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[1], Guillaume Wacquet[1,2], Alexandre Epinoux[1], Arnaud Louchart[1], Clémentine Gallot[1], Zéline Hubert[1], Florine Verghaeghe[1], Manon Ryckman[1], Pierre-Alexandre Hébert[3], Emilie Poisson-Caillault[3], P. Grosjean[4], Alain Lefebvre[2], Thomas Rutten[5]

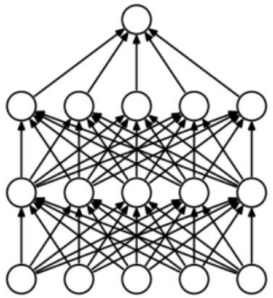
[1] Laboratory of Oceanology and Geosciences
CNRS – UMR 8187 LOG – ULCO – Wimereux (FR)
[2] LER/BL IFREMER – Boulogne sur Mer (FR)
[3] LISIC-ULCO – Calais (FR)
[4] U. Mons – Mons (BE)
[5] Thomas Rutten Projects – Middelburg (NL)

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

Expert knowledge

+

Machine learning

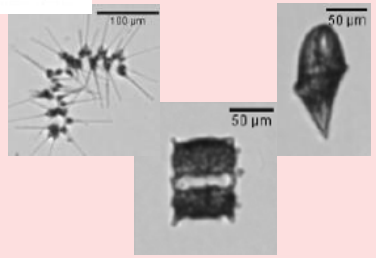


- Improving **automation of phytoplankton data analysis** from optical/imaging sensors
- Adapting training sets to previously-characterized phytoplankton communities in the studied area ("**adaptive learning**")
- Guiding classification steps by introducing some pairwise constraints ("**constrained clustering**")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

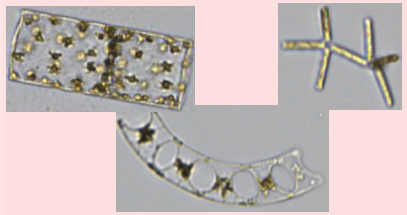
Data acquisition

Imaging in flow devises

YOKOGAWA
FLUID IMAGING TECHNOLOGIES, INC.



FlowCam VS Series



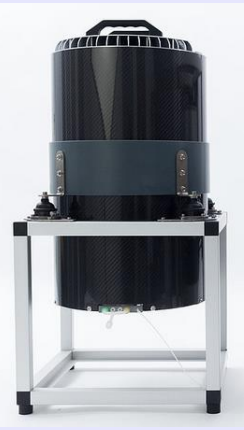
FlowCam 8100

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

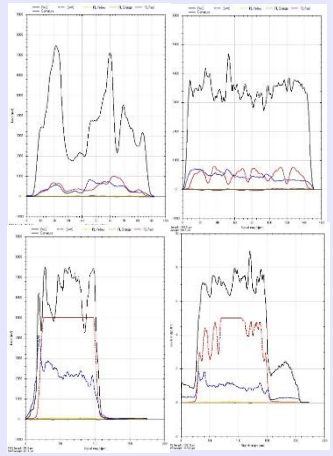
Pulse shape-recording flow cytometers

flow cytometry solutions
CytoBuoy

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** »

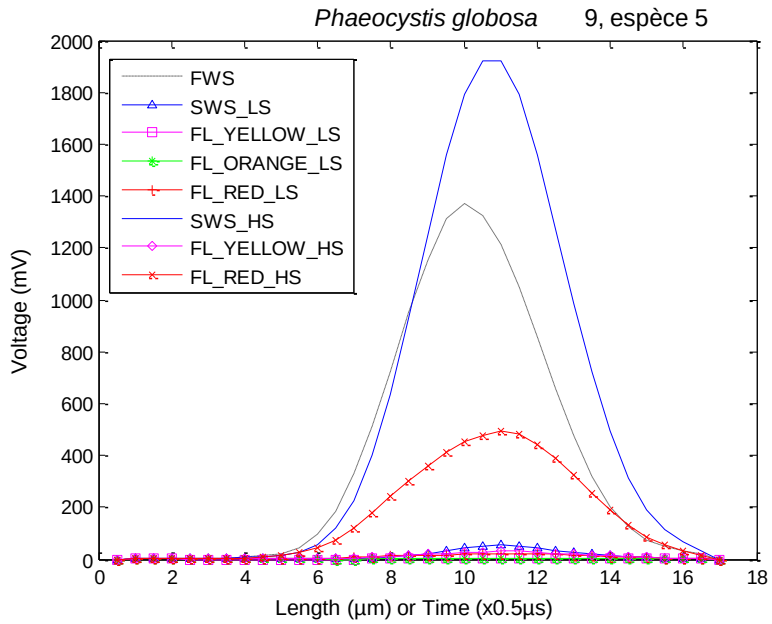
Phytoplankton species recognition: an automated classification problem

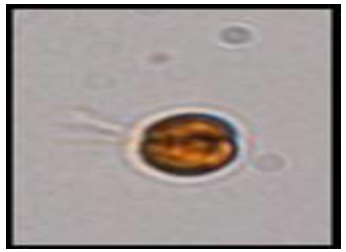
Cytometric pulse-shapes

- 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





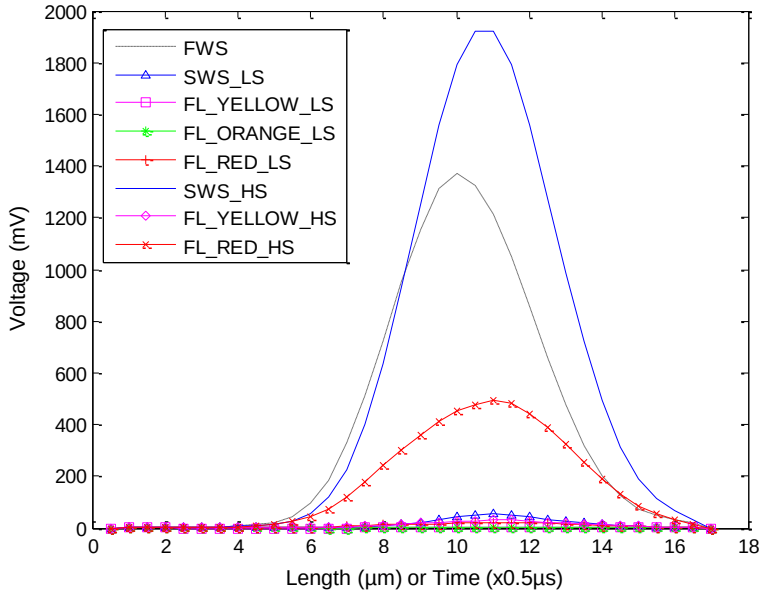
microscopic photo

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 :

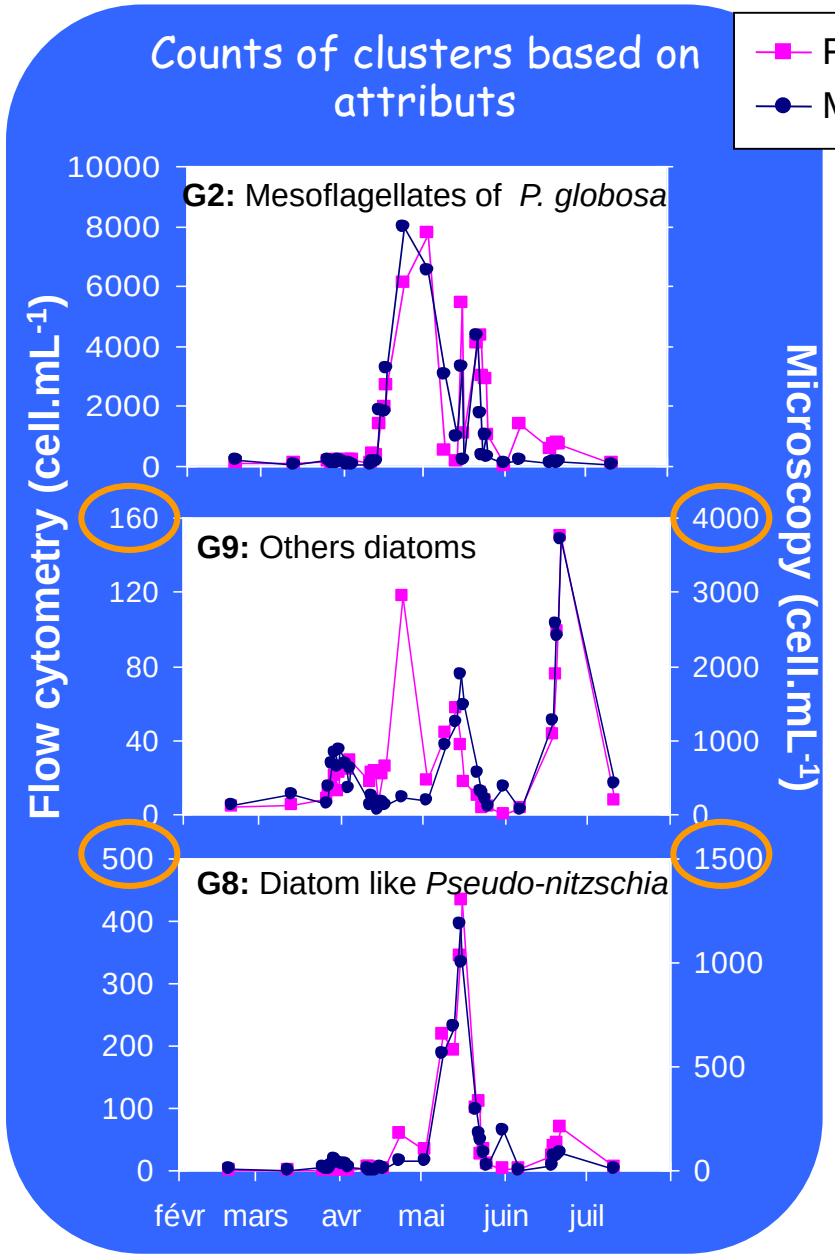
- Make the species recognition process automatic.
- Build a comparison measure between cytometric profiles robust to:
  - the intra-species variability,
  - the sensor sensibility.

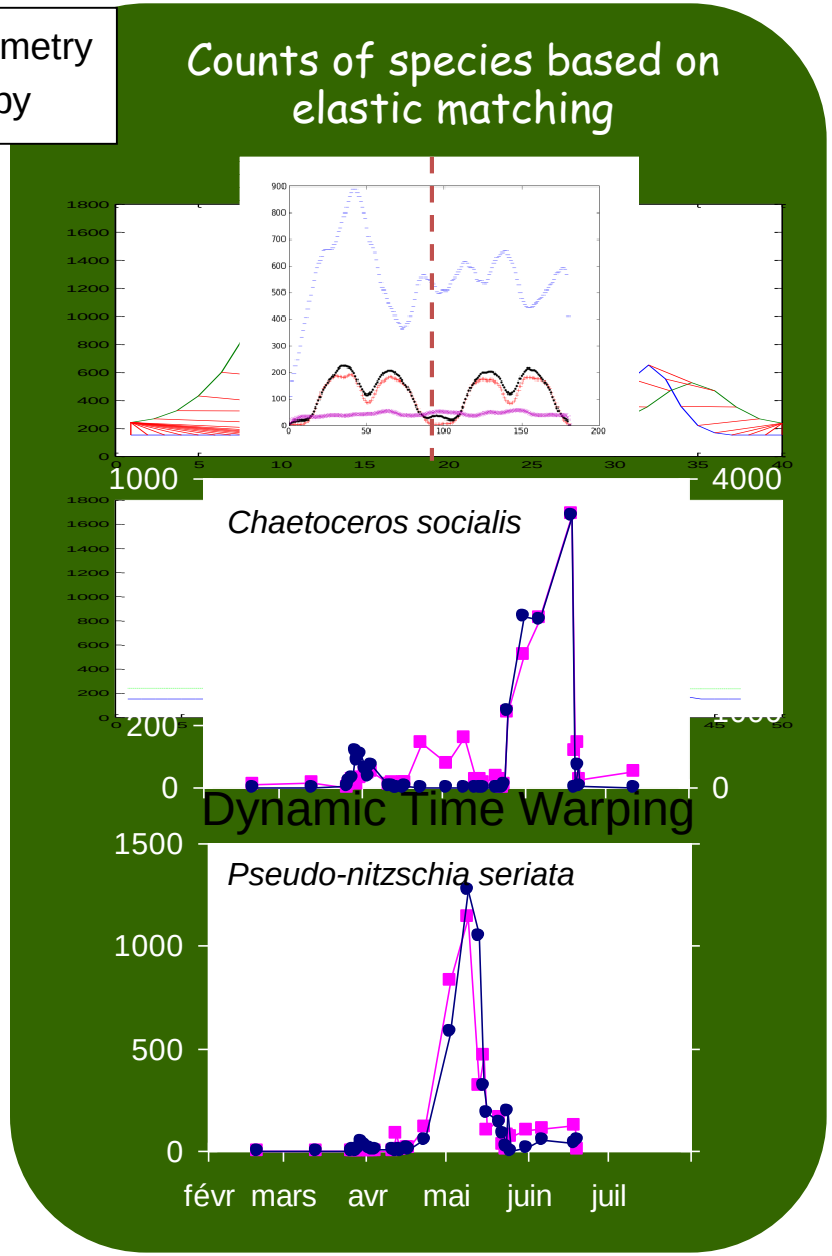
2 methods:

- Features
- Signals

Caillault et al. 2009 a,b

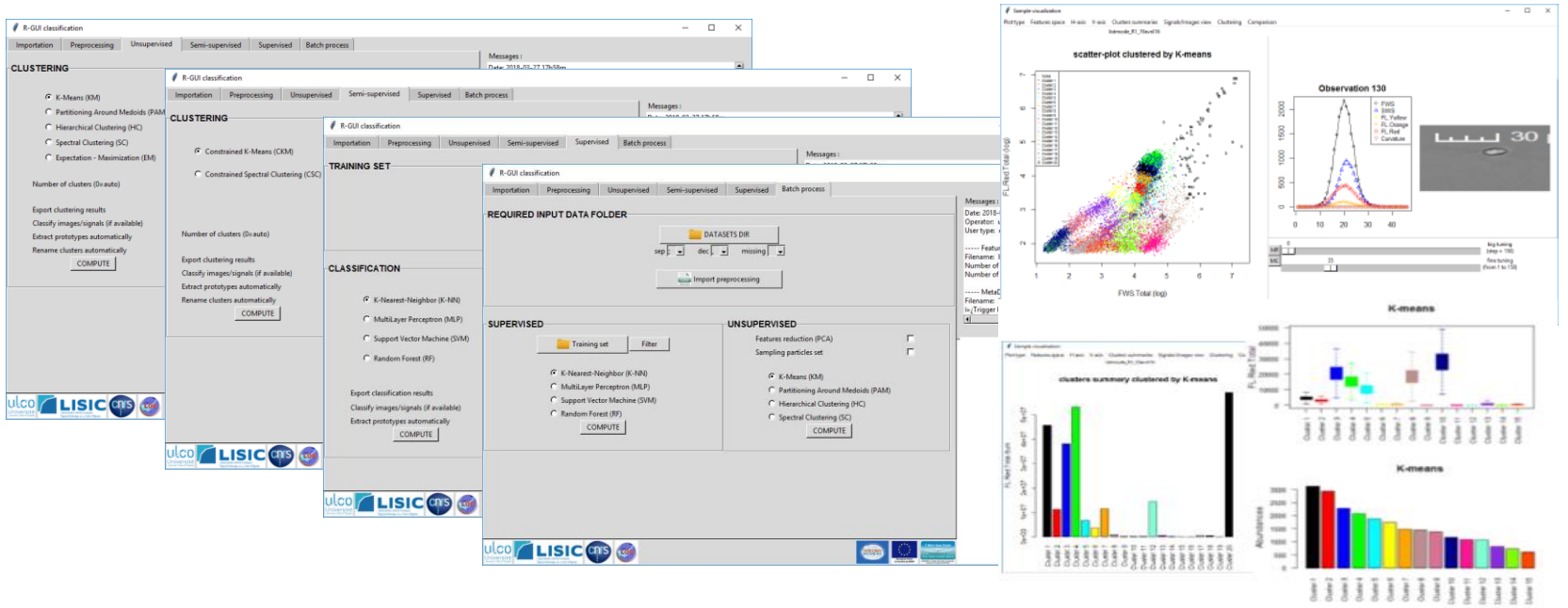
PSFM flow cytometry vs microscopy : manual and automated methods





N. Guiselin (PhD, 2010) ; G. Wacquet (PhD,2011) ; Caillault et al., 2009

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.).

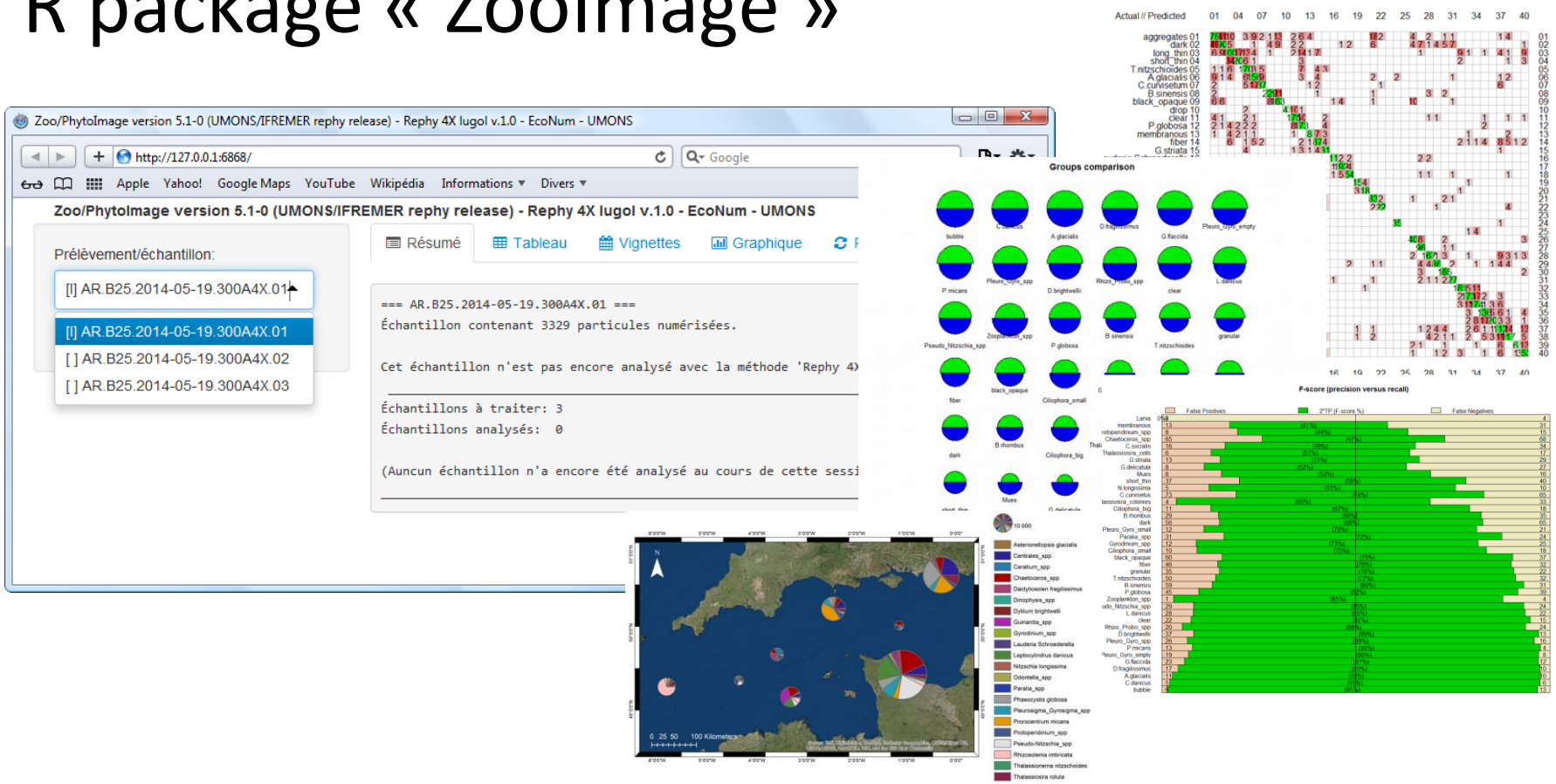
✓ **Supervised**, **unsupervised** and **semi-supervised** (with pairwise constraints) classification.

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 « ZooImage »

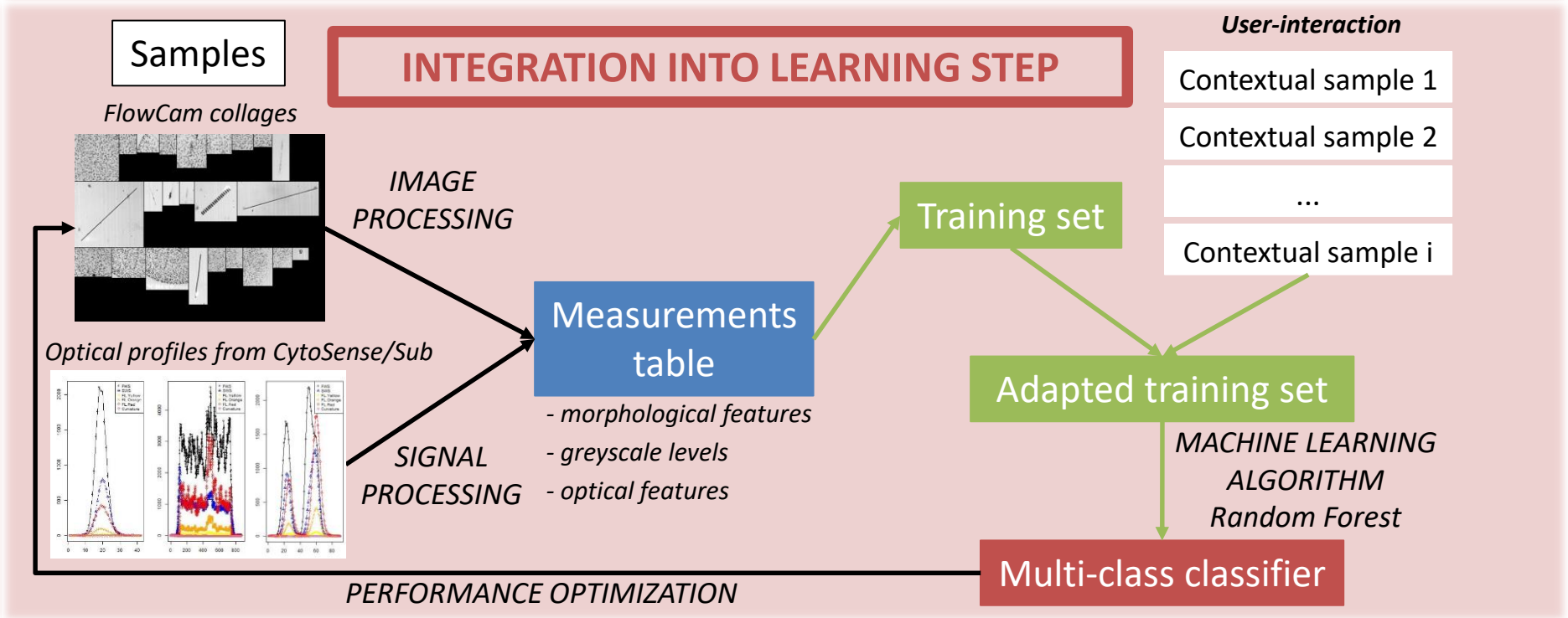


✓ Free software (open source), written in R, and specialized in the classification of digital images of zoo- or phytoplankton.

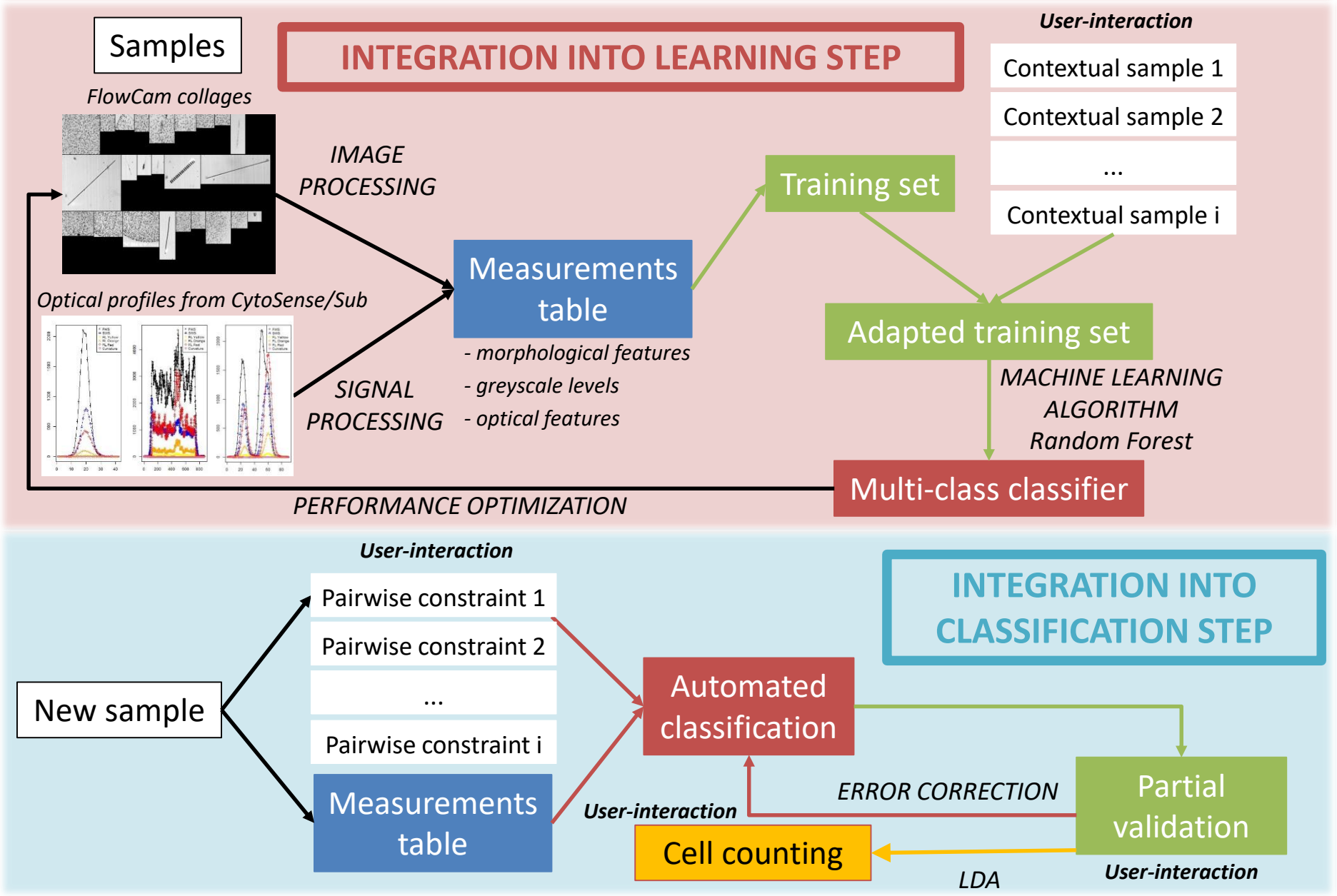
✓ For **any kinds** of plankton images, i.e. from FlowCam, Imaging FlowCytoBot, ZooScan, micro- or macrophotography, …

Download: https://cran.r-project.org/web/packages/zooimage/

Classification process



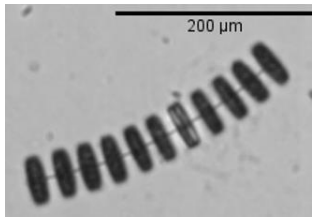
Classification process

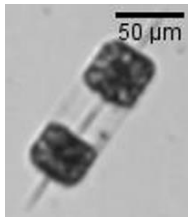


FlowCam

Device: FlowCAM (B2 series)
Magnification: 4X
Chamber depth: 300µm

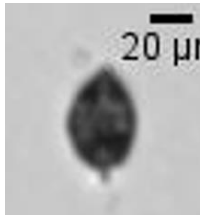
Gyrodinium spp.

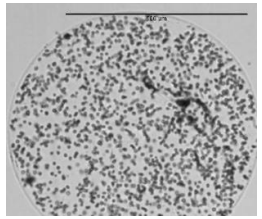




Thalassiosira spp. *Ditylum spp.*

Chaetoceros spp.

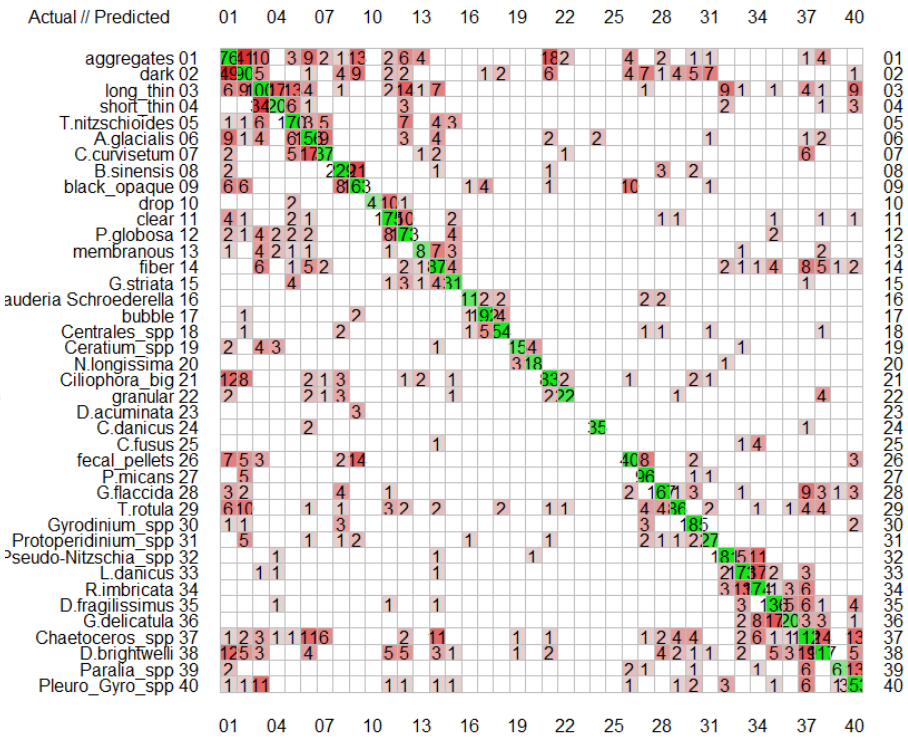




Prorocentrum spp. *Phaeocystis spp.*

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



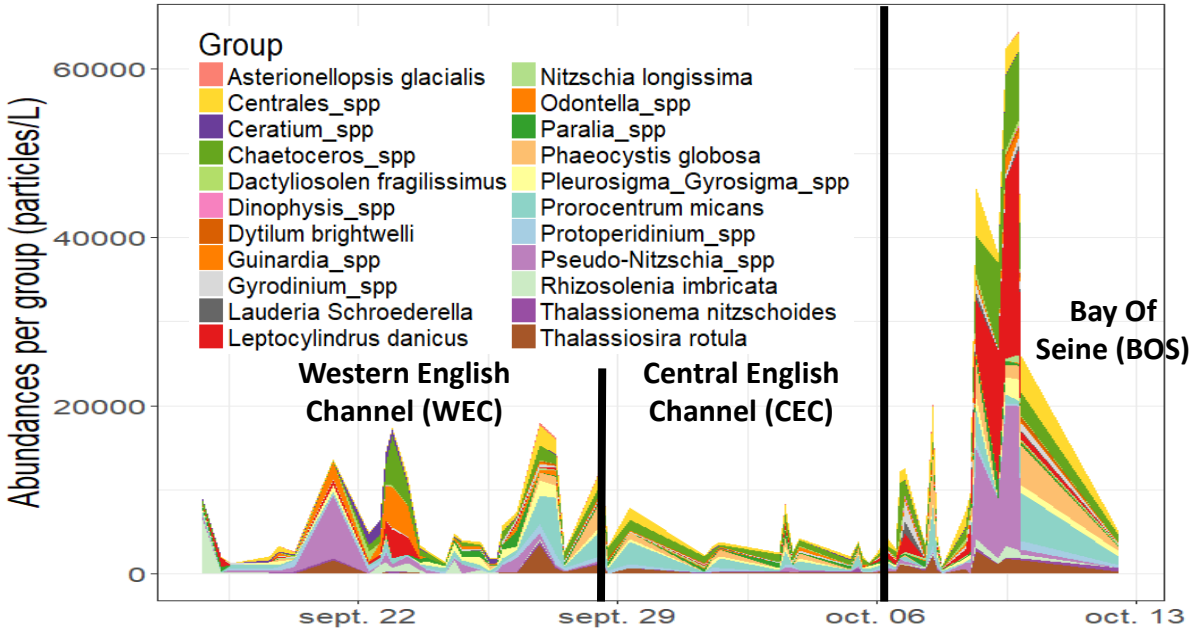
FlowCam - CAMANOC cruise

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



Discrete samples…

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.

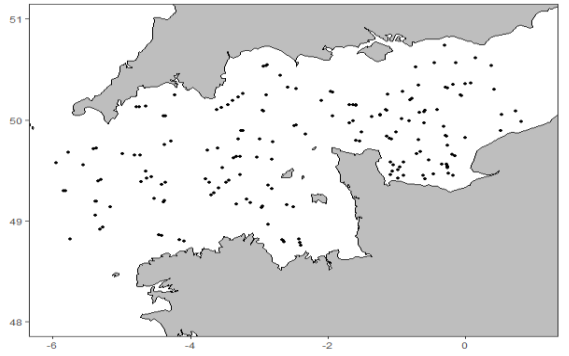
FlowCam - CAMANOC cruise

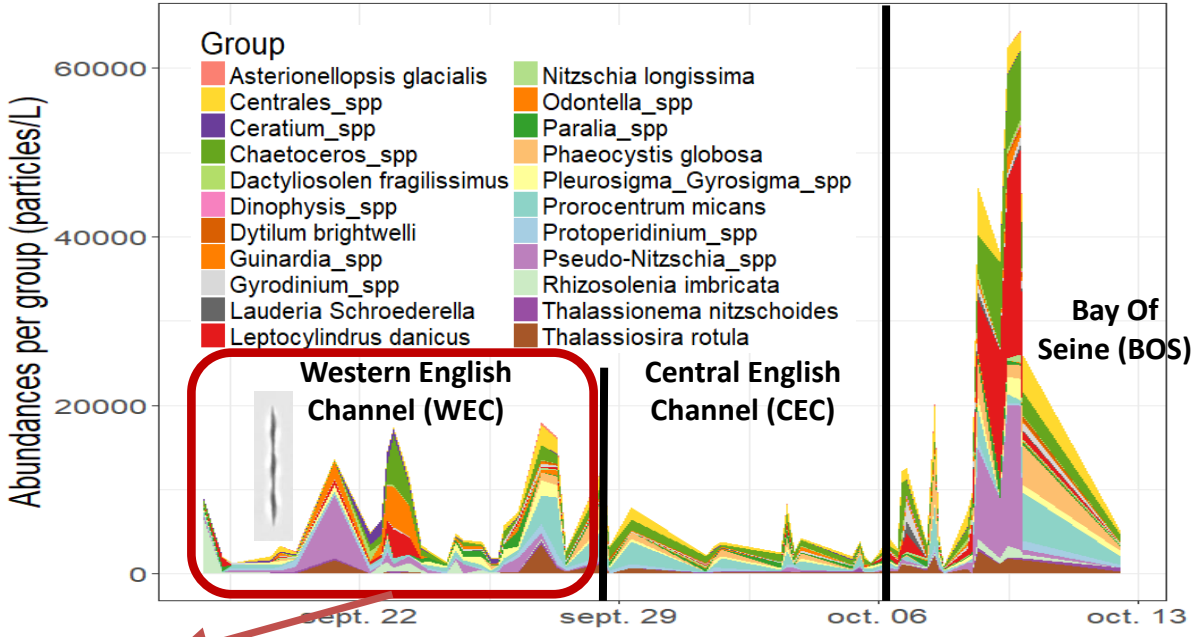
Multidisciplinary campaign: an ecosystem approach to fisheries

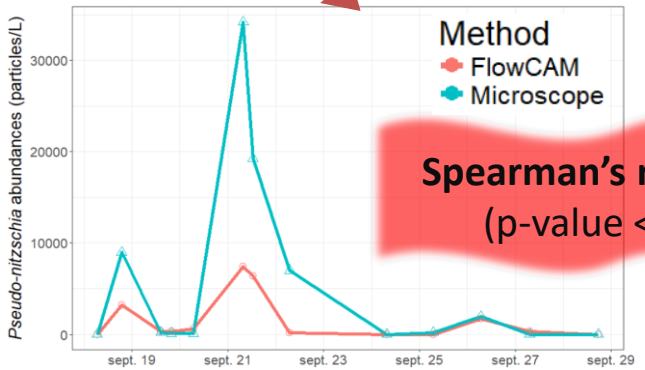
Period: 16th September-12th October, 2014 in Channel

Ship: R/V « Thalassa II » - IFREMER

Adaptive training set: 1 sample selected every 2 days







Spearman's rho = 0.74
(p-value < 0.05)

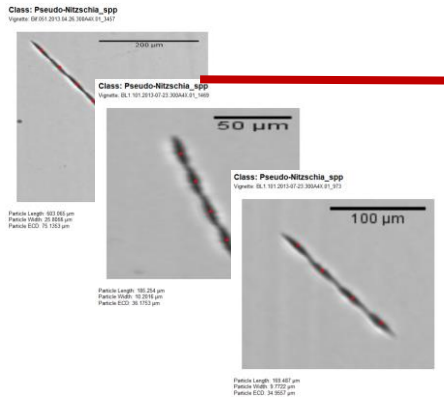
Discrete samples…

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

→ Abundance estimation for FlowCAM lower than for Microscope.

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.

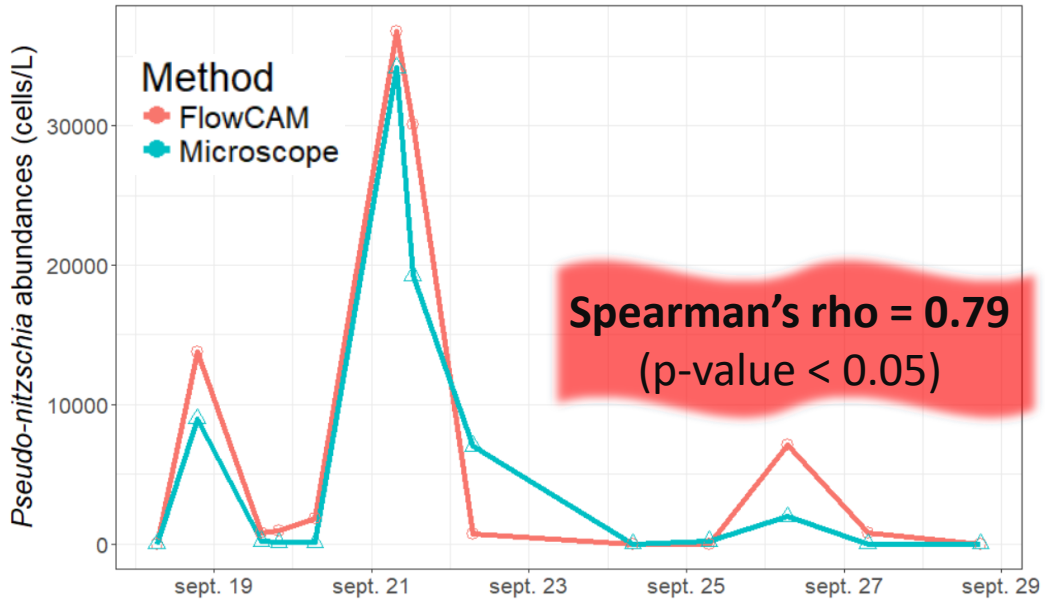
FlowCam - Counting cells in colony*



Multivariate
NbCells ~ measures

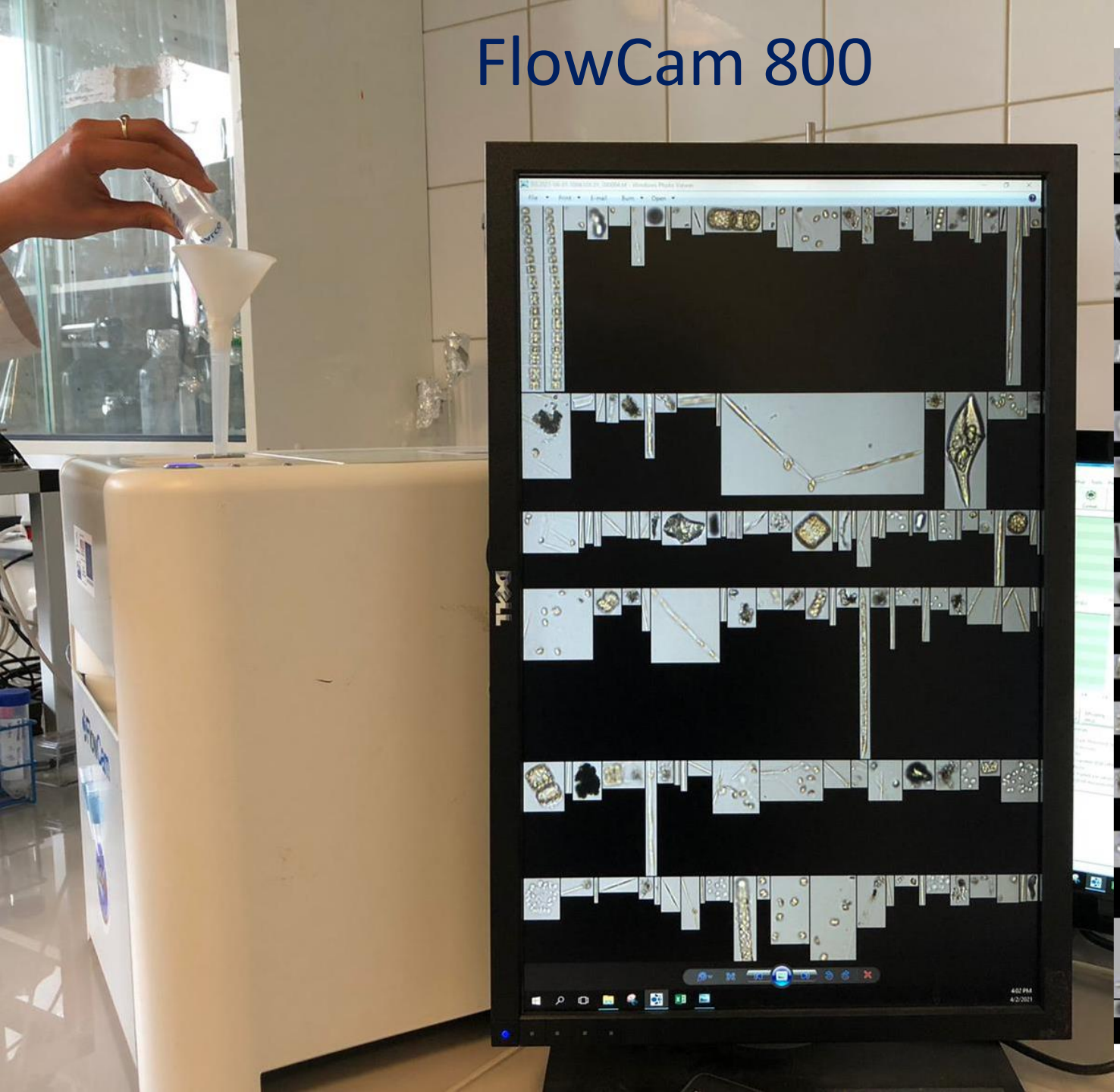
Predictive model
Linear Discriminant Analysis

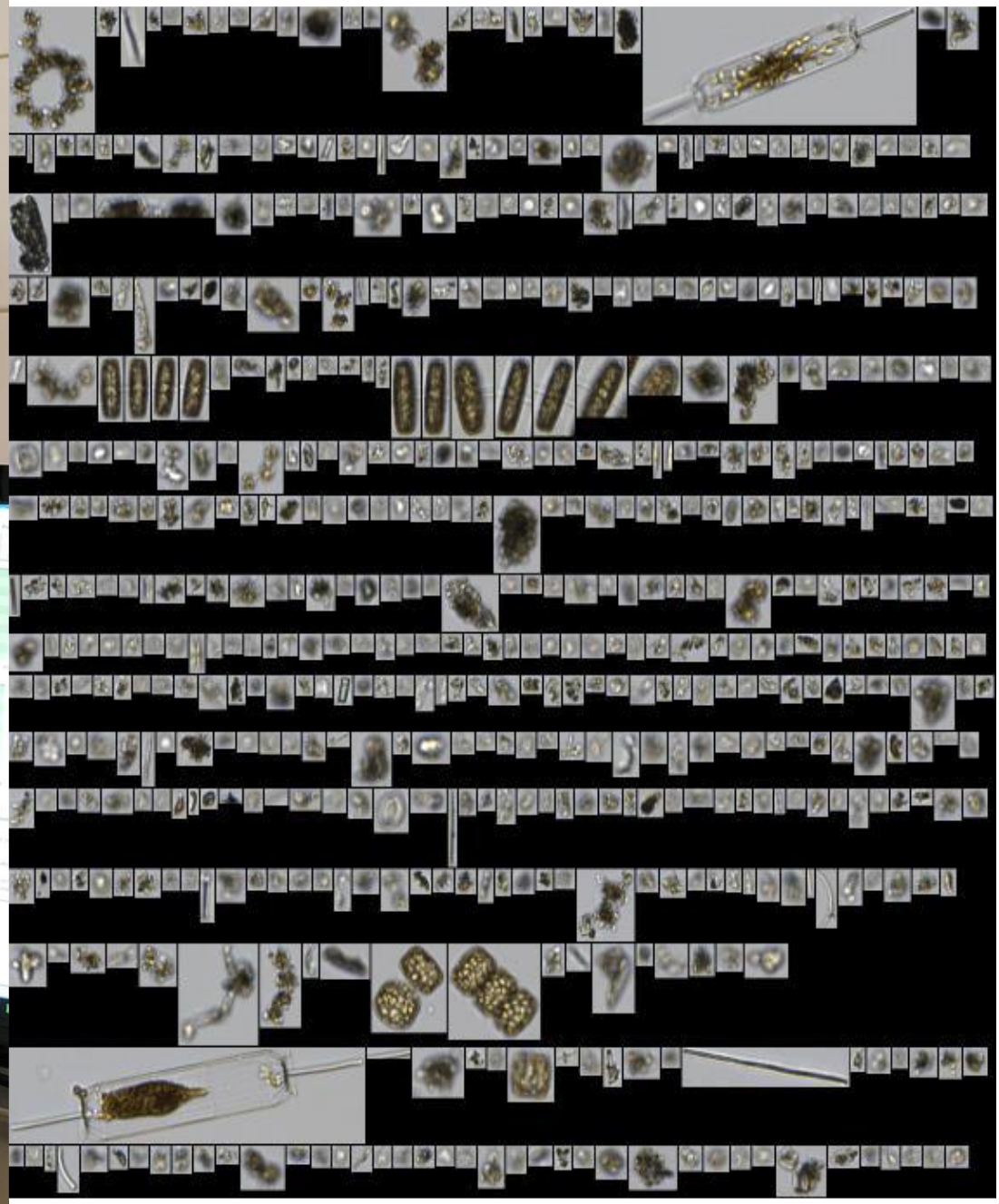
Prediction



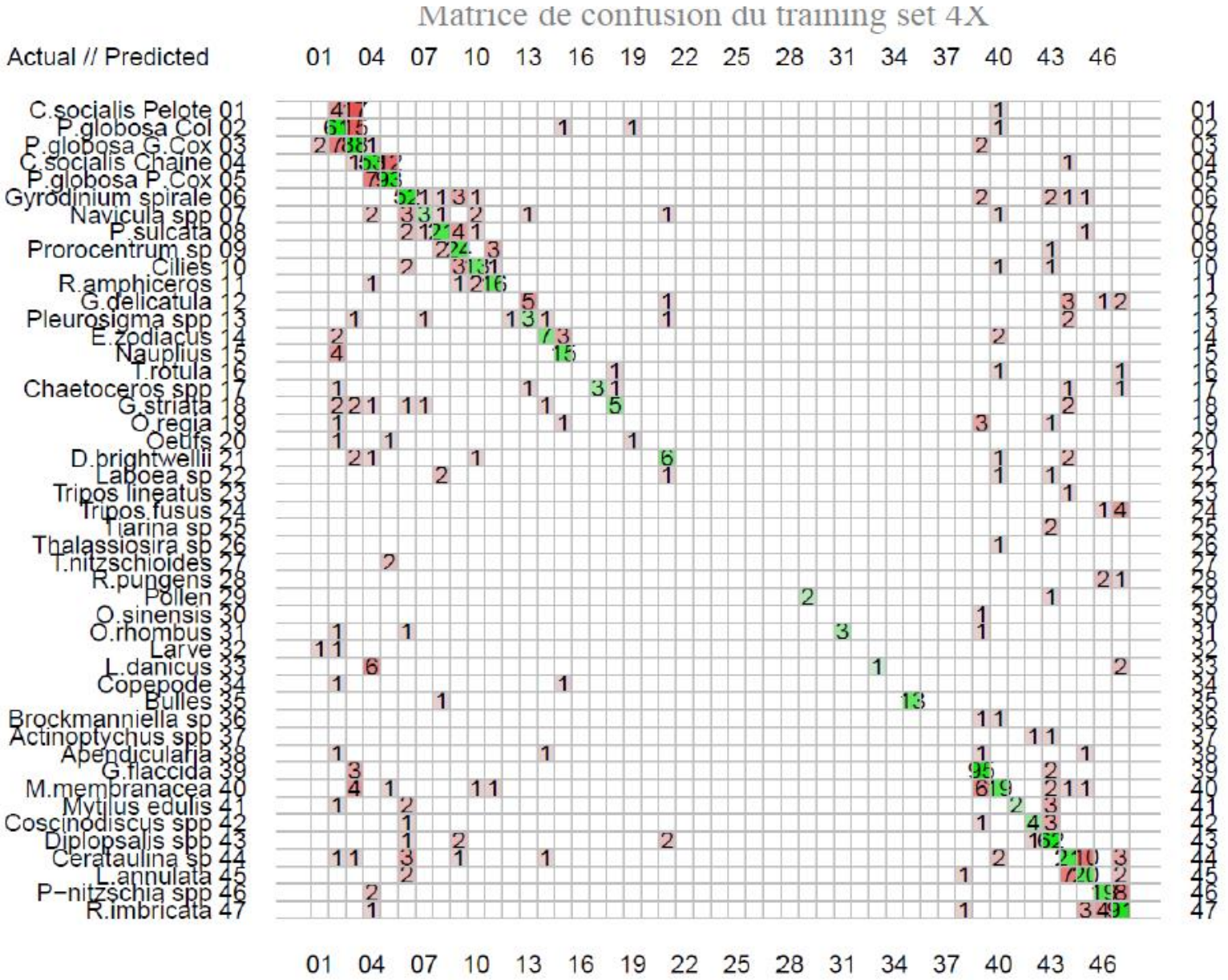
* **Wacquet *et al.*** « Combination of machine learning methodologies and imaging-in-flow systems for the automated detection of Harmful Algae ». 2020. Proceedings of the 18th International Conference on Harmful Algae, 21-26 October 2018, Nantes, France.

FlowCam 800

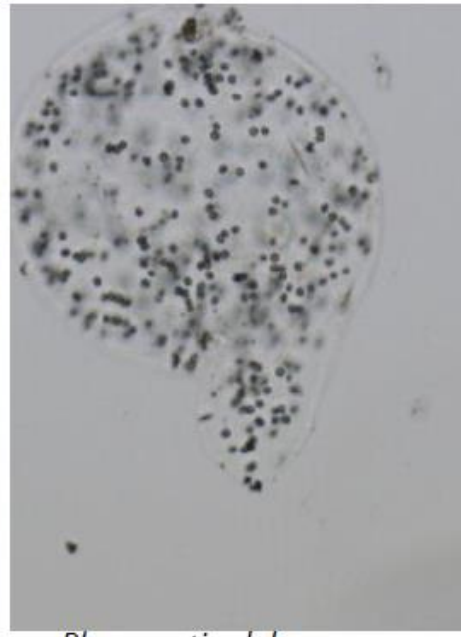




New image training set (FlowCAM 8000 Series)

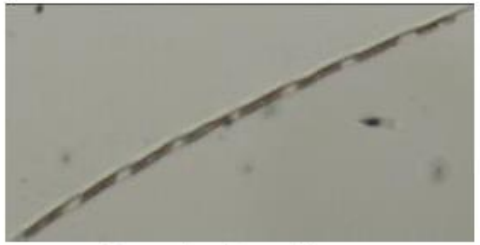


Chaetoceros socialis



Phaeocystis globosa

Rhizosolenia sp

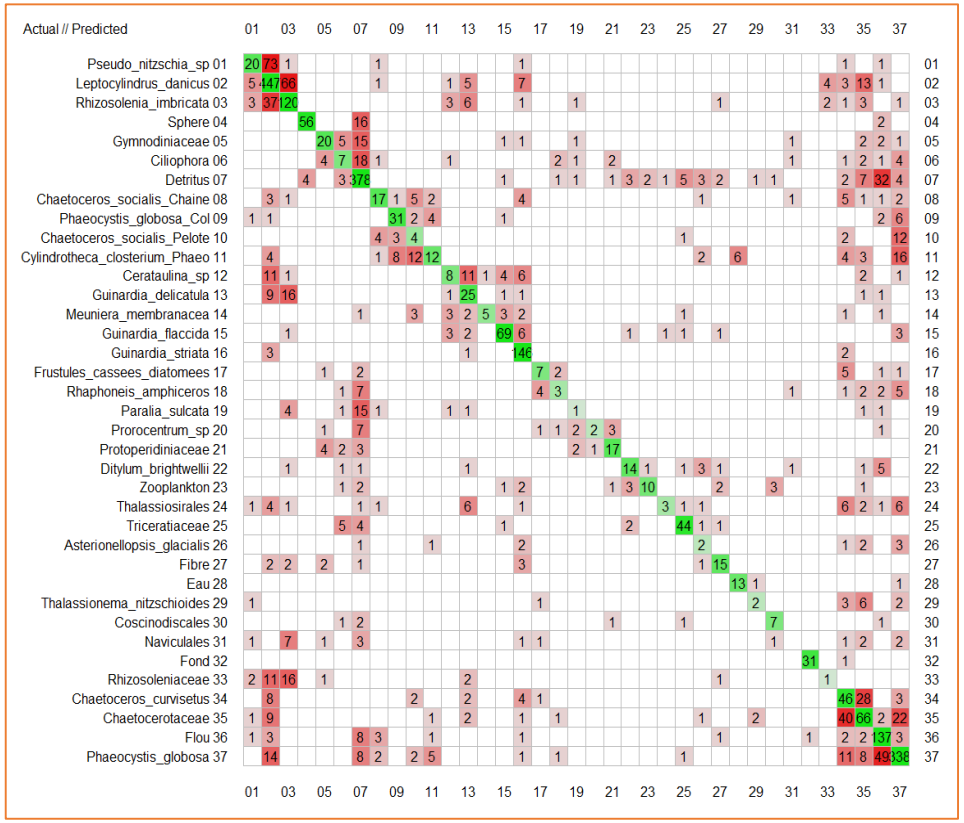


Pseudo-nitzschia sp

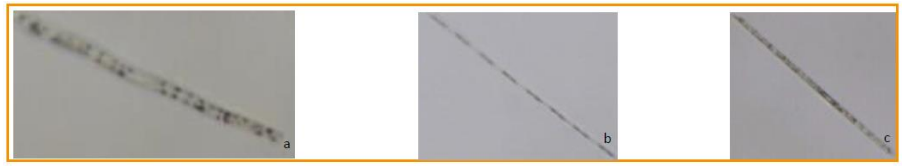
E. Delecroix, M.Sc. 2019

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

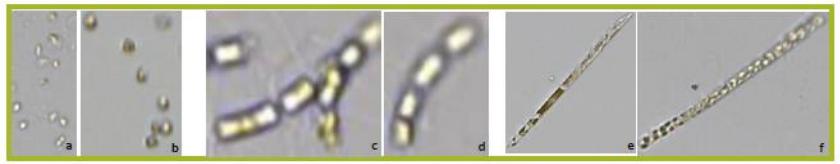
X4



X10



Vignettes X4
(A. *LEPTOCYLINDRUS DANICUS*, B. *PSEUDO-NITZSCHIA*, C.*RHIZOSOLENIA IMBRICATA*)



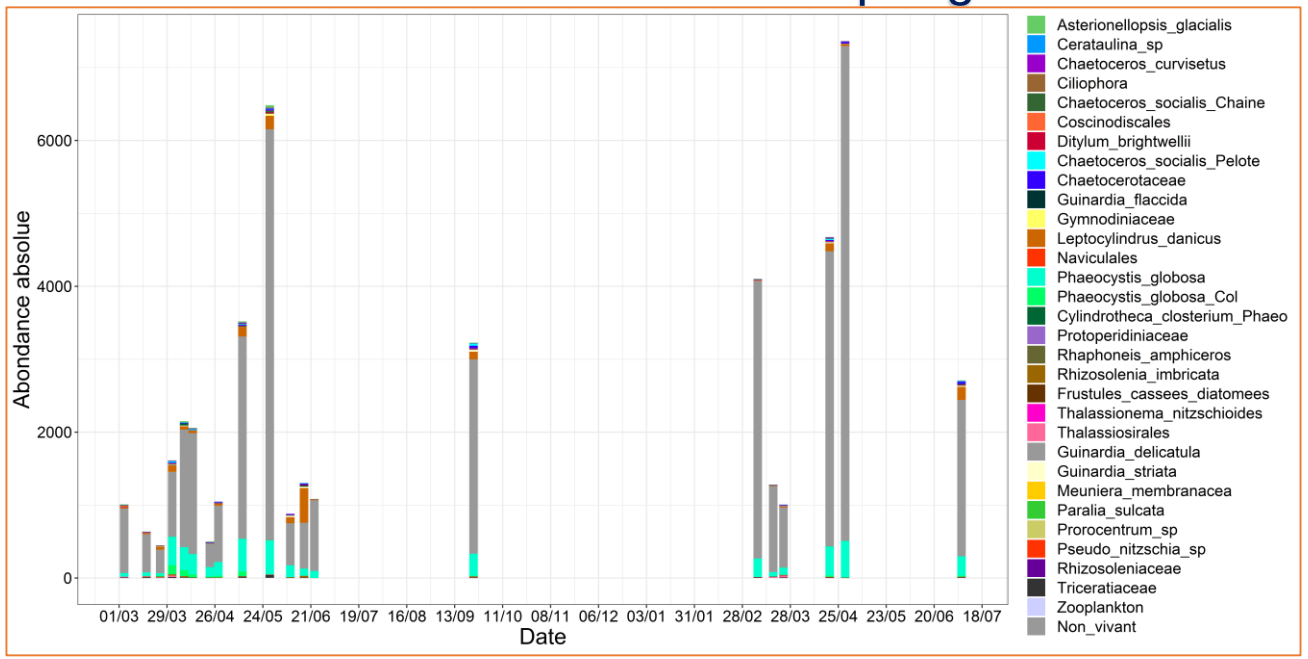
Vignettes X10
(A. *PHAEOCYSTIS GLOBOSA COX*, B. *PHAEOCYSTIS GLOBOSA, C. CHAETOCEROS SOCIALIS* , D. *CHAETOCEROS SPP, E. RHIZOSOLENIA IMBRICATA, F. LEPTOCYLINDRUS DANICUS)*

F. Verhaeghe, Eng. Int. 2022 – Wacquet & Lefebvre (2022)

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

X4

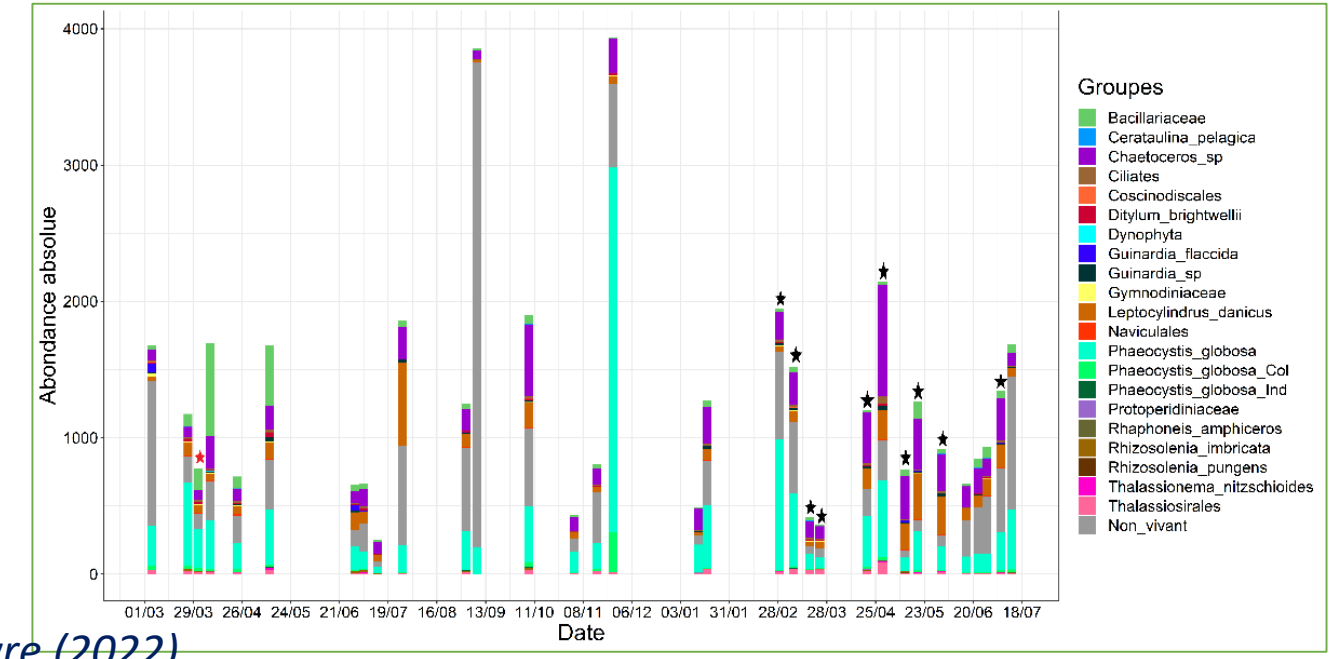
Majority of non-living particles



X10

★ 100µm filtration

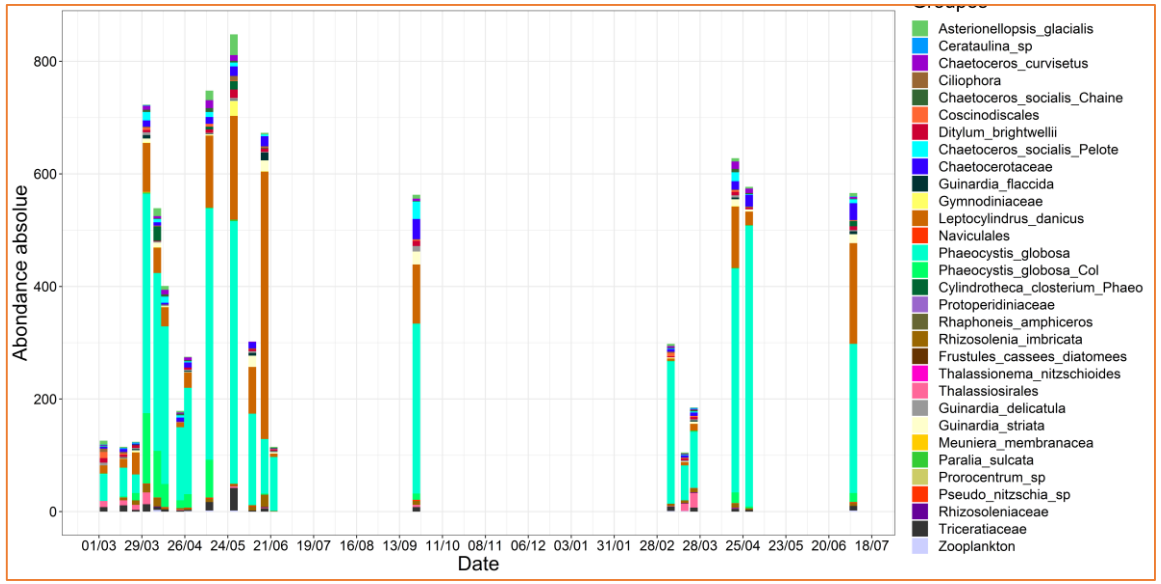
★ 150µm filtration

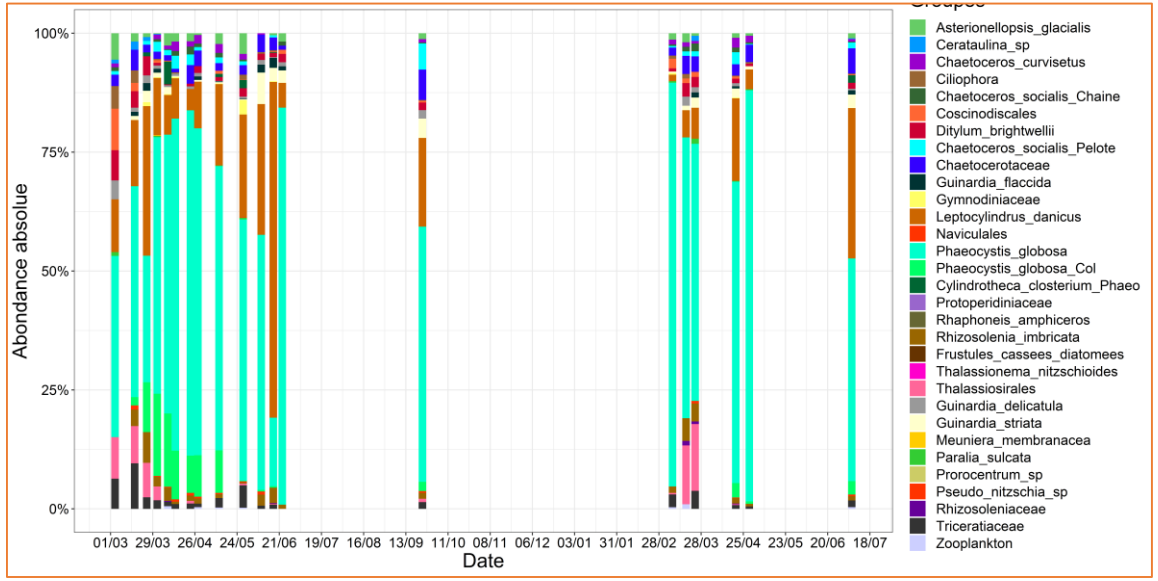


F. Verhaeghe, Eng. Int. 2022 – Wacquet & Lefebvre (2022)

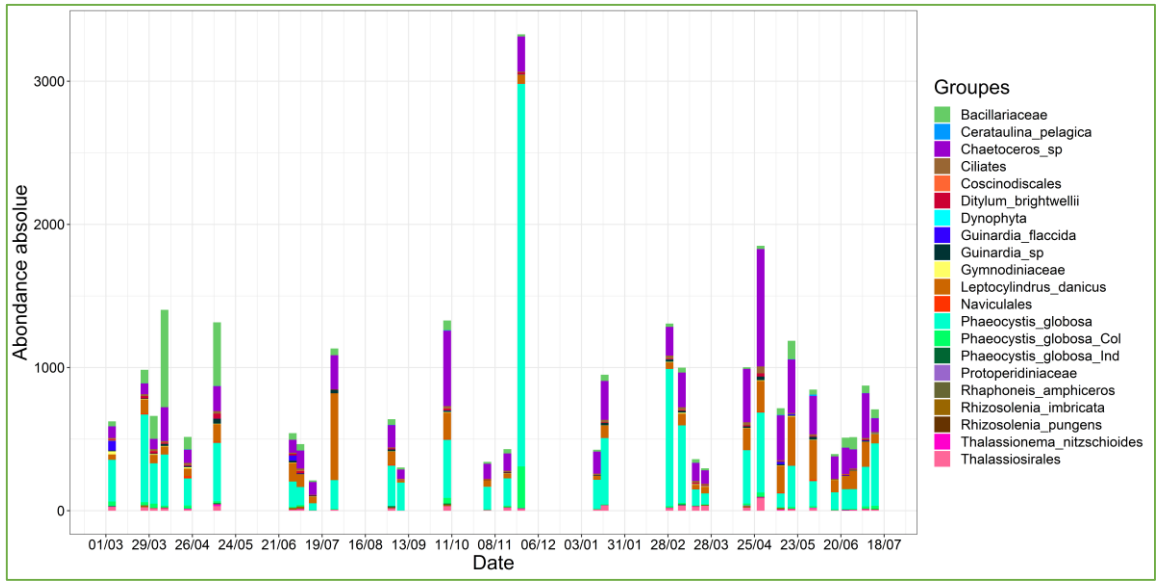
CNN Classifications results on point R1 of the DYPHYRAD transect from spring 2021 to summer 2022)

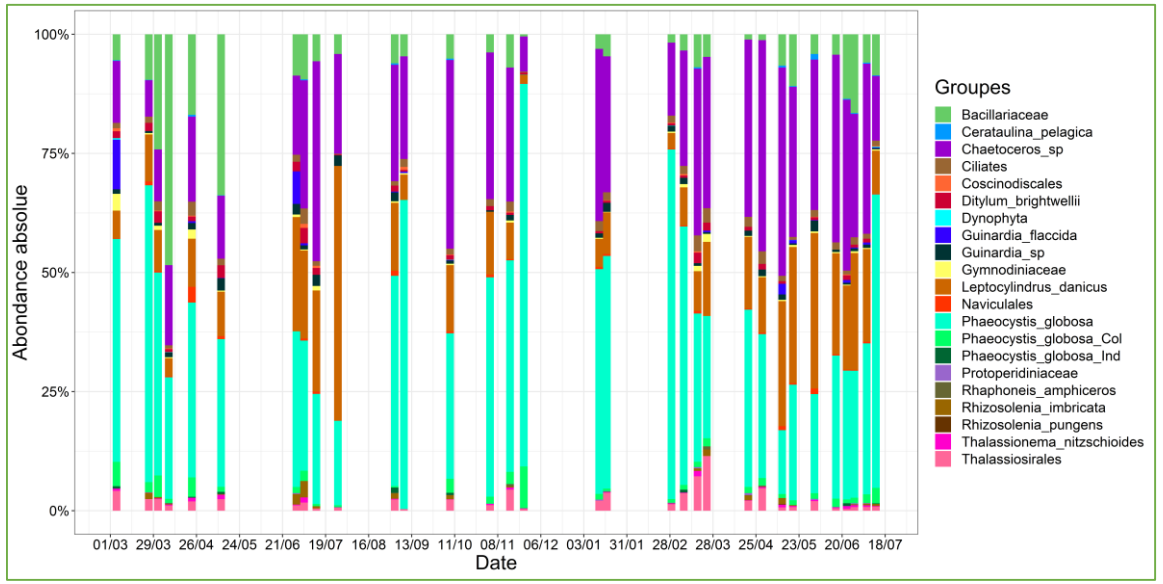
X4





X10

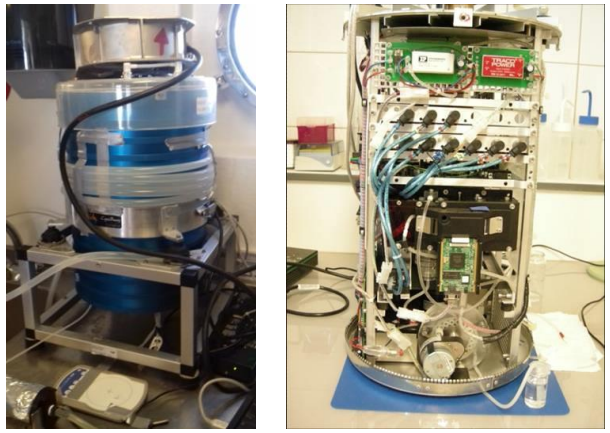




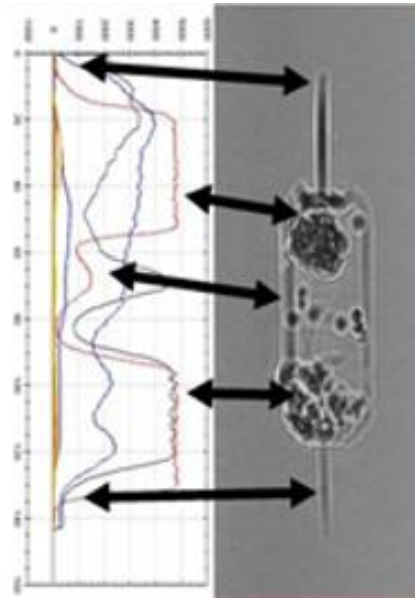
F. Verhaeghe Eng. Int. 2022 – Wacquet & Lefebvre (2022)

Flow Cytometry

Device: Pulse-Shape recording Flow Cytometer
Models: CytoSense/CytoSub
Trigger level: Red fluorescence set to 10mV

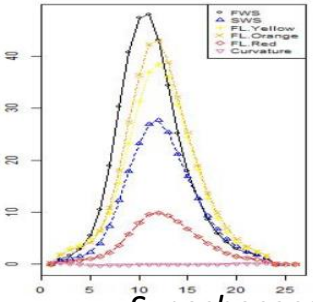


CytoSense

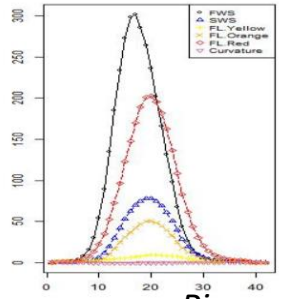


Training set

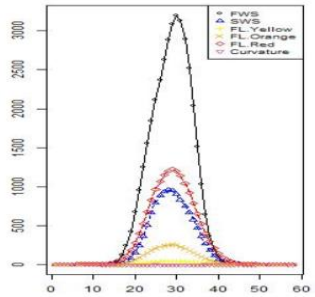
Training set: DYPHYRAD transect
Number of groups: 4 functional groups (synechoccocus, pico- & nano-eukaryotes, micro-phytoplankton) + 4 noise
Number of vignettes: 372 (~50/group)
Classifier: Random Forest (ntree = 500)
Cross-validation: 10 folds



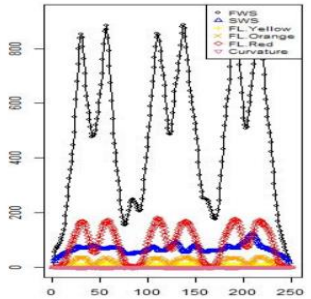
Synechococcus



Pico



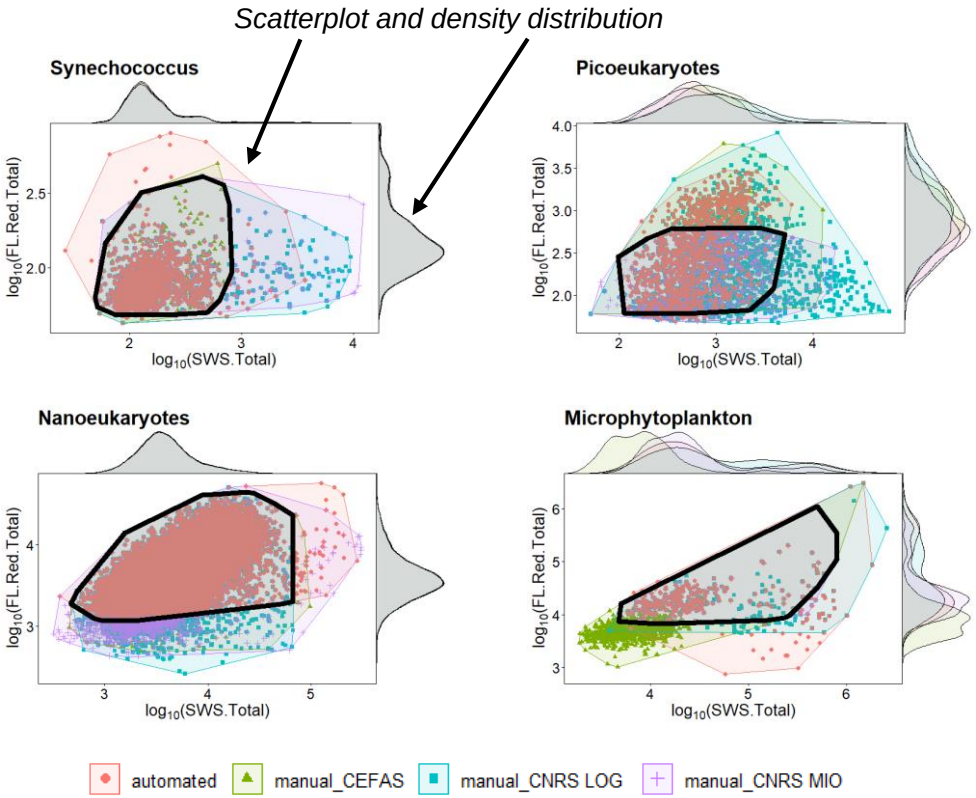
Nano



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-MOI			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)

* **Wacquet *et al.*** « Comparison between automated analysis of flow cytometry data using RclusTool package and manual clustering for *in vivo* phytoplankton recognition ». *In prep.*

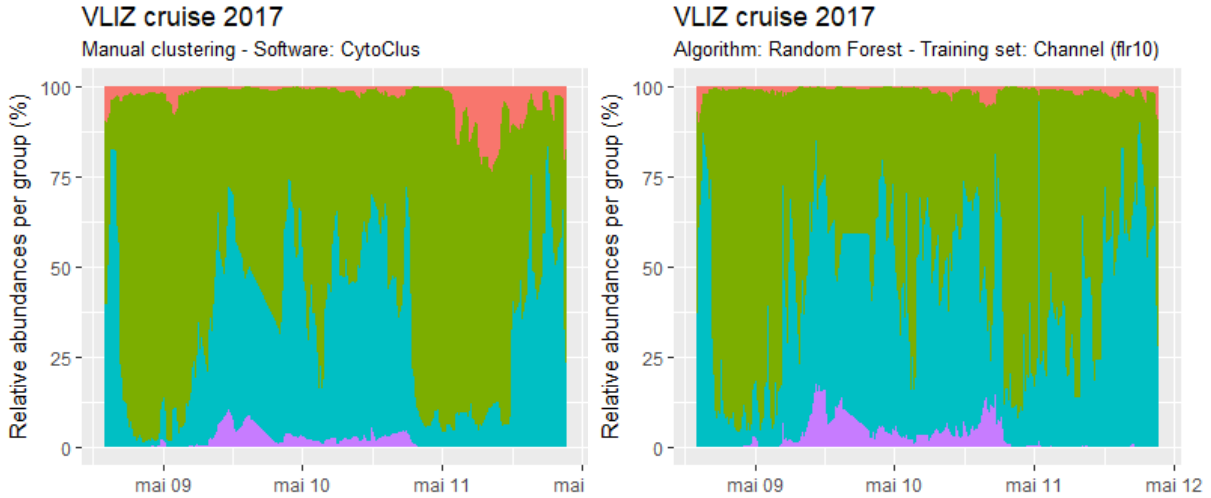
Flow cytometry - LifeWatch cruise

Scientific campaign for analysis of plankton diversity (LifeWatch/JERICO-NEXT cruise, R/V “Simon Stevin”-VLIZ).

Period: 8th May-12th May, 2017

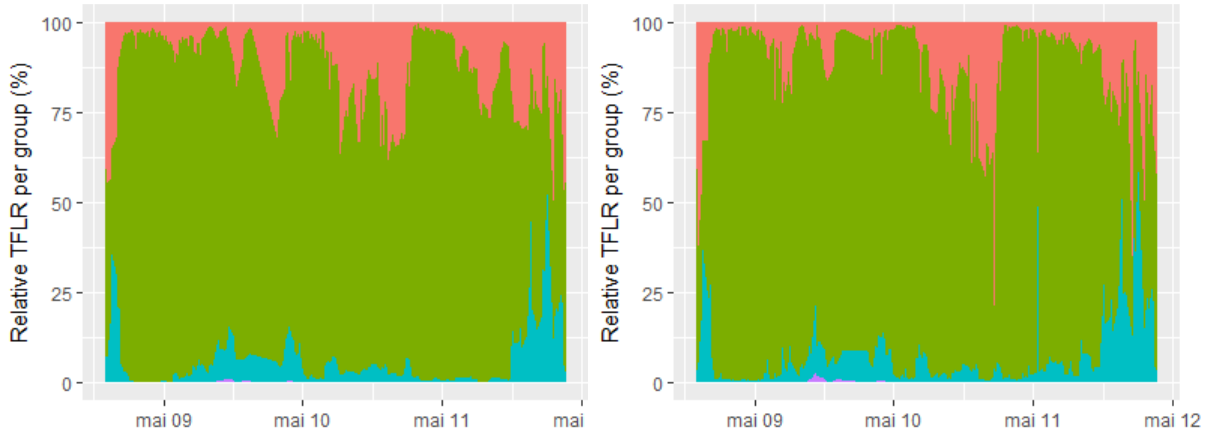
Area: North Sea

Ship: R/V « Simon Stevin » - VLIZ



Groups	ρ 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

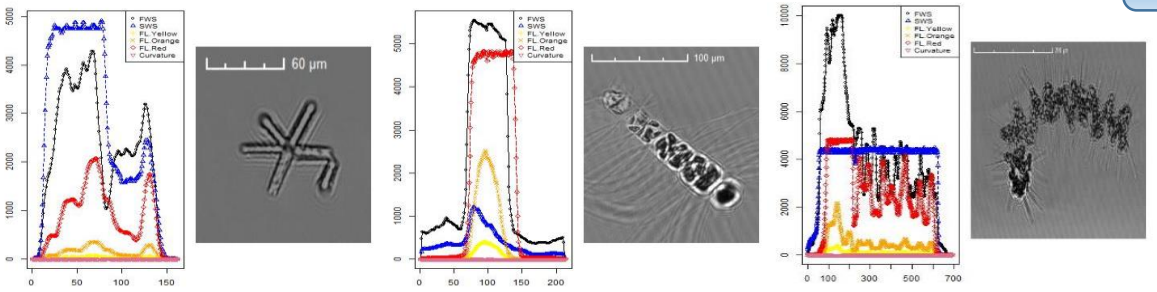


Groups	ρ tflr	pvalue
Synecho	0.95	<2e-16
Pico	0.59	<2e-16
Nano	0.95	<2e-16
Micro	0.81	<2e-16
Total	0.95	<2e-16

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

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

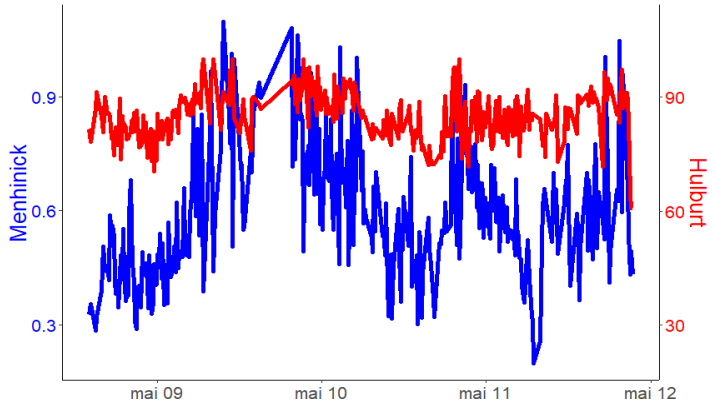


Thalassionema spp. *Chaetoceros spp.* *Asterionellopsis spp.*

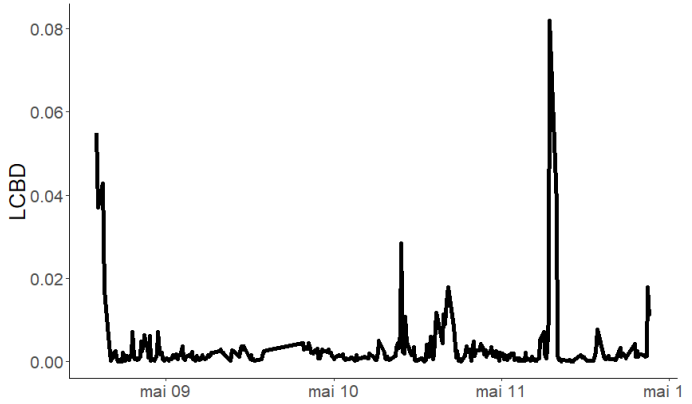
Guinardia spp. *Thalassiosira spp.* *Ditylum spp.*

Used for: - counting cells in colony
- computation of functional traits, diversity indices (MSFD), …

Functional traits



alpha diversity

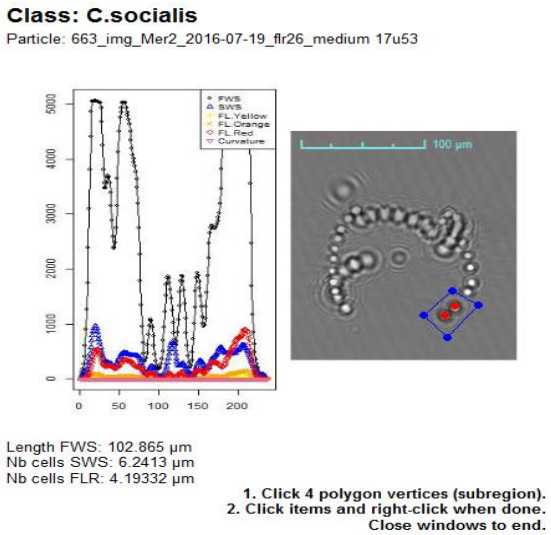


beta diversity (LCBD)

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

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

Proposed method

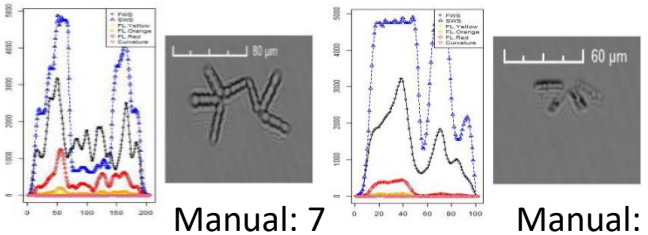


Calibration of predictive models (one per taxonomic group) for the estimation of the number of cells in colonies, based on ***manual counts*** made on the images classified in training sets.

Algorithm: **Linear Discriminant Analysis**

Some examples and results…

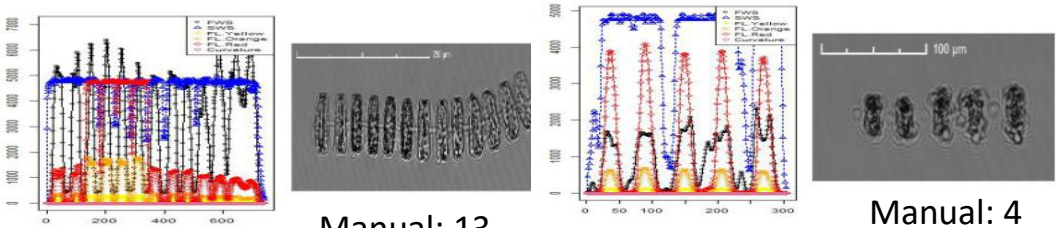
Thalassionema nitzschoides



Manual: 7
Proposed: 6

Manual: 4
Proposed: 4

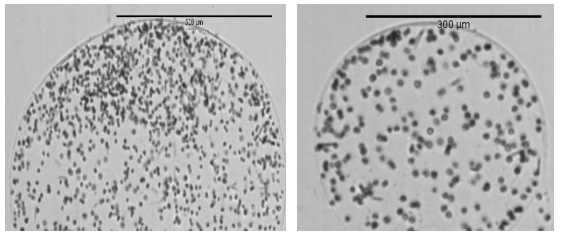
Thalassiosira rotula



Manual: 13
Proposed: 12

Manual: 4
Proposed: 4

Phaeocystis globosa

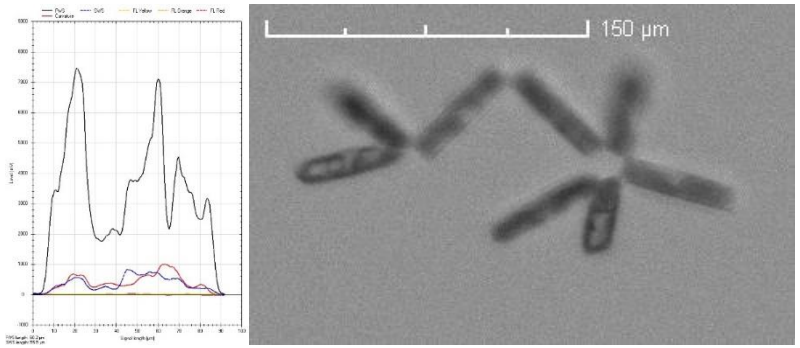


Proposed: 229

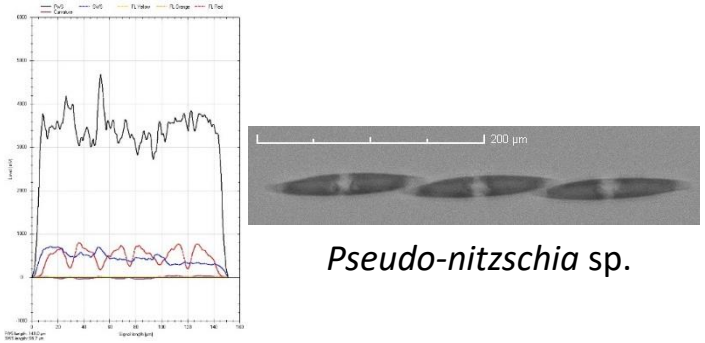
Proposed: 92

Applying new CNN classification on CytoSense images

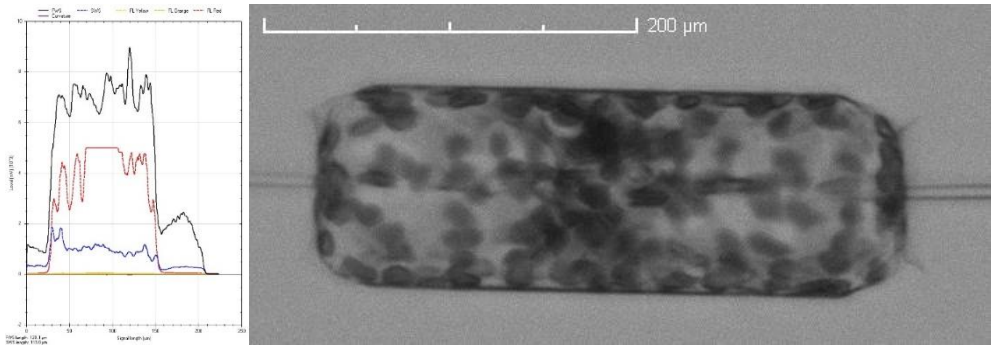
Combination of images and optical profiles



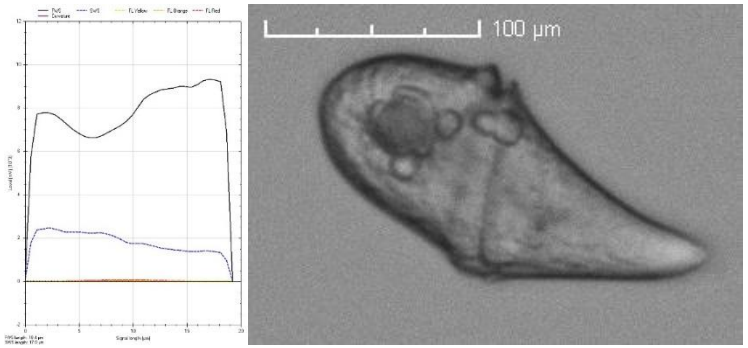
Thalassionema nitzschioides



Pseudo-nitzschia sp.

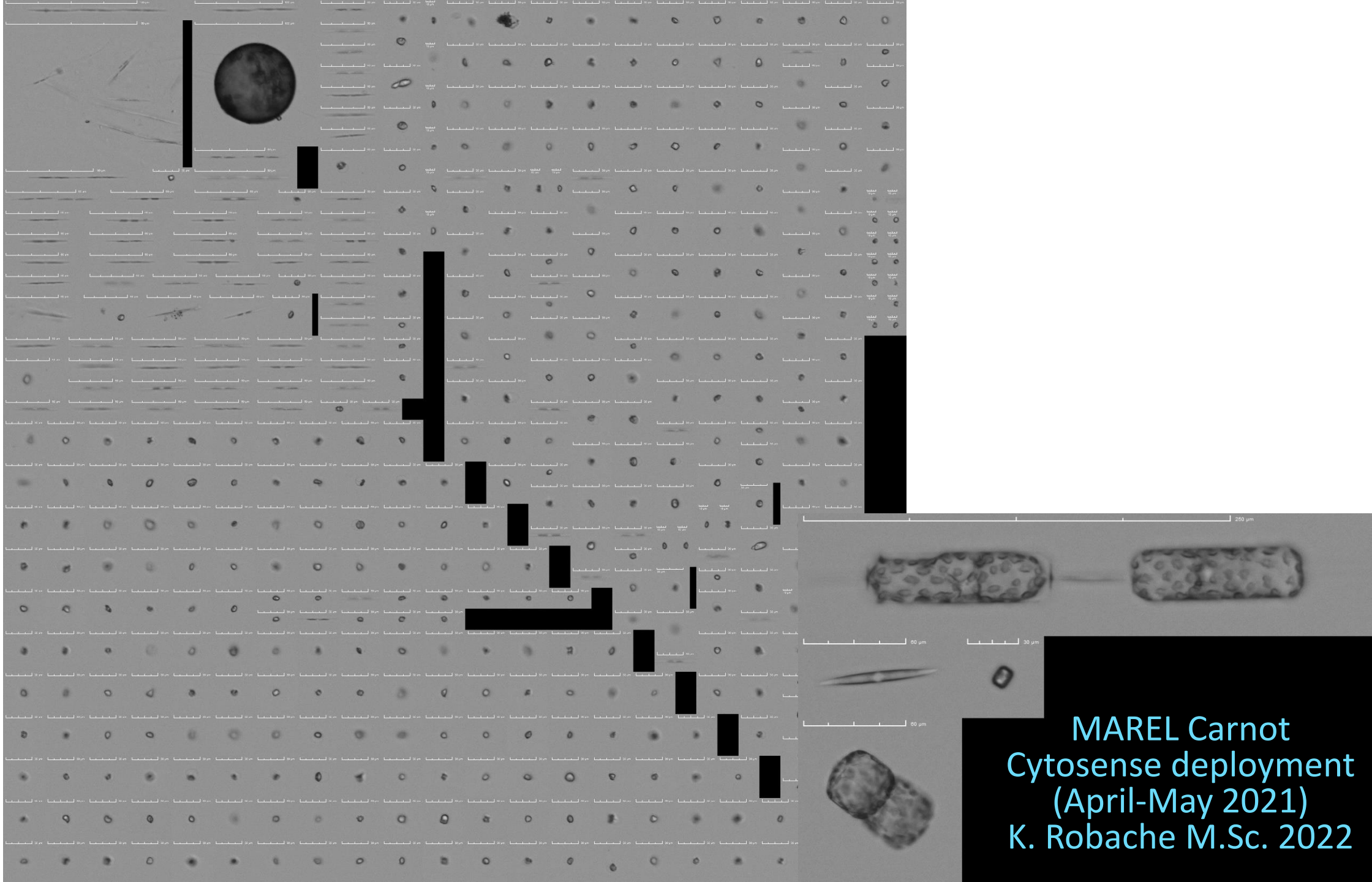


Ditylum brightwellii

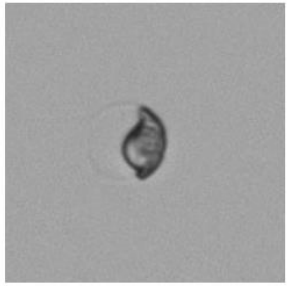


Gyrodinium spirale

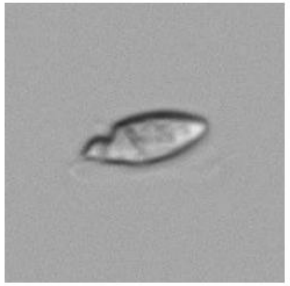
M. Ryckman M.Sc. 2022 – G. Wacquet unpubl.



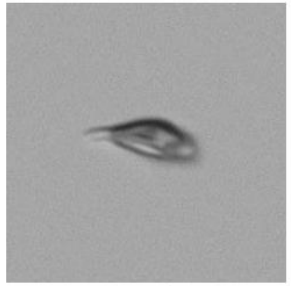
MAREL Carnot
Cytosense deployment
(April-May 2021)
K. Robache M.Sc. 2022



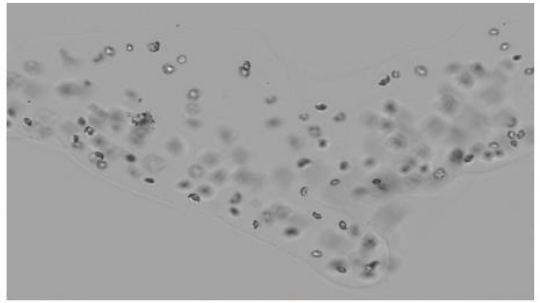
Phaeocystis globosa (cellule)



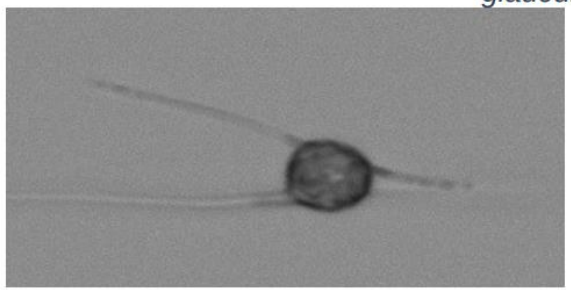
Exemple de Gymnodini Aceae (*Lebouridinium glaucum*)



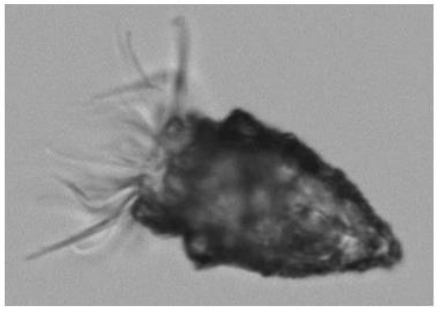
Cryptophyte



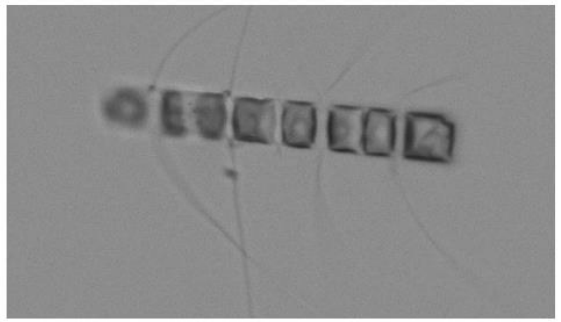
Phaeocystis globosa (colonie)



Tripos sp.

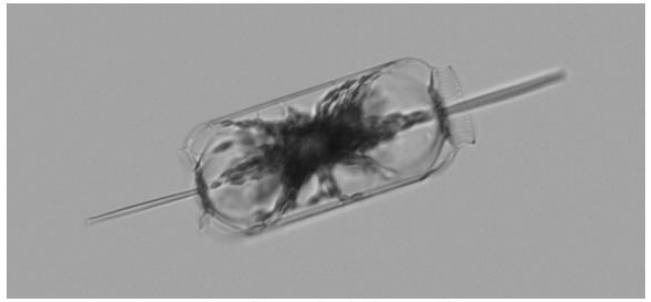


Ciliés

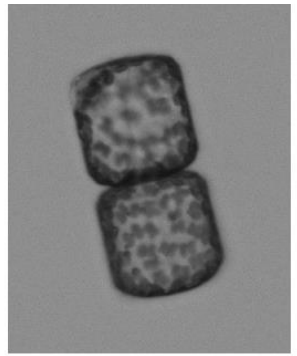


Chaetoceros sp.

Building a new training set on CytoSense images

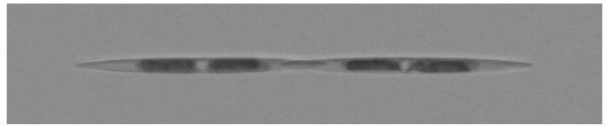


Dytilum brightwelii

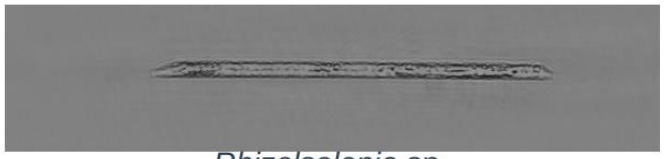


Detonula pumila

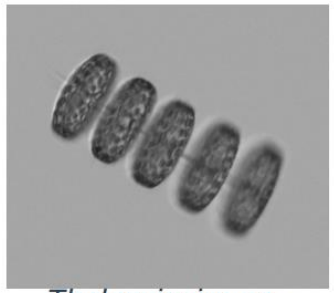
Guinardia striata



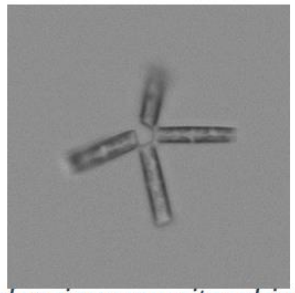
Pseudo-nitzschia sp.



Rhizolsolenia sp.



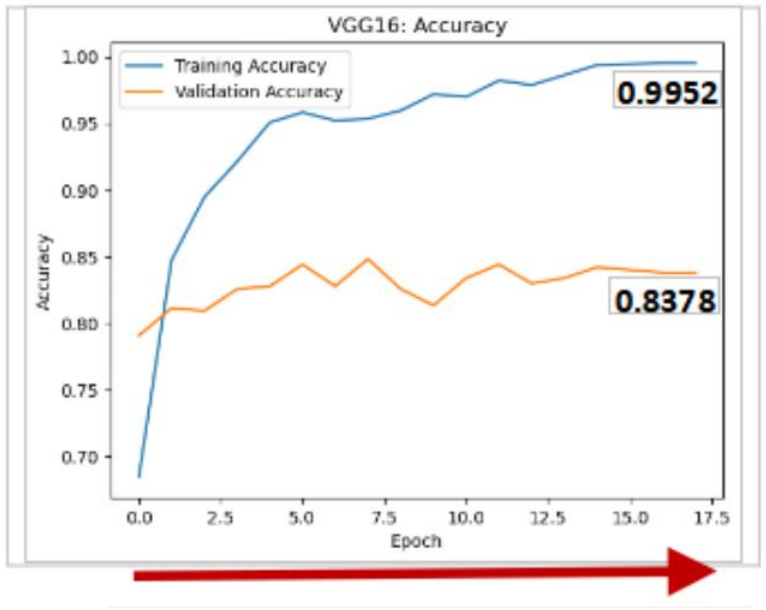
Thalassiosira sp.

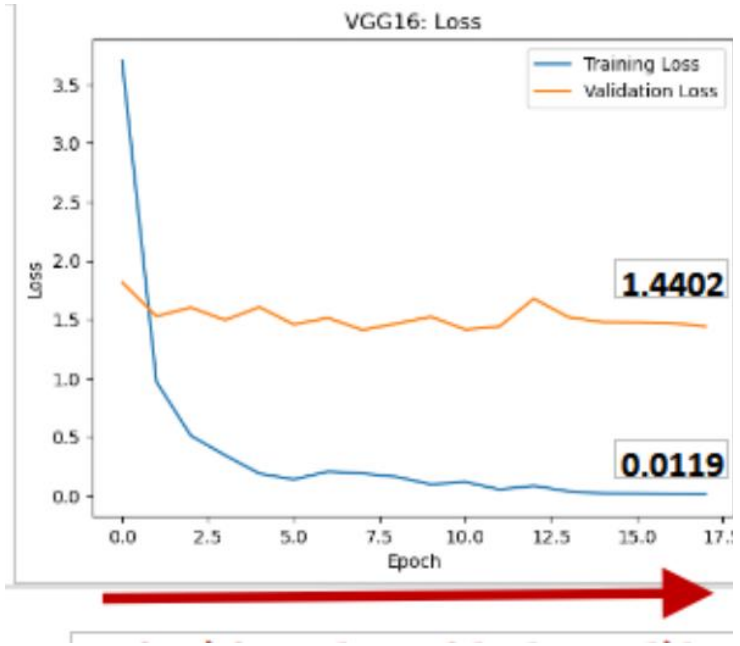


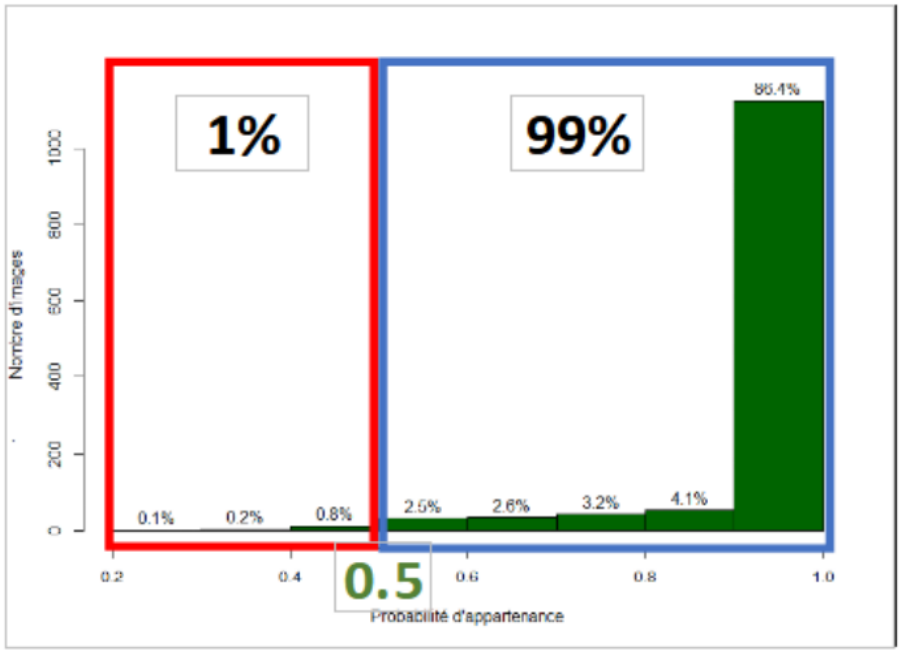
Thalassionema nitzschioides

M. Ryckman M.Sc. 2022 – G. Wacquet unpubl.

Applying new CNN classification on CytoSense images : premilinary results







M. Ryckman M.Sc. 2022 – G. Wacquet unpubl.

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

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 »

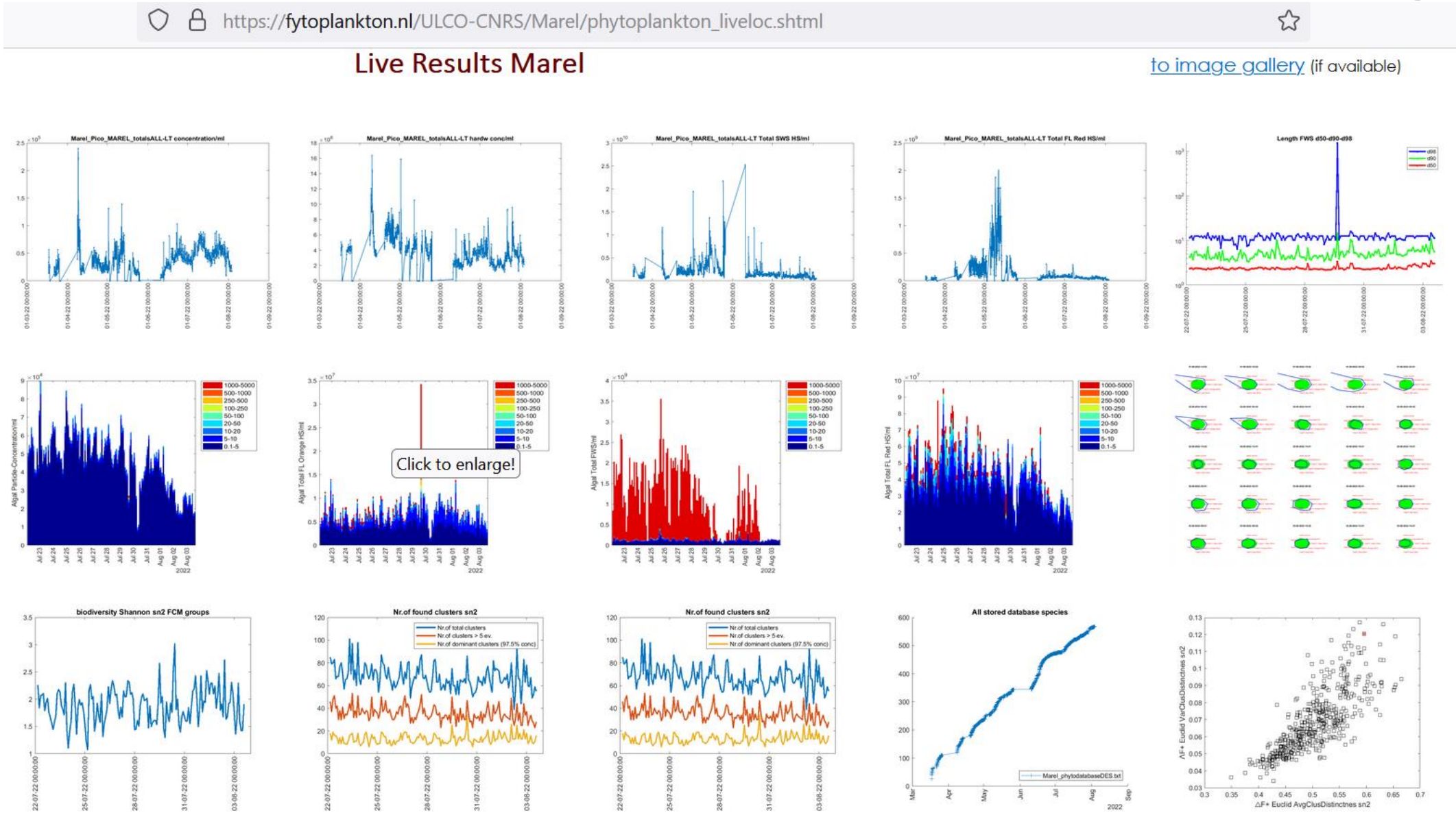
Collaborative work between ***CNRS-LOG*** and ***ULCO-LISIC*** ***(DYMAPHY (INTERREG IV A « 2 Seas », JERICO-NEXT (H2020) & MARCO (CPER HF)***

R package « ZooImage »

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

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

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



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

LIMNOLOGY
and
OCEANOGRAPHY: METHODS

ASLO

Limnol. Oceanogr.: Methods 2022

doi: 10.1002/lom3.10493

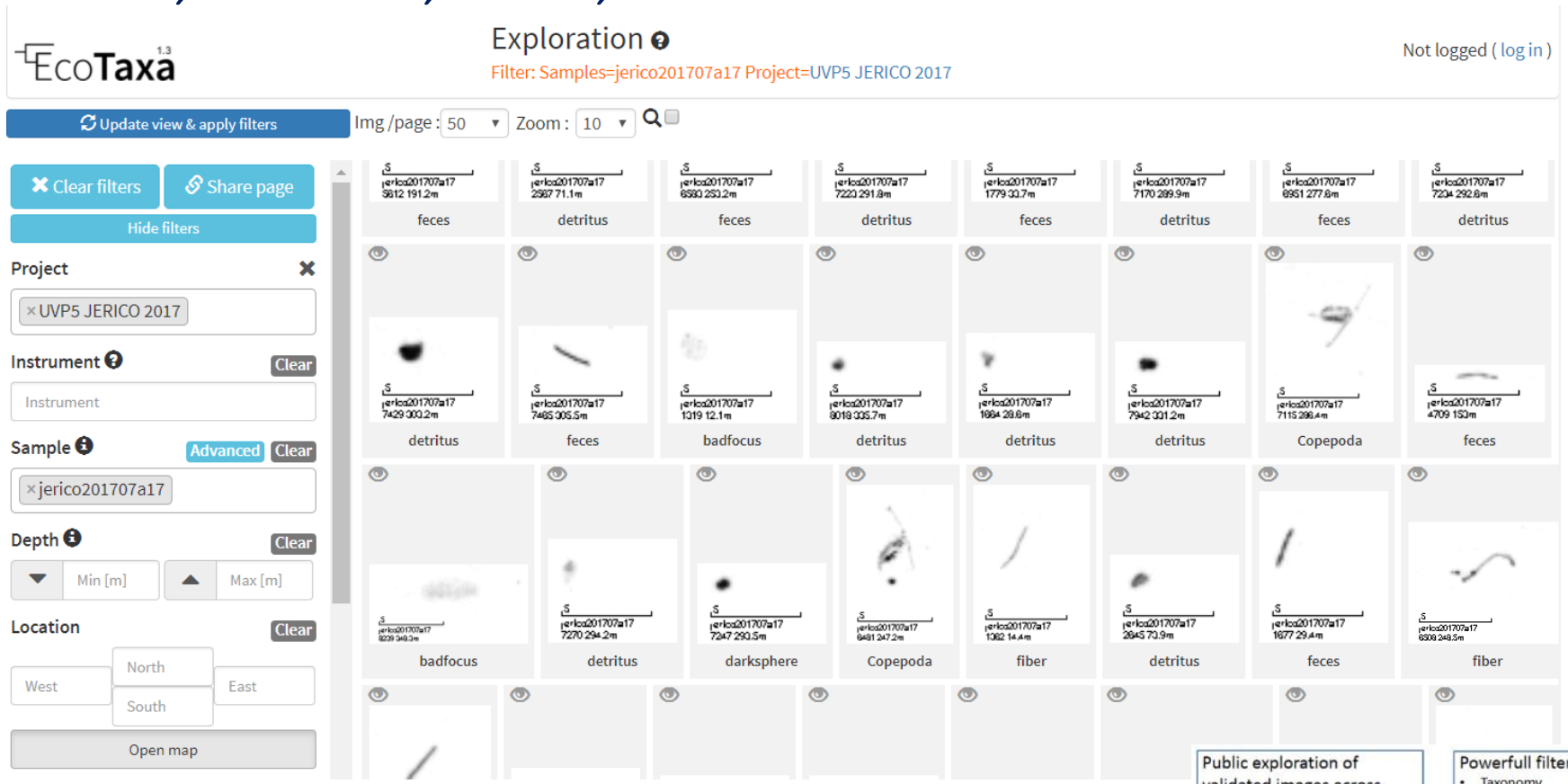
Automatic recognition of flow cytometric phytoplankton functional groups using convolutional neural networks

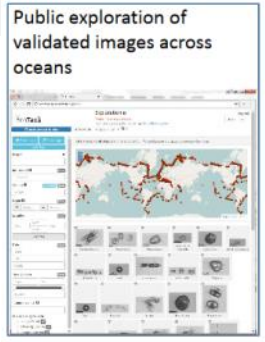
Robin Fuchs,[1,2] **Melilotus Thyssen**,[2*] **Véronique Creach**,[3] **Mathilde Dugenne**,[4] **Lloyd Izard**,[5] **Marie Latimier**,[6] **Arnaud Louchart**,[7,8] **Pierre Marrec**,[9] **Machteld Rijkeboer**,[10] **Gérald Grégori**,[2] **Denys Pommeret**[1,11,12,13]

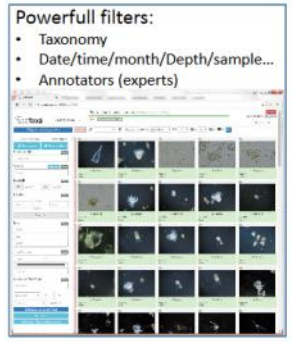
[1]Aix Marseille Univ, CNRS, I2M, Marseille, France
[2]Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France
[3]Cefas, Suffolk, UK
[4]Department of Oceanography, University of Hawai'i at Manoa, Honolulu, Hawai'i
[5]Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN-IPSL), Paris, France
[6]IFREMER, DYNECO PELAGOS, Plouzane, France
[7]Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, Italy
[8]IFREMER, Laboratoire Environnement and Ressources, Boulogne-sur-Mer, France
[9]Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island
[10]Laboratory for Hydrobiological Analysis, Rijkswaterstaat (RWS), Lelystad, The Netherlands
[11]Université Claude Bernard Lyon 1, Villeurbanne, France
[12]ISFA, Lyon, France
[13]Laboratoire de Sciences Actuarielle et Financière (SAF), Lyon, France

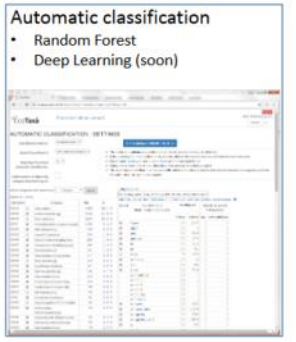
Processing and storage of thousands of data (raw data, processes data, images and automated classification)

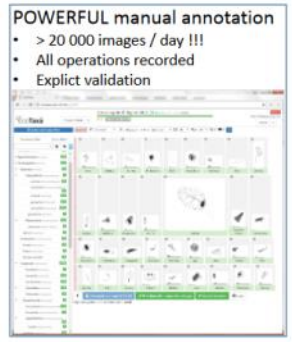
ZooPhytoImage(U Mons-Ifremer-CNRS/LOG), RClusToolBox (LISIC/ULCO-CNRS/LOG), EasyClus (TRP), EMODNET, SeaDataNet, EcoTaxa, other databases…











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

Picheral M., Stemman L., MMV 2017

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.