Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories - **JERICO-NEXT**	
Deliverable title	Novel methods for automated *in situ* observations of phytoplankton diversity
Work Package Title	WP 3
Deliverable number	D3.1
Description	Synthesis report after developments dedicated to the observation of the phytoplankton diversity
Lead beneficiary	SMHI
Lead Authors	Bengt Karlson, Felipe Artigas, Veronique Créach, Arnaud Louchart, Guillaume Wacquet and Jukka Seppälä
Contributors	Hedy Aardema, Michael Brosnahan, Reinhoud de Blok, Pascal Claquin, Florent Colas, Klaas Deneudt, Gérald Grégori, Jacco Kromkamp, Soumaya Lahbib, Alain Lefebvre, Fabrice Lizon, , Klas Möller, Emilie Poisson-Caillault, Machteld Rijkeboer, Thomas Rutten, Suvi Rytövuori, Lars Stemmann, Melilotus Thyssen, Lennert Tyberghein, and Pasi Ylöstalo.
Submitted by	Bengt Karlson
Revision number	9
Revision Date	4 October 2017
Security	Public

History			
Revision	**Date**	**Modification**	**Author**
1	18 May 2017	Draft 1 with outline and preliminary author list	Bengt Karlson
2	7 June 2017	Draft 2 with updated authors list	Bengt Karlson
3	15 Sep. 2017	Draft 3 with some of the content added	Bengt Karlson
4	21 Sep. 2017	Draft 4 Added sections compiled by Guillaume Wacquet, Jukka Seppälä and Veronique Créach.	Bengt Karlson
5	21 Sep. 2017	Draft 5 Compilation and revision by Felipe Artigas	Felipe Artigas
6	22 Sep. 2017	Draft 6 Compilation and revision by Guillaume Wacquet	Guillaume Wacquet
7	26 Sep. 2017	Added executive summary, submitted to WP-leaders.	Bengt Karlson
8	1st Oct. 2017	Added recommendation and accepted revision	Felipe Artigas
9	4 October 2017	Finalised revised version, submitted to WP-leaders	Bengt Karlson

Approvals				
	Name	**Organisation**	**Date**	**Visa**
Coordinator	P. Farcy	IFREMER	4/10/2017	PF
WP Leaders	G. Petihakis, L. Delauney	HCMR, IFREMER	4/10/2017	GP LD

Table of contents

1. Executive Summary

Novel methods for automated in situ observations of phytoplankton diversity

This is a summary of activities and results from months 1-24 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 a synthesis report after developments dedicated to the observation of the phytoplankton diversity by applying novel techniques on automated platforms. The work has been carried out in close connection with task 4.1 Biodiversity of plankton, harmful algal blooms and eutrophication. The partners involved are CNRS, SYKE, SMHI, HZG, RWS, VLIZ CEFAS, Ifremer and AZTI. Subcontractors in WP4, task 4.1 are WHOI, Scanfjord AB, Tomas Rutten b.v., CytoBuoy b.v. and UGent - PAE.
The work has been carried out mainly in the field with activities in the Baltic Sea, the Kattegat-Skagerrak, the 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 include research vessels, Ferrybox systems on merchant vessels, instrumented oceanographic buoys/fixed platforms and land based systems. Ocean observatories, i.e. multi-sensor and multi-platform systems, have been used in some locations, allowing inter comparison of techniques and sensors between at least two partners. In addition, work on developing and testing new algorithms have been carried out in offices and laboratories. Some work on microalgae cultures has also been carried out. Two international workshops have been successfully arranged, one in Wimereux, France (June 2016 – organised by CNRS-LOG) and one in Gothenburg, Sweden (September 2016, organised by SMHI) in which partners presented, discussed and were also able to inter compare the sensors and techniques that were or will be implemented in the field. 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 includes evaluating instruments and developing algorithms for automated identification of phytoplankton from 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. In addition, the CytoSense and CytoSub were used to collect images. The *in situ* video system UVP5 was implemented during a cruise in the Baltic Sea-Skagerrak-Kattegatarea, 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 includes using a subset of images of organisms for training the systems. Existing software were improved (as the PhytoZooImage) 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 and on research vessels 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.

Future work in task 3.1

Most field work has ended but some will continue, e.g. at the Utö observatory in the Baltic Sea. The data collected during months 1-24 and the new data will be processed further and used for improving the discrimination of phytoplankton taxa or functional groups by inter-comparison of techniques and continued algorithm development, as well as for preparing JERICO-NEXT delivery 3.2. In addition, scientific publication of results is in progress or being planned. A special issue in the open access journal Diversity (MPI) is being discussed. Some results and strategy will be presented during a symposium in Hannover, Germany, in October 2017 and during the FerryBox workshop on board the ship Colour Fantasy later in October 2017. Results will also be presented during the third JERICO-NEXT plankton workshop to be arranged in Marseille in March 2018, during the International Conference on Harmful Algae in Nantes in October 2018, and in other meetings to be determined.

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 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*-a and is also useful for discriminating between phytoplankton groups based on pigment content.
7. Variable 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 coupling with other optical and also biogeochemical analysis.

2. Introduction

By Bengt Karlson, Felipe Artigas, Véronique Créach and Jukka Seppälä

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 µm 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 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.

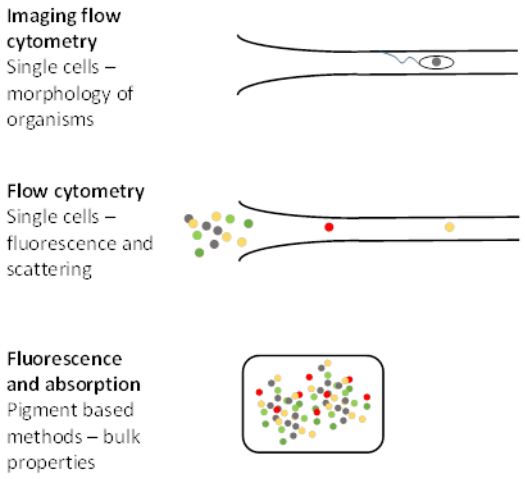


Figure 2.1. Overview of three methods for observing phytoplankton.

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 an preservation in development	Low-medium
Chlorophyll a analyses (water sampling)	-	Medium	-	Low-Medium	Low (automated water sampling is available)	Low-Medium
HPLC-analysis of photosytnhtetic 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 conditons)

3. Main report

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

Lead authors: Bengt Karlson, Guillaume Wacquet, Felipe Artigas, Florent Colas, Machteld Rijkeboer, Lars Stemann.

3.1.1. Overview of methodology and instruments commercially available

3.1.1.1. Imaging FlowCytoBot (McLane Research Laboratories)

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 <10 to 150 µm (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, ~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 2µm to 2000µm (depending on the combination « magnification/flow cell » used for the optical system, ie. 2X/600µm, 4X/300µm, 10X/100µm or 20X/100µm).

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

« AutoImage » 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.

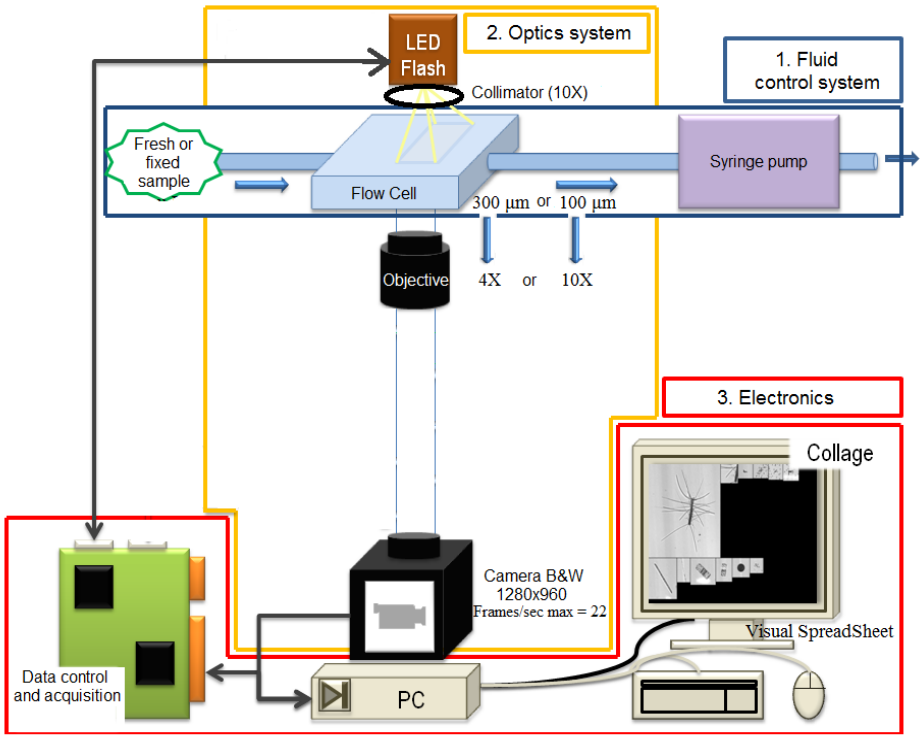
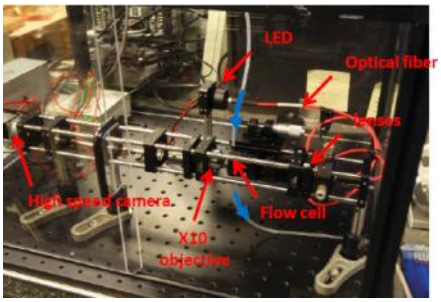


Figure 3.1.3. Overview of the FlowCAM principle.

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).

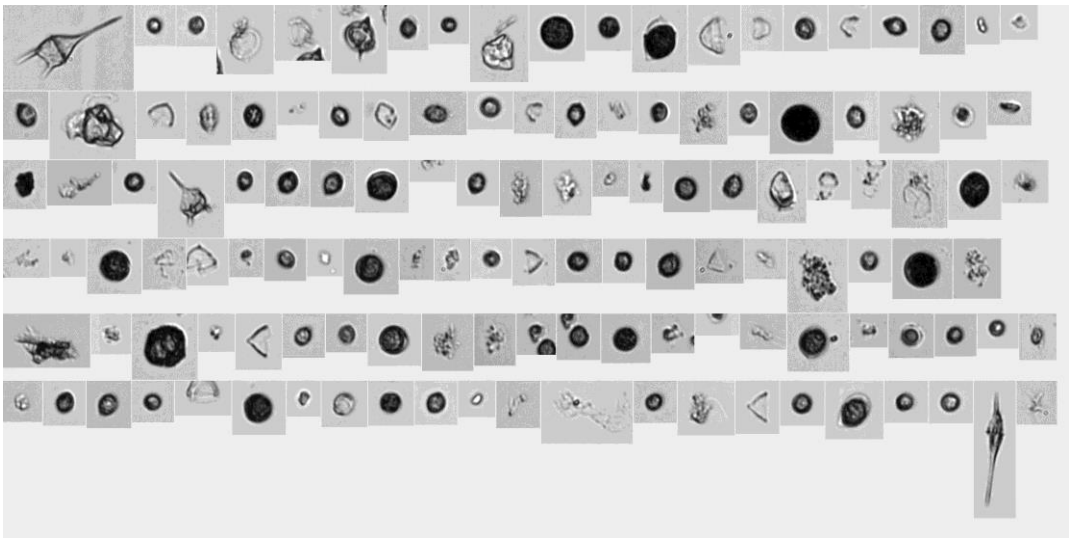


Figure 3.1.4. Thumbnails generated by the FastCAM during the JERICO-NEXT Phytoplankton workshop in Gothenborg, 27-30 September 2016.

An LED driven by a control box emits light pulses of 5 µs 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 µm. 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 « ZooImage » (R environment).

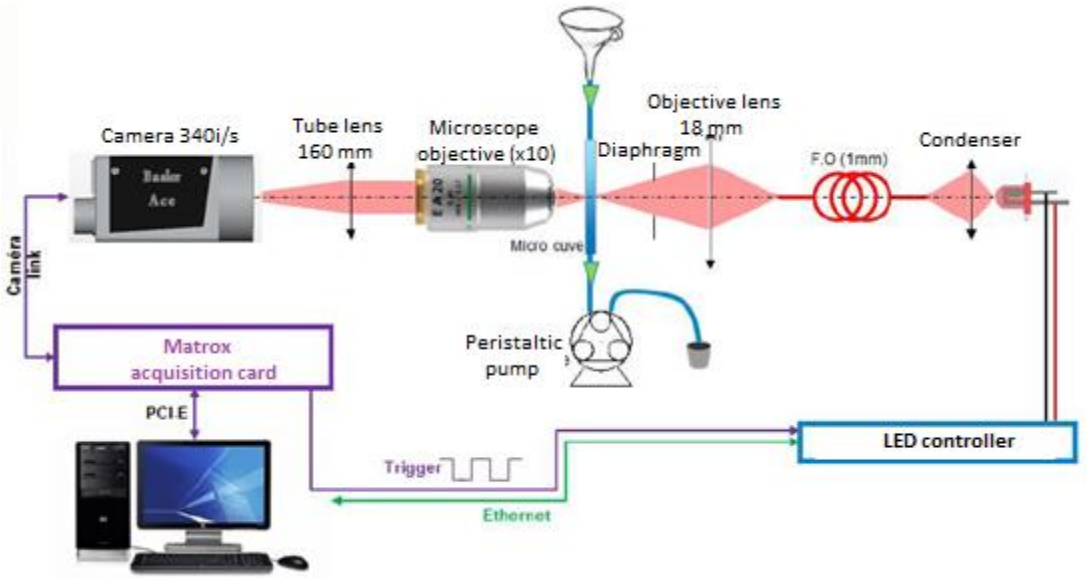


Figure 3.1.5. Overview of the FastCAM principle.

3.1.1.4. Underwater Vision Profiler UVP5 (Hydroptic)

Instrument. The UVP5 images large plankton (equivalent spherical diameter, ESD >600 µm; 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 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 µm (total number ~1 million 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.)

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 µm but able to collect pictures of 2 µm 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 (Oceanological Observatory of Villefranche-sur-mer 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, ISIIS, IFCB, automated microscope, etc.).

The application already hosts over 30 million images of plankton, 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.

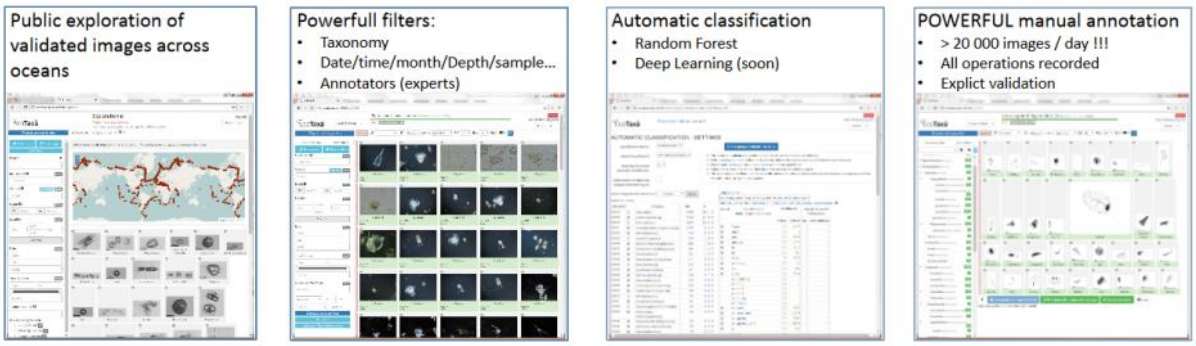


Figure 3.1.8. Overview of the ECOTAXA application.

3.1.1.7. ZooImage software (University of Mons)

« ZooImage » 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. ZooImage 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).

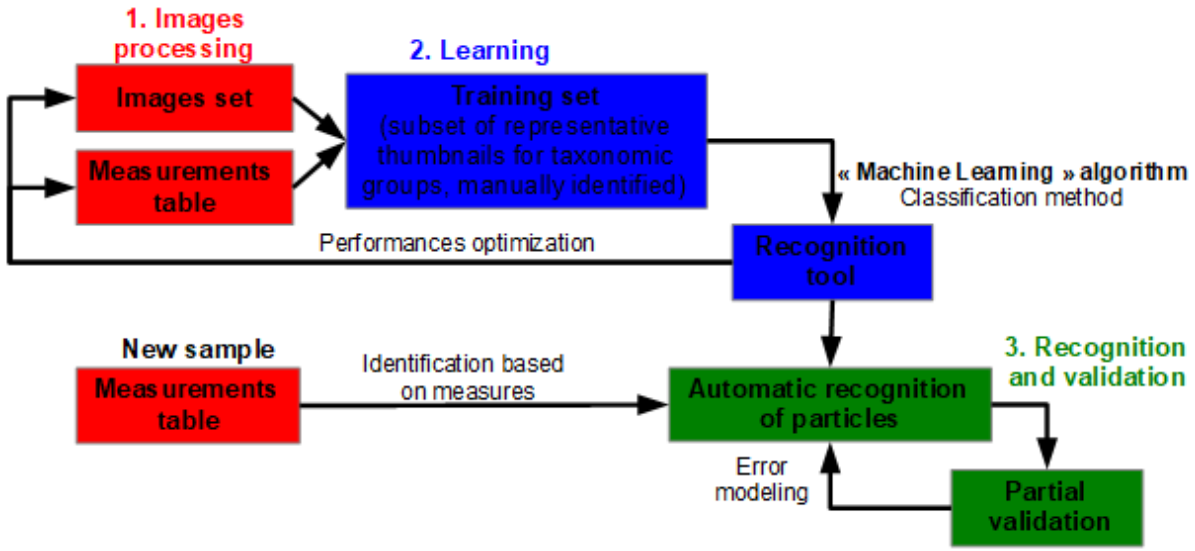


Figure 3.1.9. Overview of the classification process in ZooImage.

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).

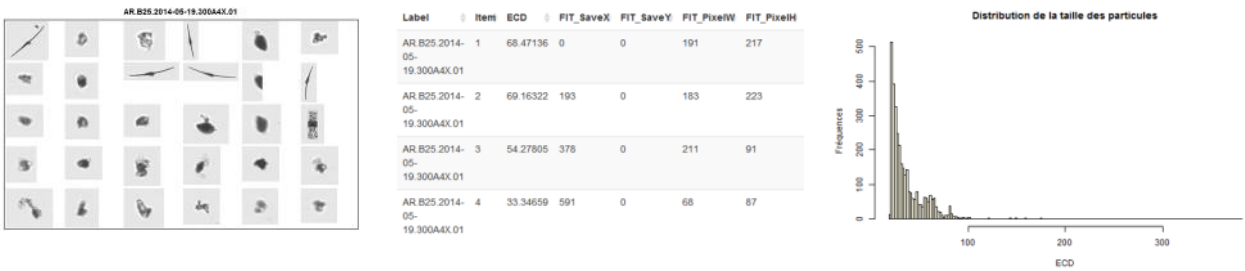


Figure 3.1.10. Examples of sample description tools in ZooImage.

During the classification process in ZooImage, 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 ZooImage.

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
IFCB	<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
FlowCAM	<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
FastCAM	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
UVP5	>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)
CytoSense/Sub (old IIF system)	>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
CytoSense/Sub (new IIF system) Since 2017	>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

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.

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

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 respected 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.

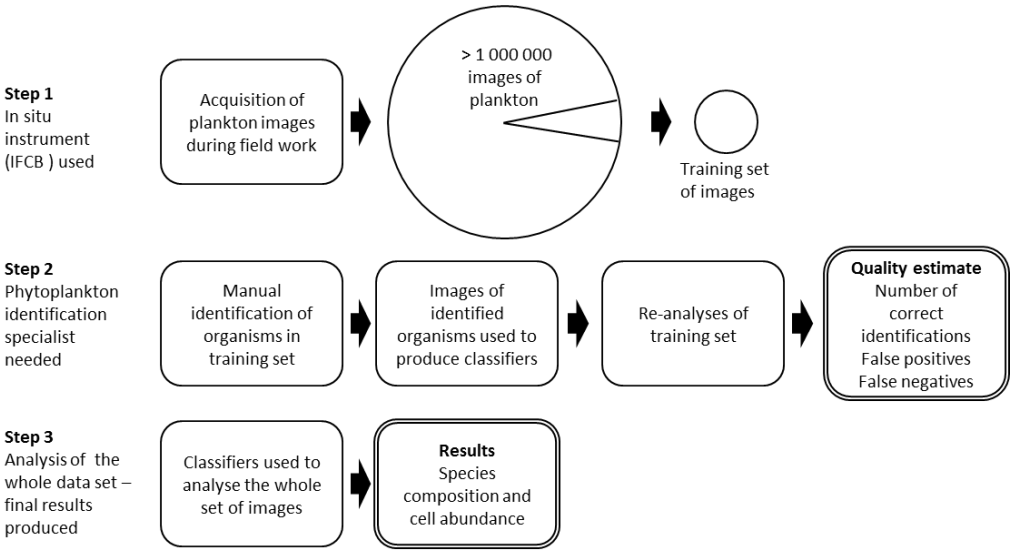


Figure 3.1.13. An overview of the steps needed for automated phytoplankton analyses using the Imaging Flow Cytobot.

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.

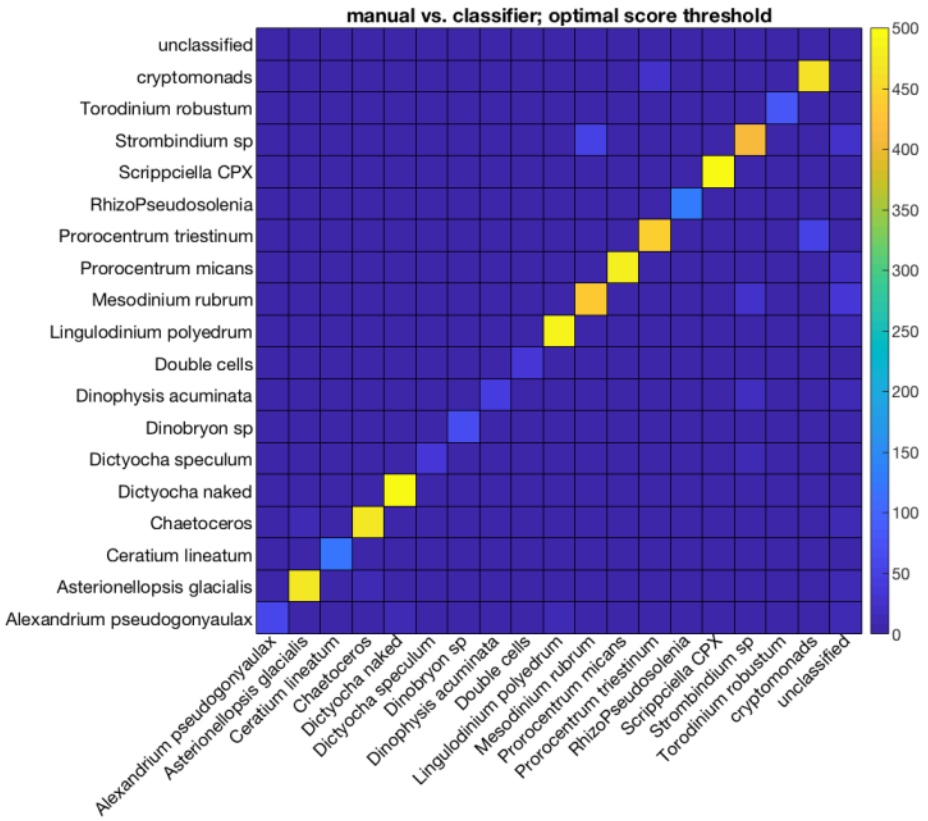
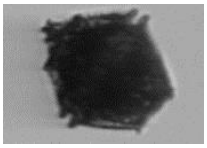


Figure 3.1.14. The graph illustrates the accuracy of automated identification of selected phytoplankton species using classifiers developed using the Tångesund data set. A higher value of optimal score threshold, i.e. closer to 500, indicates a more accurate detection.



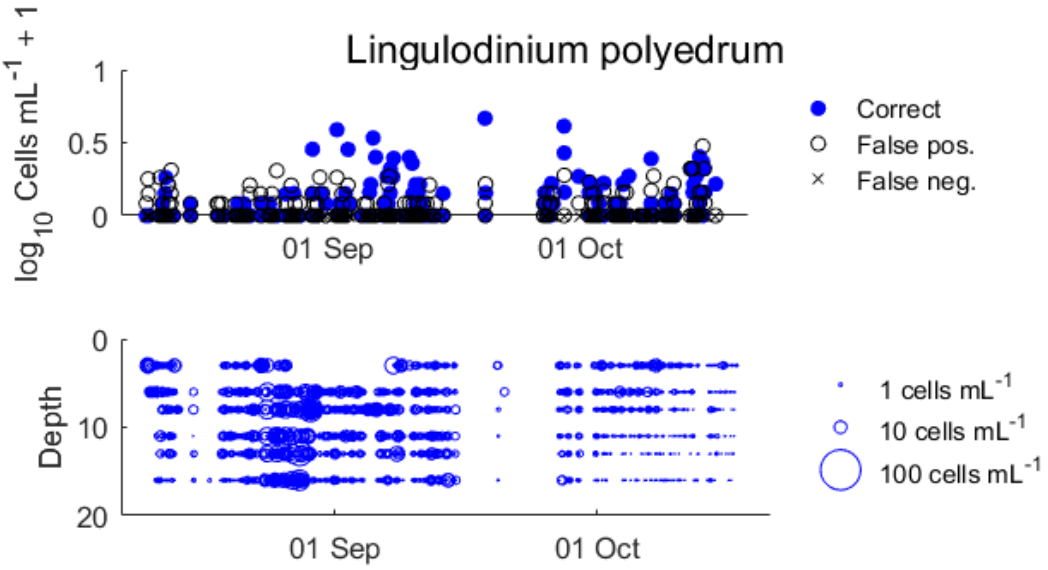


Figure 3.1.15. An example of results using automated identification and quantification of a harmful algae during the Tångesund study in 2016. Lingulodinium polyedrum is a producer of yessotoxins that may accumulate in shellfish. Upper panel: The cell abundance of L. polyedrum as estimated using the IFCB.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

3.1.2.2. FlowCAM-ZooImage training set

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 Monitoring network) 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-1. A 4X objective (40X overall magnification) coupled with a 300µm-depth flow-cell was used and samples were run in "AutoImage" operation mode. The analysis process, from raw images processing to statistical analysis, was carried out using the ZooImage 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.

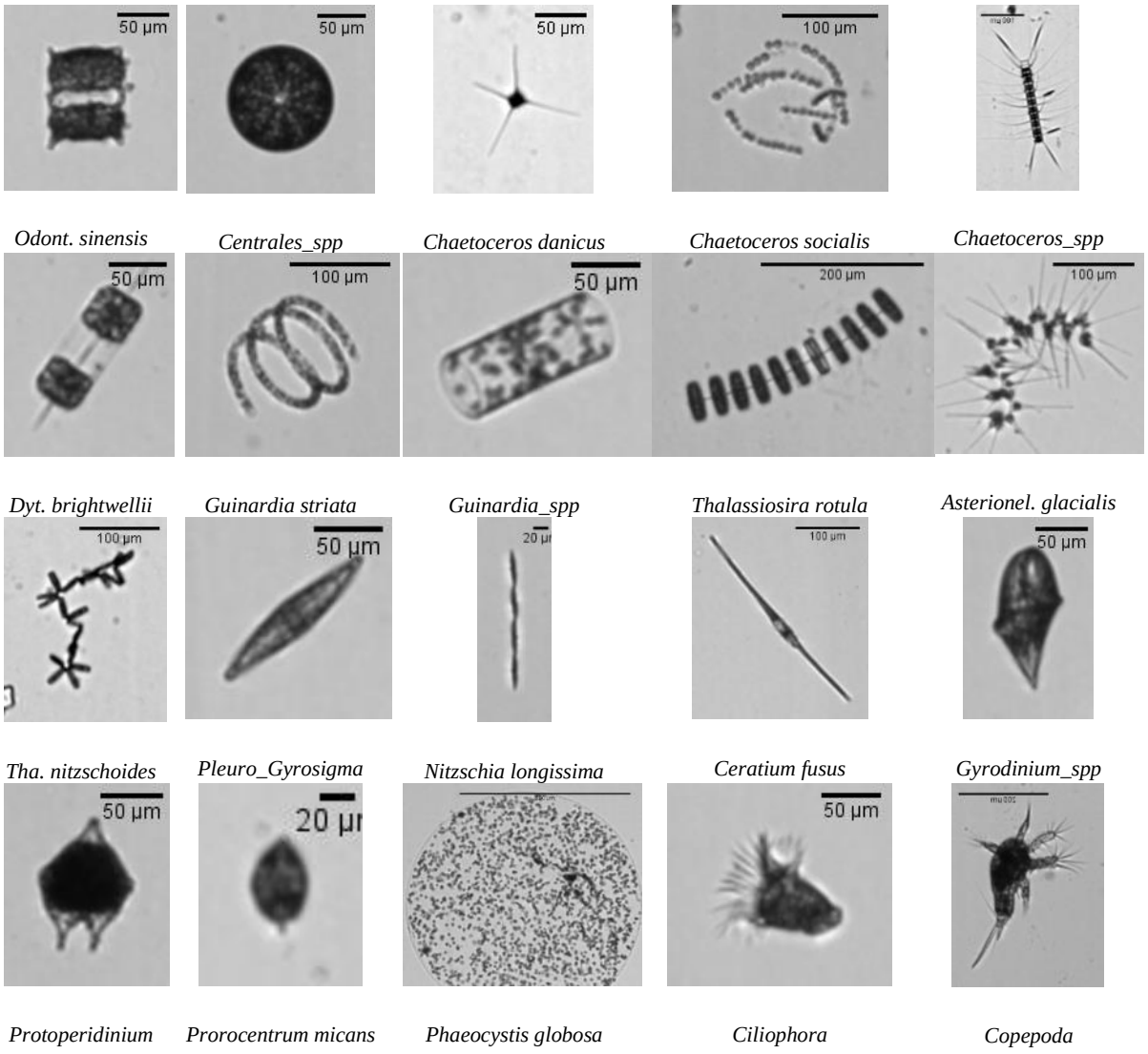


Figure 3.1.16. *Examples* of images in some taxonomic groups in the FlowCAM training set.

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

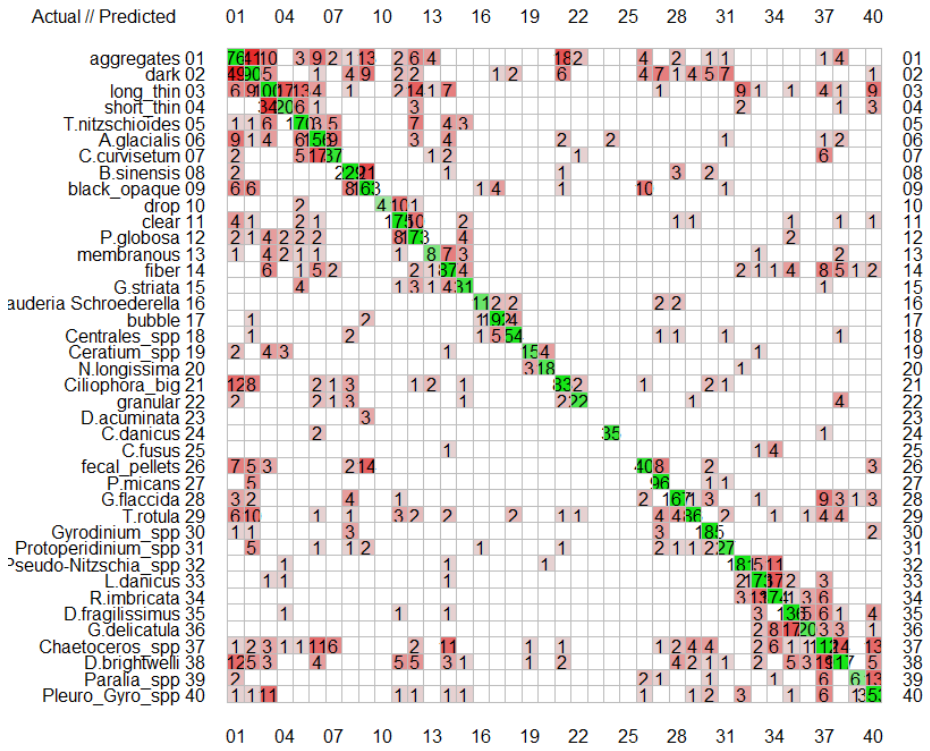


Figure 3.1.17. Confusion matrix associated to the FlowCAM recognition tool built with Random Forest.

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 nitzschoides*, *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

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 must be 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 Monitoring network, the CNRS LOG DYPHYRAD transect and the SOMLIT program. 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.

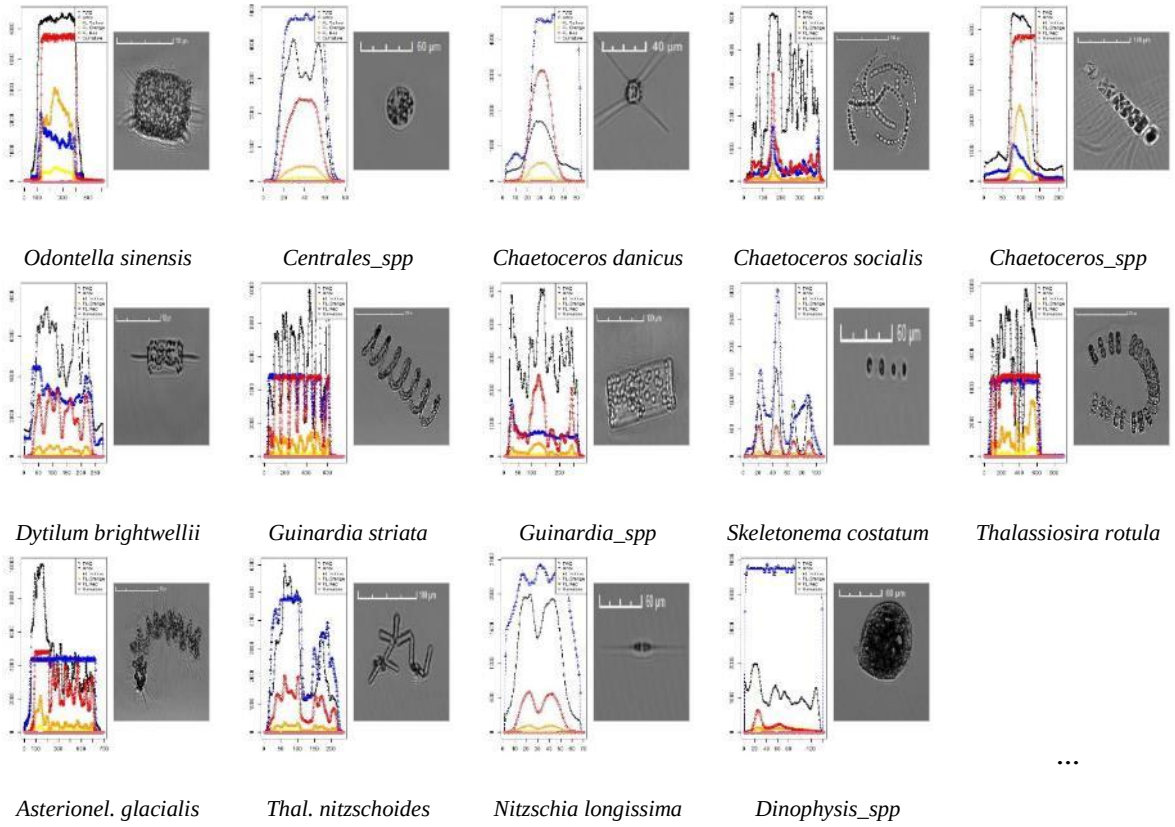


Figure 3.1.18. Examples of images in some taxonomic groups in the CytoSense training set.

Results. Abundances 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.

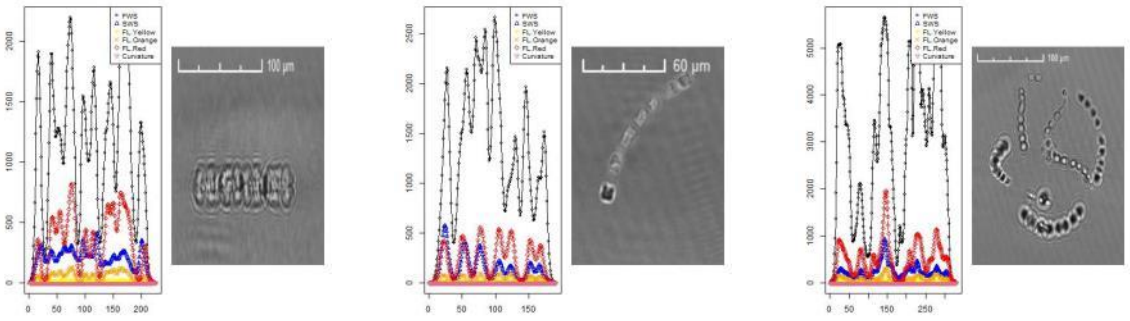
Table 3.1.3. Abundances and error rates per group for the CytoSense recognition tool built with Random Forest. The red font indicates the species or genus for which the percentage of misclassification goes over 50%.

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

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).

Actual // Predicted	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	
Guinardia_spp 01	2	12	15													01
C.socialis 02		55	27													02
Chaetoceros_spp 03		15	96		1	1							1			03
N.longissima 04				71	1											04
T.nitzschioides 05		4	3	4	67	1									1	05
C.danicus 06		3	1			18										06
Centrales_spp 07					1	1	2	1								07
Ciliophora 08		1	2			1	1	1		1						08
Dinophysis_spp 09										2						09
A.glacialis 10			1	1	2			1		3	2				1	10
D.brightwellii 11										2					2	11
S.costatum 12												21				12
G.striata 13		5	2												2	13
B.sinensis 14										1					2	14
T.rotula 15				1	2	1									17	15
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	

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



Guinardia spp *Chaetoceros spp* *Chaetoceros socialis*

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.

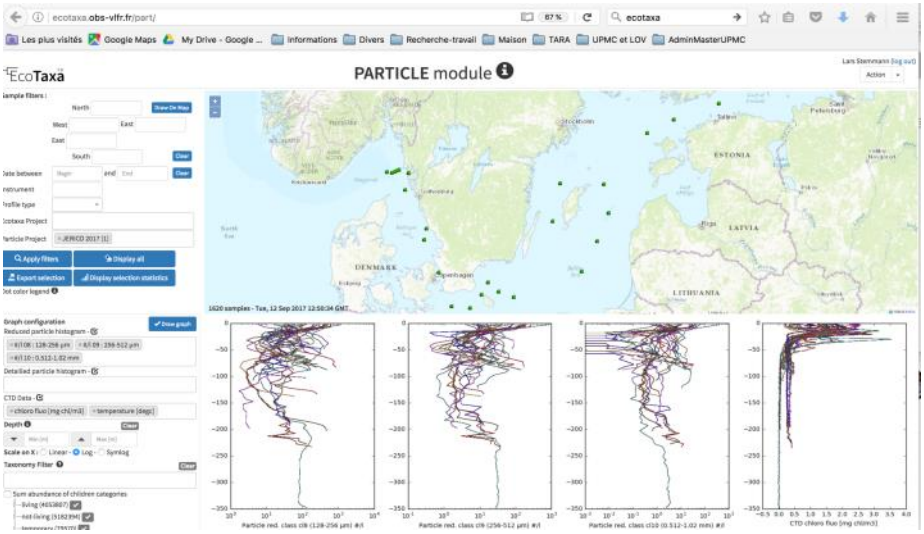
3.1.2.4. FastCAM deployment

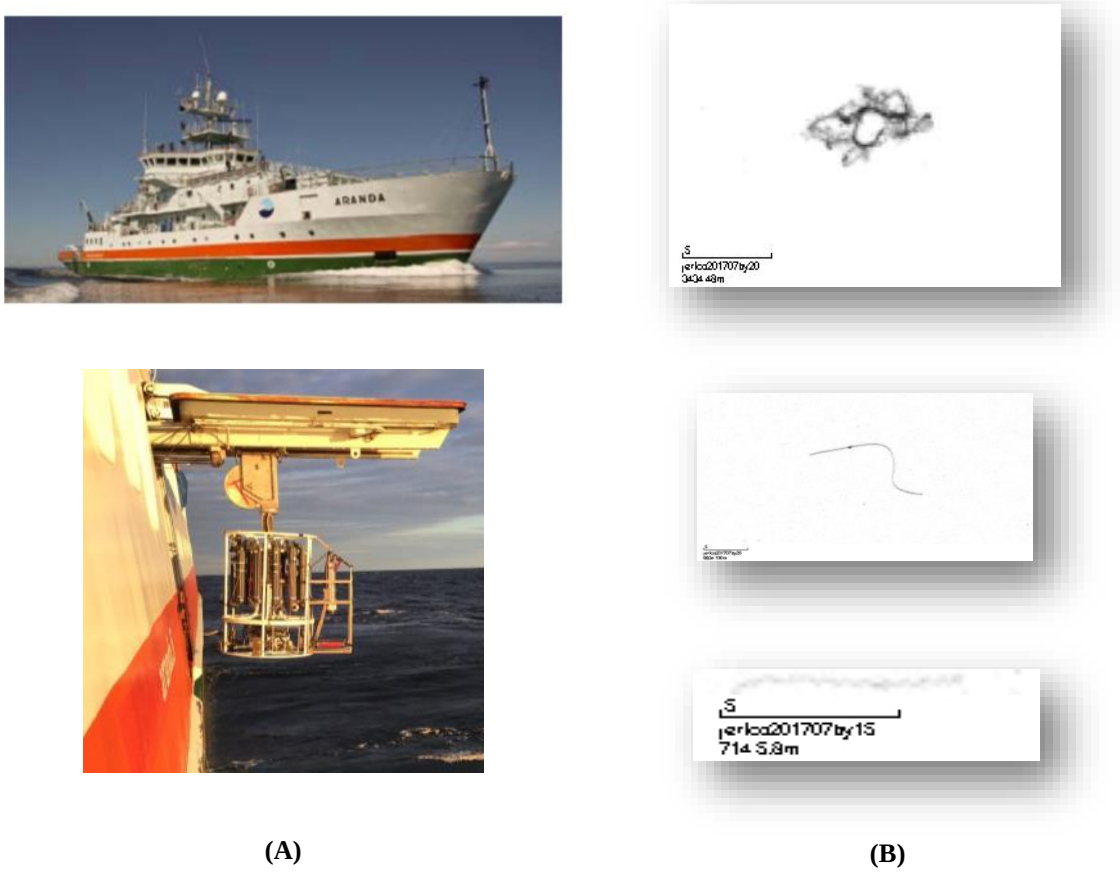
Past achievement. The FastCAM was deployed during a technological campaign in Brest bay and Iroise sea (IPARO) in 2015. It participates to the JERICO-NEXT Phytoplankton workshop in Gothenborg, 27-30 September 2016.

Next development and experiments. The FastCAM is currently being developed to make fully automatic analyses. An automatic sampling unit will be designed in 2018. The aim of make fully automatic analysis from a seawater inlet. The module will drive the process from the sample injection to the taxonomic classification, without operator.

3.1.2.5. UVP5 deployment

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).





(A) **(B)**

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 first 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. Results and recommendations will be then integrated in Deliverable D3.2.

3.2. Single-cell optical characterization

Lead authors: Véronique Créach, Guillaume Wacquet, Arnaud Louchart, Felipe Artigas, Melilotus Thyssen

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 Types (PFT). By providing large volumes of comparative data on the abundance and distribution of all PFT, 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 PFT, 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 PFT 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 µm but is able to collect pictures of 2 µm 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.

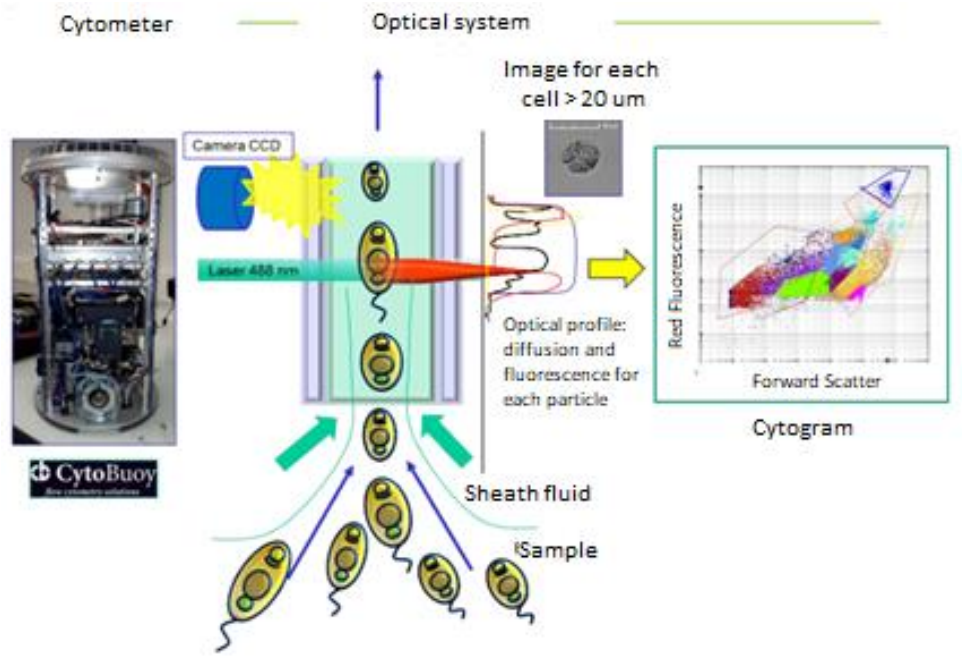


Figure 3.2.1. Overview of the flow cytometry principle.

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).

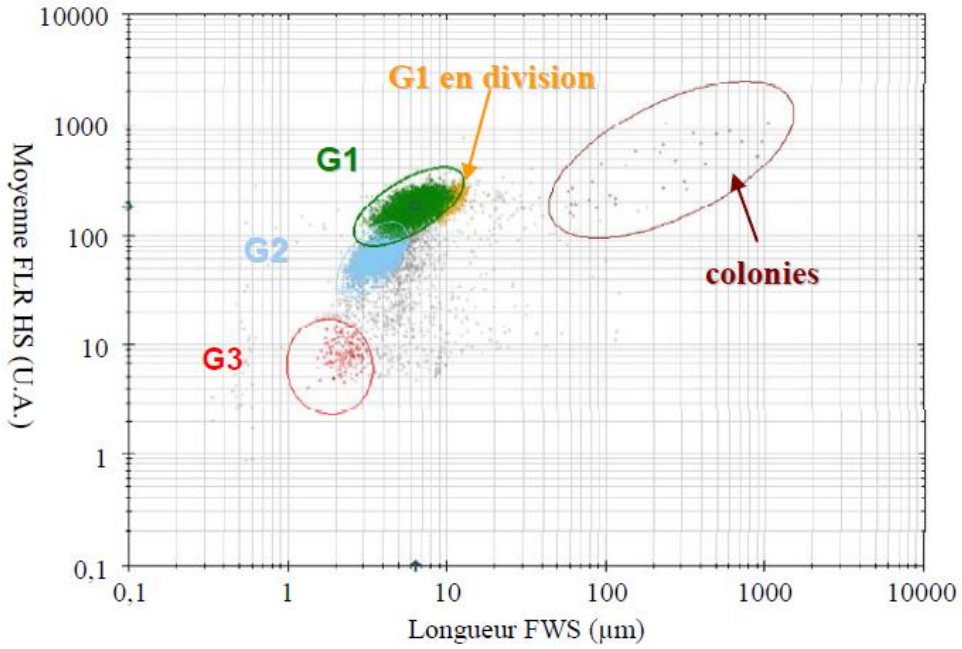


Figure 3.2.2. Discrimination of Phaeocystis globosa groups from a culture. G1, G2 and G3 groups were single-cells and colonies clusters were composed by G1 cells bulk (Guiselin, 2010).

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 µm) with the highest fluorescence content per cell. Finally, the G2 and G3 groups showed approximately the same size (G2: 3 to 6 µm and G3: 1.5 to 4.5 µm) 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 parameters * 9 traits) + 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.

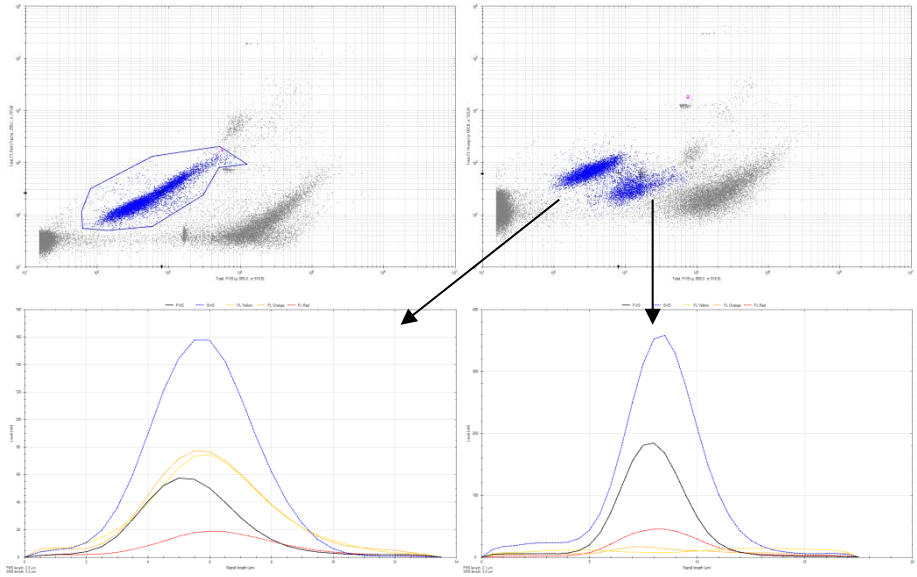


Figure 3.2.3. The use of multiple combinations is very useful and sometimes the operator can discriminate several groups in one plot that could not be seen in one other. In the above figure, the first combination (Total forward scatter vs. Total FLR) highlights only one blue group. A second combination (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, 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 Le Laboratoire d'Informatique du Signal et de l'Image de la Cote d'Opale (LISIC/ULCO, France). This tool is designed both to automatically cluster the phytoplankton functional types, 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-processings. 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:

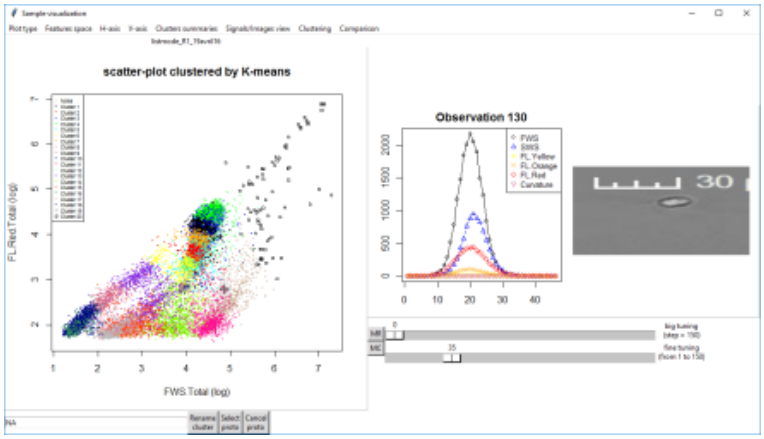


Figure 3.2.4. Scatter plot (by default): a global 2D representation of all observations contained in the dataset, each one being plotted as a point, whose coordinates match its measurements according to the 2 selected features. The colour used for each point matches its corresponding cluster.

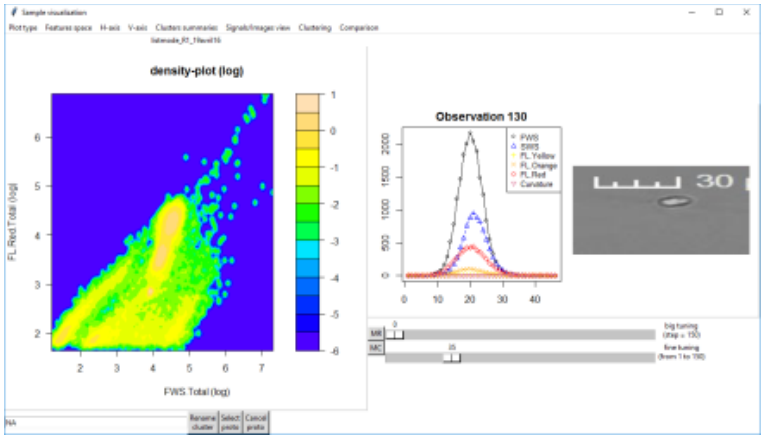


Figure 3.2.5. Density plot: a density estimation of the dataset in the selected features space, allowing a better visualisation of the concentrations than the 2D scatter plots. The colours correspond to the densities of particles with the warmest colour for the highest concentration of particles.

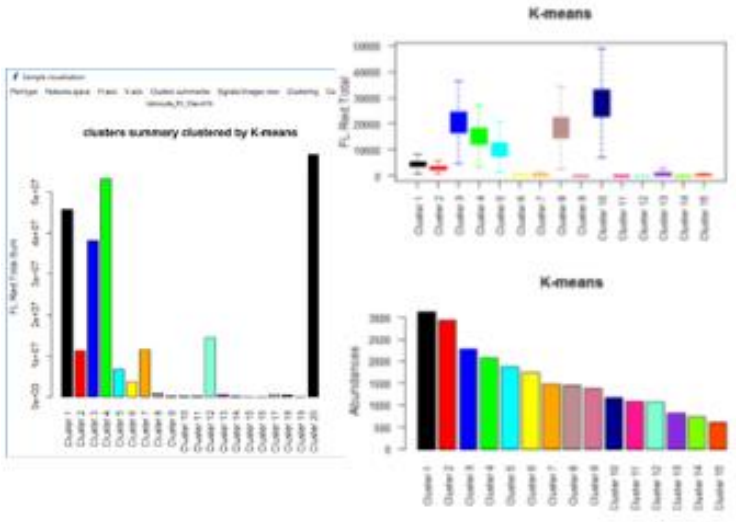


Figure 3.2.6. Cluster summary: in the menu "Clusters summaries", by-cluster synthetic graphics proposed histograms of abundance, features min, max, sum, average or std) or observation features boxplots.

Data pre-processing (B).

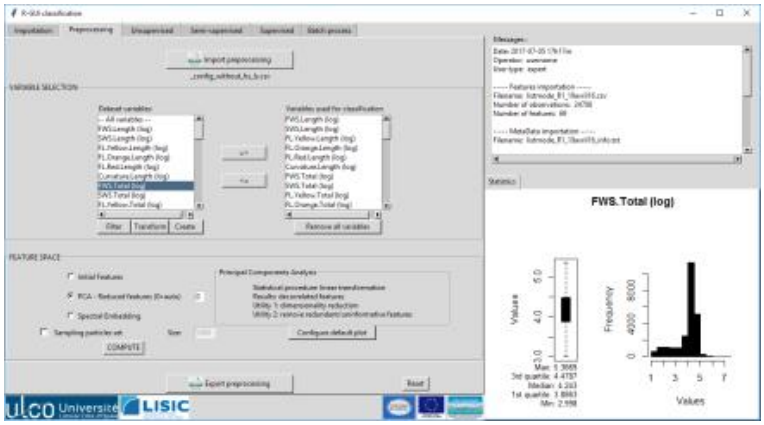


Figure 3.2.7. Processing page in RClusTool

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) 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-phycocyanin) 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 analys 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.

3.2.2.2. Standardisation of flow cytometry measurement and data

Beside 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 controlled 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 types 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), 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.

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 is 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 types: 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 PFTs

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 is to use CytoClus software to cluster the PFTs and the results will be compared with the RclusTool and EasyClus® outputs to estimate the discrepancy between the different mode of clusterisation. Data will be reported in D3.2 (M42).

Inter calibration between Cytosenses/Cytosub

Analyses of fixed samples for inter comparison

During some JERICO-NEXT cruises, samples were collected and fixed with glutaraldehyde following Marie et al. (2014). Eight samples were sent to VLIZ, Cefas, CNRS-LOG./ULCO, RWS, and CNRS-MIO to be analysed. The analyses were performed in July 2018 and the raw data are on the VLIZ FTP. The results will be used to compare the performance of the different flow cytometers.

Analyses of natural community during cruises for inter comparison

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. Data from RWS and ULCO-LOV will be compared and reported in D3.2 (M42).
- 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 Channel/ south North Sea in May 2017: the VLIZ spring cruise (May 2017, 8th - 12th) 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 were analysing the phytoplankton community in continuous mode. Data will be compared and reported in D3.2 (M42).

3.2.2.4. Next steps

When	What
Whole period	– Continued analyses of data and preparation of D3.2 and manuscripts for scientific articles
October 2017	– Some activities. will be presented in the *Symposium high throughput methods in marine time series* in Hannover
6-8 December 2017	– Workshop for discussing the results of the activity 3.1.2. Location: not decided yet
19-21 March 2018	– Third JERICO-NEXT phytoplankton workshop, Marseille

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 IVA"2 Seas", 2011-2014), intercalibration exercises were performed wth 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 will be summarised in D3.2. Exemples from different calibration exercises performed during the first 30 months of JERICO-NEXT will be demonstrated. Because the flow cytometer analysis is user and machine dependant, a single analytical procedure is not recommended today. However, by setting a control quality procedures for sample and result analysis, and intercalibration exercises in regular base, we will reach the degree of confidence needed for the data to be reported in European data portal.

3.3. Bio-optical Instrumentation

Lead authors: Jukka Seppälä, Fabrice Lizon, Felipe Artigas, Arnaud Louchart, Alain Lefebvre, Pascal Claquin, Bengt Karlson, Jacco Kromkamp, Suvi Rytövuori, Pasi Ylöstalo and Klas Möller

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 phycoeryhtrin. Chlorophyll-a is used as a proxy chlorophyll-a concentration, while phycocyanin is used as a proxy for filamentous cyanobacteria and phycoeryhtrin as a proxy for some filamentous cyanobacteria as well as picocyanobactreria, 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 fluorescence and FRRF technique used in JERICO-NEXT, 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 photoquenching 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 photoquenching 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 photoquenching, 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.

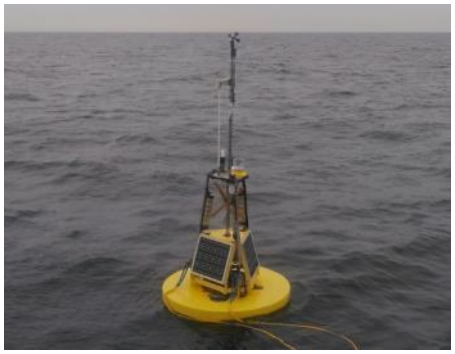


Figure 3.3.1. The SMHI Huvudskär E. oceanographic buoy.

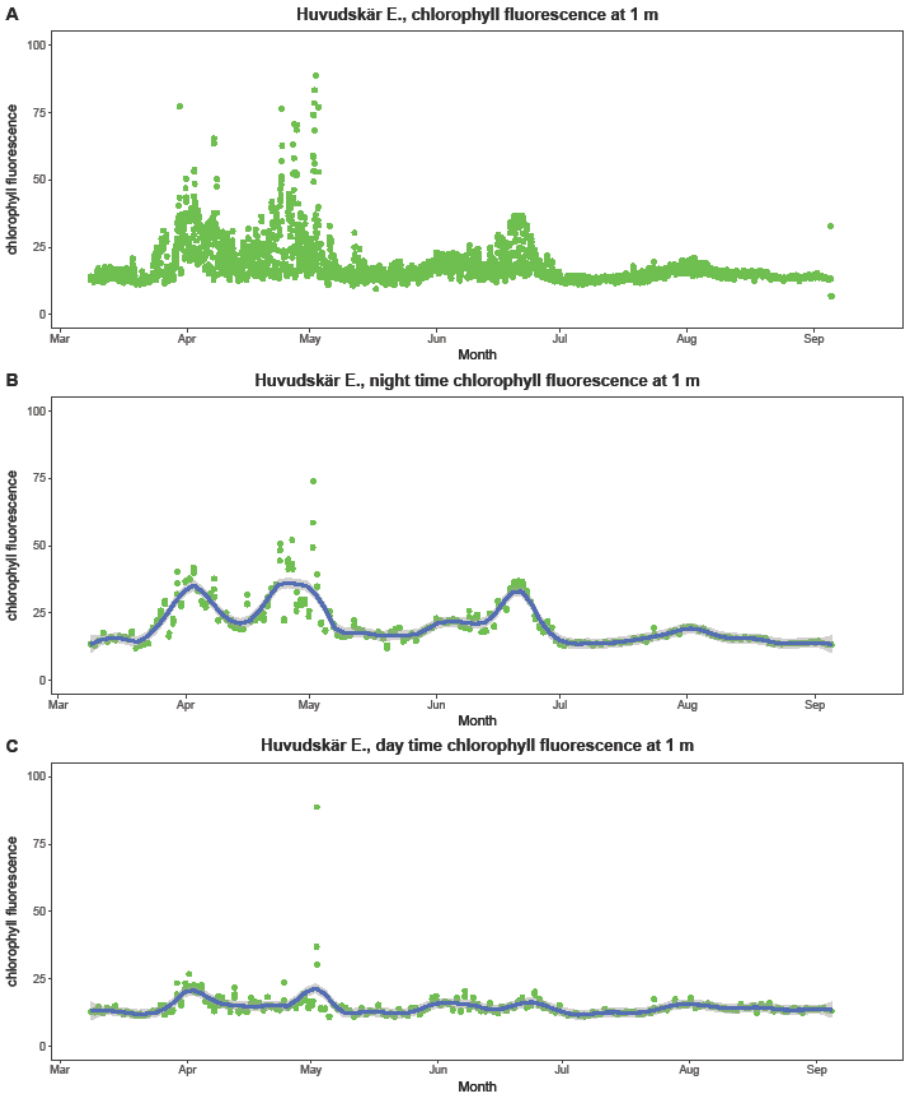


Figure 3.3.2. Chlorophyll fluorescence measured at 1 m depth in the NW part of the Baltic proper in year 2017. The y-axis shows arbitrary units. The graphs illustrate preliminary data. A: data collected every hour, B. data collected at 2200, 2300 and 0000 UTC, i.e. around midnight local time, C data 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 onboard ferry Finnmaid, as part of algaline 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.

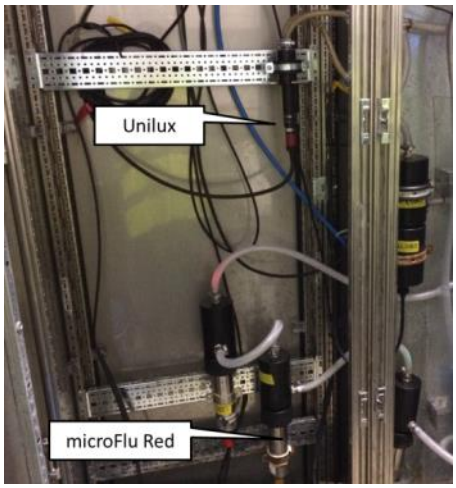


Figure 3.3.3. Phycoerythrin fluorometers installed onboard Finnmaid.

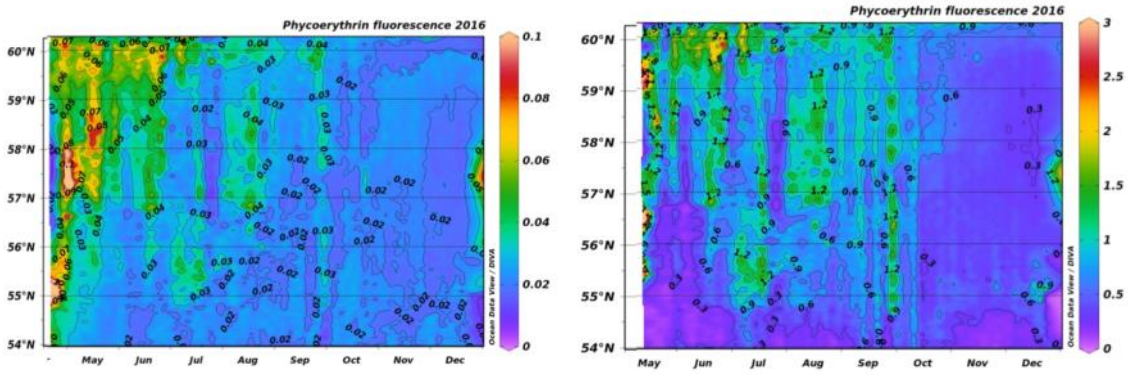


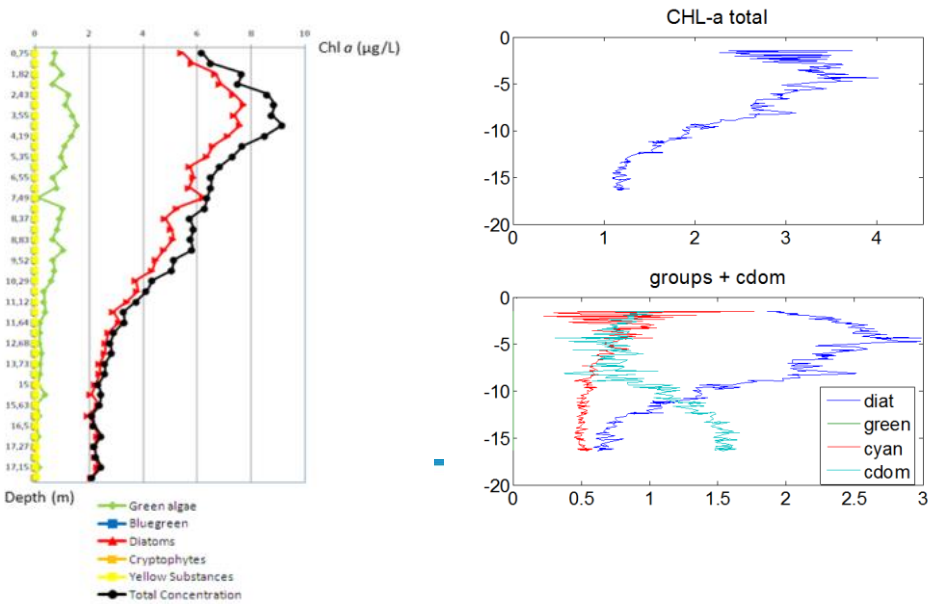
Figure 3.3.4. Variability of phycoerythrin fluorescence measured with microFlu Red (left) and unilux (right fluorometers). Limits of quantification (LoQ) for instruments were 0.025 V and 0.5 fluorescence units, respectively. The values below these LoQ threshold are to be considered as background noise. 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 fluctutations are most likely related to variation in picocyanobactreia 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 the deeper layers high PE readings most likely indicate the uplift of these organisms. MSc thesis has been written based on the results (Rytövuori 2017). More detailed data analyses is carried out during 2017-18.

Deployments of spectral fluorometers

In the JERICO-NEXT frame, the FluoroProbe was deployed by the CNRS-LOG and MultiExciter by SYKE. The instruments can be configured as a profiler or can be mounted for continuous measurements. These two configurations depend most of the time on the objectives of the cruise or deployment.

First of all, the FluoroProbe and Multiexciter were tested as a profilers during the JERICO-NEXT plankton workshop in Gothenburg (27 – 30 Sept. 2016). One of the objectives was to use our instrument to detect a potential bloom of dinoflagellates (and more specifically Dinophysis sp.), a harmful algae. The profile acquired during the workshop showed a vertical gradient of the estimated biomass with a maximum at 3.5 meters depth (Figure 3.3.5) with a dominance of the diatoms-like group. The detection of the diatoms-like group could also be interpreted by the presence of dinoflagellates (which were abundant according to microscopic observations) because of their similar pigmentary composition. This technique has the main advantage to give a quick estimate of the proportion of different pigmentary groups and their contribution to total chlorophyll-*a* fluorescence (proxy of phytoplankton biomass). Additional comparison of the two spectral fluorometer model need to be carried out.



t. 2017*.*

Figure 3.3.5. Profiles at the Tångesund station measured with FluoroProbe (left) and MultiExciter (Right); The diatoms-like group is clearly the dominant group in the water column with a maximum at the sub-surface (3.5m depth). Differences in the concentrations and abundance of the different groups are partly due to different fingerprints used.

In 2017, four cruises were sampled by Fluoroprobe profiles to characterize phytoplankton distribution along the water column (Table 3.3.1.) For each profile, settings were adjusted to a LED flash each 10 seconds. So the number of acquisitions per profile depended on the speed of the winch profile and also of the depth. The profiles revealed heterogeneity along the water column but also between studied areas, both on total fluorescence (A.U.) and concerning the relative contribution of each pigmentary group to total chlorophyll-*a* estimation, when using the company's fingerprints (Figure 3.3.6.). These results must be compared to automated flow cytometry estimations as well as to image analysis and microscopic observations. The preliminary observations for the Baltic Sea indicate that groups "Cryptophyta" and "Cyanophyta" are not clearly separated. Moreover, haptophytes fingerprint (data not shown) revealed that *Phaeocystis globosa* was the dominant group in the Eastern Channel during PHYCO and VLIZ cruises in 2017.

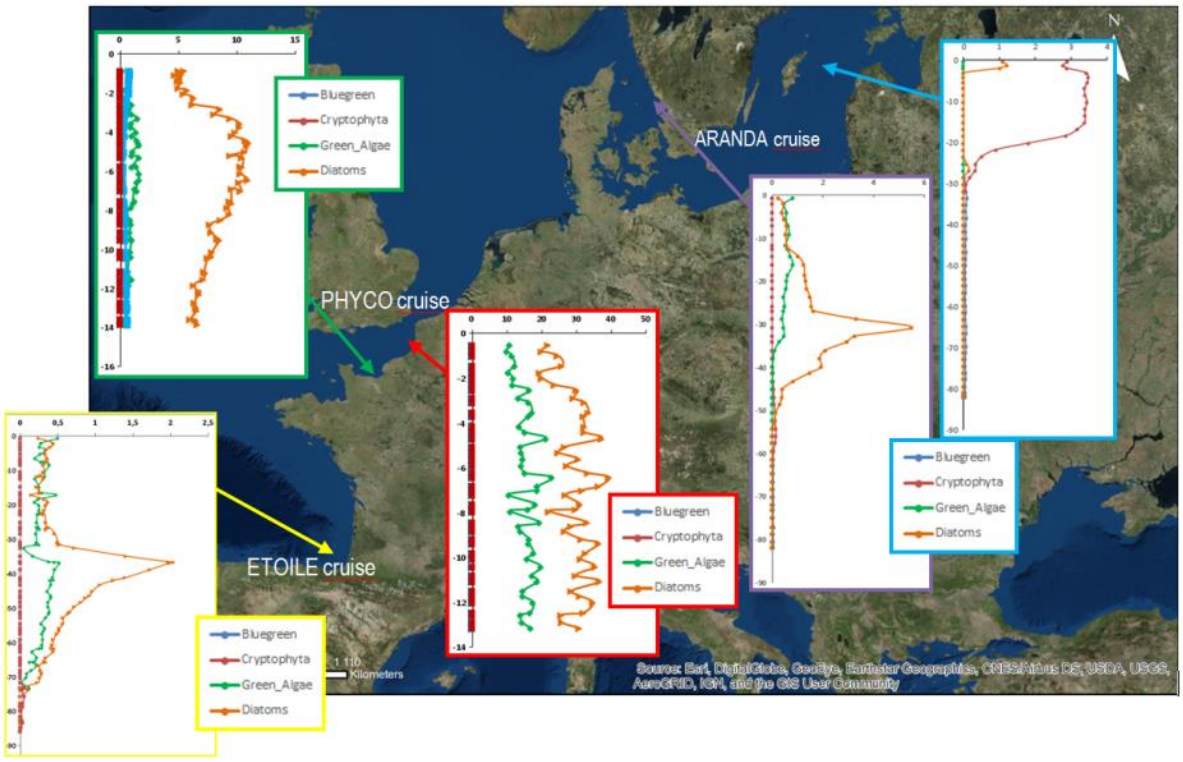


Figure 3.3.6. Some FluoroProbe profiles gathered during JERICO-NEXT cruises.

The examples of FluoroProbe in continuous recording systems, in the Baltic Sea, North Sea and Eastern channel are shown in figures 3.3.7.-3.3.9. During the Baltic Sea cruise, high spatial variability at sub-mesoscale was recorded and PE-dominant and PC-dominant Cyanobacteria were recorded in the Baltic, whereas near the straits "diatoms" and "green algae" were alternatively dominant (Figure 3.3.7.). For English Channel and South North

Sea, in mid-spring, *Phaeocystis globosa* is still the principal contributor of the phytoplankton biomass and this is reflected by the use of specific fingerprints. Diatoms are the second contributor. Within the INTERREG IVA "2 Mers Seas Zeen" DYMAPHY project (www.dymaphy.eu, Lizon et al. 2015), two fingerprints of common species/group of the eastern English Channel at the period of the spring bloom were recorded: *Phaeocytis globosa* (Houliez et al., 2012) and *Thallassiosira sp.*, making it possible to discriminate both Diatoms from Haptophytes. During the PHYCO cruise, a portable flow-through automatic measuring system called Pocket Ferry Box (4H-Jena) was implemented in the eastern English Channel during the period of the bloom of the prymnesiophyceae *Phaeocystis globosa* along the French coast. The RV « Côtes de la Manche » sails at 6 knots during the cruise, parameters were recorded at a frequency of once per minute so the spatial resolution was about 185 m. Because of a relatively brief transit time of water from the water intake to the PFB, the observations are representative of sub-surface conditions. The system used during this project was assembled with sensors for salinity and temperature (AANDERA conductivity sensor 3919), pH (Meinsberg MV 3010), oxygen concentration (AANDERA oxygen optode 3835), CDOM (AANDERA Cyclops 7 sensor) and spectral groups of phytoplancton using an Algae Online Analyser (AOA) (bbe, Moldaenke). Unfortunately, bad quality of pH and CDOM data did not allow data processing as planned. The PFB was coupled with a multiple-fixed-wavelength spectral fluorometer, the AOA, which continuously measures the chlorophyll fluorescence of microalgae in real time (results expressed in eq. µg Chl-*a* l-1). The chlorophyll-a content of green algae, blue-green algae, cryptophyceae and brown algae (i.e. diatoms, dinoflagellates) (in AOA parlance) was then assessed. First results are presented on figure 3.3.9.. Further investigations considering low resolution reference data and complementary parameters are needed to interpret data (action in progress).

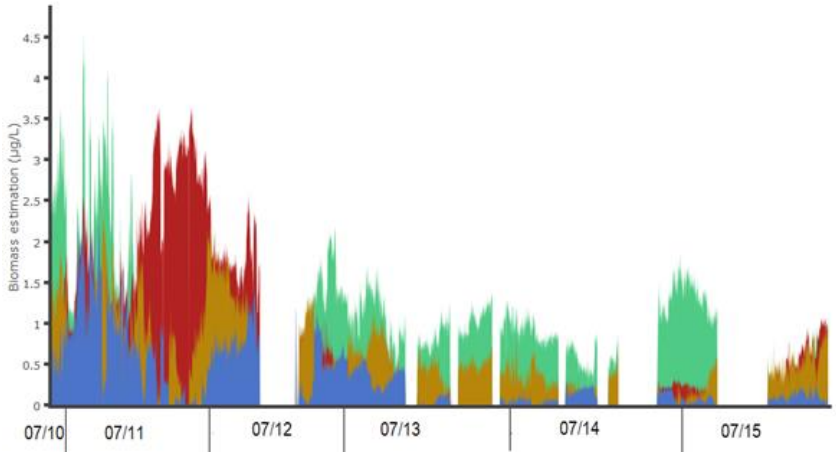


Figure 3.3.7. Spatio-temporal variability of phytoplankton pigmentary groups addressed by the FluoroProbe continuous recording in surface waters during the ARANDA cruise (Baltic Sea, Kattegat and Skagerrak, 10 – 15 July 2017).

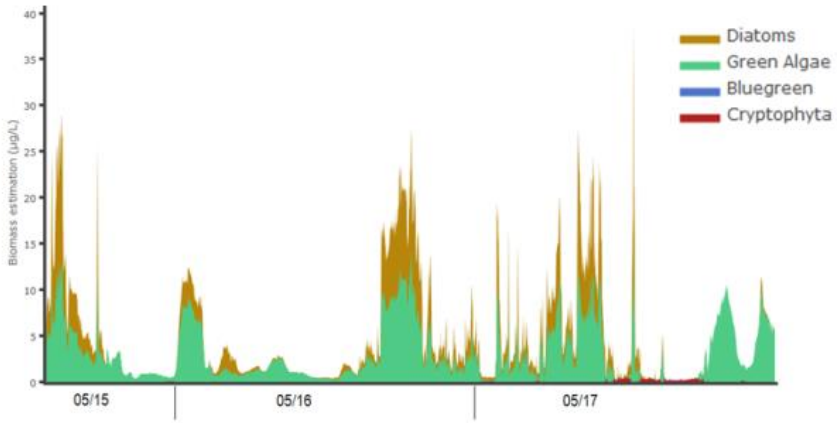


Figure 3.3.8. Spatio-temporal variability of phytoplankton pigmentary groups addressed by the FluoroProbe continuous recording in sub-surface waters during part of the RWS cruise (North Sea, 15 – 18 May, 2017). Top: Manufacturer fingerprints are used here for a comparison with the ARANDA cruise (Figure 3.3.7). Bottom: CNRS-LOG fingerprints are used for analyses, showing the high abundace of P. globosa.

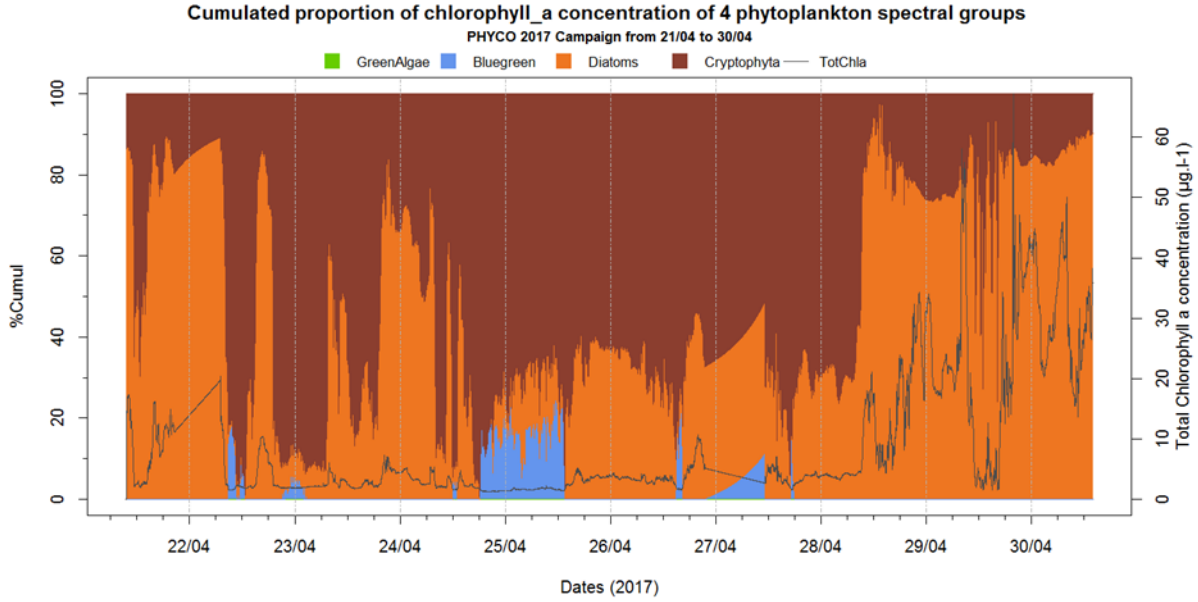


Figure 3.3.9. Evolution of the total chlorophyll-a concentration (black line, in eq. µg Chl-a l-1) and of the relative proportion of the spectral groups as assessed by the Algae Online Analyzer from the 21th to the 30th of April 2017 in the eastern English Channel.

Automated data analysis for spectral fluorescence

Work was initated 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 within the JERICO-NEXT project. 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 laboratory 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.

Developing algorithms for spectral fluorescence

Different methods to analyse spectral fluorescence data have been experimented using data collected with multiwavelength fluorometer FluoroProbe (bbe Moldaenke) during a cruise in 2014 in Benguela current (Figure 3.3.10). Similar analysis will be carried out for spectral fluorescence data collected during JERICO-NEXT WP4, JRAP#1 activities in Utö, Baltic Sea, once the campaign is over in spring 2018.
Fluoroprobe measures chlorophyll-*a* excitation using 5 LEDs. Primary output of the instrument is fluorescence intensities in relative units. Taxonomically different phytoplankton groups have different accessory pigments (chlorophylls, carotenoids, phycobilins), and thus differences in fluorescence spectra (fingerprints). Such fingerprints may be determined using phytoplankton cultures representing each taxonomic group (so called factory calibration). Typically, for a given study area, 3-4 pigment groups can be identified, including Green algae, brown algae (Diatoms + Dinoflagellates), Cryptophytes and Cyanophytes with different colours. Coloured dissolved organic matter (CDOM) need to be included as separate group if present in concentrations influencing background fluorescence. Typical application of spectral fluorometry is decomposing sample spectra into spectral components, using fingerprint spectra, and thereby estimating the concentrations of different spectral end-members. As FluoroProbe has 5 wavelengths, mathematically it is possible to solve max. 5 compounds. The key problem of the methods is that obtained fingerprints are not stable, but vary between species and physiological conditions. Thus factory-build (or any) fingerprints are not fully representative for diverse phytoplankton communities we have in nature. In this task, we look also alternative ways of analysing the spectral fluorescence data.
Vertical fluorescence profiles were collected at 34 stations along Namibian coast. An example of FluoroProbe data obtained in the Benguela Current dataset is shown in Figure 3.3.10. Fluorescence is analysed using factory calibration, i.e. using fluorescence spectra of algae cultures normalized by chlorophyll-a concentrations and as measured by manufacturer, resulting in a profile of chlorophyll-a concentration in different algae groups and sum of all groups. The profile is useful in assisting sampling different, vertically separated, phytoplankton communities and seeing the overall changes in biomass. The taxonomic information and chlorophyll-*a* concentrations need to be treated with care due to following reasons. 1) The fingerprints do not necessarily match those of the species present, like in the case of our dataset the fingerprint for cyanobacteria is measured for phycocyanin rich strain, while in the oceanic waters different phycoerythrin-rich cyanobacteria-forms dominate. 2) chlorophyll-*a* estimation relies directly on fingerprint spectra, i.e. has a fixed fluorescence to chlorophyll-a ratio. In real life, this ratio is known to be highly variable. Partly to overcome these issues, factory calibration may be overridden by calibration with local species.

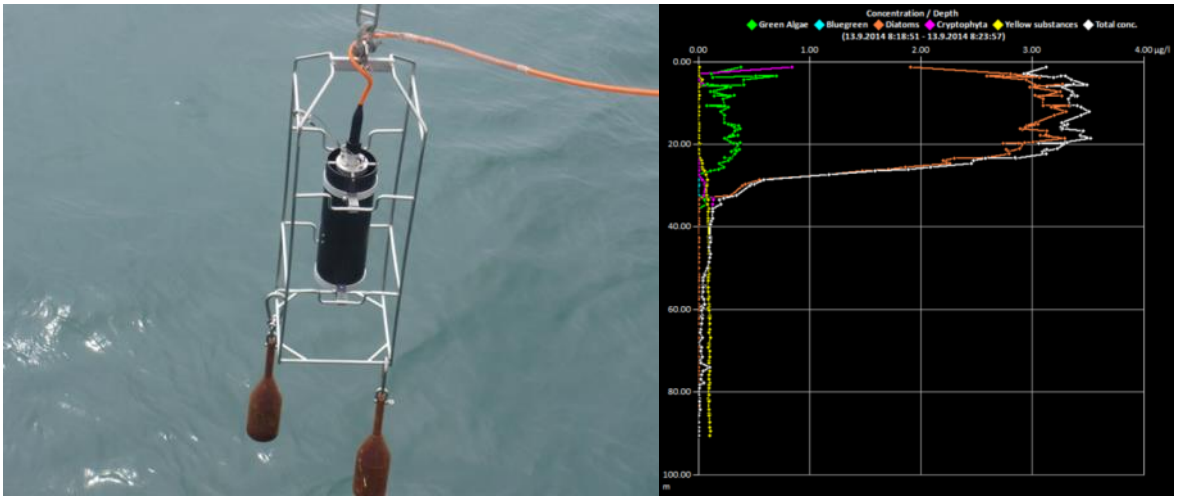


Figure 3.3.10. Left, FluoroProbe to be deployed. Right, example of profile showing concentration of different algae groups as derived by instrument software and factory calibration.

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

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

where
SFS = sample spectra at 6 wavelengths
[Chl-a]= total Chl-a concentration of sample
C_x=concentration of algal class x (µg 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 3.3.11.), 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 3.3.11.). 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 3.3.11.). 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 3.3.11.). 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.

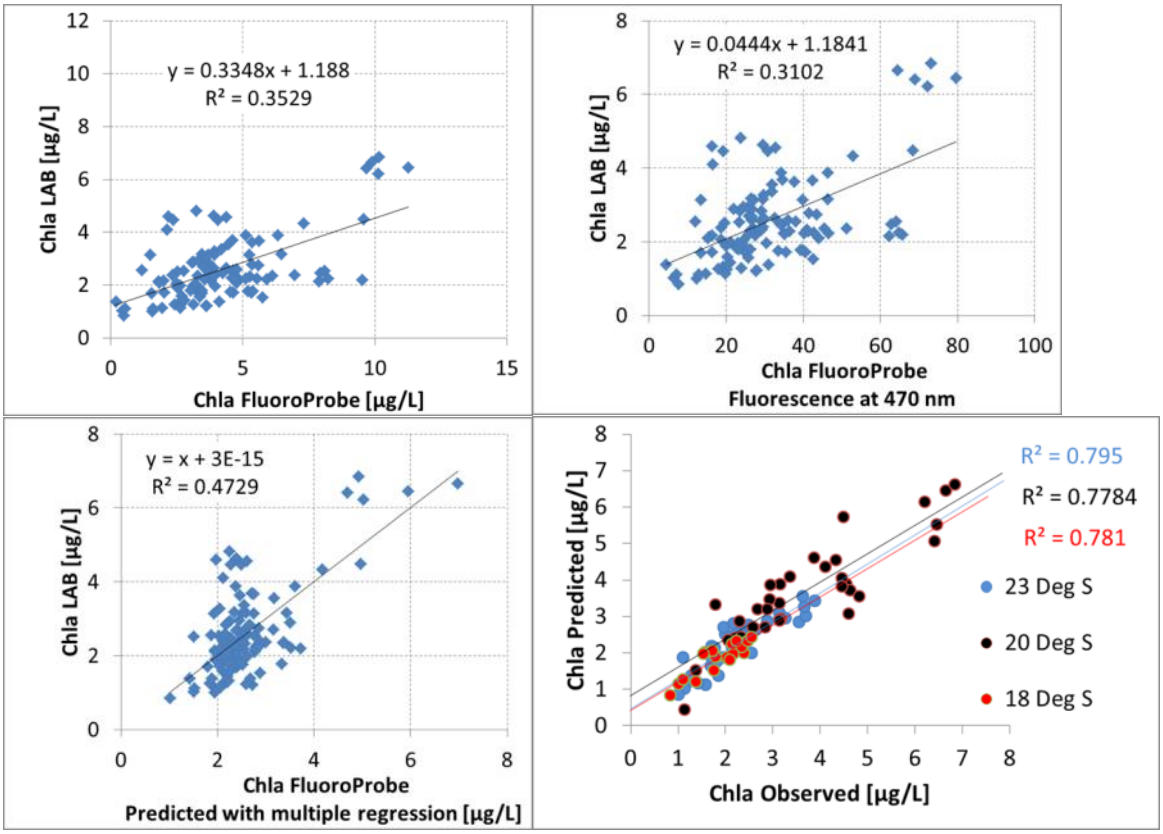


Figure 3.3.11. Estimation of chlorophyll-a concentration with FluoroProbe data. Upper-left: Default analysis using FluoroProbe software versus laboratory analyses, Upper-right: relationship between fluorescence at 470 nm waveband and chlorophyll concentration, Lower-left: relationship between chlorophyll 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.

Next, we analysed which are the major spectral components in the dataset, using principal component analysis (PCA). During the cruise, we collected altogether 3848 spectra. Three first components of PCA (PC1, PC2 and PC3) contributed 99.9 % of the total spectral variability (Figure 3.3.12). 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" phycoeryhtrin rich cyanobacterial group spectra in the Benguela system, unlike the fingerprint spectra showing phycocyanin.

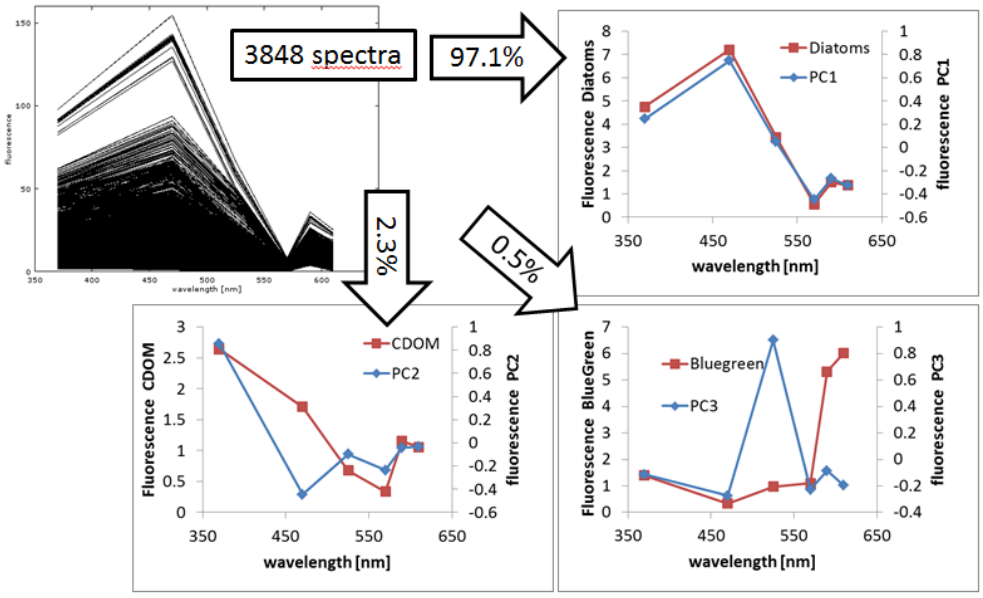


Figure 3.3.12. Principal component analyses of spectral data. Upper-left: All spectra measured with FluoroProbe, Upper-right: Principal Component 1, showing similarity with fingerprint spectra of diatoms, Lower-left: Principal Component 2, showing similarity with fingerprint spectra of CDOM, Lower-right: 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 3.3.13). 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. The method will be studied also further using other datasets emerging during JERICO-NEXT project.

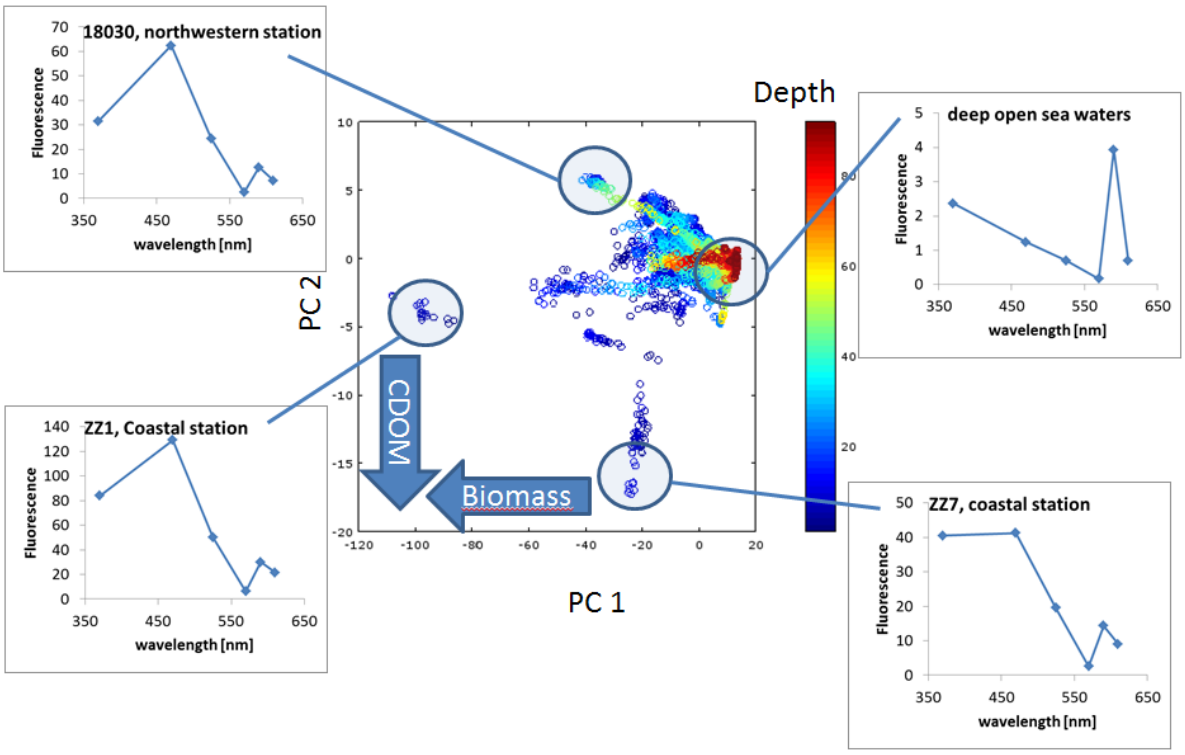


Figure 3.3.13. Relationship between principal components 1 and 2 of the principal component analysis of FluoroProbe spectral data. Different endmembers may be identified regarding the CDOM and algae biomass.

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 commercial 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.
Both optical sensors were employed and tested (Figure 3.3.14) during two cruises on board the research vessel "Heincke". The first cruise was performed in the German Bight area of the North Sea from 23.05.17 to 04.06.17

while the second cruise covered the coastal and open North Sea, the Norwegian Coast as well as two different fjord systems (Sognefjord and Trondheimfjord) from 08.07.2017 to 28.07.2017.

Figure 3.3.14: Sensor set-up during both cruises. From left to right: FerryBox connected to water inlet, HyAbS inside Zarges-box, Oscar G-2 inside black pressure housing

Cruise 1: Both integrating cavity instruments, the OSCAR (TriOS) and the HyAbS (HZG) were operated in flow-through mode and data were obtained continuously in time intervals of 20 and 5 seconds, respectively. The OSCAR was calibrated manually once a day using nigrosin dye solution. The HyAbS was running completely autonomous, performing solid standard calibrations automatically approximately every 60 min. Only the container with the reference solution (purified water) had to be refilled once a day.
As proposed during the planning of the validation, the data from OSCAR were stored internally and will be processed after the cruise. Also, the HyAbS data were stored within the instrument itself, but the biological relevant parameters (chl-a, total suspended matter and phytoplankton group information) that were derived from the calculated absorption coefficient spectra were in addition successfully transferred to the FerryBox. From there, these data together with the other FB data were transferred as 3 min averages in real-time to the FerryBox database at HZG via satellite communication.
To the time of this report, the data of the cruise are not yet evaluated. This includes both, the data of the integrating cavity instruments itself as well as the data going to be used for comparison (discrete absorption coefficient measurements, chl-a determination by HPLC, gravimetrically determined total suspended matter concentrations, and microscopic data). Thus, this report includes only observations made on the non-quality controlled data directly transferred to the database as well as personal observations. As visible in Figure 3.3.15, the absorption-based estimation of chl-a was in qualitative accordance with fluorescence-based measurements and oxygen measurements from sensors mounted in the FerryBox, at least in areas with high biomass.
In lower biomass areas, the absorption-based chl-a concentrations were occasionally negative. This was the result of an unexpected error in the HyAbS software, which took effect especially at low concentrations of absorbing material. However, since the raw data are still available, a post-processing of the data is possible and will be performed.
The high biomass areas were according to preliminary microscopic investigation often dominated by the colony-forming haptophyte *Phaeocystis* sp.. At least in the core areas of the blooms, the phytoplankton identification algorithm implemented in the HyAbS software also estimated this phytoplankton group. Although a real validation of the performance of the algorithm requires a comparison with the microscopic data, these preliminary results can be considered as promising.

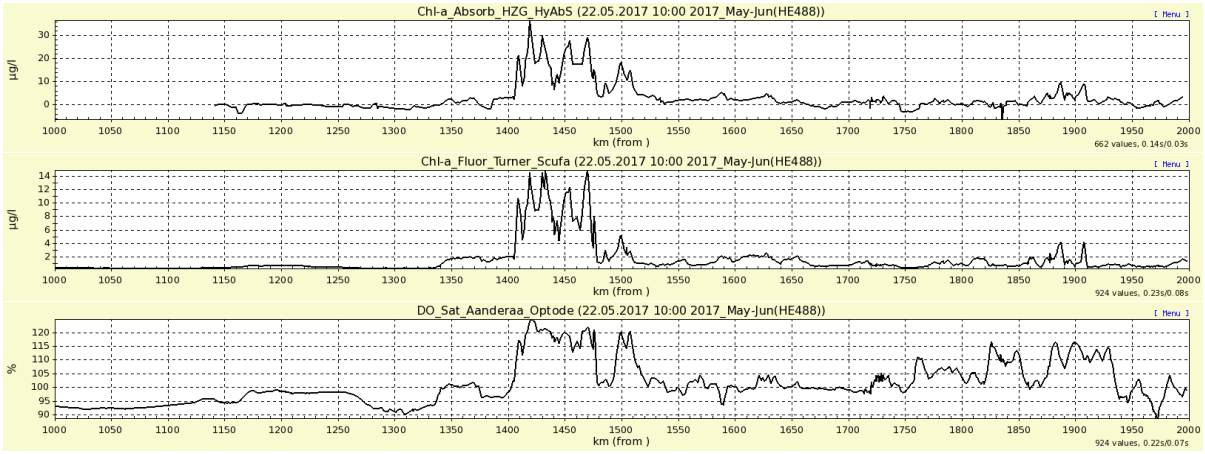


Figure 3.3.15. Chlorophyll-a estimated on the basis of absorption spectra measured by the HyAbS (upper panel) and chlorophyll-a fluorescence (mid panel). The lower panel shows the saturation of the water with oxygen, indicating also the presence of primary producer biomass.

Cruise2: During the second cruise a variety of different water masses were covered with different hydrodynamic conditions (Figure 3.3.16). For example, surface salinity values covered a range from 35 (open sea) to 2 (inner fjords) PSU. Despite these variable conditions, biomass amount and variability was generally low in the surface waters which were measured by the sensors. However, slight increases of biomass were visible close to the Sognefjord and the Trondheimfjord.

Figure 3.3.16. Sampling area with cruise stations and idealized cruise track.

Despite software-related problems that caused occasional malfunctions of the HyAbS, data were obtained for nearly the whole cruise duration. Figure 3.3.17 shows the preliminary, not quality controlled data for the absorption-based chl-a and suspended matter estimates as well as the values for the similarity index which is an indicator for the quality of the results for the phytoplankton group identification. It is visible that the qualitative development of the chl-a estimates is comparable to those obtained by the FerryBox. The slight increase of surface biomass in the Sognefjord (approx. 11.07 – 18.07) is also covered as the larger one in the Trondheimfjord (approx. 23.07). Regarding the suspended matter concentration, it can be seen that a considerable percentage of values is negative thus obviously incorrect. The reason for this are spectra which shape is distorted due to the presence of small air bubbles in the cavity, either during sample or reference measurement. These distortions take place primarily in the red region of the spectrum, where also the proxy wavelength for the determination of the suspended matter

value is located. Although also the wavelengths used for chl-a estimation are located in this region, they are less affected, because the chl-a concentration is estimated from a ratio of two coefficients, not from an absolute value like the suspended matter concentrations. Potentially, the data can be corrected to a certain degree by removing corrupted reference measurements. Nevertheless, at least also the suspended matter data show the increase observed in the area of the Trondheimfjord towards the end of the cruise probably caused by an increased presence of phytoplankton.

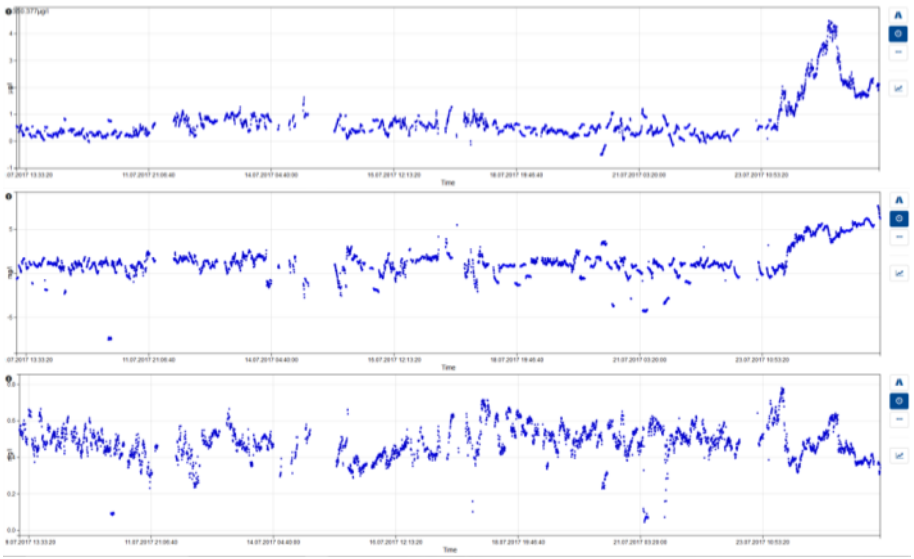


Figure 3.3.17. Data examples from the HyAbS. First panel: Absorption-based chl-a estimates. Second panel: Absorption-based suspended matter estimates. Third panel: Similarity index values for the phytoplankton group estimates.

The values for the similarity index resulting from the phytoplankton identification algorithm are generally relatively low (below 0.8). This means that no spectrum measured by the HyAbS during the cruise had a more than 80 % similarity to one of the spectra included in the reference database. Thus, the estimations can be considered of relatively limited reliability. However, it is important to note that this is not necessarily related to the quality of the data itself. Although of course the distortions mentioned above can have an effect, because the identification is based on the evaluation of the shape of the spectra, there are also other reasons. The most basic explanation is that the database simply does not contain the species or phytoplankton group that is present in the field. However, preliminary microscopic observations showed mainly the abundance of diatoms and dinoflagellates, and both groups are well represented within the database. A second explanation is that the amount of phytoplankton pigments in the sampled water is too low to create the characteristic features in the spectrum that are necessary for a reliable identification. Previous investigations suggest a tendency to lower similarity index values at chl-a concentrations below 3 µg/l (which is the case for most of the measurements made on this cruise). Although this might be the most likely explanation, a further evaluation of the data is necessary before drawing final conclusions.

Evaluation of variable fluorescence methods

The FastOcean sensor on an APD system has been used in 2017 in eight main cruises involving several laboratories of JERICO-NEXT (Table 3.3.1). From the eastern English Channel to the Baltic Sea, more than 90 vertical profiles have been done to measure photosynthesis in the water column between April and July. (Figures 3.3.18.- 3.3.19.).

Table 3.3.1. Variable and spectral fluorescence measurements during JERICO-NEXT research cruises; 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)
CNRS- Côte de la Manche	English Channel	April 20-30	X	37	46/48	X	47X
VLIZ-S Stevin	North Sea & E. Channel	May 8-12		19	25/25		27
RWS-Zirfea	North Sea	April 10-14 May 15-18 June 19-22 August 14-17		17	18/15	X	
CNRS-PELRAD	E. Channel, Strait of Dover	July 4	X	6		X	17
SMHI-Aranda	Baltic Sea, Skagerrak & Kattegat	July 10-17	X	16	19	X	16

The current mode of the FastOcean profiler deployment in joint sampling cruises was relatively time consuming. 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 probably 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 "quality" factor of measurement fits around 0.045. To obtain such values of quality factors, we must use the auto-gain function of the sensor photomultipliers (PMT of the two sensors) with a high start value of 550 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. 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 cruise 3-6 ^{13}C-uptake experiments were performed daily in order to measure the electron requirement for C-fixation. Most of these samples are still to be analysed.

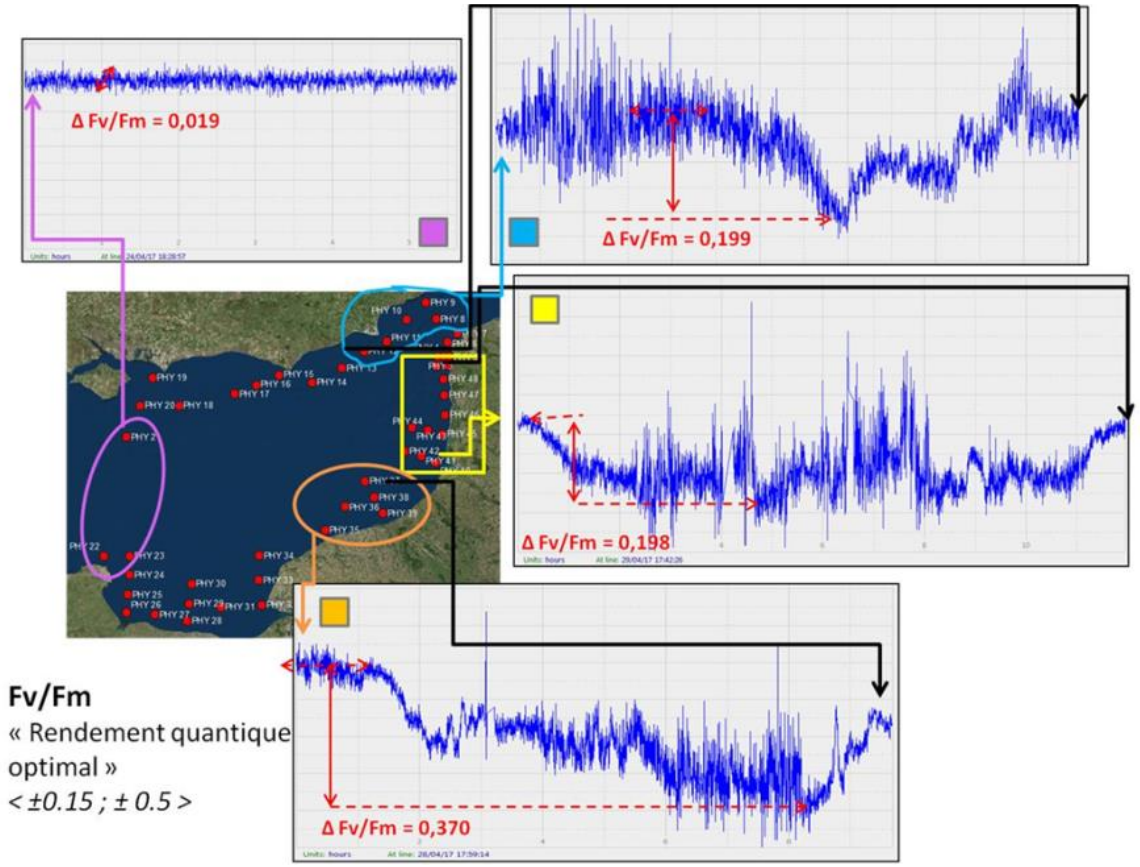


Figure 3.3.18. An example of horizontal (sub-surface continuous recording) and vertical (profiles) measurements of variable fluorescence and associated physiological parameters for the Phyco cruise in the eastern English Channel in April 2017. At this moment, the data files are in depth analysis in order to compare all the different data set of hydrological (salinity, temperature), physical (light, extinction coefficient...), biological (biomass, species…) and physiological (photochemical processes) parameters.

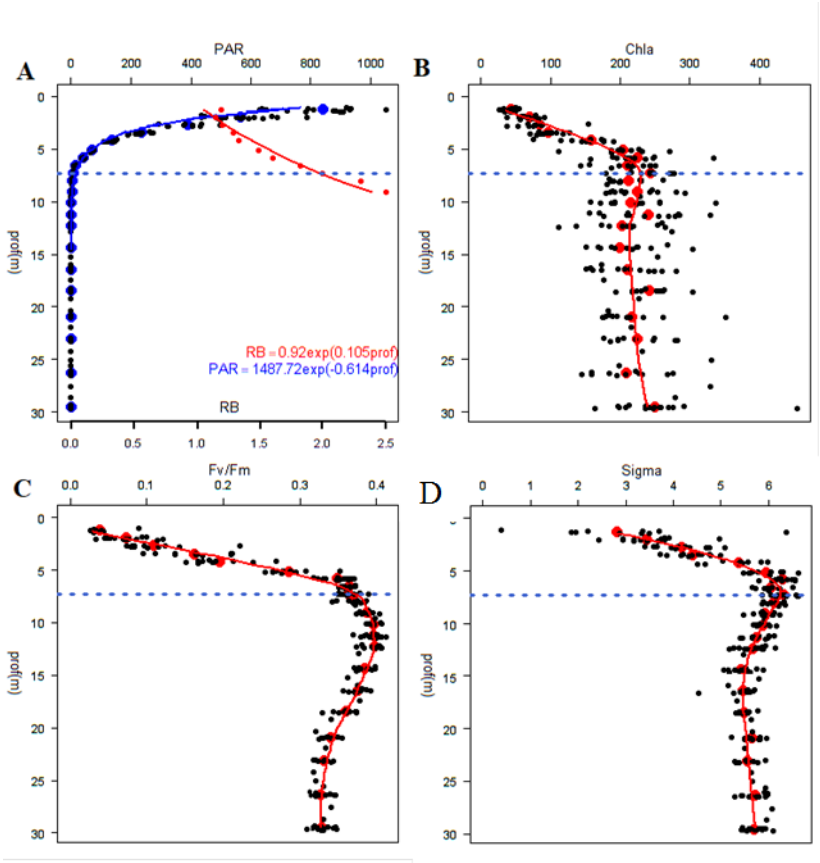


Figure 3.3.19 The vertical variation of F'v/F'm are in the same order of magnitude than horizontal variations (figure 3.3.18.) of this parameter, the maximum photochemical efficiency measured during the daylight period with the FRRF profiler and with the ambient light chamber (blue dashed line is the limit of the euphotic layer).

In Utö, 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 3.3.20.). FastOcean has been used continuously at Utö since April 2017.

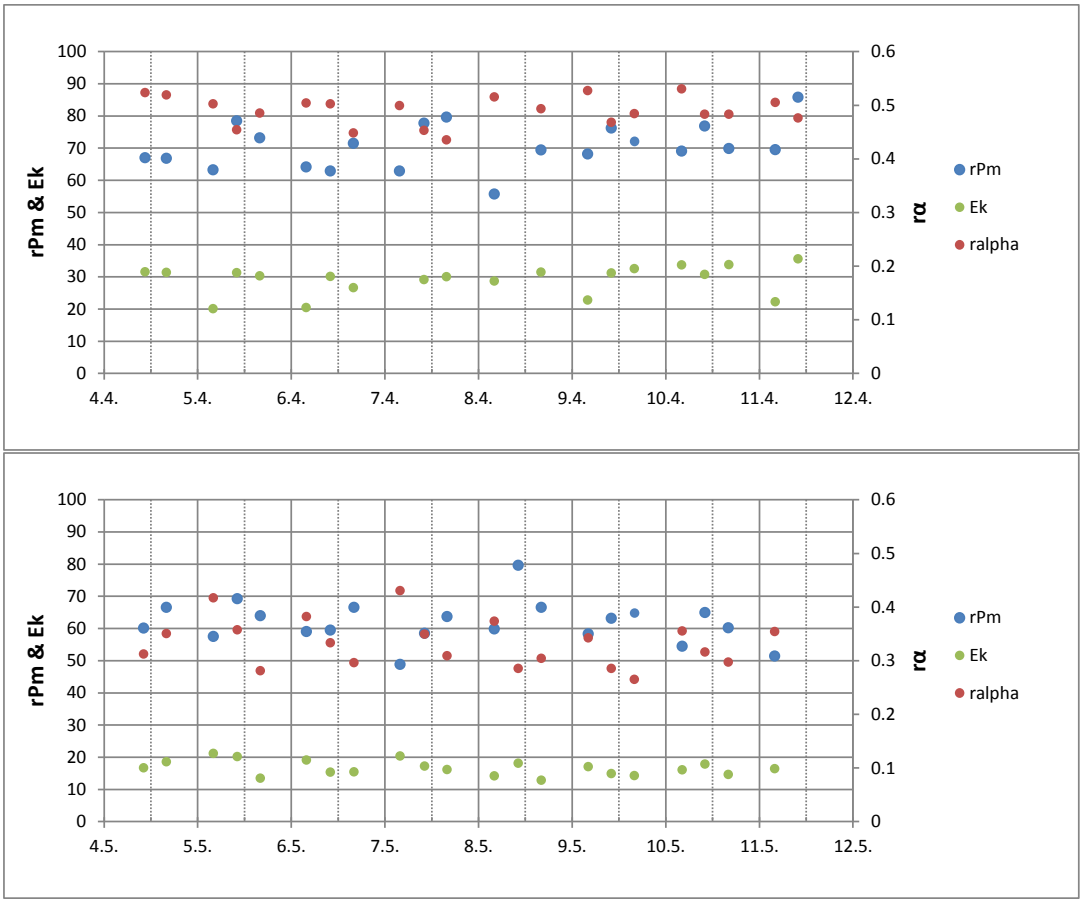


Figure 3.3.20: Example of the variability of photosynthetic parameters, maximum light utilization coefficient (ralpha), light saturation (Ek) and maximum production capacity (rPm), measured at Utö for one week periods in April and May 2017. For these periods, additional measurements of ^{14}C fixation exist and will be analysed later.

An instrumented buoy called SMILE was installed in the bay of Seine (English Channel - France) 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 3.3.21).

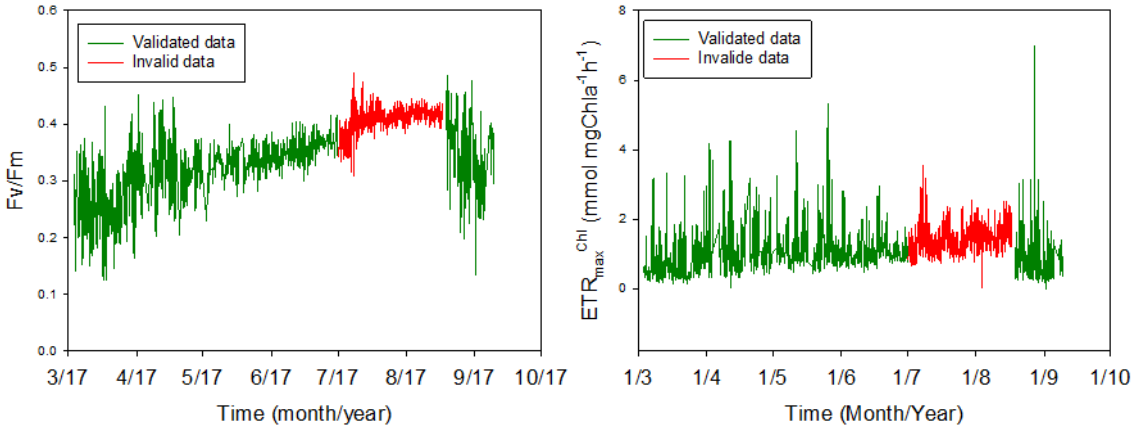


Fig. 3.3.21. Left: Fv/Fm – Maximum quantum yield of the PSII measured on natural assemblage of phytoplankton between March and September 2017 from the SMILE buoy. The data obtained from 1 July to 15 August 2017 were not validated because of biofouling development in the water jacket which regulates temperature. We observed and 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.

ETRmaxChl was determined by using the absorption algorithm proposed by Oxborough et al., 2012) after using the following equation of variable fluorescence data (Oxborough et al., 2012 ; Silsbe et al., 2015)

$$\frac{F_v{}'}{F_m{}'} = \left(\alpha \cdot \mathrm{E_K} \cdot \left[1 - e^{-E/\mathrm{E_K}}\right] - \beta \cdot \mathrm{E_{K\beta}} \cdot \left[1 - e^{-\{E-\mathrm{E_K}\}/\mathrm{E_{K\beta}}}\right]\right) \cdot E^{-1}$$

With α (photosynthetic efficiency), EK (light saturation parameter), et β (photoinhibition parameter), E (PAR in µmol photon m-2 s-1). P/E curves are performed every two hours even during the dark period. 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 3.3.22.)

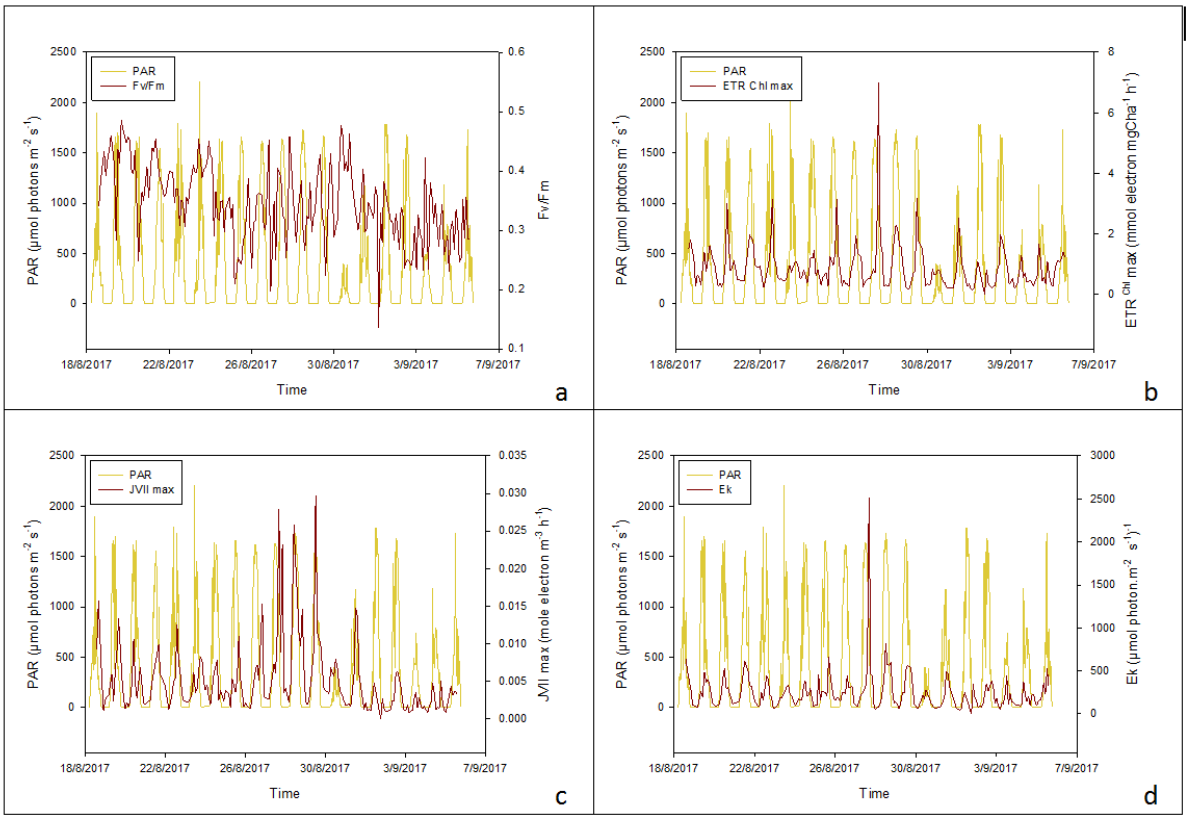


Figure 3.3.22. Examples of parameters estimated on the natural assemblage of phytoplankton by using FRRf measurements from the SMILE buoy. Fv/ Fm is Maximum quantum yield of the PSII, ETRmaxChl corresponds to the photosynthetic capacity, JVII corresponds to the maximal potential primary production, Ek corresponds to the light saturation parameter which gives to the phytoplankton photoacclimation capacities.

ETRmaxChl and JVIImax 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.

For all our JERICO-NEXT cruises, data collection single turnover (ST) fluorescence measurements were obtained with a classical acquisition protocol (100 saturation flashes, flash duration of 2 µs 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 µm 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'_0 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 FFRF production model (cf 3.3.1). During the SMHI 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
Fo, F	Minimal or steady state fluorescence	dimensionless
Fm, Fm'	Maximum fluorescence in the dark resp. light	dimensionless
σ_{PSII}	Functional absorption cross section	nm^2
Fv/Fm, Fq/Fm'	Maximum or effective quantum efficiency of PSII	dimensionless (interpreted as electron $photon^{-1}$
Chl-*a*	Concentration of chlorophyll-*a*, based on calibration factor for Fo	mg m^{-3}
E	Irradiance used in the light curve	µmol photons m^{-2} s^{-1}
τ	Rate constant for Qa reoxidation	msec
p	Connectivity between antenna complexes	dimensionless
JV_{PSII}	PSII flux (electron transport) per unit volume	Mol electrons m^{-3} d^{-1}
a_{LHII}	Volumetric PSII absorption coefficient	m^{-1}
[RCII]	Concentration of reaction centre II	nmol m^{-3}

Although the FRRF datafile gives all these parameters, they should not be used as the data has to be corrected for blank values first. The following section describes how to calculated the important parameters a_{LHII}, [RCII] and n_{PSII}.

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

$$a_{LHII}(m^{-1}) = \frac{Fm \times Fo}{Fm - Fo} \times \frac{K_R}{1 \times 10^6}$$

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 Chl-a m^{-3}) = m^2/mg Chl-a).

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

$$[RCII](mol\ m^{-3}) = \frac{F_0}{\sigma_{PSII} \times 10^{-18}} \times \frac{K_R}{10^6} \times \frac{1}{6.02 \times 10^{23}}$$

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):

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

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

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

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). The incorporation of dark into light data file must be made carefully some errors can occur if the timing of the two FastOcean sensors are not well synchronized. Such timing problems have been observed after blank measurements in laboratory using only the dark sensor connected to a PC in real time deployment mode. It is possible to correct the clock of the data file with the FastPro8 software or to contact the manufacturer. Some small decay (1 or 2 sec) can also occur sometimes between the two sensors at the starting of the instrument, without reason.
In order to compare vertical profiles and horizontal ranges of variation of variable fluorescence, a FastOcean Act2 system (lent by Cefas) was also deployed on some cruises (cf Table 1) and programed to measure at high frequency, every ten sec, classical fluorescence parameters as F'v/F'm and $\sigma_{PSII\text{-}470}$. It can be noticed that the blue excitation wavelength is here 470 nm with this old sensor (instead of 450nm with the new one). Seawater was pumped by the pump of the Act2 system using a clear tube that was darkened in the last centimeter before to enter in the sensor in order to have roughly the same dark acclimation of cells as the measurements made in situ by the FRRF profiler. A PhytoPam was also deployed on all of our cruises (cf Table 1) in order to compare another classical technique (but a multiple turnover technique) with the FRRF technique (single turnover).

The Figure 3.3.23 shows a comparison made from samples obtained during the VLIZ cruise in the Channel/Southern North Sea made with the FastOcean/Act system (using the water inlet from the ship which also fed the ferry box, and from samples taken with a Niskin bottle. In general the differences were limited, but in all three cases the alpha values from the Niskin bottles are slightly, whereas the E_k-values hardly differ. As a result, the $rETR_{max}$ values are the same in station 9, but for station 14 and 21 the $rETR_{max}$ of the Niskin is higher, especially in the case of station 21. A comparison between all stations should tell us if this difference is really significant or not. 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 (NPA) processes.

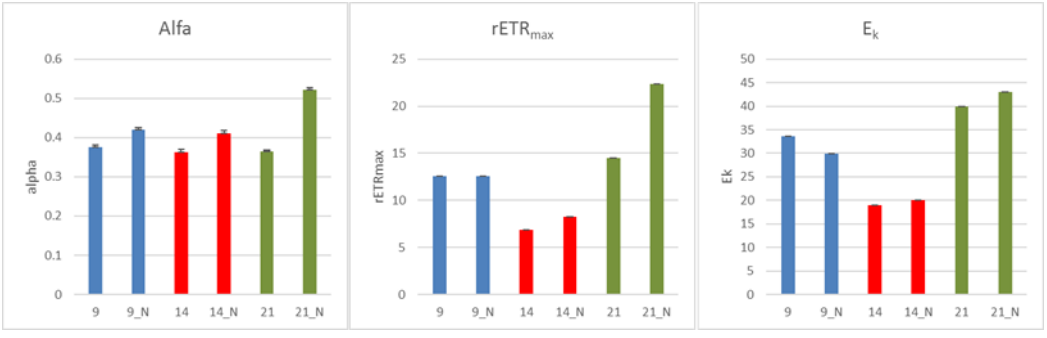


Figure 3.3.23. Results from fitted light curves from stations measured during the VLIZ cruise on the English Channel and the Southern North Sea. The numbers on the X-axis are the station numbers; the addition of _N indicates that are samples obtained from a Niskin bottle.

The figure 3.3.24. shows the result of 24 h continuous recording of fluorescence parameters measured in the first step of the light curve (dark). Up to station 12 the RV Simon-Stevin sailed along the French coast. The Channel crossing was between station 12 and 9. Station 9-21 were along the UK coast. The changes in Fo indicate that water bodies with different phytoplankton biomass were present. The decrease in Fq'/Fm' measured in the dark showed lower values around stations 21. This can be caused by photoinhibition during midday, but also by nutrient limitation. This needs further analysis. The results of the corresponding LC-parameters are shown in the figure 3.3.25.

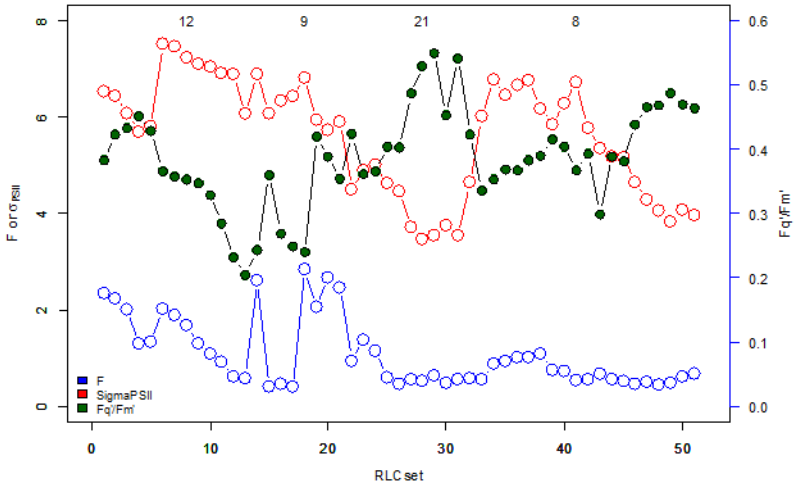


Figure 3.3.24. Fo, σ_{PSII} and Fq'/Fm' values during the 2nd day of leg 1 (09-05-2017). A number of stations is plotted at the top of the graph. RLC set means the number of the LC and corresponds to a time axis from 0:00 to 24:00 h.

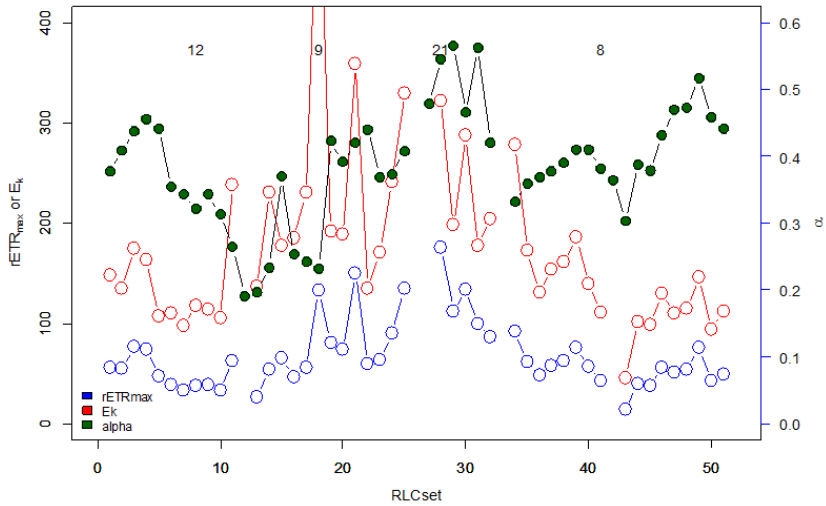


Figure 3.3.25. Data showing the LC parameters $rETR_{max}$ (maximum rate of relative electron transport), α (the initial slope of the light curve) and E_k, the light saturation parameter (µmol photons $m^{-2}s^{-1}$), i.e. the irradiance where photosynthesis switches from light limited photosynthesis to maximal rates of photosynthesis, caused by processes downstream of PSII (most likely the activity of the Calvin-Benson cycle).

Notice the diel pattern in rETRmax. It is low during the early and late hours of the day, and high during the middle of the day.
These preliminary results show that the continuous measurements of photosynthesis with the FastOcean/Act2 combination are capable of capturing spatial and temporal variability in photosynthetic activity and phytoplankton biomass. The data have to be converted to absolute rates of PSII electron transport in order to calculated rates of primary production. This also requires a comparison with the C-fixation rates, measured by the incorporation of 13C-labeled bicarbonate.
The figure 3.3.26. shows the maximum quantum efficiency of PSII measured during the 4 cruises on the North Sea this year. All the data were gathered during the 4 days cruises on the Dutch EEZ by combining the Jericho-cruises with the standard monitoring (MWTL) cruises of Rijkswaterstaat with the ship the Zirfaea. Lowest Fv/Fm 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 Fv/Fm is an indicator of P-limitation. As not all stations showed low phosphate concentrations, these stations are likely having a higher Fv/Fm, hence the spread in the data.

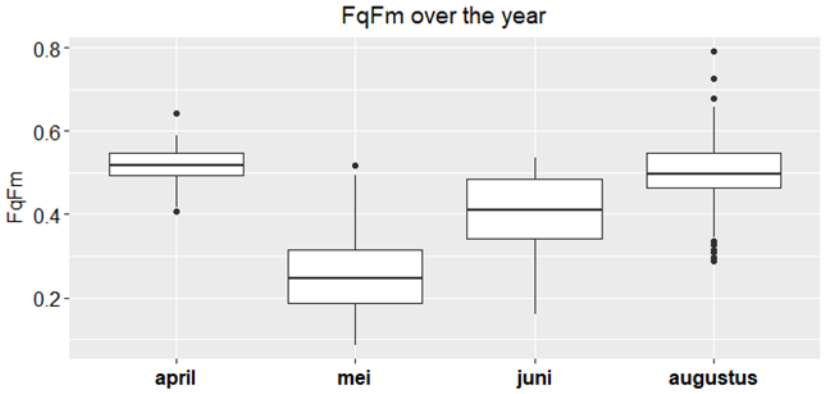


Figure 3.3.26. Box-wisker plots of maximum PSII efficiencies (Fv/Fm = Fq'/Fm') measured in the dark step during the continuous LC-measurements) in the North Sea during cruises in April (april), May (mei), June (juni) and August (augustus).

The preliminary results show the power of continuous FRRF measurements. Spatial variability and biomass, photosynthetic parameter 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 (Fo and Fm 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 Fo. 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 Fv/Fm 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 Fv/Fm 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 Fv/Fm does not exceed 0.6.
A comparison between FRRF parameters and flow cytometric analysis still need to be carried out. It is expected that phytoplankton composition will influence some of the FRRF parameters. Further analysis should demonstrate this.

3.3.3. *Conclusions and recommendations*

Photoquenching 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 provide 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 build during the DYMAPHY project to follow *Phaeocystis globosa* blooms in the eastern Channel and North Sea.
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 automatization 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 done in terms of software development and the technical set-up, but overall the above shown preliminary results can be considered as promising.
Preliminary conclusions on variable fluorescence so far are that automated FRRF is possible for continuous measurements (both on sub-surface as well as profiling the water column), and can also be applied on ships-of-opportunity. It is strongly recommended to combine FRRF measurements with Ferrybox measurements (as well as with other optical bulk or sing-cell approaches). Careful blank corrections are necessary, especially in the more oligotrophic waters. Variable fluorescence data collection and analyses have been rather well established recently, but still the different variants need to be compared. Different studies analysing the reasons for conversion factors, from electron transport rate to carbon fixation, need still to be carried out.
JERICO-NEXT WP3 has been collecting large amounts of data for evaluation of different methods. Next tasks, during the latter half of the project, should be in depth analyses of these data-set, comparing methods and algorithms, and finally providing set of recommendations for the use of different methodologies as well as providing a series research questions not resolved yet and notes for technical improvements.

4. Conclusions and future work

Main conclusions of these two years of work within 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.
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, 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 also useful for discriminating between phytoplankton groups based on pigment content.
7. Variable 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.

Most field work has ended but some will continue, e.g. at the Utö observatory in the Baltic Sea and at fixed platforms in France for both multi-spectral fluorometry, variable fluorometry and automated flow cytometry. The data collected during months 1-24 and the new data will be processed further and used for improving the discrimination of phytoplankton taxa or functional groups by inter-comparison of techniques and continued algorithm development, as well as for preparing JERICO-NEXT delivery 3.2. In addition, scientific publication of results is in progress or being planned.
A special issue in the open access journal Diversity (MPI) is being discussed. Some results and strategy will be presented during a symposium in Hannover, Germany, in October 2017 and during the FerryBox workshop on board the ship Colour Fantasy later in October 2017. Results will also be presented during the third JERICO-NEXT plankton workshop to be arranged in Marseille in March 2018, during the International Conference on Harmful Algae in Nantes in October 2018, and in other meetings to be determined.

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 Cen 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., Cornille 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., Vaulot 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, Vaulot 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

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

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 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.

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.

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, *Limnology And Oceanography-methods*, vol. 2, pp. 316-332

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.