



## GlobalHAB symposium on automated in situ observations of plankton Kristineberg Marine Research Station, Fiskebäckskil, Sweden

August 22-27, 2022  
Session 7



# Automated classification/recognition of phytoplankton from optical features, pulse shapes and images

Luis Felipe Artigas<sup>1</sup>, Guillaume Wacquet<sup>1,2</sup>, Alexandre Epinoux<sup>1</sup>, Arnaud Louchart<sup>1</sup>, Clémentine Gallot<sup>1</sup>, Zéline Hubert<sup>1</sup>, Florine Verghaege<sup>1</sup>, Manon Ryckman<sup>1</sup>, Pierre-Alexandre Hébert<sup>3</sup>, Emilie Poisson-Caillault<sup>3</sup>, P. Grosjean<sup>4</sup>, Alain Lefebvre<sup>2</sup>, Thomas Rutten<sup>5</sup>

<sup>1</sup>Laboratory of Oceanology and Geosciences  
CNRS – UMR 8187 LOG – ULCO – Wimereux (FR)

<sup>2</sup>LER/BL IFREMER – Boulogne sur Mer (FR)

<sup>3</sup>LISIC-ULCO – Calais (FR)

<sup>4</sup>U. Mons – Mons (BE)

<sup>5</sup>Thomas Rutten Projects – Middelburg (NL)

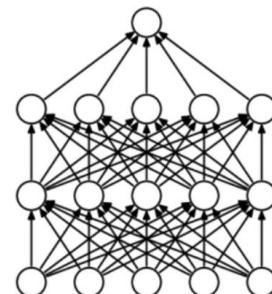


# Combination of automated data acquisition techniques and machine learning methods for the discrimination and recognition of phytoplankton

Expert knowledge



Machine learning



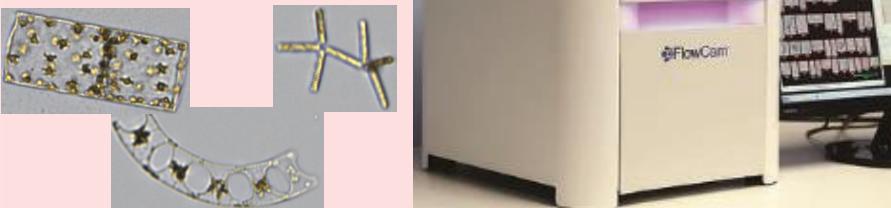
- Improving automation of phytoplankton data analysis from optical/imaging sensors
- Adapting training sets to previously-characterized phytoplankton communities in the studied area (**“adaptive learning”**)
- Guiding classification steps by introducing some pairwise constraints (**“constrained clustering”**)
- Partially validating predictions obtained by automated methods (**“error correction”**)
- Calibrating predictive models for estimating the number of cells in colonies (**“cell counting”**)

# Data acquisition

## Imaging in flow devises



*FlowCam VS Series*



*FlowCam 8100*

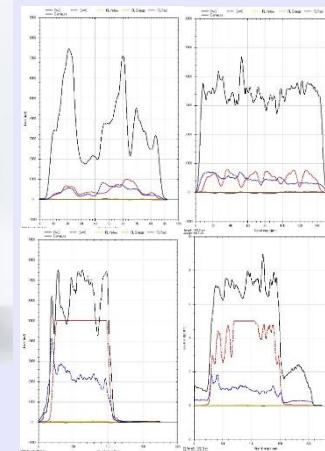
- Images (grayscale level or color)
- Measurement table (25/image)
  - 20 morphological
  - 5 textures (intensity)
- R package « **Zoolimage** »

## Pulse shape-recording flow cytometers



*CytoSense*

*CytoSub*

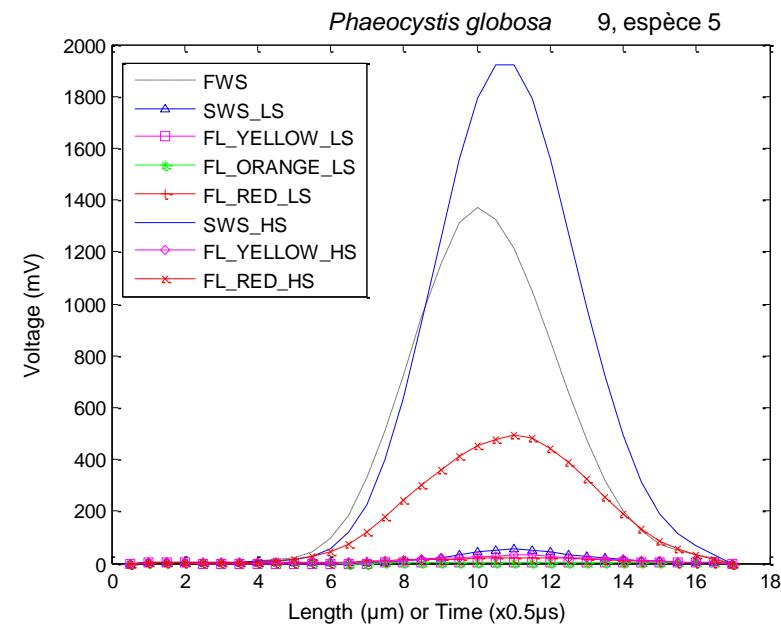


- Signals
  - 3 for scattering
  - 3 for fluorescence
- Measurement table (11 features/signal)
  - 33 for scattering
  - 33 for fluorescence
- Profiles & Images
- R package « **RClusTool** »

# Phytoplankton species recognition: an automated classification problem

## Cytometric pulse-shapes

- 8 raw time signals per cell
  - identical experimental conditions  
(same sampling rates, same detection threshold, etc.)
- one signal on forward scatter (FWS), corresponding to the cell length;  
-two signals on sideward scatter (SWS), corresponding to the internal structure;  
-two signals on red fluorescence (FLR) which characterize chlorophyll pigments;  
-one signal on orange fluorescence (FLO), ages, specific pigments;  
-two signals on yellow fluorescence (FLY), specific pigments



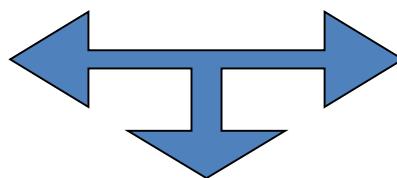
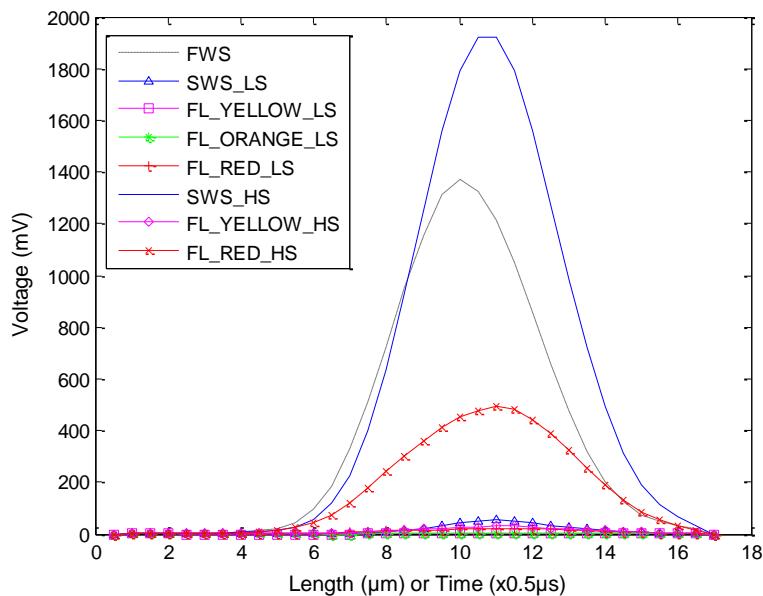
microscopic photo

Caillault et al. 2009 a,b

Ph.D. Thesis of Guillaume  
Waquet- 2011

# Phytoplankton species recognition: an automated Classification problem

## Cell characterization from their cytometric optical pulse shape



**Which one ?**

### Known profiles:

*Chaetoceros socialis*  
*Emiliania Huxleyi*  
*Lauderia annulata*  
*Leptocylindrus minimus*  
*Phaeocystis globosa*  
*Skeletonema costatum*  
*Thalassiosira rotula*.

### Purpose :

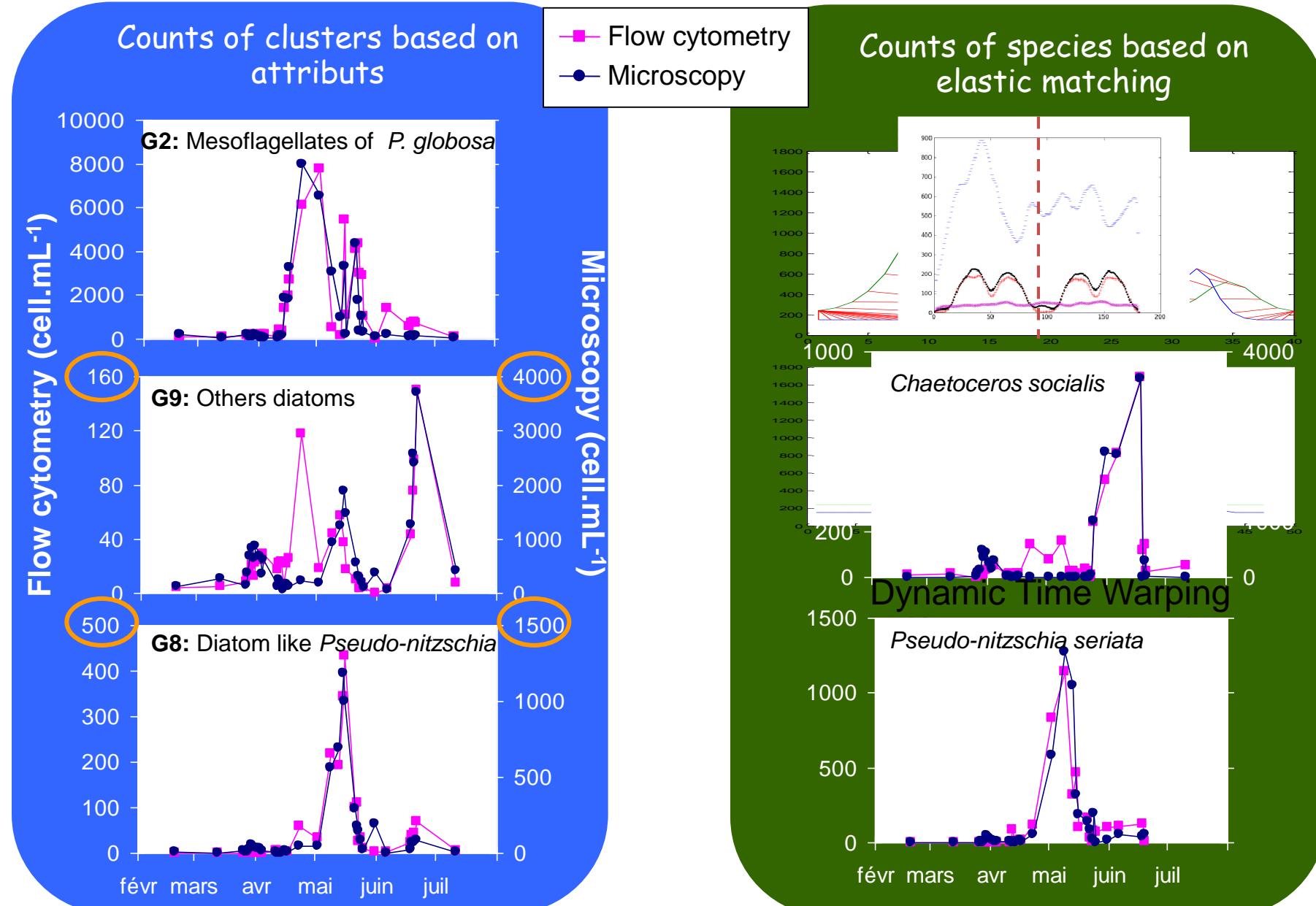
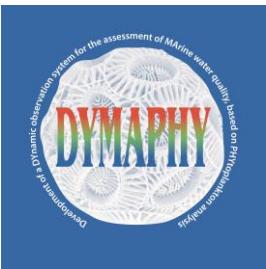
- Make the species recognition process automatic.
- Build a comparison measure between cytometric profiles robust to:
  - the intra-species variability,
  - the sensor sensibility.

### 2 methods:

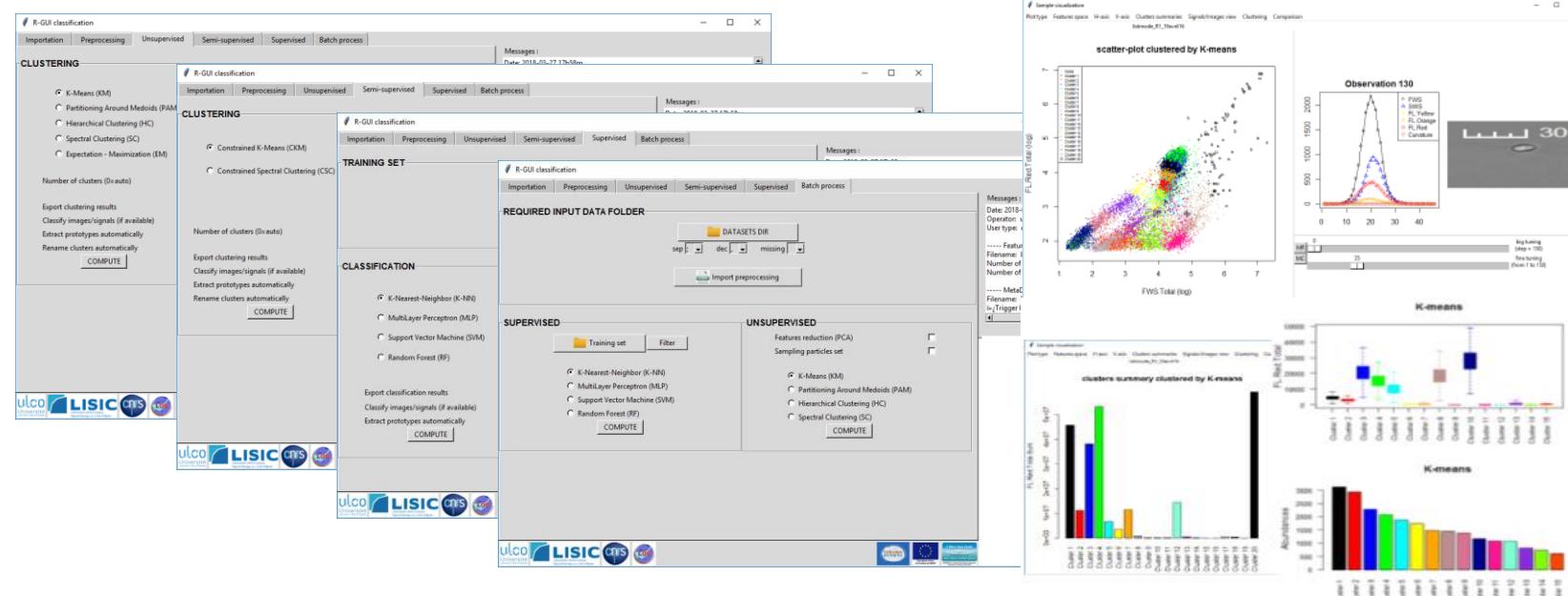
- Features
- Signals

*Caillault et al. 2009 a,b*

# PSFM flow cytometry vs microscopy : manual and automated methods



# R package « RClusTool »\*



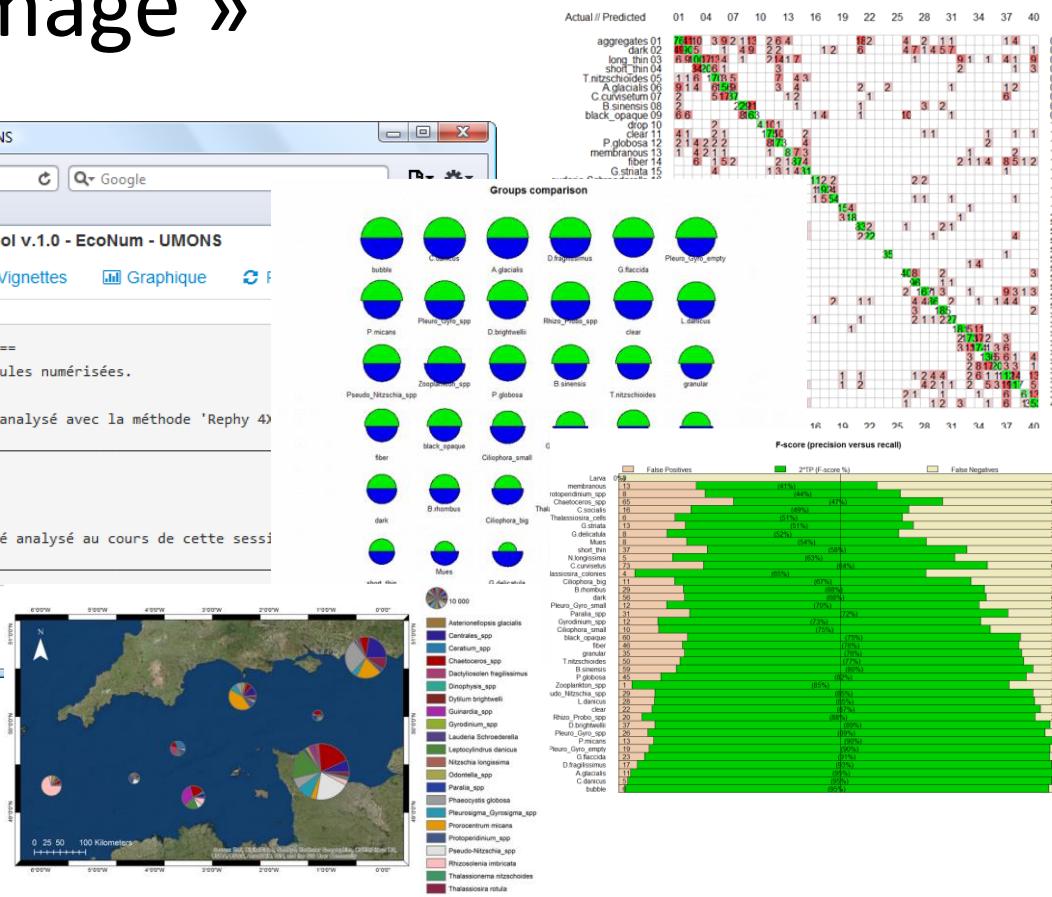
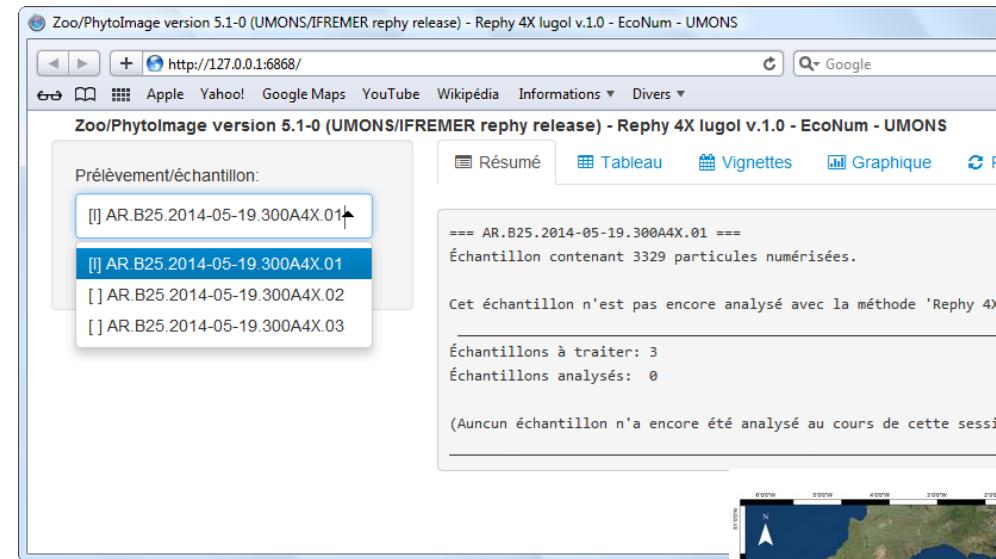
- ✓ Free software (open source), written in "R", and specialized in classification of **any kinds** of observations (with feature tables, signals, images, etc.).
- ✓ **Supervised, unsupervised and semi-supervised** (with pairwise constraints) classification.

Download: <http://mawenzi.univ-littoral.fr/RclusTool/>

Wacquet, G., Hébert, P.-A., Poisson-Caillault, E., Talon, P., 2020. RclusTool. URL <https://cran.r-project.org/web/packages/RclusTool/RclusTool.pdf>

\* Wacquet *et al.* « RclusTool: An R-based Graphical User Interface to explore and classify data interactively (case study of flow cytometry) ». In prep.

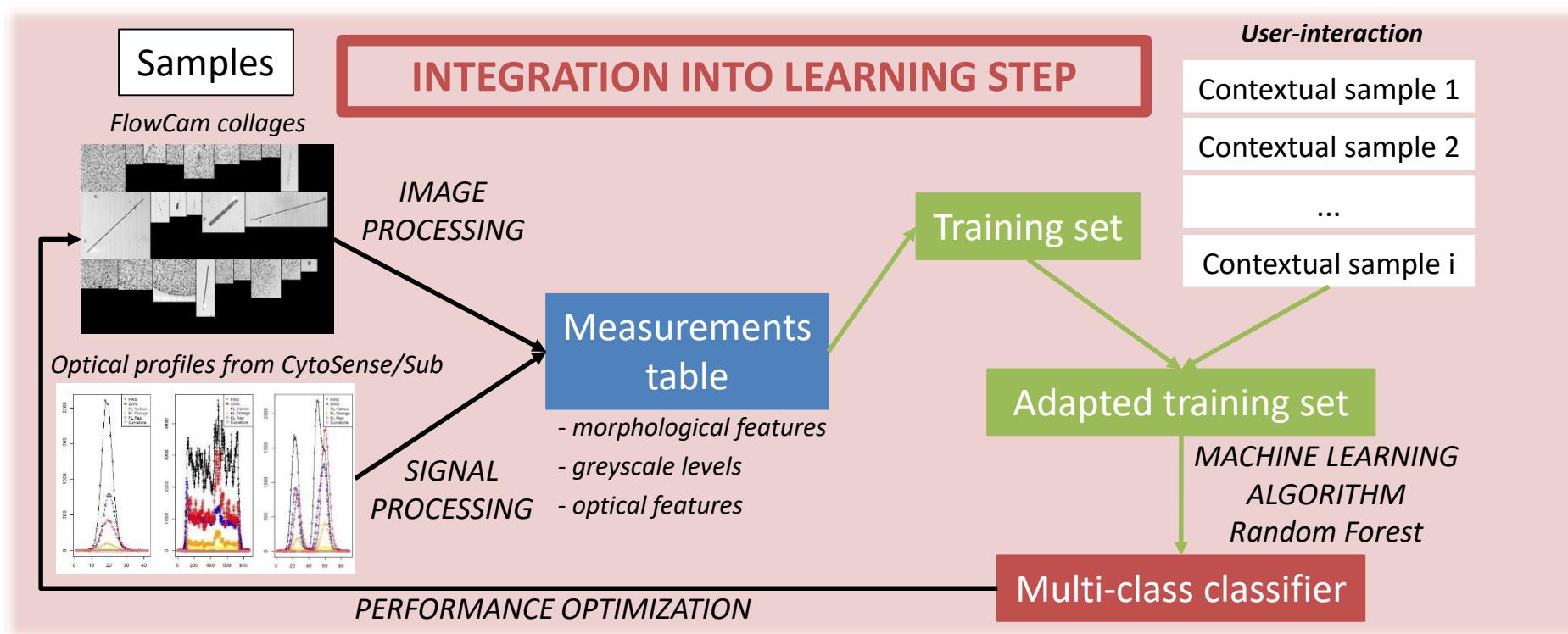
# R package « Zoolimage »



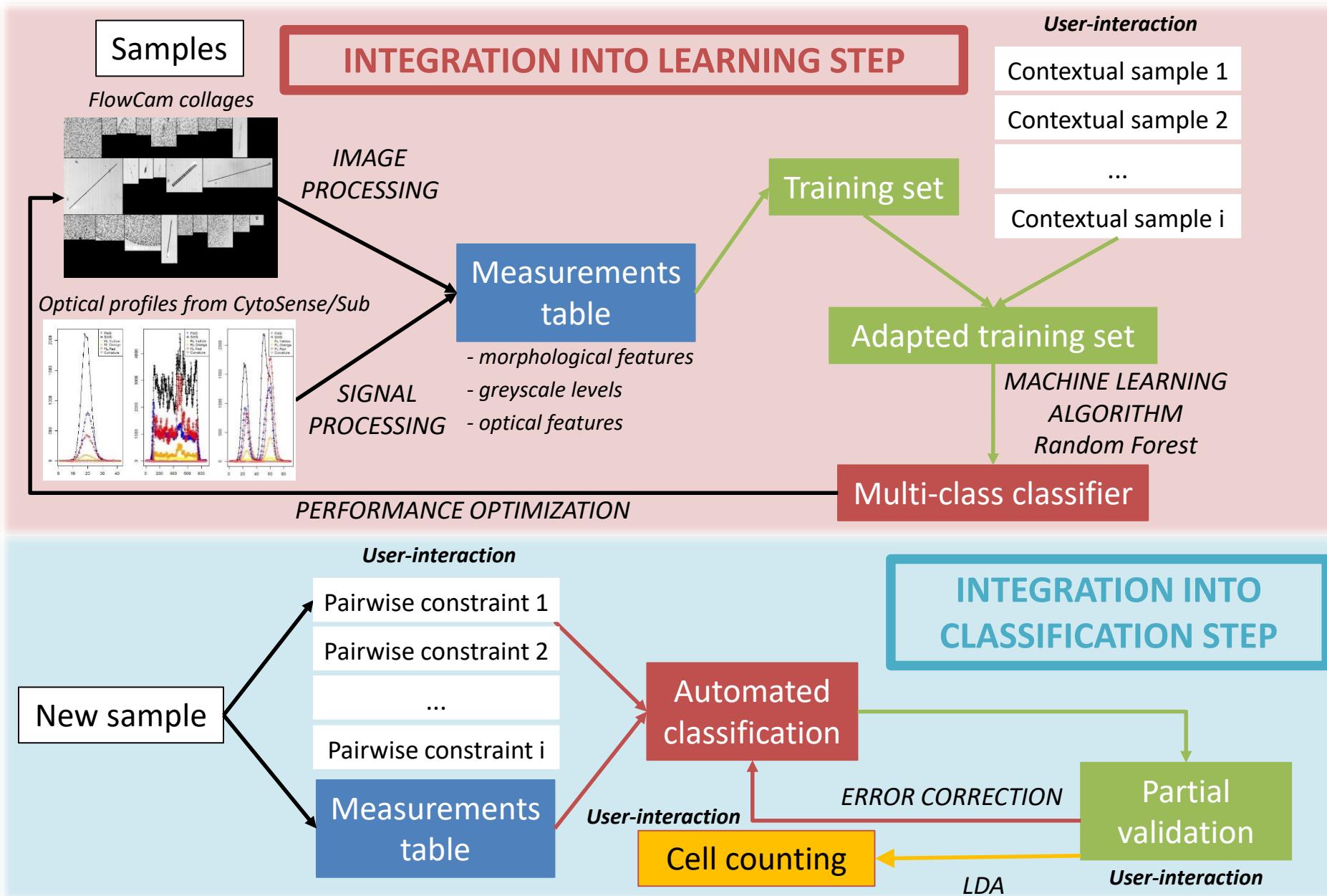
- ✓ Free software (open source), written in R, and specialized in the classification of digital images of zoo- or phytoplankton.
- ✓ For any kinds of plankton images, i.e. from FlowCam, Imaging FlowCytoBot, ZooScan, micro- or macrophotography, ...

Download: <https://cran.r-project.org/web/packages/zoolimage/>

# Classification process



# Classification process

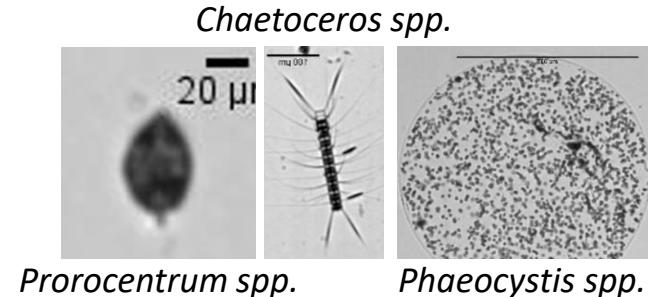
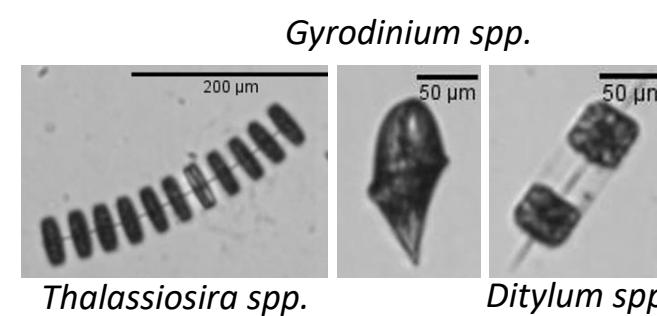


# FlowCam

**Device:** FlowCAM (B2 series)

**Magnification:** 4X

**Chamber depth:** 300µm



Actual // Predicted	01	04	07	10	13	16	19	22	25	28	31	34	37	40	
aggregates 01	76	110	3 9 2 113	2 6 4				182	4	2	11		14	1	01
dark 02	49	5	1	4 9	2 2			6	7						02
long_thin 03	6	9	0 17 134	1	2 14 17										03
short_thin 04			3426 1		3										04
T.nitzschioides 05	1 1 6	1 7 3 5			7 4 3										05
A.glaialis 06	9 1 4	6 3 9		3	4										06
C.curvisetum 07	2	5 1 7 9		1 2				2 1	2						07
B.sinensis 08	2	2 1 2 1		1				1							08
black_ opaque 09	6 6	8 6 3		1 4				1	10	3 2		1			09
drop 10		2		4 10 1											10
clear 11		2 1		1 7 10	2										11
P.globosa 12		2 1 4 2 2 2		8 7 3	4										12
membranous 13		1	4 2 1 1	1	8 7 3										13
fiber 14		6	1 5 2		2 1 3 4										14
G.striata 15		4		1 3 1 4 3 1											15
auderia Schroederella 16					112 2				2 2						16
bubble 17					10 4										17
Centrales spp 18		1		2											18
Ceratium spp 19		1			1										19
N.longissima 20		2	4 3					154		1 1	1	1			20
Ciliophora big 21					1 2 1			3 18							21
granular 22					1			32	2 2						22
D.acuminata 23					3			22							23
C.danicus 24								35							24
C.fusus 25															25
fecal_pellets 26															26
P.micans 27									40 8	2	14				27
G.flaccida 28									16 7 1 3	1	9 3 1 3				28
Gyrodinium spp 29									2 4 4 3 2	1	1 1 4 4				29
T.rotula 29									3 3 1 8 5						30
Protoperdinium spp 31									2 1 1 2 7						31
<sup>a</sup> pseudo-Nitzschia spp 32										18 5 11					32
L.danicus 33										2 1 7 1 3	3				33
R.imbricata 34										3 1 5 1 3	6 1	4			34
D.fragilissimus 35										3 2 8 1 2 0	3 3				35
G.delicatula 36										3 1 5 1 3	6 1	4			36
Chaetoceros spp 37	1 2 3	1 1 1 6	2	11	1	1	1	1	1 2 4 4	2 6 1 1 1 4	1 3				37
D.brightwelli 38	1 2 5 3	4	5 5	3 1	2	1	2	1 1	4 2 1 1	5 3 1 9 1 7	5				38
Paralia spp 39	2								2 1	1 2 3	1 6	6 1 3			39
Pleuro_Gyro spp 40	1 1 1 1		1 1	1 1					1	1 2 1 3	1 6	6 1 3 5			40

## Training set

**Training set:** REPHY network (Ifremer)  
DYPHYRAD transect (CNRS LOG)

**Number of groups:** 28 phyto + 12 noise

**Number of vignettes:** 5154 (~125/group)

**Classifier:** Random Forest (ntree = 500)

**Cross-validation:** 10 folds

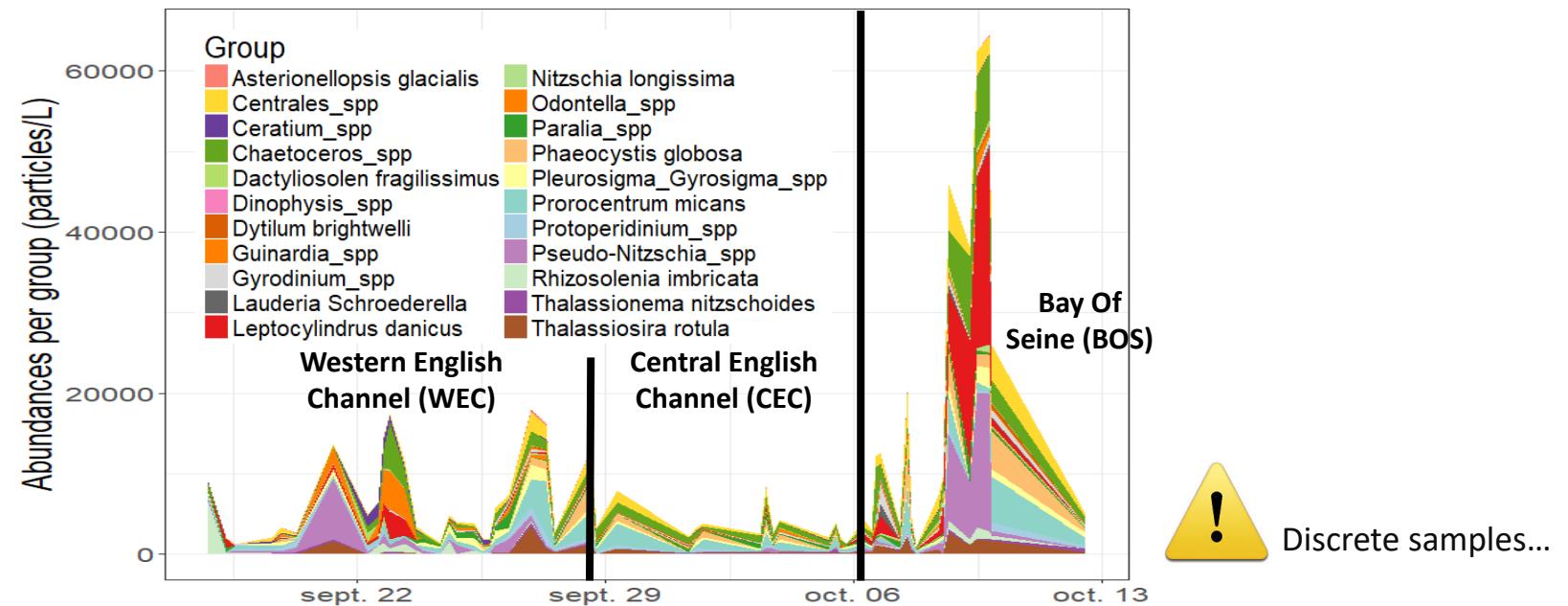
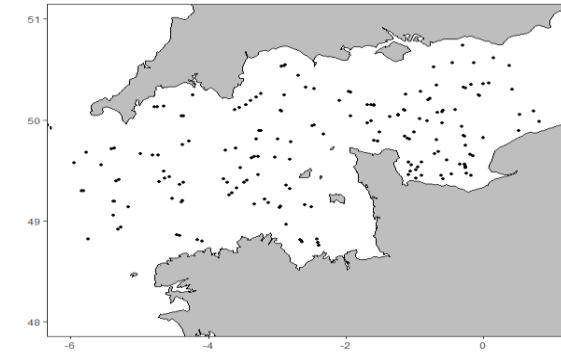
# FlowCam - CAMANOC cruise

Multidisciplinary campaign: an ecosystem approach to fisheries.

**Period:** 16th September-12th October, 2014 - English Channel

**Ship:** R/V « Thalassa II » - IFREMER

**Adaptive training set:** 1 sample selected every 2 days



**Wacquet et al.** « Combination of machine learning methodologies and imaging-in-flow systems for the automated detection of Harmful Algae ». 2020. Proceedings of the 18th International Conference on Harmful Algae, 21-26 October 2018, Nantes, France.

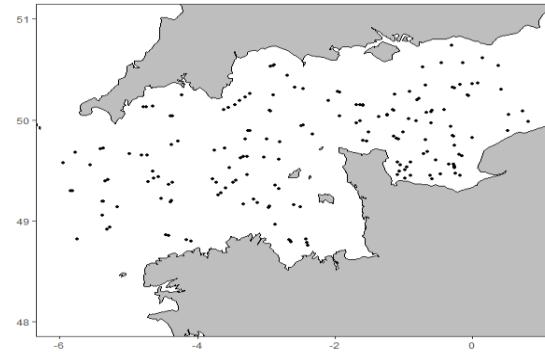
# FlowCam - CAMANOC cruise

Multidisciplinary campaign: an ecosystem approach to fisheries

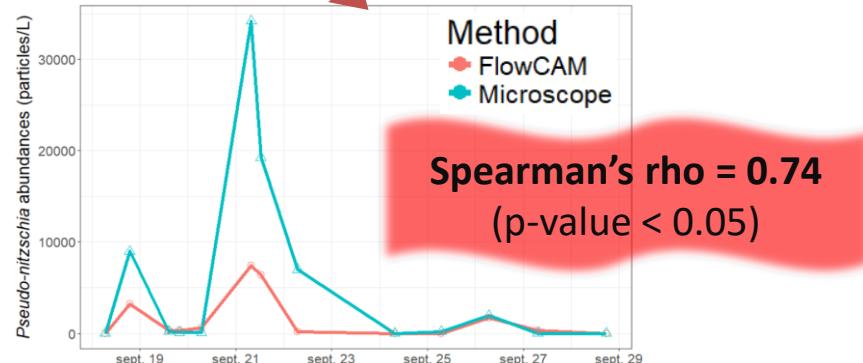
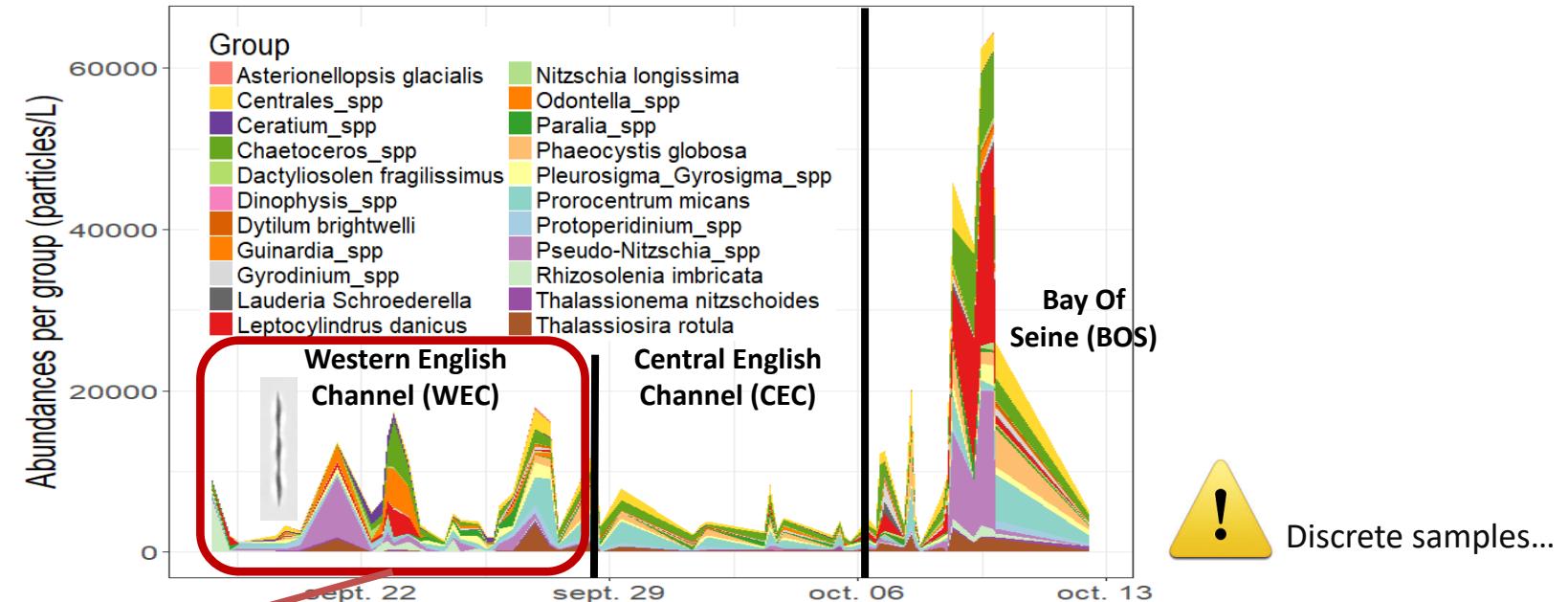
**Period:** 16th September-12th October, 2014 in Channel

**Ship:** R/V « Thalassa II » - IFREMER

**Adaptive training set:** 1 sample selected every 2 days



**Wacquet et al.** « Combination of machine learning methodologies and imaging-in-flow systems for the automated detection of Harmful Algae ». 2020. Proceedings of the 18th International Conference on Harmful Algae, 21-26 October 2018, Nantes, France.



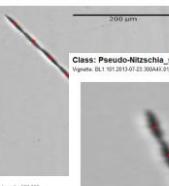
**Overall dynamics** of abundance obtained by microscopy and FlowCAM for *Pseudo-Nitzschia* genus are **similar**:

→ Abundance estimation for FlowCAM lower than for Microscope.

# FlowCam - Counting cells in colony\*

\* Wacquet *et al.* « Combination of machine learning methodologies and imaging-in-flow systems for the automated detection of Harmful Algae ». 2020. Proceedings of the 18th International Conference on Harmful Algae, 21-26 October 2018, Nantes, France.

Class: *Pseudo-Nitzschia* spp  
Vignette: 06.01.2013 14:26\_368461\_11\_347



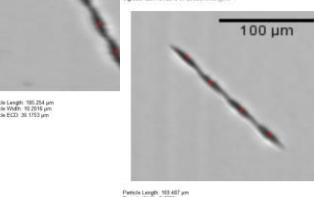
Particle Length: 301.061 μm  
Particle Width: 40.264 μm  
Particle EDD: 75.153 μm

Class: *Pseudo-Nitzschia* spp  
Vignette: 06.1.101.2013 01:25\_368461\_21\_347



Particle Length: 180.254 μm  
Particle Width: 40.264 μm  
Particle EDD: 70.152 μm

Class: *Pseudo-Nitzschia* spp  
Vignette: 06.1.101.2013 01:25\_368461\_21\_347

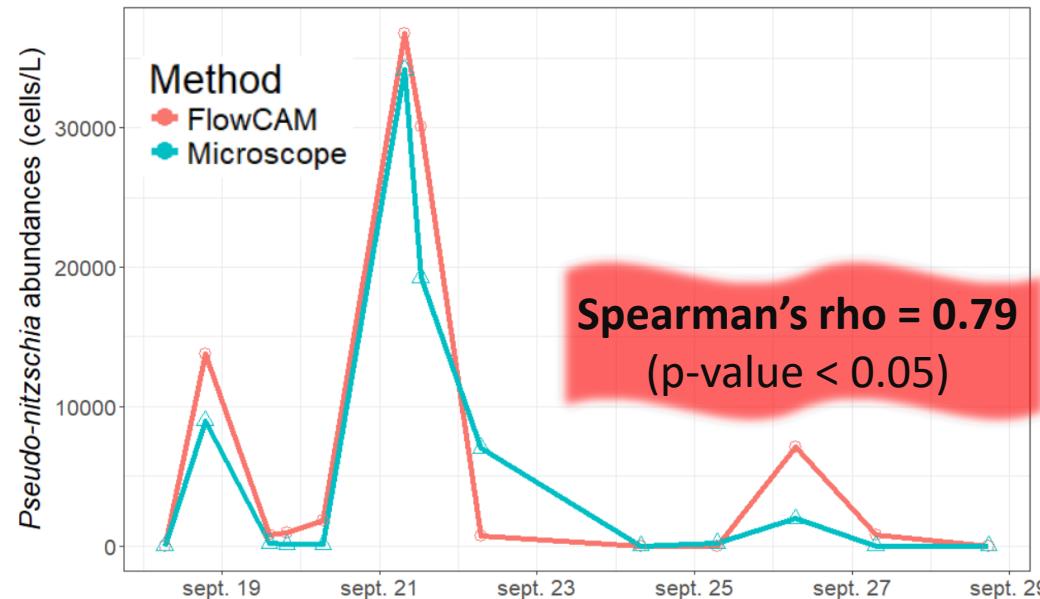


Particle Length: 180.467 μm  
Particle Width: 40.264 μm  
Particle EDD: 74.151 μm

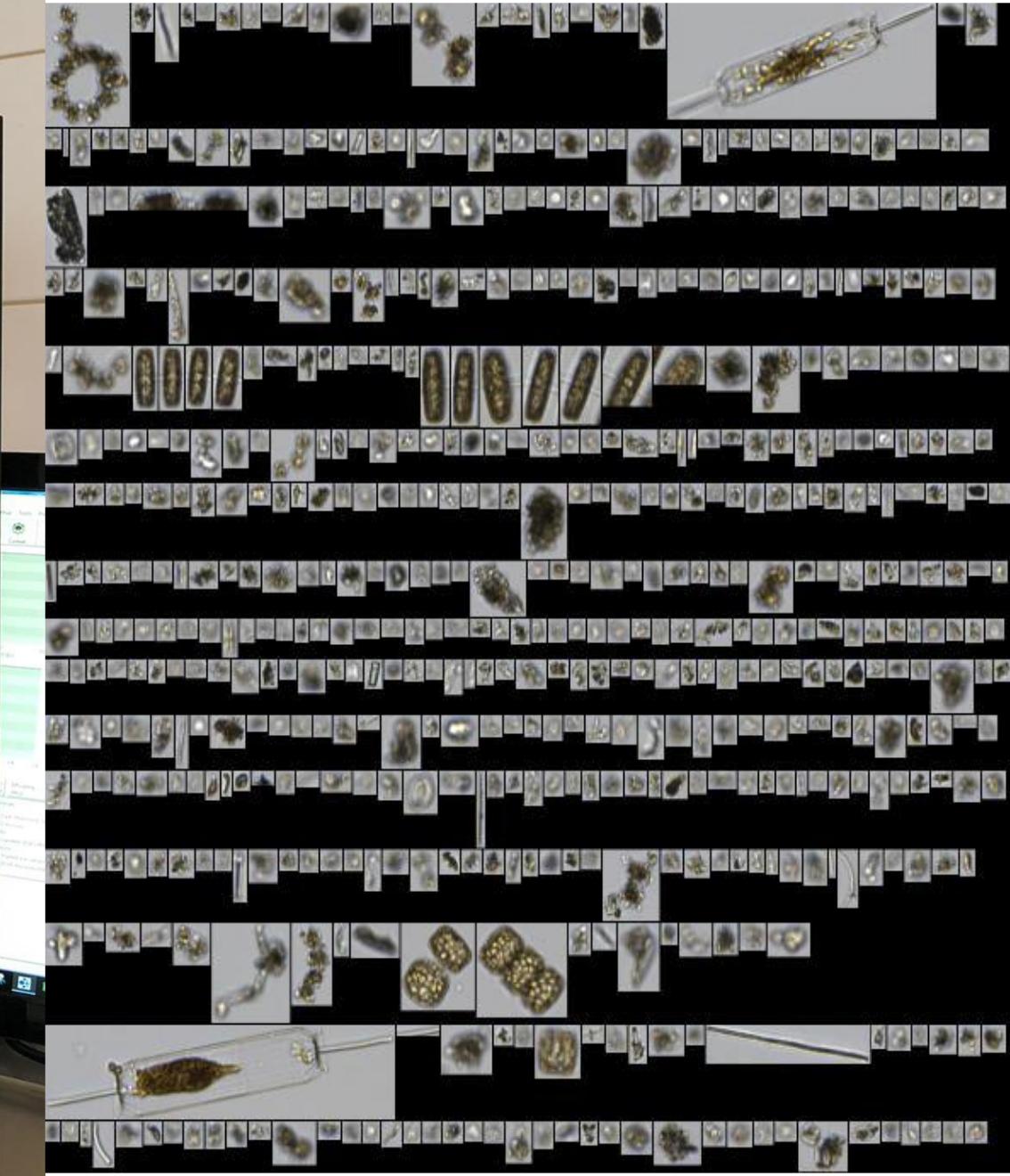
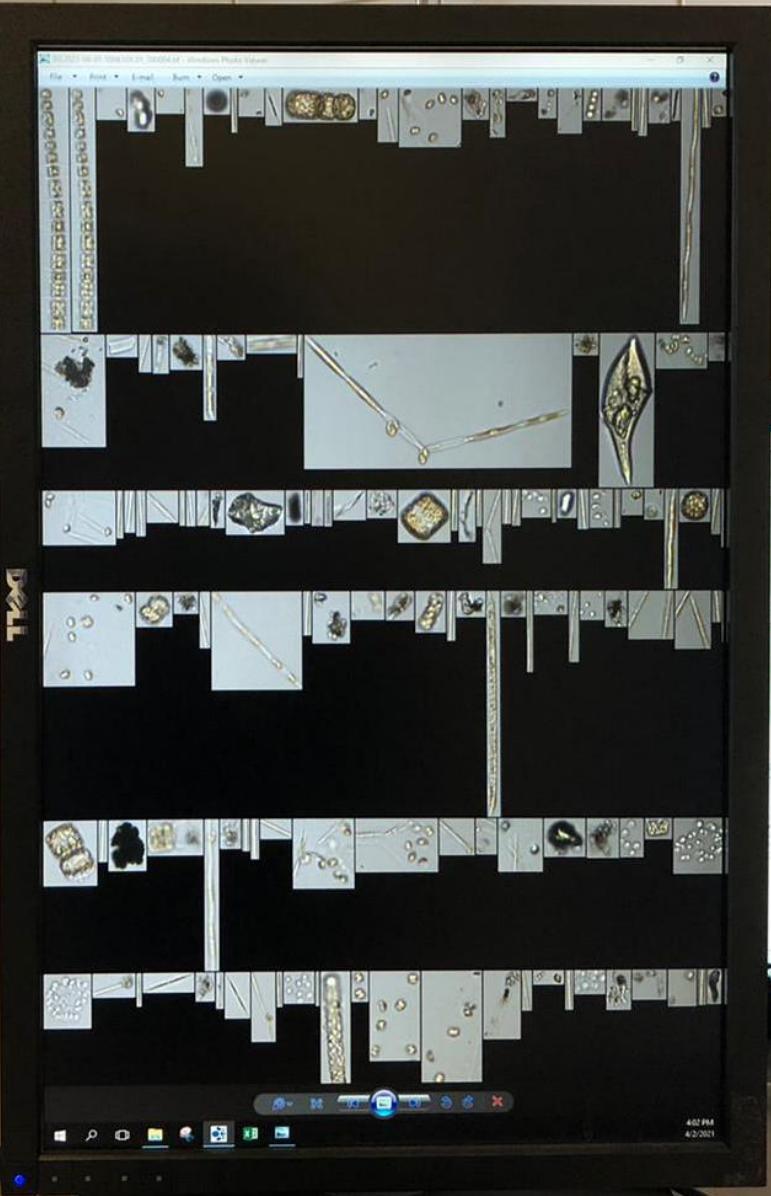
Multivariate  
NbCells ~ measures

Predictive model  
Linear Discriminant Analysis

Prediction



# FlowCam 800



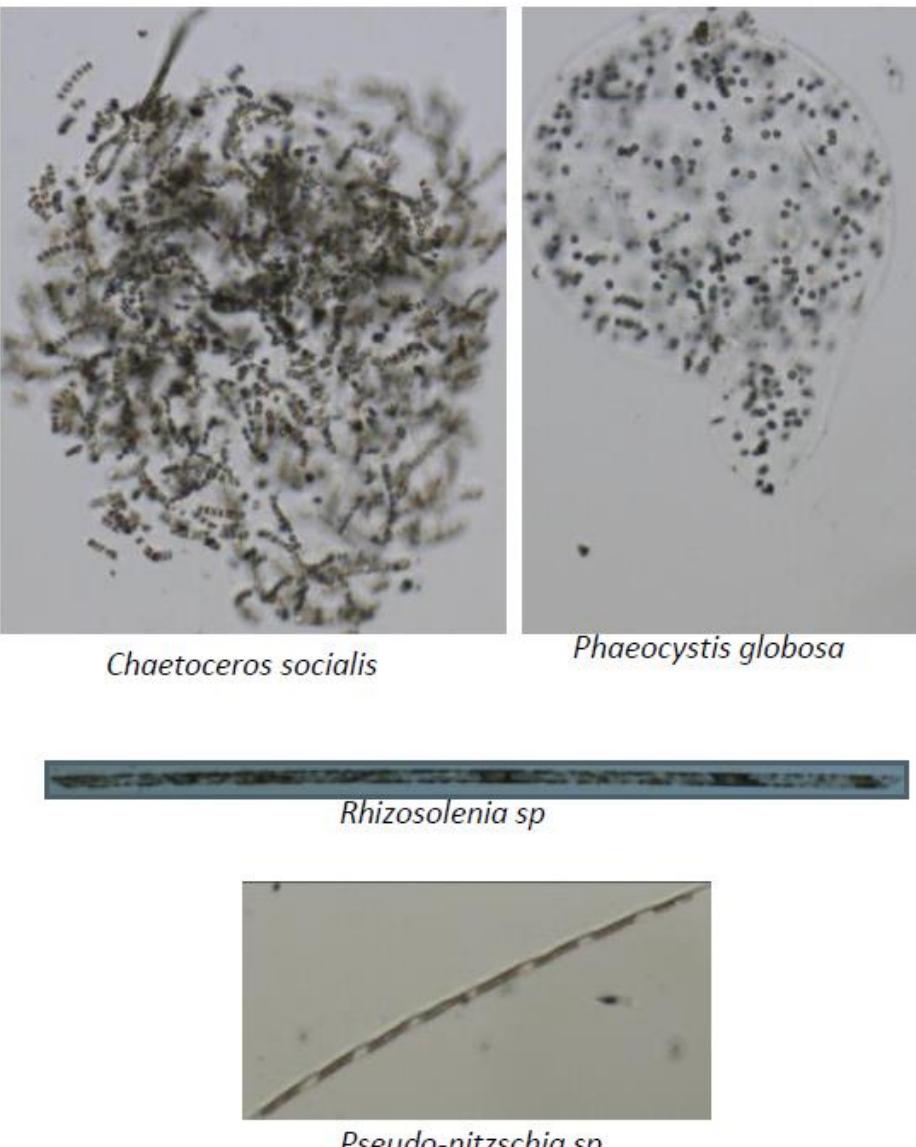
# New image training set (FlowCAM 8000 Series)

Actual // Predicted

C. socialis	Pelote	01
P. globosa	Col	02
P. globosa	G. Cox	03
C. socialis	Chaine	04
P. globosa	P. Cox	05
Gyrodinium	spirale	06
Navicula	spp	07
P. sulcata		08
Prorocentrum	sp	09
Ciliés		10
R. amphiceros		11
G. delicatula		12
Pleurosigma	spp	13
E. zodiacus		14
Nauplius		15
T. rotula		16
Chaetoceros	spp	17
G. striata		18
O. regia		19
O. ocellata		20
D. brightwellii		21
L. boea	sp	22
Tripos	lineatus	23
Tripos	fusus	24
Tiarina	sp	25
Thalassiosira	sp	26
T. nitzschioidea		27
R. pungens		28
Pollen		29
O. sinensis		30
O. rhombus		31
Larve		32
L. danicus		33
Copepode		34
Bulles		35
Brockmanniella	sp	36
Actinoptychus	spp	37
Apendicularia		38
G. flaccida		39
M. membranacea		40
Mytilus	edulis	41
Coscinodiscus	spp	42
Diplopsalis	spp	43
Cerataulina	sp	44
L. annulata		45
P-nitzschia	spp	46
R. imbricata		47

Matrice de confusion du training set 4X

Actual // Predicted	01	04	07	10	13	16	19	22	25	28	31	34	37	40	43	46
C. socialis	41													1		
P. globosa	6	15														
P. globosa	2	38	1													
C. socialis	1															
P. globosa																
Gyrodinium																
Navicula																
P. sulcata																
Prorocentrum																
Ciliés																
R. amphiceros																
G. delicatula																
Pleurosigma																
E. zodiacus																
Nauplius																
T. rotula																
Chaetoceros																
G. striata																
O. regia																
O. ocellata																
D. brightwellii																
L. boea																
Tripos																
Tripos																
Tiarina																
Thalassiosira																
T. nitzschioidea																
R. pungens																
Pollen																
O. sinensis																
O. rhombus																
Larve																
L. danicus																
Copepode																
Bulles																
Brockmanniella																
Actinoptychus																
Apendicularia																
G. flaccida																
M. membranacea																
Mytilus																
Coscinodiscus																
Diplopsalis																
Cerataulina																
L. annulata																
P-nitzschia																
R. imbricata																



# New classifier developed on CNN basis – Confusion matrix on FlowCAM images

X4

Actual // Predicted	01	03	05	07	09	11	13	15	17	19	21	23	25	27	29	31	33	35	37	
Pseudo_nitzschia_sp 01	20	73	1			1										1	1			
Leptocylindrus_danicus 02	5	44	66			1										1	2			
Rhizosolenia_imbricata 03	3	37	20			1	5	7								4	3	13	1	
Sphere 04		56		16												2	1	3	1	
Gymnodiniaceae 05		20	5	15		1		1								1	2	1	4	
Ciliophora 06		4	7	18	1				2	1	2					1	2	1	4	
Detritus 07		4				1		1	1	3	2	1	5	3	2	1	2	7	32	4
Chaetoceros_socialis_Chaine 08		3	1		17	1	5	2								5	1	1	2	
Phaeocystis_globosa_Col 09			31	2	4				1							2	6		9	
Chaetoceros_socialis_Pelote 10			4	3	4											2	12		10	
Cylindrotheca_closterium_Phao 11		4			1	8	12	12								4	3	16		
Cerataulina_sp 12		11	1													2	1		12	
Guinardia_delicatula 13		9	16													1	1		13	
Meuniera_membranacea 14			1		3		3	2	5	3	2					1	1		14	
Guinardia_flaccida 15			1					3	2	6	6					3			15	
Guinardia_strigata 16		3										1	1	1	1		2			
Frustules_casses_diatomées 17			1	2												5	1		17	
Rhaphoneis_amphiceras 18				1	7											1	2	5	18	
Paralia_sulcata 19		4			1	15	1		1							1	1		19	
Protorcentrum_sp 20			1	7						1	1	2	2	3					20	
Protoperidiniaceae 21		4	2	3															21	
Ditylum_brightwellii 22		1																	22	
Zooplankton 23			1	2															23	
Thalassiosiraes 24		1	4	1		1		6		1						6	2	1	6	
Triceratiaceae 25			5	4															25	
Asterionellopsis_glacialis 26				1		1					2					1	2	3	26	
Fibre 27		2	2	2	1						3								27	
Eau 28																			28	
Thalassionema_nitzschioidea 29		1														3	6	2	29	
Coscinodiscles 30					1	2													30	
Naviculaes 31		1	7	1	3											5	15	1	31	
Fond 32																1	1	3	32	
Rhizosoleniae 33		2	11	16	1											31			33	
Chaetoceros_curvisetus 34		8			2				4	1						46	28	3	34	
Chaetocerotaceae 35		1	9			1	2									40	66	2	35	
Flou 36		1	3		8	3	1									2	2	10	3	
Phaeocystis_globosa 37			14			8	2	5								11	8	49	38	
	01	03	05	07	09	11	13	15	17	19	21	23	25	27	29	31	33	35	37	

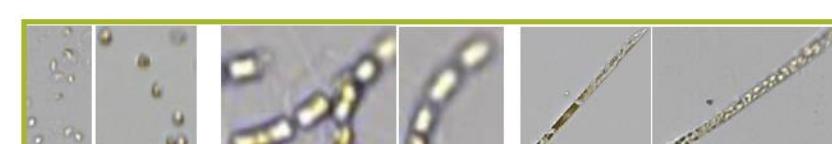
X10

Actual // Predicted	01	03	05	07	09	11	13	15	17	19	21	23	25	27	
Chaetoceros_sp 01	148	28	25			1	6	2							01
Phaeocystis_globosa 02	12	245	3			1	1	4	2						02
Phaeocystis_globosa_Col 03	6	66	6	1					4						03
Leptocylindrus_danicus 04	8	6		719	14				1	12	4	2			04
Rhizosolenia_imbricata 05	2		37	39						5		3			05
Billes 06		1				764	27	5							06
Bulles 07						186	1					6			07
Eau 08		1	1			3	28	10	3	1					08
Flou 09						10	71	2					3		09
Fond 10		2	3	1					793						10
Bacillariaceae 11										12	167	1			11
Naviculaes 12		2	2	3	1					5	15	1	1	1	12
Guinardia_sp 13			1		11	2					13	3			13
Cerataulina_pelagica 14					6	1				1	4	17	7		14
Guinardia_flaccida 15					1						1	47			15
Vide 16							1	3	1				420		16
Rhizosolenia_pungens 17							12	3			2	1			17
Thalassionema_nitzschioidea 18		2			1						18				18
Dynophyta 19											6		1	2	1
Protoperidiniaceae 20											6	4		2	2
Raphoneis_amphiceras 21		1	4	2							1	4	12	2	5
Gymnodiniaceae 22		2	1	5	1						7	1	3	15	4
Phaeocystis_globosa_Ind 23		2	2	30							2	9	100		23
Ciliates 24		3	3		1						4	3		17	8
Ditylum_brightwellii 25						1			3	1	3		4	35	3
Coscinodiscles 26		1									1	2	1	5	6
Thalassiosiraes 27		1	6		6						1		1	3	
	01	03	05	07	09	11	13	15	17	19	21	23	25	27	



Vignettes X4

(A. LEPTOCYLINDRUS DANICUS, B. PSEUDO-NITZSCHIA,  
C. RHIZOSOLENIA IMBRICATA)



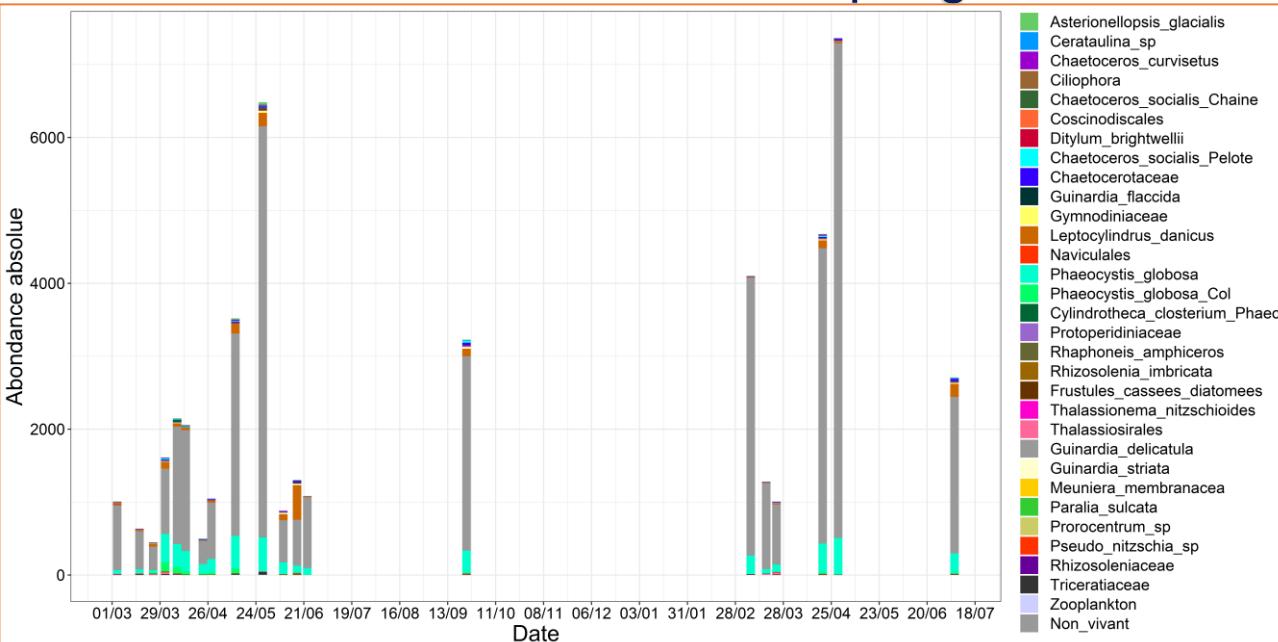
Vignettes X10

(A. PHAEOCYSTIS GLOBOSA COX, B. PHAEOCYSTIS  
GLOBOSA, C. CHAETOCEROS SOCIALIS , D. CHAETOCEROS  
SPP, E. RHIZOSOLENIA IMBRICATA, F. LEPTOCYLINDRUS  
DANICUS)

# CNN Classifications results on point R1 of the DYPHYRAD transect from spring 2021 to summer 2022)

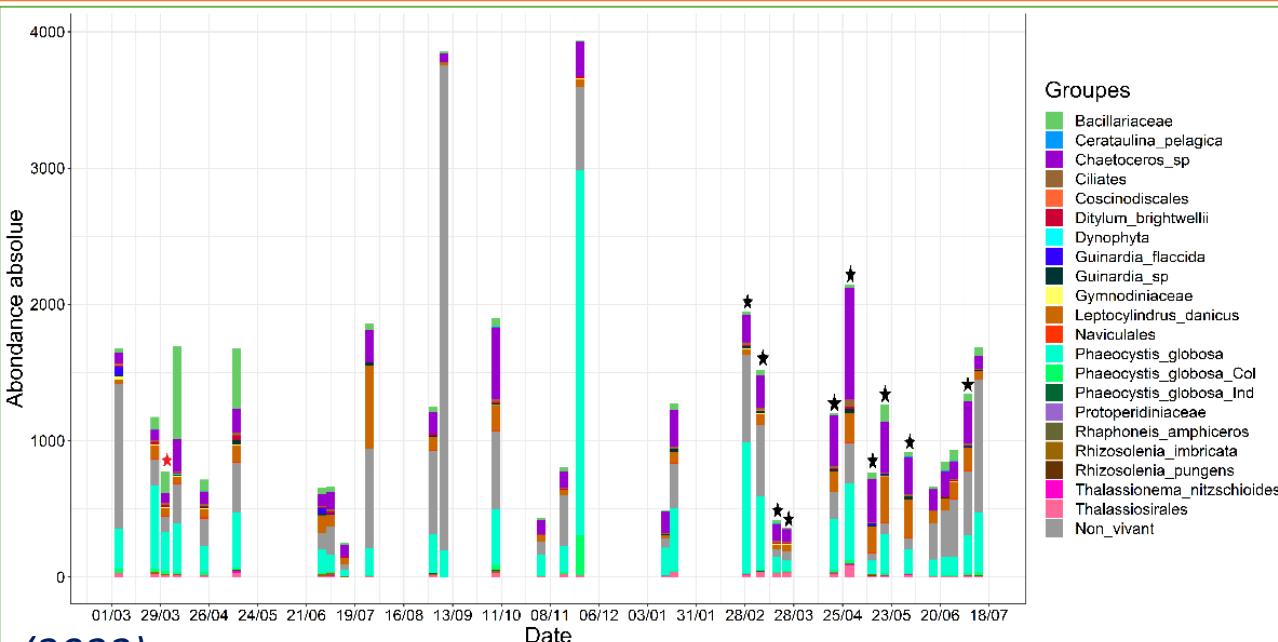
X4

Majority of non-living particles



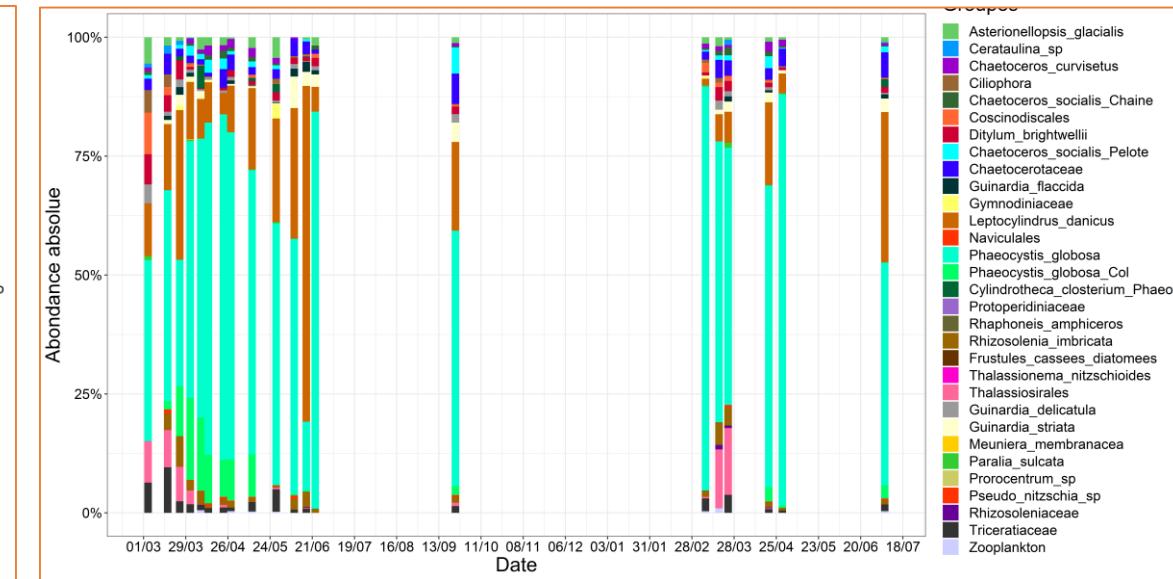
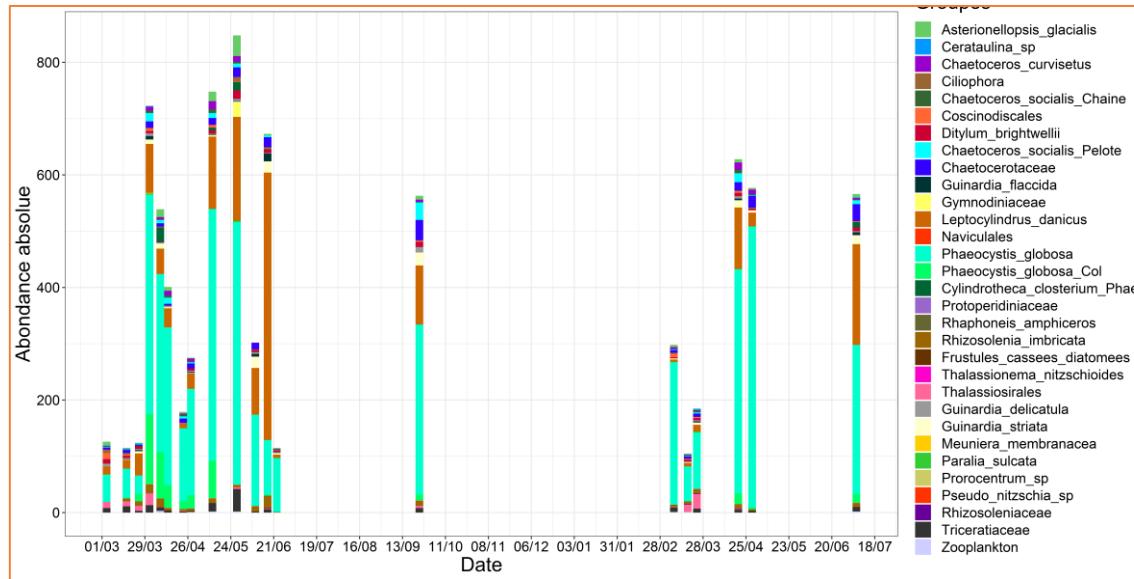
X10

- ★ 100µm filtration
- ★ 150µm filtration

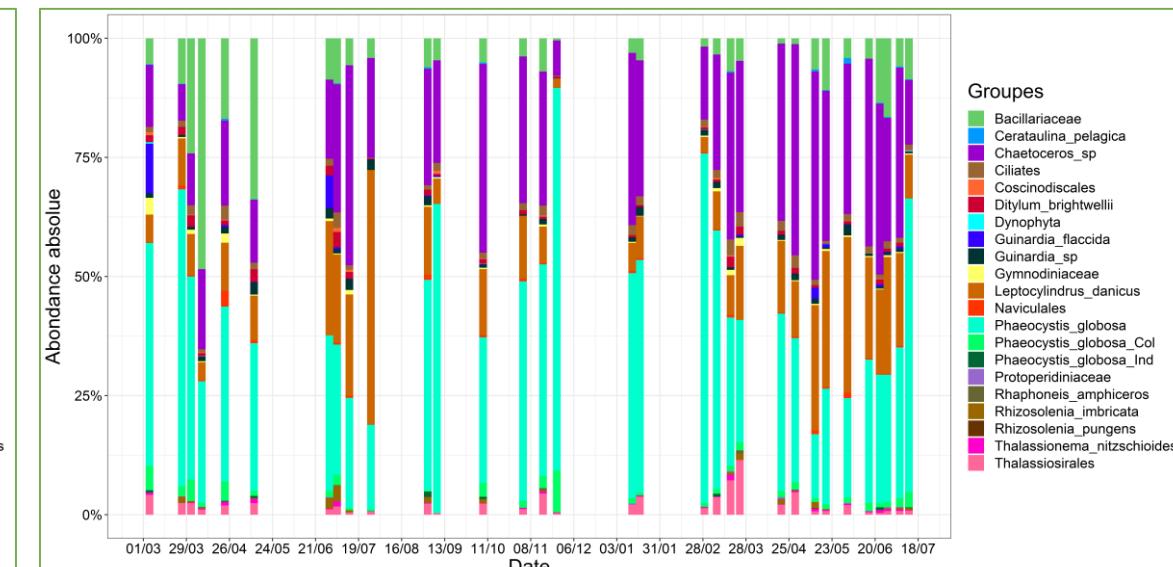
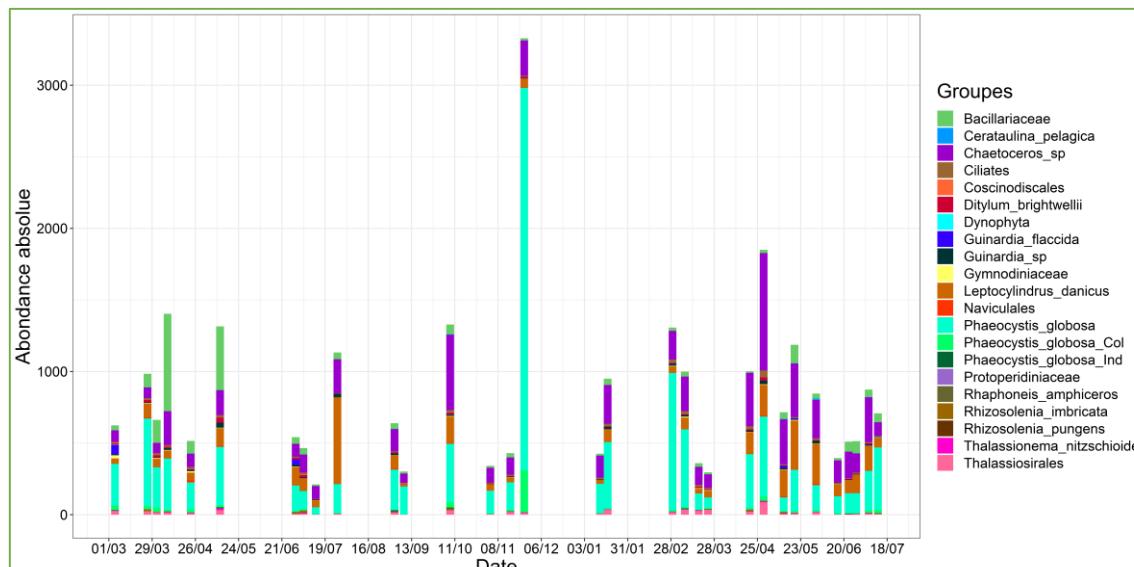


# CNN Classifications results on point R1 of the DYPHYRAD transect from spring 2021 to summer 2022)

X4



X10



# Flow Cytometry

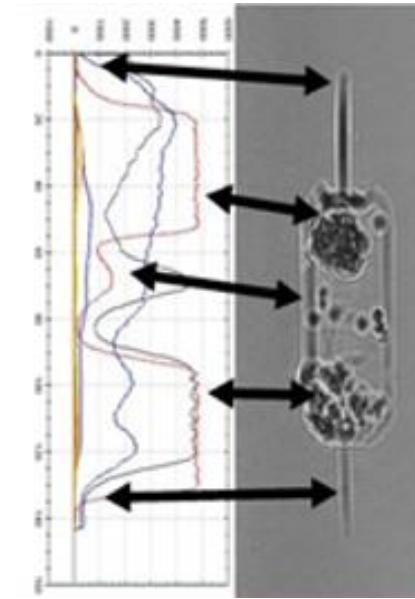
**Device:** Pulse-Shape recording Flow Cytometer

**Models:** CytoSense/CytoSub

**Trigger level:** Red fluorescence set to 10mV



CytoSense



## Training set

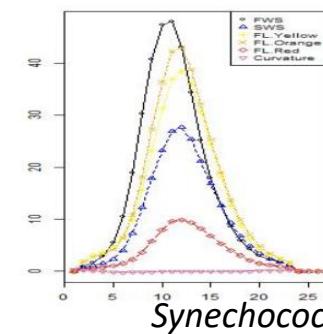
**Training set:** DYPHYRAD transect

**Number of groups:** 4 functional groups  
(synechococcus, pico- & nano-eukaryotes,  
micro-phytoplankton) + 4 noise

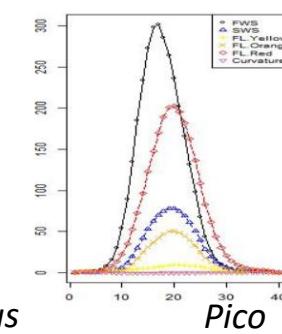
**Number of vignettes:** 372 (~50/group)

**Classifier:** Random Forest (ntree = 500)

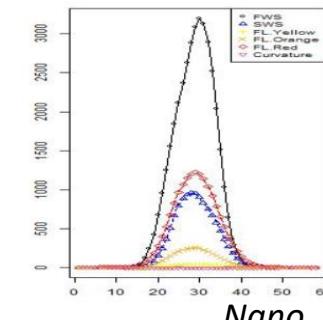
**Cross-validation:** 10 folds



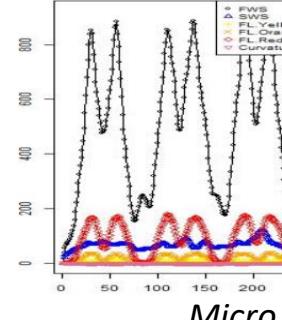
Synechococcus



Pico



Nano



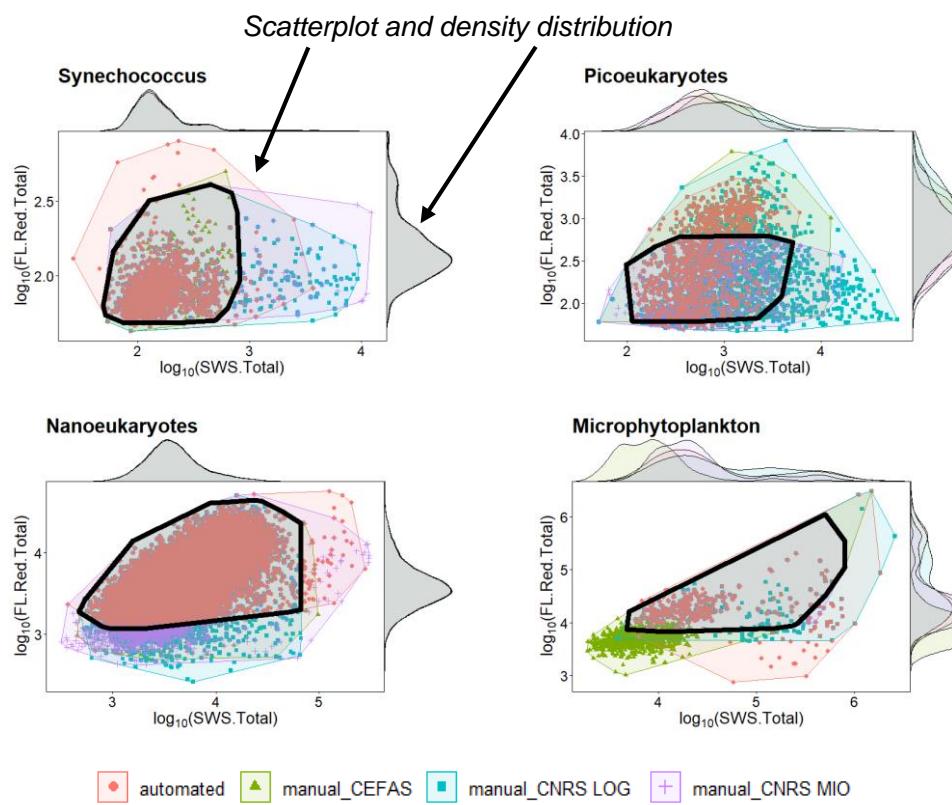
Micro

# Flow cytometry – Intercomparison exercise\*

Partners involved: CEFAS, CNRS-LOG, CNRS-MIO

Manual classification (CytoClus software): CEFAS, CNRS-LOG, CNRS-MIO

Automated classification (RclusTool package): CNRS-LOG



% common (black polygon)

	CEFAS	CNRS-LOG	CNRS-MIO	Automated
Synechococcus	98.73	91.71	95.14	96.65
Pico	56.37	68.11	90.88	63.12
Nano	97.67	97.55	97.30	99.86
Micro	32.09	80.43	97.78	83.84

Rand index (~similarity between partitions)

	CEFAS	CNRS-LOG	CNRS-MIO	Automated
CEFAS	1	0.95	0.94	0.89
CNRS-LOG		1	0.97	0.91
CNRS-MIO			1	0.90
Automated				1

Rand index = proportion of pairs of particles which are similarly classified in the 2 partitions (either in the same group or in different groups)

# Flow cytometry - LifeWatch cruise

Scientific campaign for analysis of plankton diversity

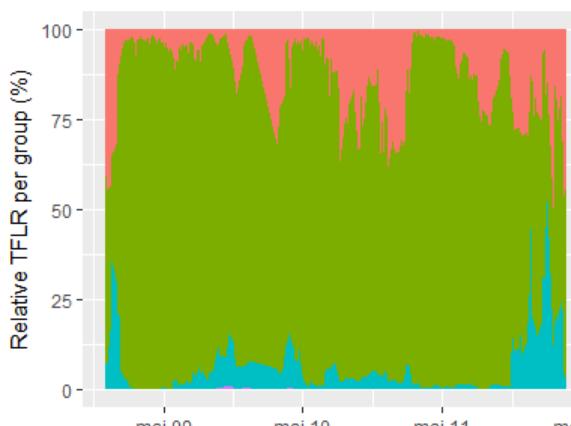
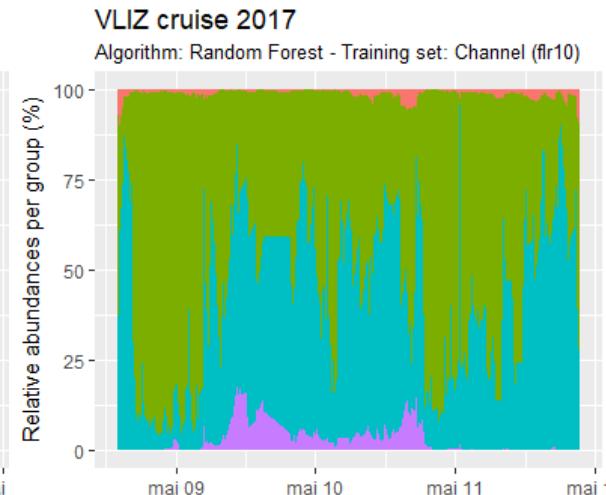
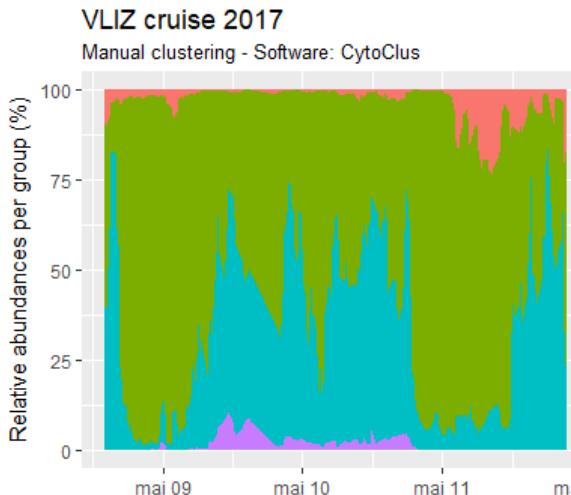
(LifeWatch/JERICO-NEXT cruise, R/V "Simon Stevin"-VLIZ).



**Period:** 8th May-12th May, 2017

**Area:** North Sea

**Ship:** R/V « Simon Stevin » - VLIZ



Groups	$\rho$ abd	pvalue
Synecho	0.99	<2e-16
Pico	0.84	<2e-16
Nano	0.89	<2e-16
Micro	0.73	<2e-16
Total	0.88	<2e-16

Spearman's rank  
correlation coefficient

Groups	$\rho$ tflr	pvalue
Synecho	0.95	<2e-16
Pico	0.59	<2e-16
Nano	0.95	<2e-16
Micro	0.81	<2e-16
Total	0.95	<2e-16

# Specific training set combining optical shapes and image

**Training set:** DYPHYRAD transect (CNRS-LOG)

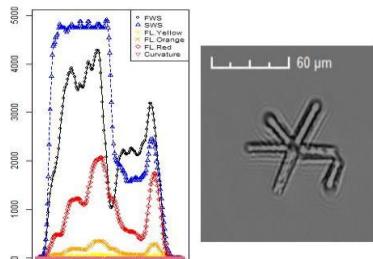
**Number of groups:** 11 phyto + 2 noise (to be completed with more groups)

**Number of vignettes:** 319 (to be completed with more images, CytoSense limitation = 150 img/sample)

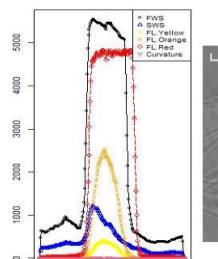
**Classifier:** Random Forest (ntree = 500)

**Cross-validation:** 10 folds

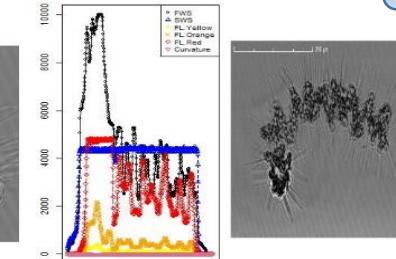
## Functional traits



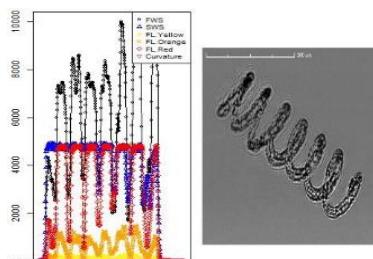
*Thalassionema* spp.



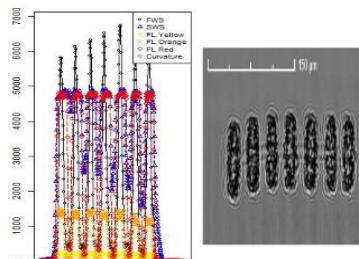
*Chaetoceros* spp.



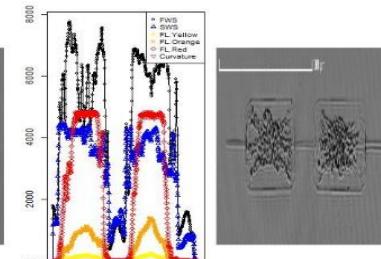
*Asterionellopsis* spp.



*Guinardia* spp.



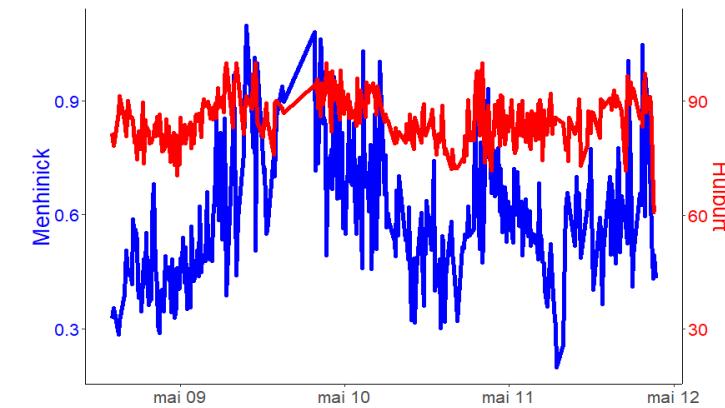
*Thalassiosira* spp.



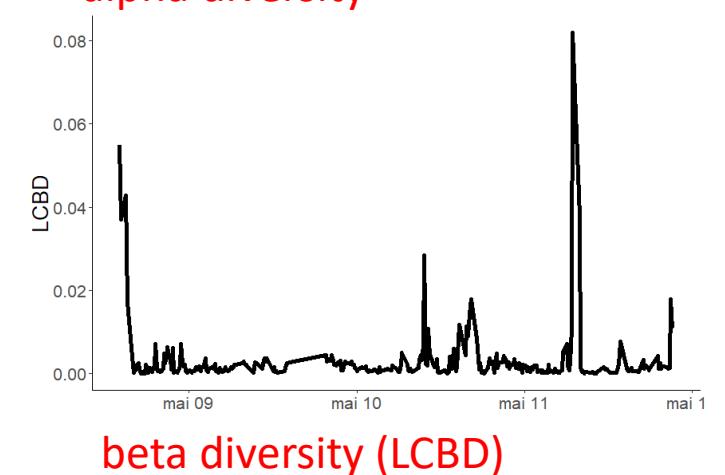
*Ditylum* spp.

**Used for:**

- counting cells in colony
- computation of functional traits, diversity indices (MSFD), ...



## alpha diversity



## beta diversity (LCBD)

# Automated estimation of cells in colonies (flow cytometry signal shapes and images)

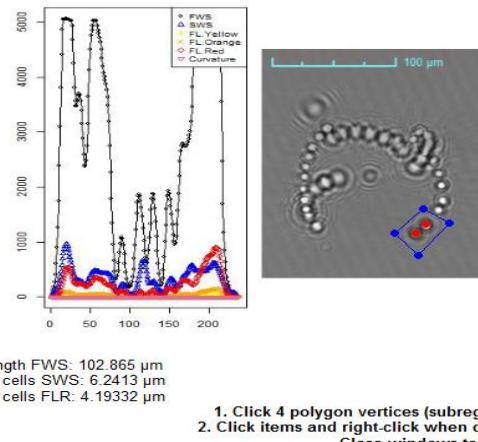
Today, for imaging-in-flow systems, 1 image = 1 particle = 1 cell

Some examples and results...

## Proposed method

Class: *C.socialis*

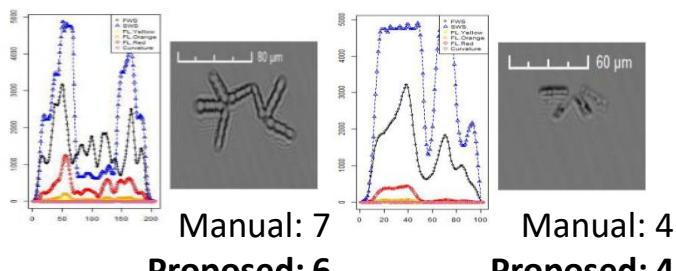
Particle: 663\_img\_Mer2\_2016-07-19\_flr26\_medium 17u53



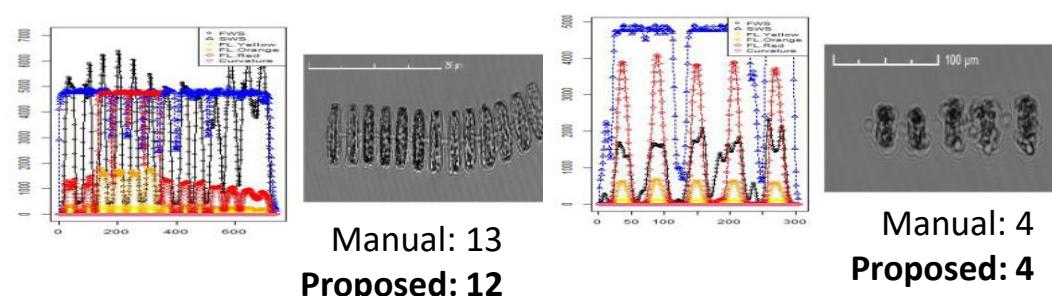
Calibration of predictive models (one per taxonomic group) for the estimation of the number of cells in colonies, based on **manual counts** made on the images classified in training sets.

Algorithm: **Linear Discriminant Analysis**

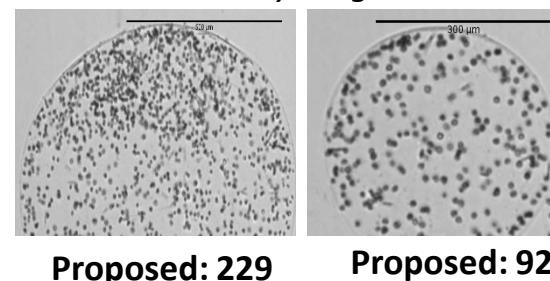
*Thalassionema nitzschoides*



*Thalassiosira rotula*

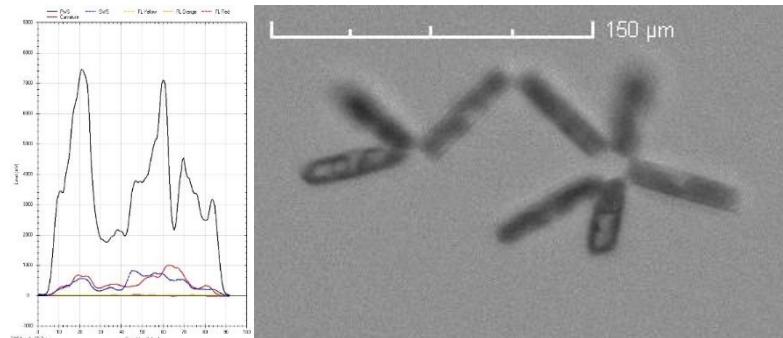


*Phaeocystis globosa*

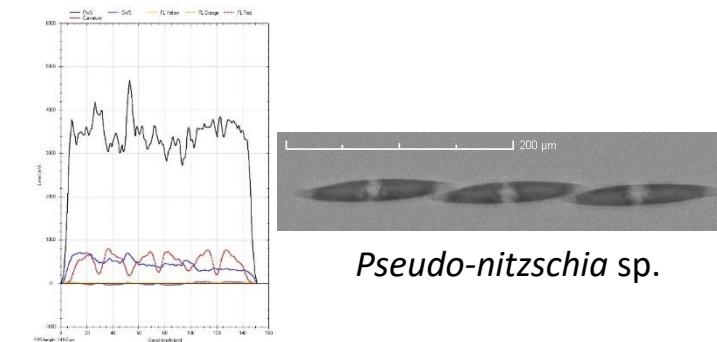


# Applying new CNN classification on CytoSense images

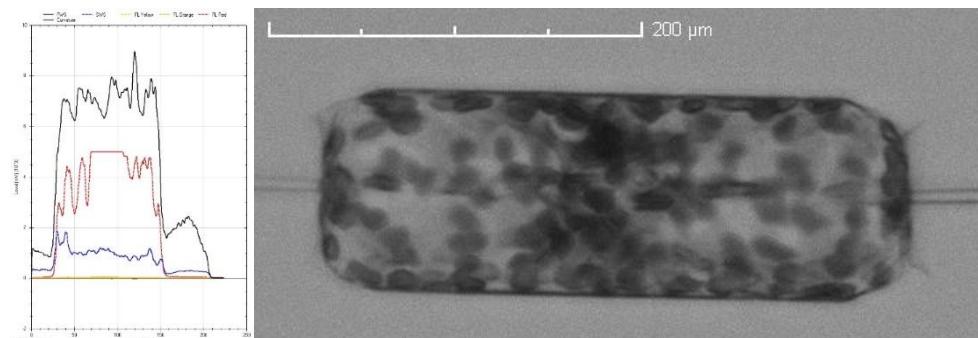
## Combination of images and optical profiles



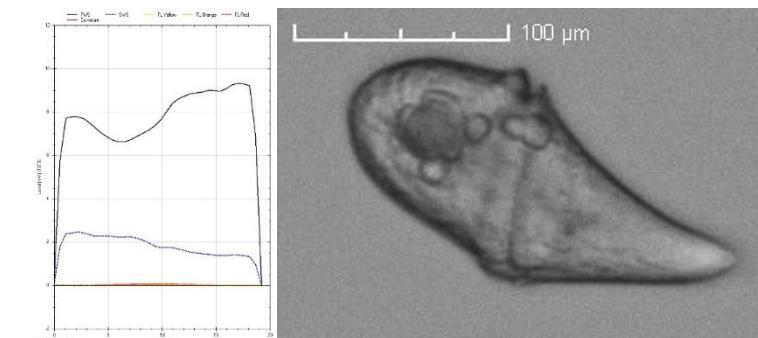
*Thalassionema nitzschiooides*



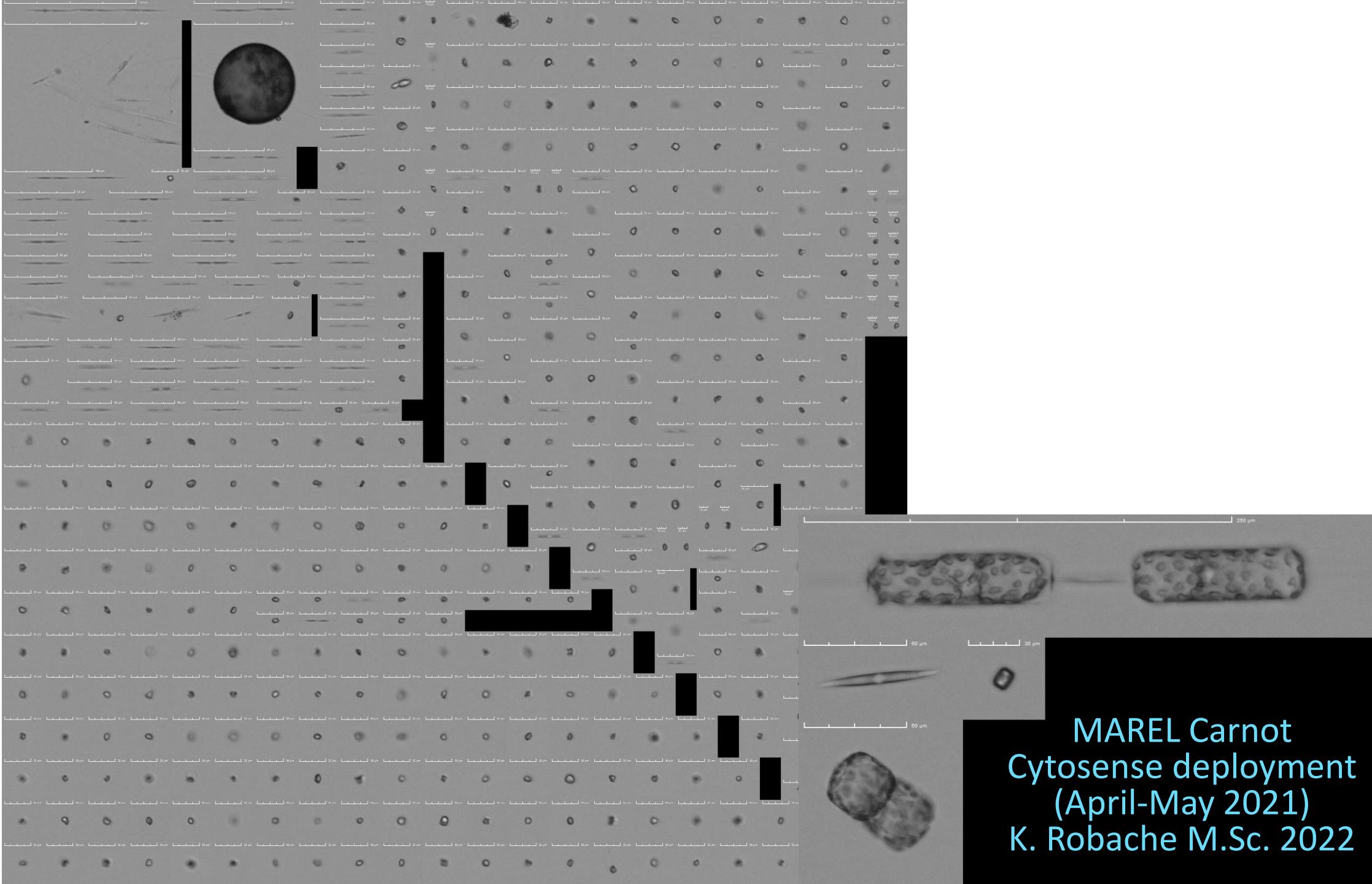
*Pseudo-nitzschia* sp.



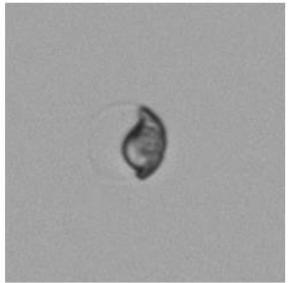
*Ditylum brightwellii*



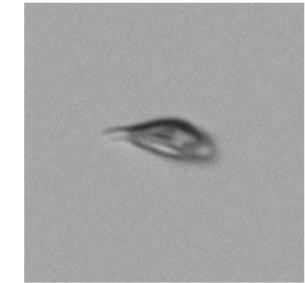
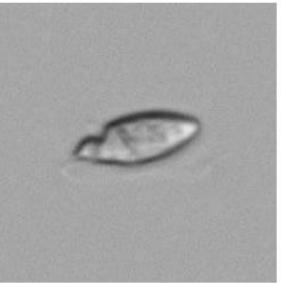
*Gyrodinium spirale*



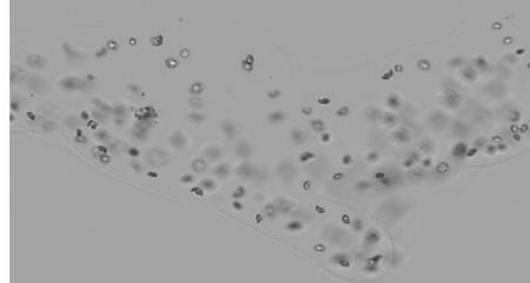
MAREL Carnot  
Cytosense deployment  
(April-May 2021)  
K. Robache M.Sc. 2022



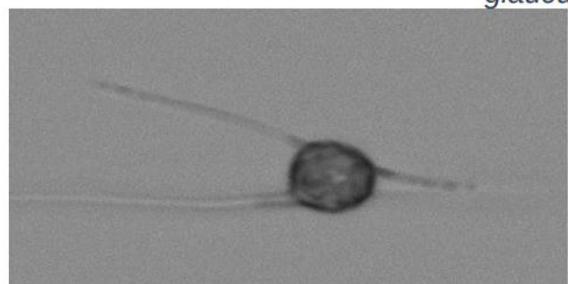
*Phaeocystis globosa* (cellule) Exemple de Gymnodini Aceae (*Lebourdinium glaucum*)



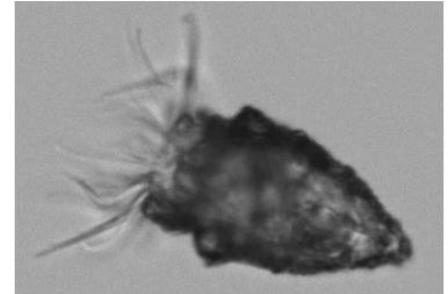
Cryptophyte



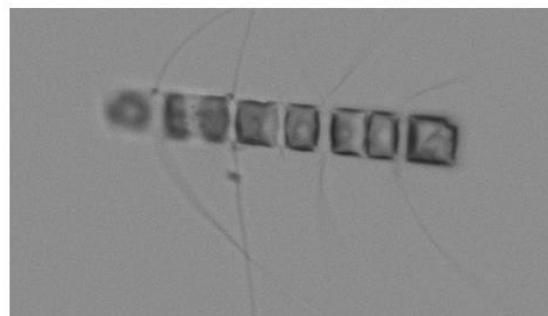
*Phaeocystis globosa* (colonie)



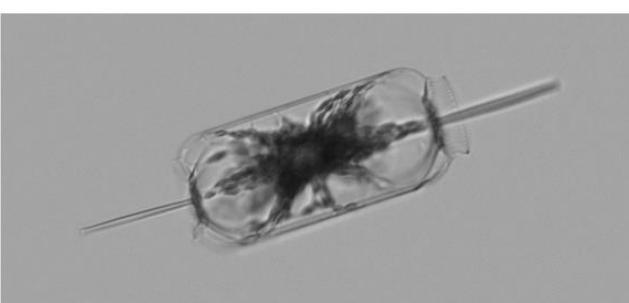
*Tripos* sp.



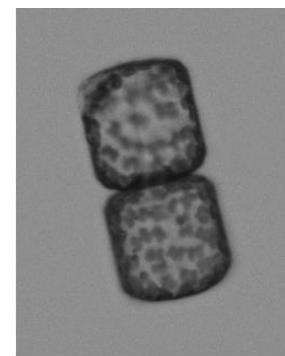
Ciliés



*Chaetoceros* sp.



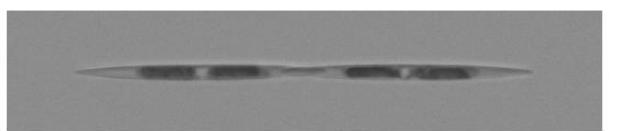
*Ditylum brightwellii*



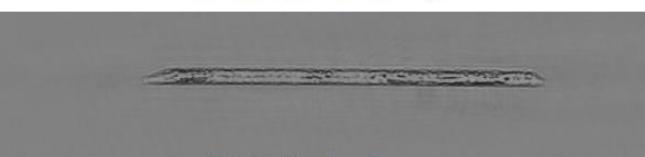
*Detonula pumila*



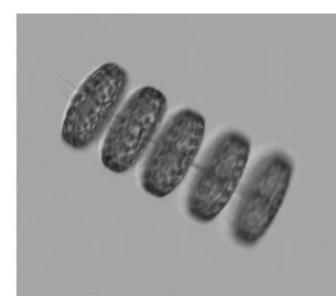
*Guinardia striata*



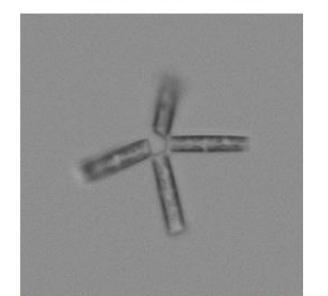
*Pseudo-nitzschia* sp.



*Rhizosolenia* sp.



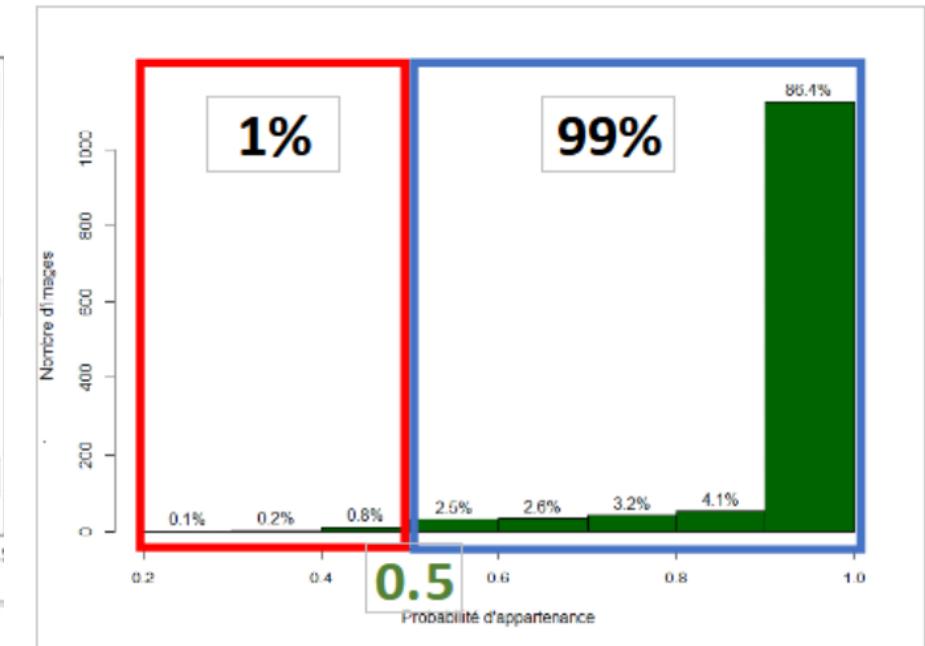
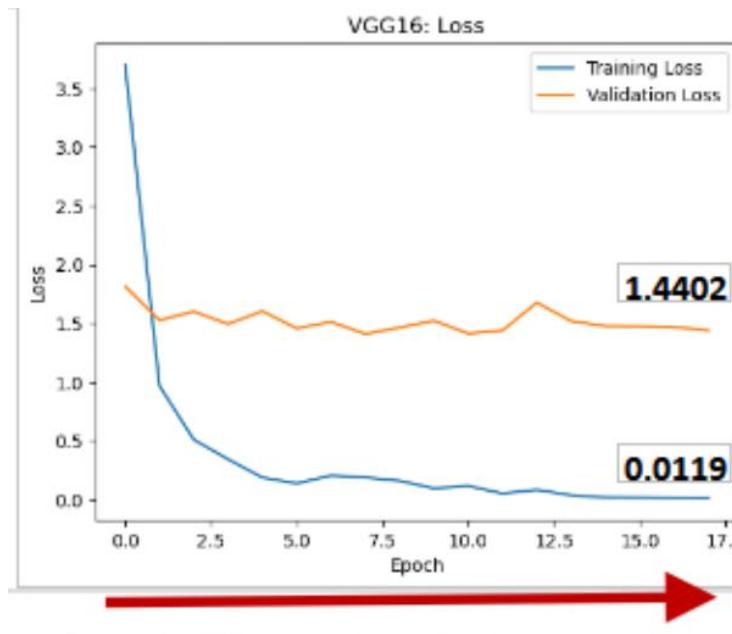
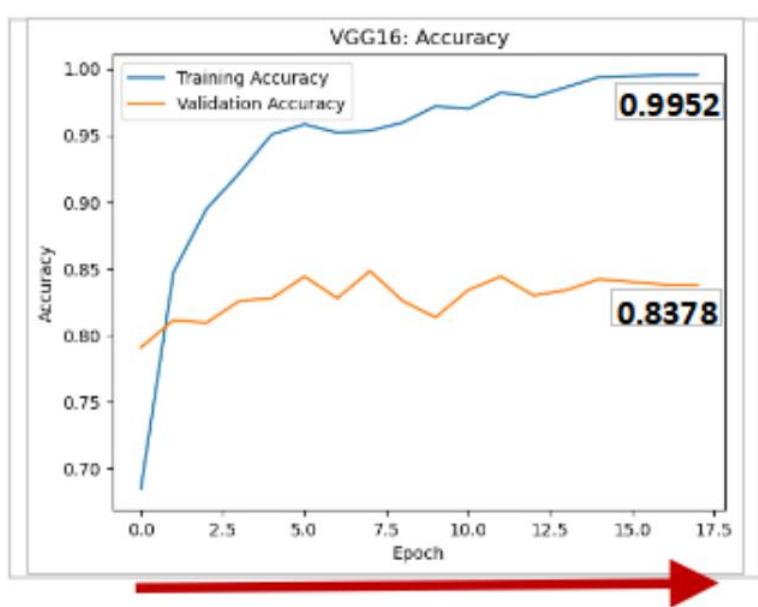
*Thalassiosira* sp.



*Thalassionema nitzschioides*

Building a new training set  
on CytoSense images

# Applying new CNN classification on CytoSense images : premiminary results



# Conclusions and perspectives

## Conclusions

- Use of expert knowledge to “orientate” the automated methods
- Integration of this knowledge at different levels
  - Learning, constraints, validation, single-cell counts
- Application on different datasets
  - Imaging-in-flow system, automated flow cytometry
- Development of two R packages
  - RClusTool, ZoolImage

# Conclusions and perspectives

## Conclusions

- Use of expert knowledge to “orientate” the automated methods
- Integration of this knowledge at different levels
  - Learning, constraints, validation, single-cell counts
- Application on different datasets
  - Imaging-in-flow system, automated flow cytometry
- Development of two R packages
  - RclusTool, zooimage

## Perspectives and ongoing work

- Application to other sensors and data acquisition systems
  - ZooScan (for zooplankton), Fluorometers (for total chlorophyll or pigmentary groups...)
- Optimization of training sets (specific to each geographical area)
- Use of deep learning to obtain more information from data
  - Convolutional Neural Network, Recurrent Neural Network, ...
- Combination of signal and image information
  - for micro-phytoplankton with flow cytometry

## R package « RClusTool »

Collaborative work between **CNRS-LOG** and **ULCO-LISIC**  
**(DYMAPHY (INTERREG IV A « 2 Seas », JERICO-NEXT (H2020) & MARCO (CPER HF)**



Université  
de Lille



**LISIC**  
Laboratoire d'Informatique  
Signal & Image de la Côte d'Opale

## R package « Zoolimage »

Collaborative work between **UMONS**, **IFREMER** and **CNRS-LOG**  
**(IFREMER, JERICO-NEXT & MARCO projects)**



Université de Mons



New « R » package tool based on CNN (Wacquet & Lefebvre, 2022)

# Web app with near real-time results Pulse Shape-Recording FCM

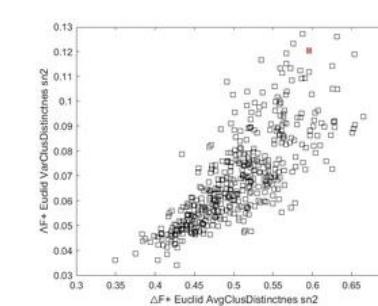
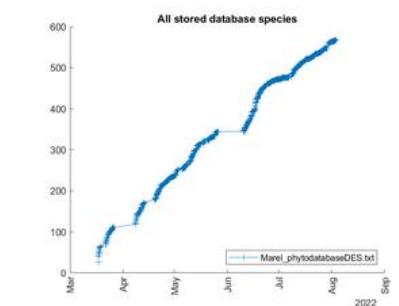
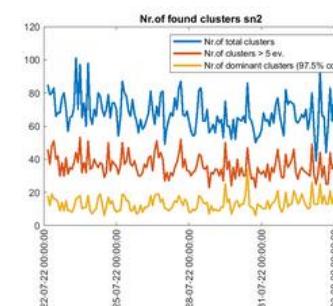
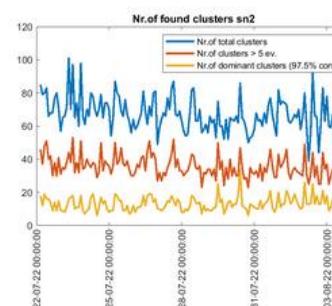
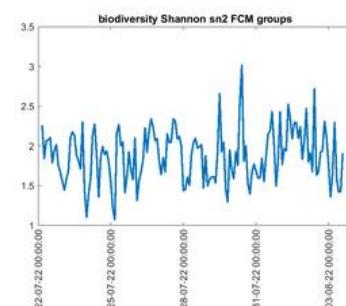
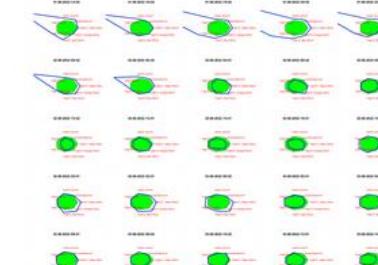
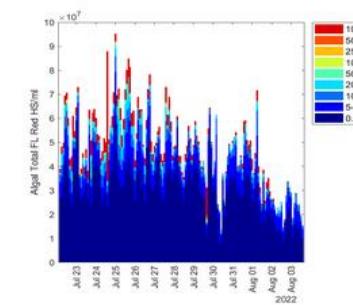
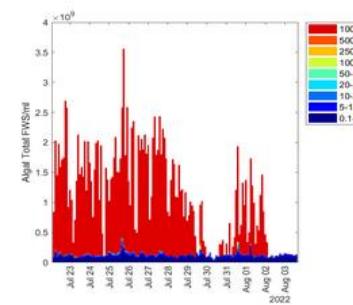
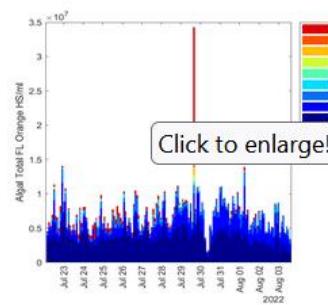
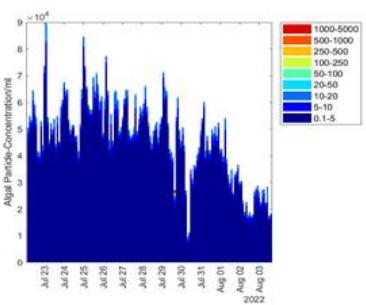
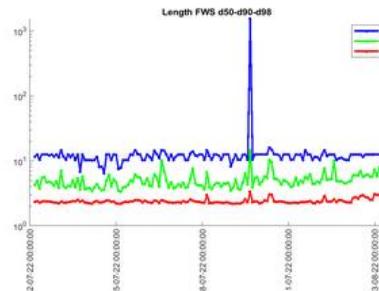
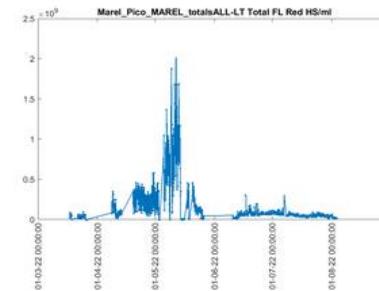
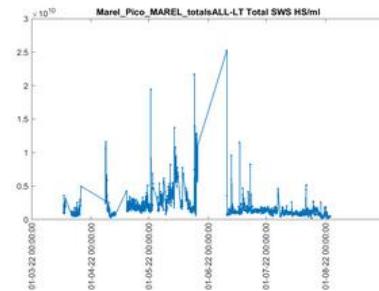
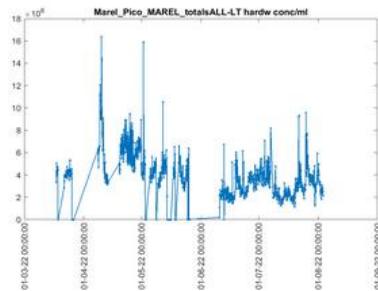
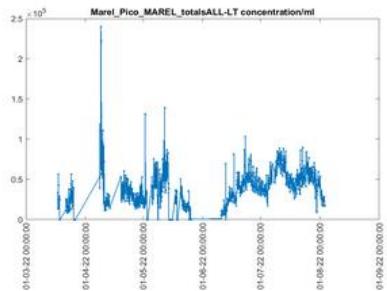


[https://fytoplankton.nl/ULCO-CNRS/Marel/phytoplankton\\_liveloc.shtml](https://fytoplankton.nl/ULCO-CNRS/Marel/phytoplankton_liveloc.shtml)



## Live Results Marel

[to image gallery](#) (if available)



Easyclus Live web app. Instrument check; Unsupervised clustering Totals per sample (abundance, chl a); Cluster plots; Biodiversity indicators - EasyClus tool : Possibility to build a classifier to perform supervised analysis - **Thomas Rutten Projects**

## Automatic recognition of flow cytometric phytoplankton functional groups using convolutional neural networks

Robin Fuchs  <sup>1,2</sup>, Melilotus Thyssen  <sup>2,\*</sup>, Véronique Creach  <sup>3</sup>, Mathilde Dugenne  <sup>4</sup>, Lloyd Izard  <sup>5</sup>,  
Marie Latimier, <sup>6</sup> Arnaud Louchart  <sup>7,8</sup>, Pierre Marrec  <sup>9</sup>, Machteld Rijkeboer  <sup>10</sup>, Gérald Grégori  <sup>2</sup>,  
Denys Pommeret  <sup>1,11,12,13</sup>

<sup>1</sup>Aix Marseille Univ, CNRS, I2M, Marseille, France

<sup>2</sup>Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France

<sup>3</sup>Cefas, Suffolk, UK

<sup>4</sup>Department of Oceanography, University of Hawai'i at Manoa, Honolulu, Hawai'i

<sup>5</sup>Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN-IPSL), Paris, France

<sup>6</sup>IFREMER, DYNECO PELAGOS, Plouzane, France

<sup>7</sup>Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, Italy

<sup>8</sup>IFREMER, Laboratoire Environnement and Ressources, Boulogne-sur-Mer, France

<sup>9</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island

<sup>10</sup>Laboratory for Hydrobiological Analysis, Rijkswaterstaat (RWS), Lelystad, The Netherlands

<sup>11</sup>Université Claude Bernard Lyon 1, Villeurbanne, France

<sup>12</sup>ISFA, Lyon, France

<sup>13</sup>Laboratoire de Sciences Actuarielle et Financière (SAF), Lyon, France

# Processing and storage of thousands of data (raw data, processes data, images and automated classification)

ZooPhytoImage(U Mons-Ifremer-CNRS/LOG), RClusToolBox (LISIC/ULCO-CNRS/LOG), EasyClus (TRP), EMODNET, SeaDataNet, EcoTaxa, other databases...

## EcoTaxa 1.3

Exploration

Filter: Samples:jerico201707a17 Project:UVP5 JERIC0 2017

Not logged ([log in](#))

Update view & apply filters Img /page : 50 Zoom : 10

[Clear filters](#) [Share page](#) [Hide filters](#)

**Project** [x UVP5 JERIC0 2017](#)

**Instrument** [Clear](#)  
Instrument

**Sample** [Advanced](#) [Clear](#)  
[x jerico201707a17](#)

**Depth** [Clear](#)  
Min [m] Max [m]

**Location** [Clear](#)  
West North South East  
[Open map](#)

feces detritus feces detritus feces detritus feces detritus

detritus feces badfocus detritus detritus detritus Copepoda feces

detritus feces badfocus detritus feces detritus fiber feces fiber

badfocus detritus darksphere Copepoda fiber detritus feces fiber

**Public exploration of validated images across oceans**

**Powerfull filters:**

- Taxonomy
- Date/time/month/Depth/sample...
- Annotators (experts)

**Automatic classification**

- Random Forest
- Deep Learning (soon)

**POWERFUL manual annotation**

- > 20 000 images / day !!!
- All operations recorded
- Explicit validation

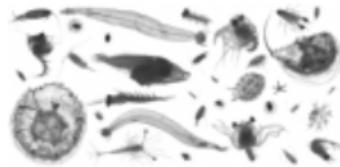
<http://ecotaxa.obs-vlfr.fr/explore/>

Picheral M., Stemman L., MMV 2017

[Home](#) > Explore Virtual Access services

## Explore Virtual Access services

### EcoTaxa



EcoTaxa is a web application where users can (i) upload and host images of individual objects, (ii) name them according to a universal taxonomy with the help of advanced machine learning algorithms, and (iii) export the resulting ecological data in standard and easy to use formats, for scientific exploitation and dissemination. It is currently massively used for images collected by quantitative plankton imaging instruments.

### LISIC clusTools



Mawenzi is a tool center with R-packages and their RShiny graphical user interfaces for data interpretation, from data completion to data prediction.

[RClusTool](#): Clustering and Classification Tool, with Visualisation and Labelling Features. R-package and its GUI are well adapted to clustering, insert expert knowledge in this clustering as label or pair constraints (these objects must be linked or not linked). Flow cytometry data with pulse signals and/or features could be used with an image visualisation to help in the expert classification step. But it is generalised to any dataset with at least feature input. This package is developed by LISIC-ULCO (Pierre-Alexandre Hébert, Emilie Poisson) in collaboration with CNRS-LOG (Luis Felipe Artigas).

*Thank you very much  
for your attention!*



Part of this work was supported by the JERICO-S3 project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n° 871153.