curve fit (ST measurements). Measured signals can be very noisy in clear waters and when light intensity increase during the day. Data may also be lost. Suggett et al (2011) report about this stating that FRRF are known to be subject to the "red light effect" in the near surface where high red light intensities can saturate the photomultiplier tube (PMT). Between the surface and 5 m deep, sometimes 10 m deep (it depends on the clarity of the water and on the light intensity of the sun), measurements are very noisy as it appears on the Figure 3.3.19. Even if the PMT is optimised in real time so that the ADC is within range, it may still be difficult to get a strong  $F_v$  signal to curve fit. To improve this, it is necessary to increase the sequence repetition, and the time of measurement spent in surface waters: the more data that can be averaged the more likely those gaps in data will be filled. However, during this time the incident light can change a lot. The second measurement difficulty is therefore to find the right compromise in the measurement time for each depth in the most enlightened part of the water column.

In this example, vertical Chla profile from FRRf (Figure 3.3.19) shows small but not clear vertical variations (noisy signal in surface) and not significant through the entire water column. In contrast, parameters related to the use of light by phytoplankton,  $a_{LHII}$  and  $\sigma_{PSII}$ , and also the number of RCII display clearer variations through the water column with two stages of variations in the first part of the water column (between 1 and 10m depth). For RCII number, we can observe through the loess fit (red line) two small steps of decline with depth and between the two, a break around 5m. For  $a_{LHII}$  and  $\sigma_{PSII}$ , we observe more clearly a first step of increase and a second step of decrease after the 5m depth. Vertical profiles of FRRf data have been reported for different ecosystems since the development of this technique (Moore et al 2003, Smyth et al 2004, Aiken et al 2004, Suggett et al 2006, Erga et al 2012, Neale et al 2012, Cheah et al 2017) using the first FRRf generation. These first works were mainly conducted across Open Ocean or frontal area with very deep waters. Here we present FRRf vertical profiles for very coastal and turbid waters, with data from 1.5 m depth, displaying complex vertical shape considering that sampled water columns are theoretically well mixed by tide.

Before talking about photoacclimation of phytoplankton in the vertical gradient of light, it is important to consider that the water column structure may also affect the shape of these physiological parameters according to depth. The rate of vertical mixing can control the light response of physiological processes, as well as the vertical gradients in community structure. As Suggett et al (2011), we can say that an homogeneous community with depth will not necessarily produce a homogeneous vertical profile because this community is composed of multiple taxa with potentially different responses to the light gradient. In the present case, measurements of functional phytoplankton groups with the bbe Fluoroprobe do not show vertical variation in phytoplankton community structure (dominated by Phaeocystis) at this sampling station. So vertical variations of our physiological parameters could be related to photoregulation and to some transient photoacclimation processes occurring separately in the upper (from the surface to the "mid-point" of the euphotic zone, located at 10m depth) and in the lower part of the water column under the vertical tidal mixing control. These processes and/or physiological parameter variations could also be linked, to a lesser extent, to the small variations of phytoplankton biomass distributions in the upper water column.

Work is still in progress to categorize and synthesize the measured FRRf profiles through the 8 cruises carried out in 2017 and after that, considering more particularly physiological parameters versus optical depths and accessory phytoplankton pigments where data are available. One of the objectives is to show the added value of a



measurement approach across the entire water column compared to a surface approach in water quality monitoring programs.

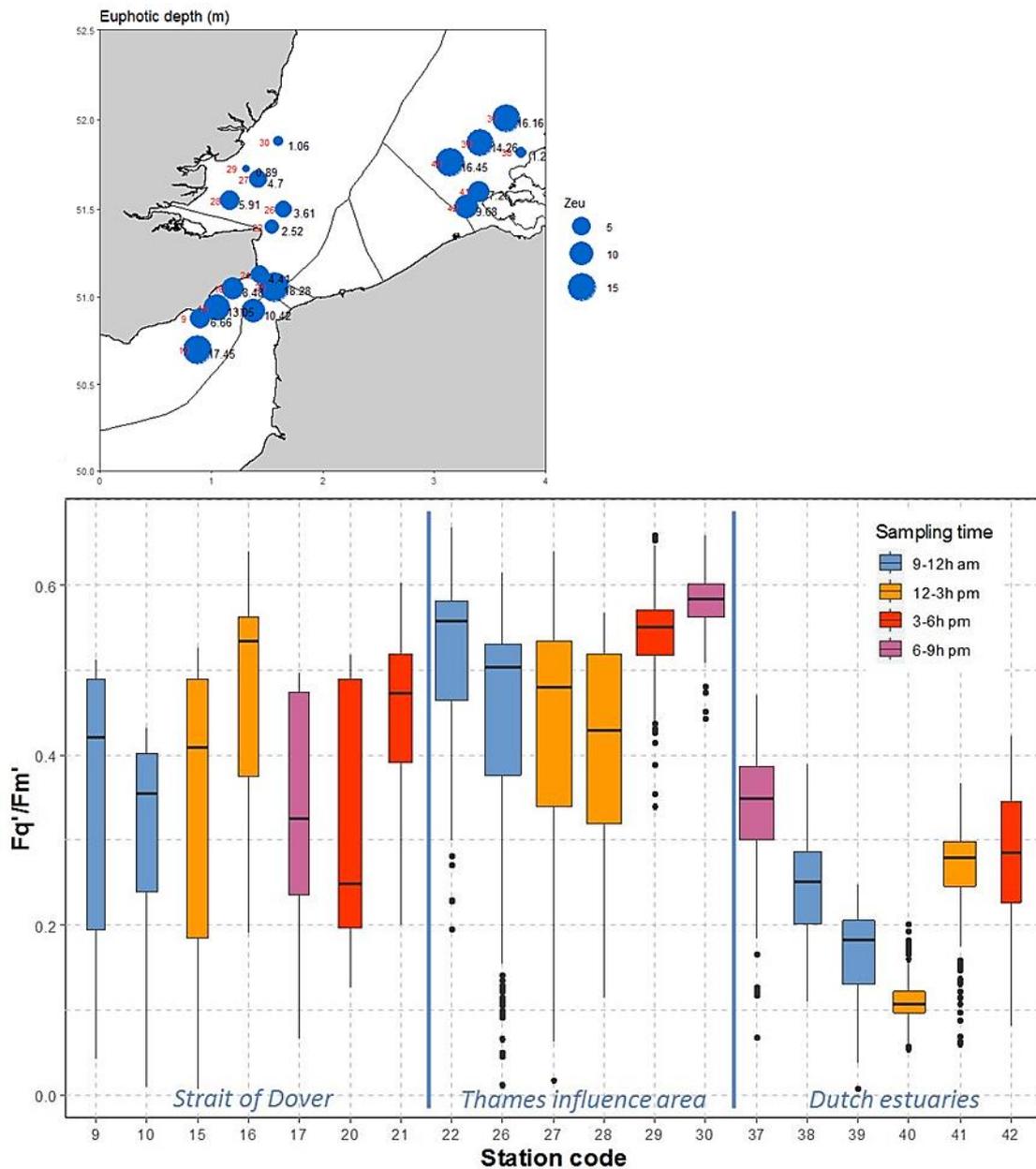


Figure 3.3.20. Map of the FRRf profiler sampling stations of the VLIZ 2017 JericoNext cruise (station code are in red and euphotic zone depth in black, and blue dots) and boxplots of effective quantum efficiency for all data measured in the water column by the FRRf APD system at 19 sampling stations, during the 2017 Vliz cruise

The Figure 3.3.20 shows that the physiological data obtained from the profiler APD system in the water column are also well suited to globally discriminate against water bodies sampled during a large-scale spatial campaign (here the VLIZ cruise in spring 2017 in the south bight of the North Sea) even if it is well known that physiological parameters display daily cycles (see further in this section).

Surprisingly, the box-plots of vertically measured data for the effective quantum efficiency ( $F_q'/F_m'$ ) display clear differences in medians and in distributions (interquartile ranges) between the waters of the Strait of Dover, the Thames influence area and the Dutch estuaries. Mean values are also significantly different between the three areas.

In addition, there is no relationship between these data distributions (box-plots) and the sampling time (cf. boxplot colour code) which is here considered as an indicator of the duration of illumination experienced by the cells. This duration could theoretically control phytoplankton physiology at daily scale and so the extent of the vertical distribution of the physiological parameters. This is probably due to the fact that these distributions incorporate all measurements from the surface to the bottom of the water column, i.e. under the euphotic layer.

**During the VLIZ 2017 cruise in the eastern English Channel and southern North Sea (R.V. (S. Stevin), rapid light-curves (FLCs)** were also measured in a continuous way on board, with the FRRF-Act2 system and some preliminary results were already presented in D3.1 (pages 61-62).. A further analysis on more data should tell us if their differences are truly significant or not. If so, it is most likely due to the fact that Niskin samples have undergone a longer dark acclimation than the flow-through samples, which causes differences in the relaxation (“disappearance”) of non-photochemical quenching (NPQ) processes.

As can be seen in Fig. 3.3.21, we encountered very different conditions as signified by the varying phytoplankton concentrations, using the minimal fluorescence ( $F_o$ ) as a proxy. In contrast to the discrete measurement taken from samples from Niskin bottles, these data were obtained from the seawater supply of the ship which was fed into the FRRF which was running automatically. Because of this the samples were not fully dark acclimated before the measurements. Passage times from the inlet of the water to the fluorometer were not known but most likely in the order of 4-5 min. Before the measurements started the samples were pre-acclimated for 105 seconds to a low irradiance ( $12 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ ) in order to relax NPQ as much as possible as low light works better and faster than dark acclimation (Goss et al. 2006, Goss & Lepetit 2015).

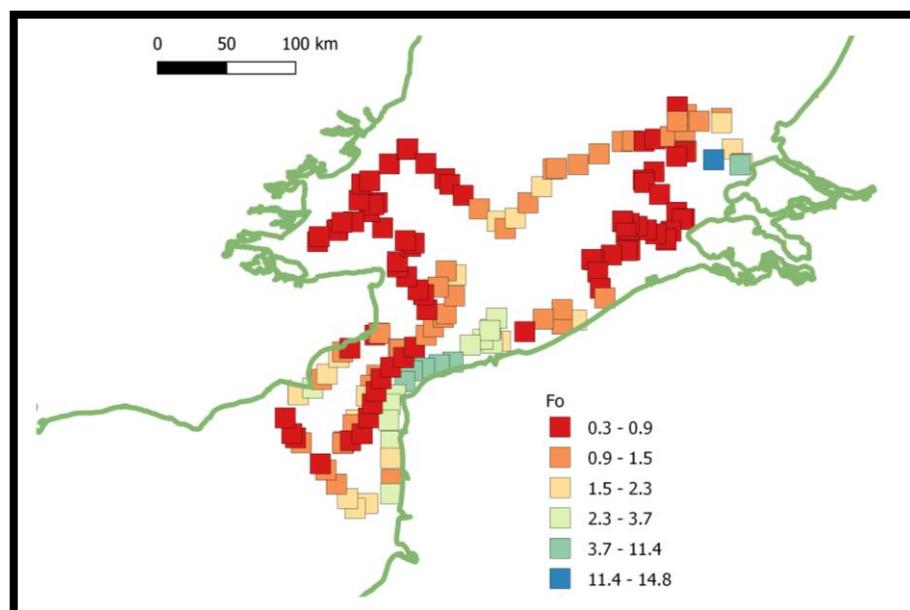


Fig. 3.3.21. Spatial distribution of phytoplankton biomass ( $F_o$  as proxy, data FRRF) of surface water phytoplankton.

Low phytoplankton biomass observed near the southern Dutch and UK coast and in the middle of the Channel (Fig. 3.3.21). Highest biomass was observed near the French coast north of Calais. The pattern obtained from the FRRF (Fig. 3.3.21) and CytoSense FCM (Fig. 3.3.22) are very similar showing that both instruments are useful tools to obtain high resolution maps of phytoplankton biomass.

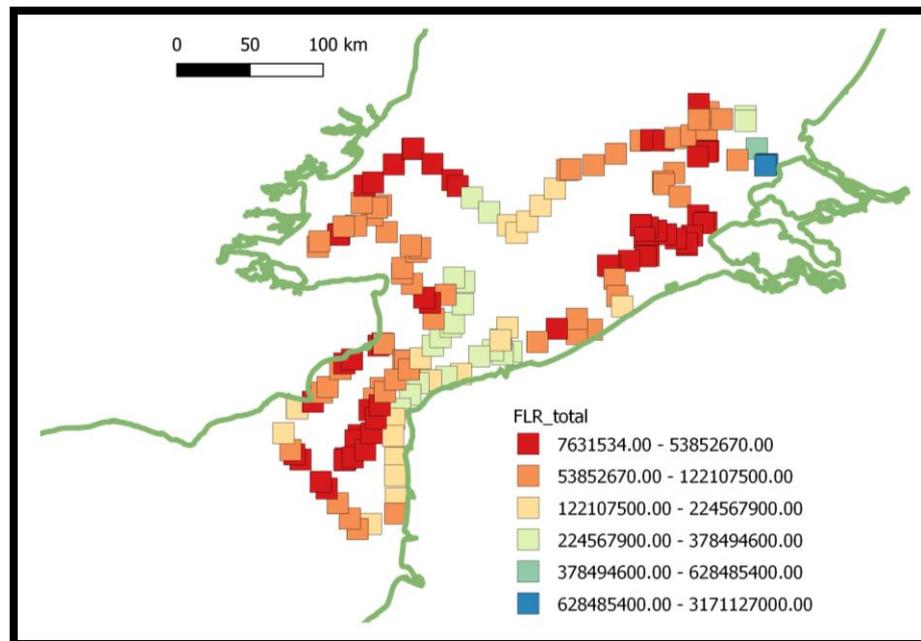


Fig. 3.3.22. Spatial distribution of phytoplankton biomass using the total red fluorescence obtained from the CytoSense FlowCytometer.

Primary production results (surface waters) are shown in Fig. 3.3.23. Highest production rates were observed in a small area near the Dutch coast with  $P_{\max}$  values between 1.7-3.7  $\mu\text{mol electrons m}^{-3}\text{h}^{-1}$ . The lowest values were found at several locations, but were mainly observed during the night (see times in middle panel Fig. 8). The only exceptions were the afternoon values off the mouth of the Westerschelde estuary.

The lowest  $E_k$  values were also observed at night or late evening (again with the exception of the some locations offshore the mouth of the Westerschelde area). In general values near the UK coast east of the Thames estuary were highest. Photosynthetic efficiencies (i.e.  $\alpha$ -values, the initial slope values of the light curves) followed the same pattern as  $P_{\max}$ , with the exception of the values near the UK coast east of the Thames estuary.

These results demonstrate that diel patterns in photosynthetic activity can mask changes occurring due to different phytoplankton populations. A clear example is detected in the middle of the eastern Channel: we sailed afternoon southward, and in the late evening and night northward following nearly the same track, and the night time values were lower for all the LC parameters shown in Fig. 3.3.23. In order to further study this effect, a necessity to quantify the effect on estimates of primary production), Langragian experiments should be performed which might allow us to model the effect of diurnal cycles on estimates of primary productivity.

Therefore, it is possible to affirm that continuous measurements of photosynthesis with the FastOcean/Act2 combination are capable of capturing spatial and temporal variability in photosynthetic activity and phytoplankton biomass.

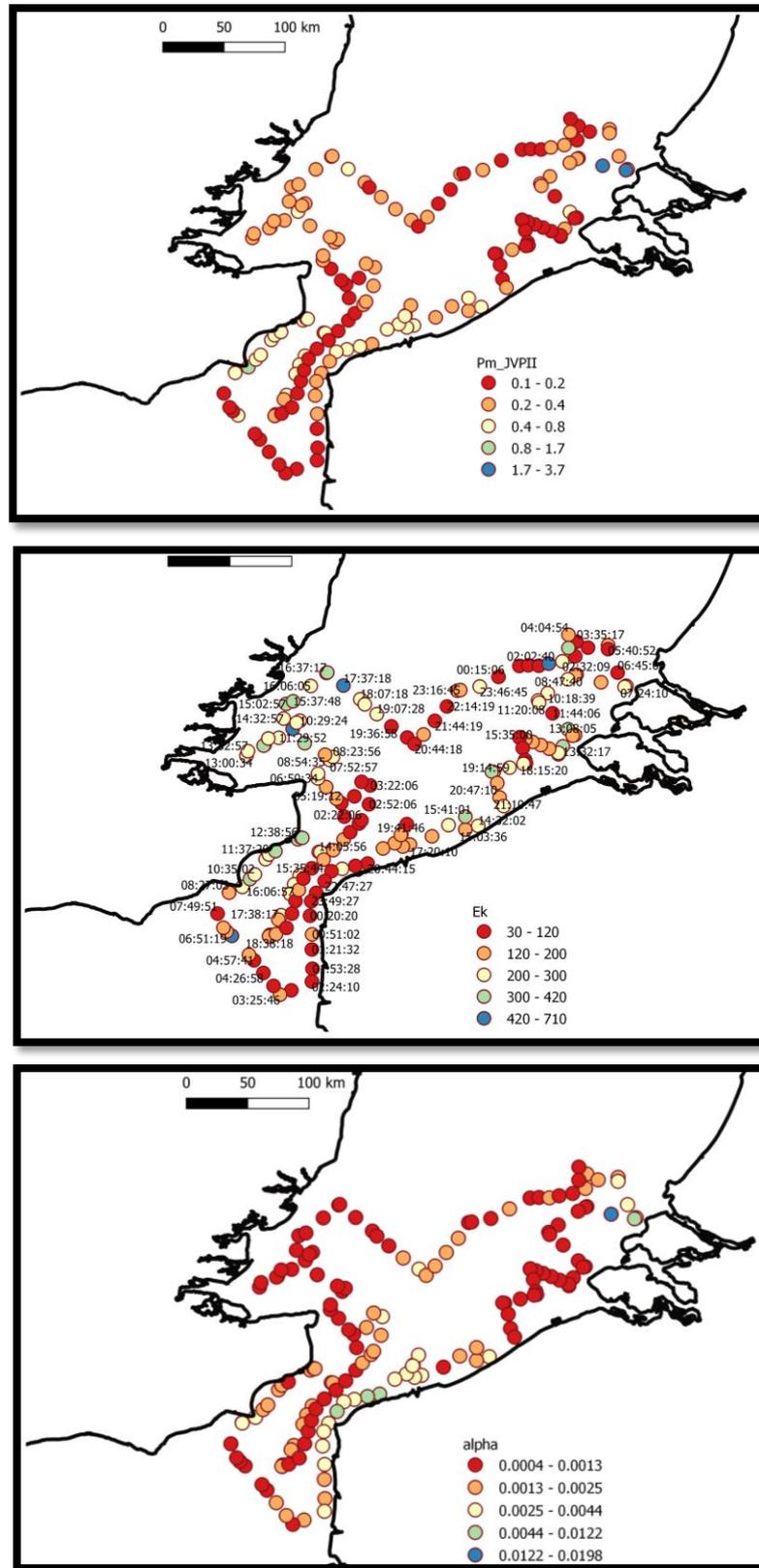


Fig. 3.3.23. Results from fitting the absolute rate of volumetric electron transport ( $JV_{PII}$ ). Top: max rate of ETR, middle:  $E_k$  and bottom: initial slope (i.e. photosynthetic affinity).

During the cruises on the Channel and on the Southern North Sea we obtained evidence that the photosynthetic activity showed a clear diurnal cycle. Daily (24 h) continuous recording of fluorescence parameters measured in the first step of the light curve (dark) were carried out during both the VLIZ cruise and during the cruises in the Dutch waters of the North Sea, and some results are shown on D3.1 (page 62).

For the North Sea cruises we looked further into this by calculating z-scores of a number of photophysiological parameters obtained from the FRRF. Basically the z-score looks at the difference of a value and the daily mean of that value and normalises it to the standard deviation:

$$Z - score = \frac{x - \text{daily mean}(x_0 \dots x_{24})}{\text{Daily standard deviation}(x_0 \dots x_{24})}$$

Fig. 3.3.24 shows the result of the z-scores as a function of time of the day for  $F_v/F_m$ ,  $E_k$ ,  $\sigma_{PSII}$  and  $1/\tau$ .

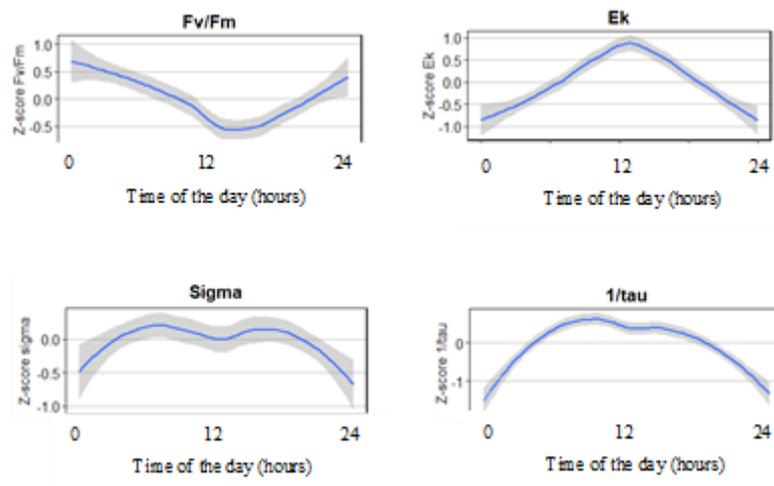


Fig. 3.3.24 Standardized z-scores (i.e. daily anomalies) of  $F_v/F_m$ ,  $E_k$ ,  $\sigma_{PSII}$  and  $1/\tau$  to show the diurnal trends in photosynthetic parameters. On the x-axis the time of the day and on the y-axis the z-score.

The data show a clear diurnal trend in all the parameters plotted. This suggests that diurnal patterns can mask changes occurring when entering new phytoplankton populations (although this will be picked up by the CytoSense and needs to be checked). Whether the changes are driven by endogenous clocks (a circadian oscillator, Cohen & Golden 2015) or by photoacclimation to varying light regimes, or both, is not clear at the moment, but likely they will both play a role and the photoacclimation will most likely be more pronounced on clear and sunny days than on cloudy days.

The data have to be converted to absolute rates of PSII electron transport in order to calculate rates of primary production. This also requires a comparison with the C-fixation rates, measured by the incorporation of  $^{13}C$ -labeled bicarbonate. The Figure 3.3.26 shows the maximum quantum efficiency of PSII measured during the 4 cruises on the North Sea in 2017. All the data were gathered during the 4 days' cruises on the Dutch EEZ by combining the JERICO-NEXT -cruises with the standard monitoring (MWTL) cruises of Rijkswaterstaat with the Zirfaea R.V.. Lowest  $F_v/F_m$  values were obtained during the May cruise, and are likely caused by the low phosphate concentrations measured during this period, hence in this case the decrease in  $F_v/F_m$  is an indicator of P-limitation. As not all stations showed low phosphate concentrations, these stations are likely having a higher  $F_v/F_m$ , hence the spread in the data.



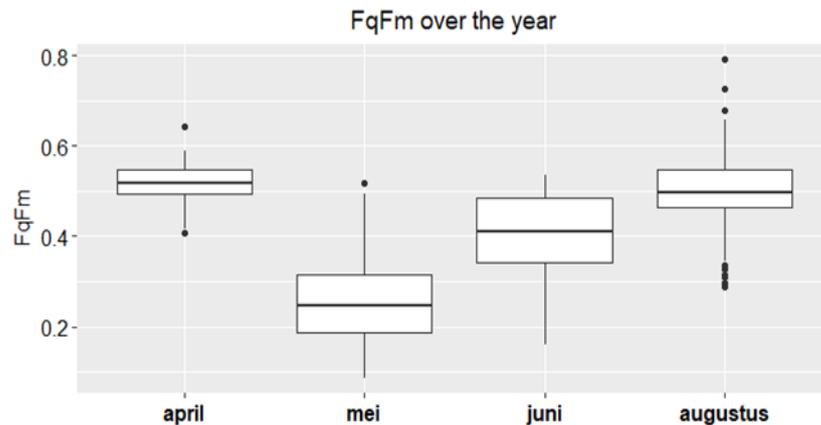


Figure 3.3.26. Box-wisker plots of maximum effective quantum efficiency of PSII measured in the dark step during the continuous LC-measurements) in the North Sea during cruises in April (april), May (mei), June (juni) and August (augustus).

The preliminary results show the power of continuous FRRF measurements. Spatial variability and biomass, photosynthetic parameters and physiological conditions are detected giving valuable information about the state of the phytoplankton populations. If the electron requirement for C-fixation is measured, the ETR rates can be converted to C-fixation rates. Calculation of primary production is then possible if the light attenuation coefficient is also known. These can in principle be obtained from the turbidity measurements of the ferrybox, or from remote sensing, although in the latter case good validation is required.

A point of concern is the blank correction. In filtered water, some compounds contribute to the fluorescence signal, and the FRRF-data ( $F_0$  and  $F_m$  and all data based on these) have to be corrected for this. This was done by measuring filtered samples. Although spatial variability in blank values seems limited, careful blanking the data is required if the phytoplankton biomass is low. We have obtained situation where the blank values >50% of  $F_0$ . This point need further attention. A possible recommendation is to investigate if the blank value corresponds to CDOM measurements, and CDOM can be obtained from automated sensors as well.

In general, our  $F_v/F_m$  measurements do not seem to exceed a value of 0.55 to 0.6, which corresponds to most literature values on nutrient replete waters. If after blank correction  $F_v/F_m$  exceeds these values, it is likely that the blank correction was too strong and that blank values were overestimated. It is suggested to correct the blank values in these cases so that  $F_v/F_m$  does not exceed 0.6.

A comparison between FRRF parameters and flow cytometry is still in process as it is expected that phytoplankton composition would influence some of the FRRF parameters. Extended results are shown in D4.4 (North Sea-English Channel section) and D4.5 (synthesis of JRAP#1 ) and were published by Aardema et al. (2018).

### **Time series.**

In the bay of Seine (English Channel - France), an instrumented buoy called SMILE was installed in June 2016. This project is conducted by the CNRS-BOREA (University of Caen-CREC) and IFREMER. Beside traditional high-frequency measurements performed in oceanography, a Fast Rate Fluorometer Act2 Chelsea Instrument which allows to estimate primary production of phytoplankton was installed on the buoy. Several technical issues had to be solved to allow such type of measurements from a buoy. The energy is supplied by wave energy and sun with a hydrogen battery as a backup. This system was developed by GEPS techno. In order to optimize production of wave energy, the SMILE buoy is anchored by using an intermediated smaller buoy which allow more efficient movement for energy production. The FRRf was implanted in a watertight compartment on the buoy. The FRRf is controlled by a PC using a software developed by NKE instrument which control the energy supplied of the FRRf and send data after each measurement by using GPRS transmission. The measuring chamber of the FRRf is automatically cleaned by HCl at low concentration every 24h. During the last six months we have not observed any biofouling within the measuring chamber but in July 2017 biofouling appeared in the water jacket of the measuring chamber. The distilled water of the water jacket was contaminated by microalgae, since then, the water is acidified. Data has been produced using the FRRf on SMILE buoy since the 3rd March 2017 (Figure 3.3.27).



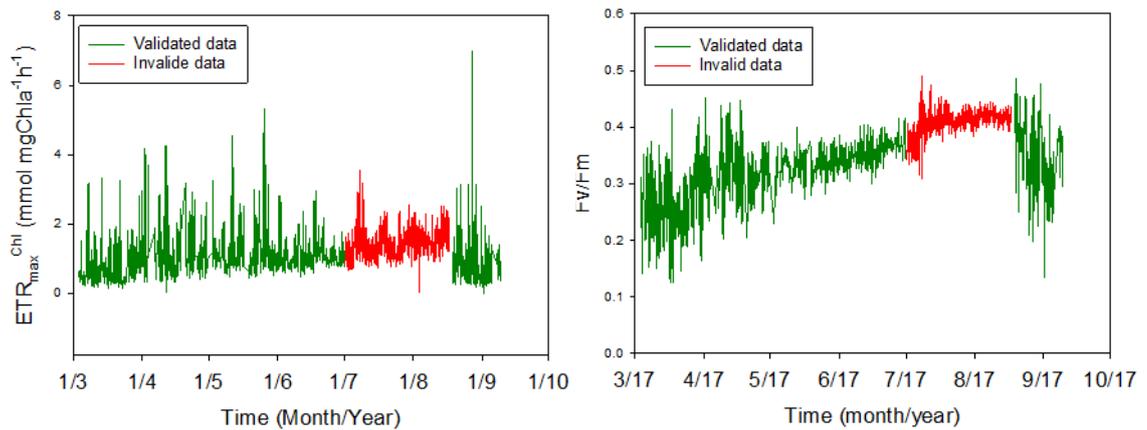


Figure 3.3.27. Left: Fv/Fm – Maximum quantum yield of the PSII measured on natural assemblage of phytoplankton between March and September 2017 from the SMILE buoy. The data obtained from 1 July to 15 August 2017 were not validated because of biofouling development in the water jacket which regulates temperature. We observed an increase of the Fv/Fm ratio during spring time. More data are required to deal with the annual dynamics of this parameter over the year. Right: ETRmaxChl (Electron transport rate from the PSII per Chl-a unit) corresponding to the photosynthetic capacity of the natural assemblage of phytoplankton between March and September 2017 measured from SMILE buoy.

In order to observe the photosynthetic parameters dynamics as a function of daily PAR variations we present a focus on a shorter period (3 weeks) (Figure 3.3.28).

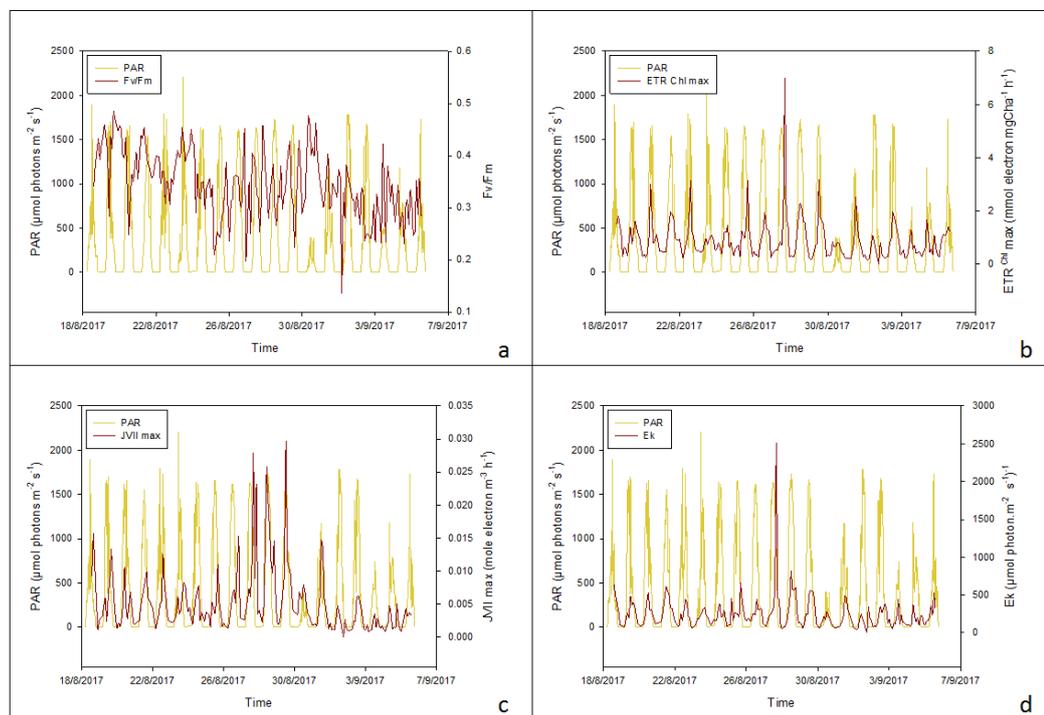


Figure 3.3.28. Examples of parameters estimated on the natural assemblage of phytoplankton by using FRRf measurements from the SMILE buoy. Fv/ Fm is Maximum quantum yield of the PSII, ETRmaxChl corresponds to the photosynthetic capacity, JvII corresponds to the maximal potential primary production, Ek corresponds to the light saturation parameter which gives to the phytoplankton photoacclimation capacities.





Both ETR<sub>maxChl</sub> and J<sub>VII</sub>max are highly variable and the variations are largely related to the PAR dynamics. By using all these photosynthetic parameters, we will be able to estimate the primary production on this site and to explore the dynamics of the primary production over the year in contrasted situations.

**In the Utö fixed station in the Baltic**, a FastOcean connected to FastAct or Act2 has been used in 2015-16 to study short term variability in the conversion factors from electron transport rate to C-fixation. SYKE performed three 3-day measuring campaigns with 4 hour intervals between samplings in 2015-16 and has carried out so far five 7-day campaigns with three daily sampling events in 2016-17 (Figure 3.3.29). FastOcean has been used continuously at Utö since April 2017.

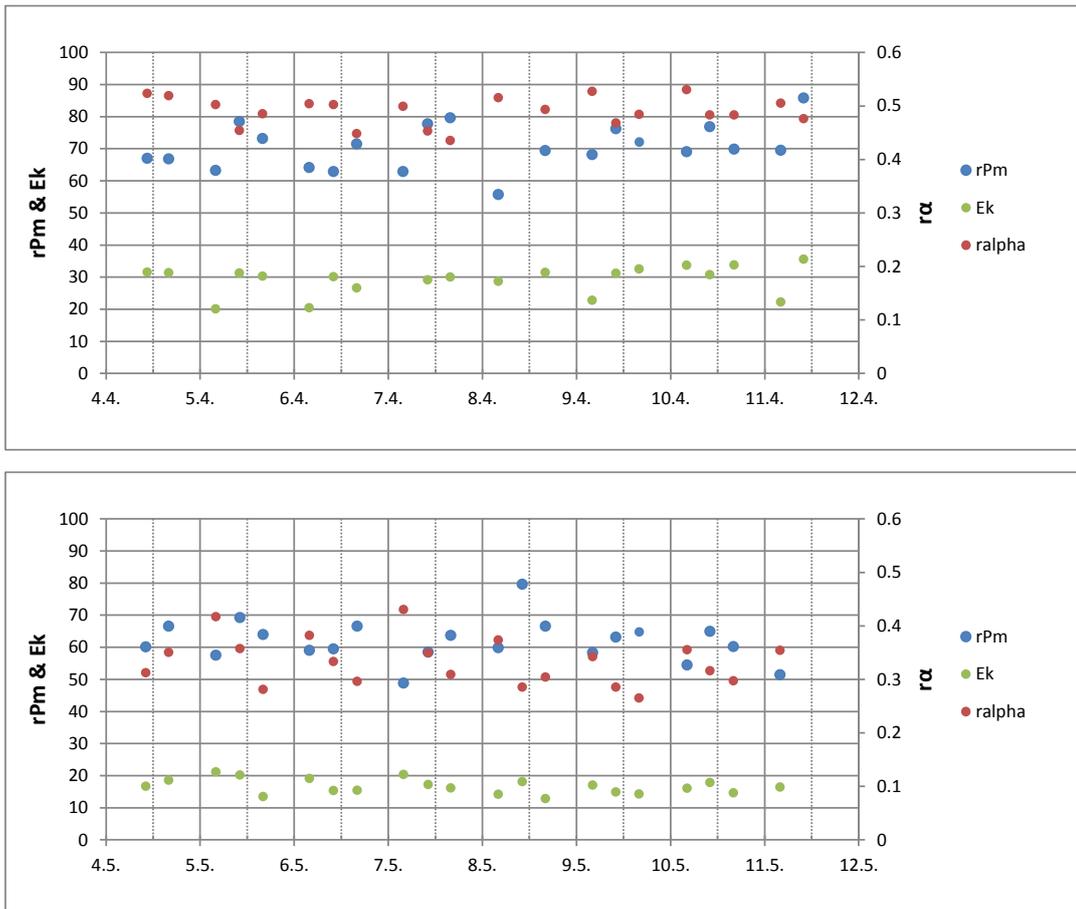


Figure 3.3.29 Example of the variability of photosynthetic parameters, maximum light utilization coefficient (ralpha), light saturation (Ek) and maximum production capacity (rPm), measured at Utö for one week periods in April and May 2017. For these periods, additional measurements of <sup>14</sup>C fixation exist and will be analysed later.



**Appendix of FRRf parameter calculations**

Calculation of the optical cross section of PSII using the following equation ( $K_R=11800$ , see manual FastPro8, Oxborough et al 2012 and Silsbe et al 2015)

$$\alpha_{LHII}(m^{-1}) = \frac{F_m \times F_o}{F_m - F_o} \times \frac{K_R}{1 \times 10^6}$$

If the chlorophyll concentration in  $mg/m^3$  is known the optical cross section of PSII (the absorption coefficient of PSII) can be obtained by dividing the above value by the Chl-a concentration ( $m^{-1}/(mg \text{ Chl-a } m^{-3}) = m^2/mg \text{ Chl-a}$ ).

Second, you can calculate the concentration of reaction centers II:

$$[RCII](mol \ m^{-3}) = \frac{F_o}{\sigma_{PSII} \times 10^{-18}} \times \frac{K_R}{10^6} \times \frac{1}{6.02 \times 10^{23}}$$

from this you can calculate the number of PSII per mg chlorophyll:

$$n_{PSII} (mol \ RCII \ mol^{-1} \ cha) = \frac{[RCII]}{[chla]/893.5/1000}$$

Here 893.5 is the molecular weight of Chl-a, and 1000 is a conversion from mg to gram.

With these equation, absolute photosynthetic electron transport can be calculated as follows using the absorption algorithm (Oxborough et al, 2012):

$$ETR_V = J_{V_{PSII}} (\mu mol \ electrons \ L^{-1} \ h^{-1}) = \frac{F_m \times F_o}{F_m - F_o} \times \frac{\Delta F}{F_m'} \times \frac{K_R}{E_{LED}} \times E_{PAR}$$

With knowledge of the optical absorption cross section ( $\alpha_{LHII}$ ) the sigma algorithm can be used to calculate absolute rates of ETR

$$ETR (\mu mol \ electrons \ (mg \ Chl - a)^{-1} \ h^{-1}) = \frac{\Delta F}{F_m'} \times E \times \alpha_{PSII}$$

The only other parameter needed to convert ETR to C-fixation rates is knowledge of the electron requirement for C-fixation ( $\Phi_{e,c}$ ) (Kromkamp et al. 2008, Lawrenz et al. 2013), and this will be obtained from the comparison of C-fixation measurements (carried out with the  $^{13}CO_2$  (as  $NaH^{13}CO_2$ ) incubation experiments and the FRRF light curves. The data will also be used to improve the predictive equations developed by Lawrenz et al. (2013).

PSII electron flux on a volume basis, can be calculated with FastPro8 according to two different algorithms: the Sigma algorithm (Suggett et al., 2001, 2004 and Moore et al., 2006) and the Absorption algorithm (Oxborough et al., 2012). This parameter allows computing primary production but some comparisons with other approaches or techniques must be made in a near future in the JERICO-NEXT context, before evaluation and comment (cf D4.3).





### 3.3.4. Conclusions and recommendations

**Photo-quenching** is the major issue limiting the use of chlorophyll-*a* fluorescence in estimation of chlorophyll-*a* concentrations. We need to study more in detail possibilities in bringing the light levels and possibly also variable fluorescence into the validation equations. As well fully statistical approaches describing the variability in chlorophyll-*a* fluorescence yield need to be elaborated further. Primary calibration of instrumentation should be agreed, to allow instrument-to-instrument comparison.

**Phycobilin fluorescence** provides high resolution information on the distribution of phycobilin containing organisms. Variability in phycobilin content of the cells, little known variations in phycobilin fluorescence yield and lack of methods to quantify phycobilins needs to be studied further, as they limit the usability of the method.

**Multi-spectral fluorometry** provides proxies for abundance of spectral phytoplankton groups, but comes with methodological limitations. Spectral libraries of group specific shapes and chlorophyll-*a* specific fluorescence yields are known to cause biases in the results. However, some specific fingerprints can reveal the distribution and dynamics of a target phytoplankton group, as the haptophyte fingerprint built during the DYMAPHY project to follow *Phaeocystis globosa* blooms in the eastern Channel and North Sea. Moreover, the PCA analysis approach based on raw data (LEDs) is promising as it makes it possible to get free of any calibration biases and algorithm specificities and to deal with any type of ecosystem and phytoplankton community.

On the other hand, results from two test cruises verify that **hyper-spectral absorption measurements** with the HyAbS provide reliable proxies of phytoplankton biomass, suspended particles in the water column and the phytoplankton species distribution. Using a solid standard calibration procedure instead of a liquid dye allows a fully automation of the measurement. This allows connecting the sensor to FerryBox systems on ships of opportunity without supervision and obtain data in real-time in the future. However, further effort needs to be carried out in terms of software development and the technical set-up, but overall the above shown preliminary results can be considered as promising.

The results of this work clearly show the potential of **in-line automated measurements of primary production** based on FRRF or similar **active fluorescence based measurements**. Nevertheless, there are still some hurdles which needs to be taken. Automated FRRf continuous measurements of photosynthesis with the FastOcean/Act2 combination are capable of capturing spatial and temporal variability in photosynthetic activity and phytoplankton biomass (both on sub-surface as well as profiling the water column). It is strongly recommended to combine FRRf measurements with FerryBox measurements (as well as with other optical bulk or sing-cell approaches as automated FCM). When phytoplankton biomass is low, the **background fluorescence** (i.e. the fluorescence of material passing a 0.22  $\mu\text{m}$  filter water sample) can be a substantial part of the total fluorescence signal. This causes an underestimate of the true photochemical quantum efficiency and an overestimate of the chl *a* concentration and estimates of the optical absorption coefficient of PSII (aLHII), used in the absorption algorithm to estimate volumetric rates of electron transport. Thus carrying out good blanks is a pre-requisite, but difficult to do on automated systems.

In underway measurements, full dark (or better low light) acclimation is not possible because that will reduce the high spatial resolution of the dataset. To reduce the possible artefacts (caused by remaining **Non Photochemical Quenching**) after transport through the ships water inlet) samples can be pre-acclimated to a low irradiance for ~ 2 min. In addition, we employed 90 sec light steps, and results show that during this period *F* and *F<sub>m</sub>'* did not reach steady state values, which might result in erroneous estimates of the photosynthetic quantum efficiency. Hence, some improvements of the FRRF protocol might be necessary. This is one of the topic of the SCOR working group 156 on "Active Chlorophyll fluorescence for autonomous measurements of global marine primary productivity".

Moreover, different studies need still to be carried out for understanding the reasons of variability of conversion factors for converting **volumetric electron rates (ETR<sub>v</sub> or JVP<sub>II</sub>)** estimated from FRRf measurements **into carbon fluxes** based on ecosystem and phytoplankton community specificities. This is the topic of a recent and important review by Hughes et al (2018) who proposed a roadmap for correctly measuring, calculating and interpreting joint FRRf and carbon flux measurements.

Strengthening links between members of the FRRf user's community could generate a high quality-controlled data sets, allow us to better understand and model the variability of the "electron requirement for C-fixation" ( $\Phi_{e,C}$ ) at great spatial and temporal scales as "photosynthetic electron transfert is a fundamental driver of aquatic biogeochemical cycles" (Hughes et al., 2018). Two years ago, an UE report by Kromkamp et al (2017) already





presented the need to estimate primary production with a common measuring approach, through 3 different scenarios including or not national monitoring program, with the implementation of the FRRF technique. It is indeed time to act now in 2020, before the centennial anniversary of Gaader and Gram in 2027.

#### 4. Conclusions and future work

Main conclusions within Task WP3.1 are:

1. The methods used are reliable for automated observation of phytoplankton biodiversity (functional groups, size classes, taxa when possible) and biomass (complementing manual methods for sampling and microscope analyses), as well as photosynthetic physiology and productivity.
2. Operating the equipment and interpreting the results still need a lot of knowledge and time. Even though some operational procedures can be established, the standardization of analytical and data processing as well as data management need more development. The degree of automation varies depending on the method considered. For FRRf measurements, filtered seawater (0.2  $\mu\text{m}$ ) is still required to manually measure background fluorescence and to correct spectral light quality. Moreover, complementary measurements such spectral measurements of light and chl<sub>a</sub> absorption coefficient ( $a^*\text{phy}$ ) in the water column is recommended to correct the use of an artificial light source with a different spectrum from that of the natural environment; more particularly with the Act2 system in order to compute a vertical primary production profile.
3. Imaging in flow and *in situ* imaging provide means for identifying and counting phytoplankton at the genus or species level. Also, biomass based on cell volume of individual cells can be estimated. Development of classifiers for automated identification of organisms is time consuming and requires specific skills on signal analysis and on taxonomy.
4. Flow cytometry has proven to be a useful tool for counting phytoplankton and for describing the phytoplankton community as size based classes and functional groups. There was an agreement to report the phytoplankton count in four groups for inter comparison purposes: *Synechococcus* (pico-cyanobacteria), pico-eukaryotic organisms, nano-eukaryotes and microplankton. Protocols for appropriate ferry box implementation were produced as well as inter comparison results by comparing different machines and analytical methods (manual and automated analytical tools commercially available and/or developed within the frame of the JERICO-NEXT project).
5. Single and multi-wavelength fluorometry makes it possible to estimate phytoplankton biomass (at a chlorophyll-*a* basis) and to differentiate, to some extent, phytoplankton based on photosynthetic pigments. Sunlight induced photoquenching is a problem for estimating chlorophyll-*a* from fluorescence. For instruments mounted buoys or vessels, night time data can be used to minimize the problem.
6. Multi-wavelength absorption is a useful tool for estimating chlorophyll-*a* and is a useful tool for discriminating between phytoplankton groups based on pigment content, by using algorithms based on fingerprints of main species and groups occurring in the system studied, or by working directly on the raw data (LED response to excitation) and characterizing the main fluorescence characteristics of the whole community based on the major contributors (phytoplankton pigmentary groups).
7. Variable (active) fluorescence is available for addressing phytoplankton physiology, photosynthetic parameters and to estimate primary productivity on both continuous sub-surface recording and water column profiles, mediating careful use and coupling with other optical and also biogeochemical analysis. It is now possible to assess *in situ* phytoplankton production in vertical and horizontal dimensions with the same sensor (profiler and Act2 system of Chelsea). Dedicated cruises to measure primary production (i.e. solving the constraints of spatial-temporal sampling of this particular biological process) should still be implemented in an international context, dealing more specifically with conversion factors.



## 5. References

### References section 3.1

Biard T, Stemman L, Picheral M, Mayot N, Vandromme P, Hauss H, Gorsky G, Guidi L, Kiko R, Not F (2016). In situ imaging reveals the biomass of giant protists in the global ocean. *Nature* 532: 504-507.

Breiman LB (2001). Random forests. *Machine Learning* 45: 5-32.

Fernandez-Delgado M, Cernadas E, Barro S, Amorim D (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?. *Journal of Machine Learning Research* 15: 3133-3181.

Grosjean P, Denis K (2014). Supervised classification of images, applied to plankton samples using R and zooimage. *Data Mining Application with R, Yanchang Z and Gen J (Eds.), Elsevier, Oxford, UK, pp 331-365.*

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J.R., Coelho, L.P., Espinoza, J.C.I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., Stemmann, L., Not, F., Hingamp, P., Speich, S., Follows, M., Karp-Boss, L., Boss, E., Ogata, H., Pesant, S., Weissenbach, J., Wincker, P., Acinas, S.G., Bork, P., de Vargas, C., Iudicone, D., Sullivan, M.B., Raes, J., Karsenti, E., Bowler, C., Gorsky, G., Coordinator, T.O.C., 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature*: 532, 465.

Olson R, Sosik H (2007). A submersible imaging-in-flow instrument to analyse nano-and microplankton: Imaging FlowCytobot. *Limnology and Oceanography Methods* 5: 331-365.

Picheral M, Guidi L, Stemman L, Karl D, Iddaoud G, Gorsky G (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography Methods* 8: 462-473.

Sosik HM, Olson RJ (2007) Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography-Methods* 5:204-216

Zarauz L, Irigoien X, Urtizberea A, Gonzalez M (2007). Mapping plankton distribution in the Bay of Biscay during three consecutive spring surveys. *Marine Ecology Progress Series* 345: 27–39.

### References section 3.2

Bonato S., Breton E., Didry M., Lizon F., Comille V., Lecuyer E., Christaki U., Artigas L.F. (2016) Spatio-temporal patterns in phytoplankton assemblages in inshore-offshore gradients using flow cytometry : A case study in the eastern English Channel. *Journal of Marine Systems*, 156:76-85

Bonato S., Christaki U., Lefebvre A., Lizon F., Thyssen M. & Artigas L.F. (2015) High spatial variability of phytoplankton assessed by flow cytometry, in a dynamic productive coastal area, in spring: the eastern English Channel. *Estuarine Coastal and Shelf Science*, 154: 214-223.

Campbell L, Olson RJ, Sosik HM, Abraham A, Henrichs DW, Hyatt CJ, Buskey EJ (2010) First harmful *inophysis* (Dinophyceae, Dinophysiales) bloom in the U.S. is revealed by automated imaging flow cytometry. *J Phycol* 46:66–75

Collier J (2000) Flow cytometry and the single cell in phycology. *J Phycol* 36:628– 644





- Courties C, Vaquer A, Troussellier M, Lautier J (1994) Smallest eukaryotic organism. *Nature* 370:255
- Dubelaar G.B.J, Geerders P.J. F., Jonker R. R. (2004) High frequency monitoring reveals phytoplankton dynamics. *Journal of Environmental Monitoring*, 2004, 6, 946-952
- Dubelaar GBJ, Gerritzen PL (2000) CytoBuoy: a step forward towards using flow cytometry in operational oceanography. *Sci Mar* 64:255–265
- Hébert P-A., Poisson-Caillault E., Wacquet G., Artigas L.F., Creach V., Bonato S., Thyssen M., 2014. The DYMAPHY FCMclustering R-Toolbox. INTERREG IVA “2 Mers Seas Zeeën”, DYMAPHY project, Scientific report Activity 2, 28 pages.
- Houliez, E., Lizon, F., Thyssen, M., Artigas, L.F., Schmitt, F.G.. (2012) Spectral fluorometric characterization of Haptophyte dynamics using the FluoroProbe: an application in the eastern English Channel for monitoring *Phaeocystis globosa*. *J. Plankton Res.* 34: 136-15
- Jonker R, Groben R, Tarran G, Medlin L, Wilkins C, Garcia L, Zabala L, Boddy L (2000) Automated identification and characterisation of microbial populations using flow cytometry: the AIMS project. *Sci Mar* 64:225–234
- Katano T, Nakano S (2006) Growth rates of *Synechococcus* types with different phycoerythrin composition estimated by dual-laser flow cytometry in relationship to the light environment in the Uwa Sea. *J Sea Res* 55:182–190
- Li WKW (2009) From cytometry to macroecology: a quarter century quest in microbial oceanography. *Aquat Microb Ecol* 57:239–251
- Marie D., Rigault-Jalabert F., Vaultot D. (2014) An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples. *Cytometry A* 85 (11): 962–968
- Ng A., Jordan M., Weiss Y., 1971. On spectral clustering: Analysis and an algorithm. NIPS14, Neural Information Processing Systems, pp. 849-856.
- Olson R, Frankel S, Chisholm S (1983) An inexpensive flow cytometer for the analysis of fluorescence signals in phytoplankton: chlorophyll and DNA distributions. *J Exp Mar Bio Ecol* 68:129–144 257
- Olson R, Vaultot D, Chisholm S (1985) Marine phytoplankton distributions measured using shipboard flow cytometry. *Deep Sea Res* 32(10):1273-1280
- Poisson-Caillault E., Creach V., Hébert, P-A., Bonato S., Artigas L.F., 2014. Comparison between manual and automated classification of flow cytometry data in aquatic samples. INTERREG IVA “2 Mers Seas Zeeën”, DYMAPHY project, Scientific report Activity 2, 10 pages.
- Rand W., 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, pp. 846-850.
- Rutten T., Rijkeboer M., Veen A., Thyssen M., Bonato S., Créach V, Artigas L.F. (2013). Operational Common Protocol for Pulse-shape recording Flow Cytometry, Scientific report, Activity 1, Dymaphy, 10 pages.
- Thyssen M, Garcia N, Denis M (2009) Sub meso scale phytoplankton distribution in the North East Atlantic surface waters determined with an automated flow cytometer. *Biogeosciences* 6:569–583





Thyssen M., Alvain S., Lefebvre A., Dessailly D., Rijkeboer M., Guiselin N., Creach V., Artigas L.F. (2015) High-resolution analysis of a North Sea phytoplankton community structure based on in situ flow cytometry observations and potential implication for remote sensing *Biogeosciences*, 12, 4051–4066, 2015

Veldhuis MJW, Kraay GWK (2000) Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. *Sci Mar* 64:121–134

Veldhuis MJW, Kraay GWK (2004) Phytoplankton in the subtropical Atlantic Ocean: towards a better assessment of biomass and composition. *Deep Sea Res Part I Oceanogr Res Pap* 51:507–530

Wacquet G., Hébert P.-A., Caillault-Poisson E. and Hamad D., 2011. Semi- Supervised K-Way Spectral Clustering using Pairwise Constraints. *NCTA, International Conference on Neural Computation Theory and Applications*, pp. 72-81.

Yentsch CS, Yentsch CM (1979) Fluorescent spectral signatures: the characterisation of phytoplankton populations by the use of excitation and emission spectra. *J Mar Res* 37:471–483

Zubkov M, Tarran G, Fuchs B (2004) Depth related amino acid uptake by *Prochlorococcus* cyanobacteria in the Southern Atlantic tropical gyre. *FEMS Microbiol Ecol* 50:153–161

### References section 3.3

Aardema, H. M., Rijkeboer, M., Lefebvre, A., Veen, A., and Kromkamp, J. C.: High resolution in situ measurements of phytoplankton photosynthesis and abundance in the Dutch North Sea, *Ocean Sci. Discuss.*, <https://doi.org/10.5194/os-2018-21>, accepted, 2018.

Aiken J., Fishwick J., Moore G. and Pemberton K. (2004). The annual cycle of phytoplankton photosynthetic quantum efficiency, pigment composition and optical properties in the western English Channel. *J. Mar. Biol. Ass. U.K.*, 84: 301-313.

Cohen SE, Golden SS (2015) Circadian Rhythms in Cyanobacteria. *Microbiol Mol Biol Rev* 79 : 373-385

Cheah W., Soppa M.A., Wiegmann S., Ossebaar S., Laglera L.M., Strass V.H., Santos-Echeandía J., Hoppema M., Wolf-Gladrow D., Bracher A. (2017) Importance of deep mixing and silicic acid in regulating phytoplankton biomass and community in the iron-limited Antarctic Polar Front region in summer. *Deep-Sea Research Part II*, 138: 74–85.

Erga S.R., Ssebiyonga N., Hamre B., Frette O., Hovland E., Hancke K., Drinkwater K. and Rey F. (2012) Environmental control of phytoplankton distribution and photosynthetic performance at the Jan Mayen Front in the Norwegian Sea. *Estuarine, Coastal and Shelf Science*, 97, 20: 91-103.

Gaarder, T. and Gram, H. H. (1927) Investigations of the production of phytoplankton in the Oslo Fjord. *Rapp. P.V. Cons. Int. Explor. Mer* , 42, 1–48.

Goss R, Lepetit B (2015) Biodiversity of NPQ. *J Plant Physiol* 172:13-32

Goss R, Pinto EA, Wilhelm C, Richter M (2006) The importance of a highly active and Delta pH-regulated diatoxanthin epoxidase for the regulation of the PSII antenna function in diadinoxanthin cycle containing algae. *J Plant Physiol* 163:1008-1021





- Harrison, J.W., Howell, E.T., Watson, S.B. and Smith, R.E.H., 2016. Improved estimates of phytoplankton community composition based on in situ spectral fluorescence: use of ordination and field-derived norm spectra for the bbe FluoroProbe. *Can. J. Fish. Aq. Sci.* 73, 1472-1482. <https://doi.org/10.1139/cjfas-2015-0360>
- Houliez, E., Lizon, F., Thyssen, M., Artigas, L.F., Schmitt, F.G.. (2012) Spectral fluorometric characterization of Haptophyte dynamics using the FluoroProbe: an application in the eastern English Channel for monitoring *Phaeocystis globosa*. *J. Plankton Res.* 34: 136-15
- Houliez, E., Simis, S., Nenonen, S., Ylöstalo, P., Seppälä, J. (2017) Basin-scale spatio-temporal variability and control of phytoplankton photosynthesis in the Baltic Sea: The first multiwavelength fast repetition rate fluorescence study operated on a ship-of-opportunity. *Journal of Marine Systems* 169 40–51
- Hughes D. J., Campbell D.A, Doblin M.A., Kromkamp J.C., Lawrenz E., Moore M., Oxborough K., Prášil O., Ralph P.J., Alvarez M.F. and Suggett D.J. (2018) Roadmaps and detours : Active chlorophylla assessments of primary productivity across marine and freshwater systems. *Env. Sc. and Technology*, 52, 12039-12054
- Karlson, B., Axe, P., Funkquist, L., Kaitala, S. and Sørensen, K., 2009. Infrastructure for marine monitoring and operational oceanography, Reports Oceanography no. 39, Swedish Meteorological and Hydrological Institute, 101 pp. <http://www.smhi.se/publikationer/infrastructure-for-marine-monitoring-and-operational-oceanography-1.2072>
- Kolber, Z.S., Prasil, O., Falkowski, P.G. (1998) Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochim. Biophys. Acta Bioenerg.* 1367, 88–106.
- Kromkamp J., Capuzzo E. and Philippart C.J.M. (2017). Measuring phytoplankton primary production: a review of existing methodologies and suggestions for a common approach. *EcApRHA Deliverable WP 3.2*. [S.n.]: [s.l.]. ISBN 9781911458272. 24 pp.
- Kromkamp JC, Dijkman NA, Peene J, Simis SGH, Gons HJ (2008) Estimating phytoplankton primary production in Lake IJsselmeer (The Netherlands) using variable fluorescence (PAM-FRRF) and C-uptake techniques. *Eur J Phycol* 43:327 – 344
- Lawrenz E, Silsbe G, Capuzzo E, Ylöstalo P, Forster RM, Simis SGH, Prášil O, Kromkamp JC, Hickman AE, Moore CM, Forget M-H, Geider RJ, Suggett DJ (2013) Predicting the Electron Requirement for Carbon Fixation in Seas and Oceans. *PLoS ONE* 8:e58137
- Lizon, F., Artigas, L.F., Bonato, S., Cornille, V., Créach, V., Degros, N., Didry, M., Houliez, E., Lefebvre, A., Rijkeboer, M., Thyssen, M., Veen, A. (2015) Spectral fluorometry: A technique for detecting phytoplankton pigmentary groups – DYMAPHY Project. Activity 1. Report 1.7, 24 pp.
- Milligan, A.J., Aparicio, U.A., Behrenfeld, M.J., 2012. Fluorescence and nonphotochemical quenching responses to simulated vertical mixing in the marine diatom *Thalassiosira weissflogii*, *Mar. Ecol. Prog. Ser.*, 448, 67–78, doi:10.3354/meps09544.
- Moore, C.M., Suggett, D.J., Hickman, A.E., Kim, Y., Tweddle, J.F., Sharples, J., Geider, R.J. and Holligan, P.M. 2006. Phytoplankton photoacclimation and photoadaptation in response to environmental gradients in a shelf sea. *Limnology And Oceanography*, vol. 51, no. 2, pp. 936-949.
- Müller, P., Li, X.P., Niyogi, K.K., 2001. Non-photochemical quenching. A response to excess light energy, *Plant Physiol.*, 125, 1558–1566.





- Neale P.J., Sobrino C. and Gargett A.E. (2012) Vertical mixing and the effects of solar radiation on photosystem II electron transport by phytoplankton in the Ross Sea Polynya. *Deep Sea Research Part I*, 63, 118-132.
- Oxborough, K., Moore, M.C., Suggett, D., Lawson, T., Chan, H.G., Geider, R.J. (2012) Direct estimation of functional PSII reaction centre concentration and PSII electron flux on a volume basis: a new approach to the analysis of Fast Repetition Rate fluorometry (FRRf) data. *Limnol Oceanogr Meth* 10: 142–154.
- Petersen, W., Schroeder, F., Bockelmann, F.D., 2011. Fe7rryBox - Application of continuous water quality observations along transects in the North Sea. *Ocean Dynamics* 61 (10), 1541-1554.
- Röttgers, R., Schönfeld, W., Kipp, P.R., Doerffer, R., 2005. Practical test of a point-source integrating cavity absorption meter: the performance of different collector assemblies. *Applied Optics* 44 (26), 5549-5560.
- Rytövuori S. (2017) Detection of picocyanobacteria and other algae containing phycoerythrin pigment in the Baltic Sea. In Finnish - MSc thesis, University of Helsinki . 84 p. <http://urn.fi/URN:NBN:fi:hulib-201705174155>
- Seppälä, J., Ylöstalo, P. Kuosa H., 2005. Spectral absorption and fluorescence characteristics of phytoplankton in different size fractions across the salinity gradient in the Baltic Sea. *Int. J. Rem. Sens.* 26, 387-414.
- Silsbe G.M., and Kromkamp J.C. 2012. Modeling the irradiance dependency of the quantum efficiency of photosynthesis. *Limnol. Oceanogr.: Methods* 10, 645–652
- Silsbe, G., Oxborough, K.D.J.S., Suggett, D.J., Forster, R., Ihnken, S., Komàrek, O., Lawrenz, E., Pràsil, O., Röttgers, R., Sicner, M., Simis, S.G.H., Van Dijk, M.A., Kromkamp, J. (2015) Toward autonomous measurements of photosynthetic electron transport rates: an evaluation of active fluorescence-based measurements of photochemistry. *Limnol. Oceanogr. Methods* 13, 138–155.
- Smyth T.J., Pemberton K.L., Aiken J. and Geider R.J. (2004). A methodology to determine primary production and phytoplankton photosynthetic parameters from Fast Repetition Rate Fluorometry. *J. Plankton Res.* 26, 11, 1337–1350.
- Suggett, D.J., Kraay, G., Holligan, P., Davey, M., Aiken, J. & Geider, R. 2001. Assessment of photosynthesis in a spring cyanobacterial bloom by use of a fast repetition rate fluorometer. *Limnology And Oceanography*, vol. 46, no. 4, pp. 802-810.
- Suggett, D.J., MacIntyre, H.L. and Geider, R.J. 2004. Evaluation of biophysical and optical determinations of light absorption by photosystem II in phytoplankton, *Limnol. Oceanogr.: Methods*, vol. 2, pp. 316-332
- Suggett D., Maberly S.C. and Geider R.J. (2006) Gross photosynthesis and lake community metabolism during the spring phytoplankton bloom. *Limnol. Oceanogr.*, 51(5), 2064–2076.
- Wollschläger, J., Röttgers, R., Petersen, W., Wiltshire, K.H., 2014. Performance of absorption coefficient measurements for the *in situ* determination of chlorophyll-a and total suspended matter. *Journal of Experimental Marine Biology and Ecology* 453 (0), 138-147.
- Wollschläger, J., Röttgers, R., Petersen, W., Zielinski O., 2019. Stick or Dye: Evaluating a Solid Standard Calibration Approach for Point-source Integrating Cavity Absorption Meters (PSICAM). *Front. Mar. Sci.* 5:534.

