APPLICATION DU MACHINE LEARNING À L'IMAGERIE DU PHYTOPLANCTON

DE LA RECONNAISSANCE AU DÉNOMBREMENT

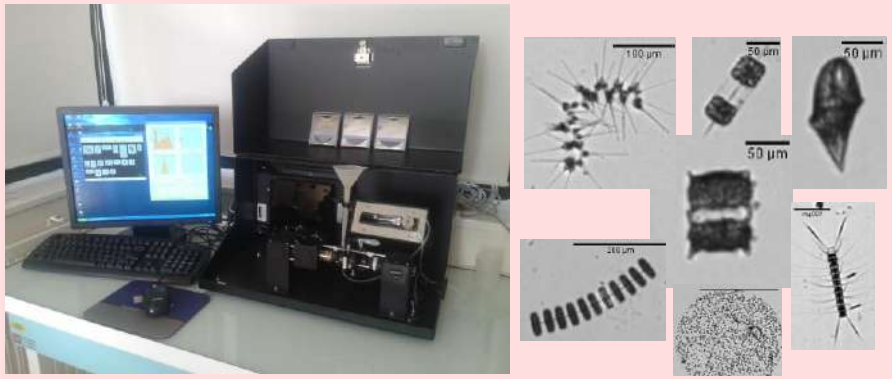
Guillaume WACQUET, Alain LEFEBVRE

Ifremer, Unité COAST
Laboratoire Environnement Ressources
Boulogne-sur-Mer, France
Prenom.Nom@Ifremer.fr

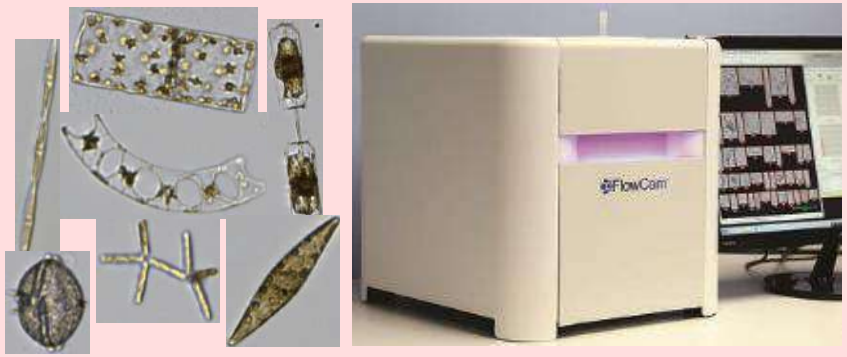
Atelier commun CYBER/ODATIS/ILICO
04-05 juin 2024 – Paris (France)
Utilisation de l'IA pour l'analyse de données issues de séries longues

Imagerie du phytoplancton au LER-BL

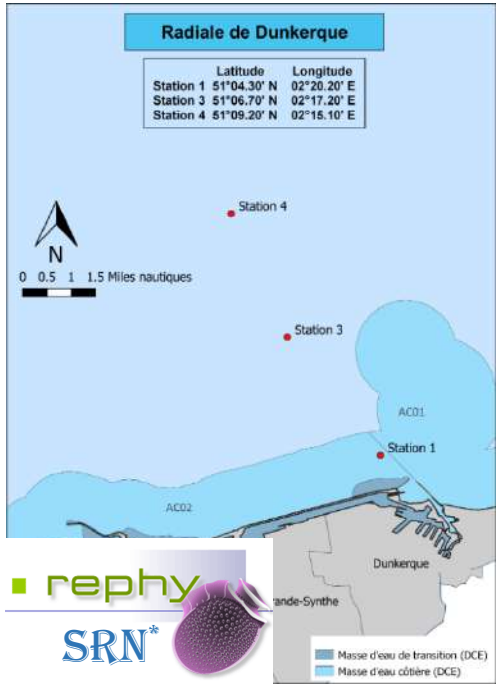
FlowCam®
Yokogawa Fluid Imaging Technologies, Inc.



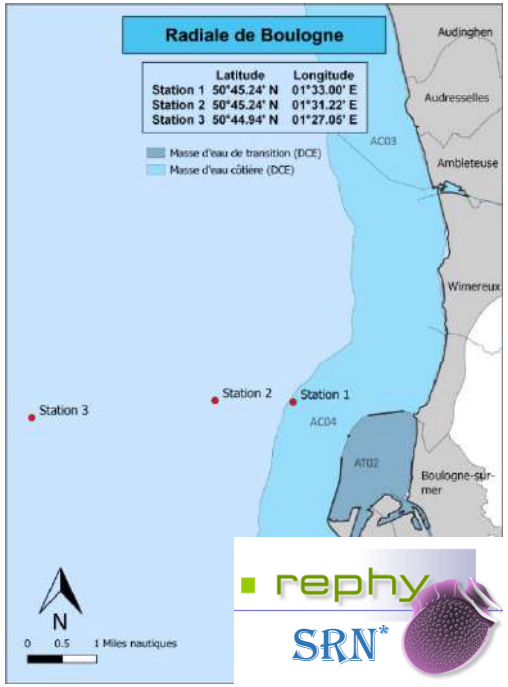
FlowCam VS Series



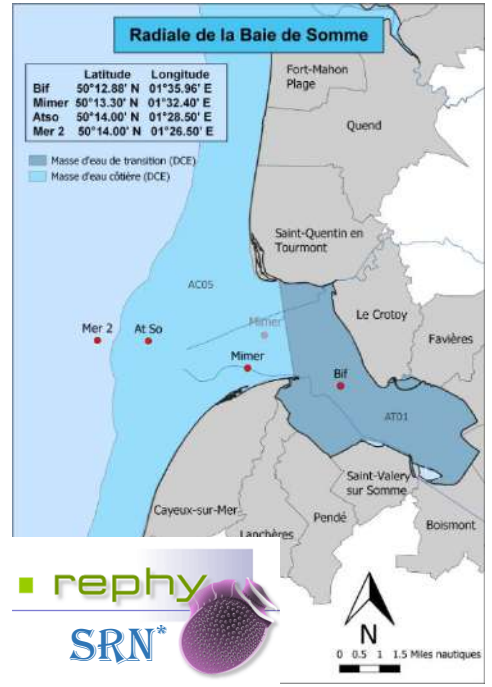
FlowCam 8100



Dunkerque
3 points
(DK1, DK3, DK4)



Boulogne-sur-Mer
3 points
(BL1, BL2, BL3)



Baie de Somme
4 points
(Bif, Mimer, AtSo, Mer2)

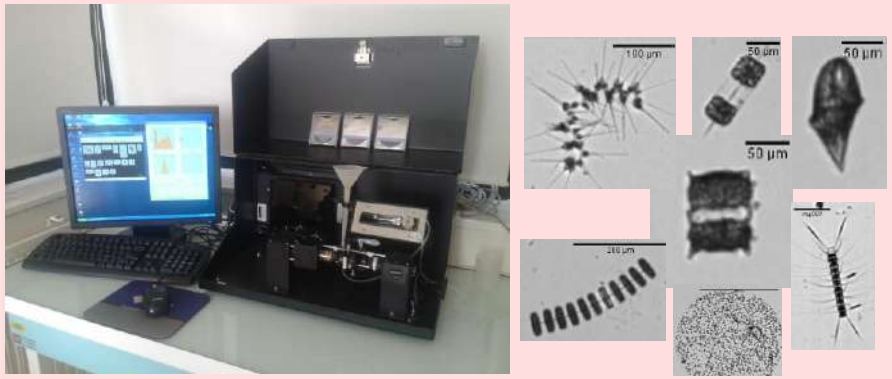
<u>Fréquence</u> : (bi-)mensuelle
<u>Début de la série</u> : 2012

+ campagnes

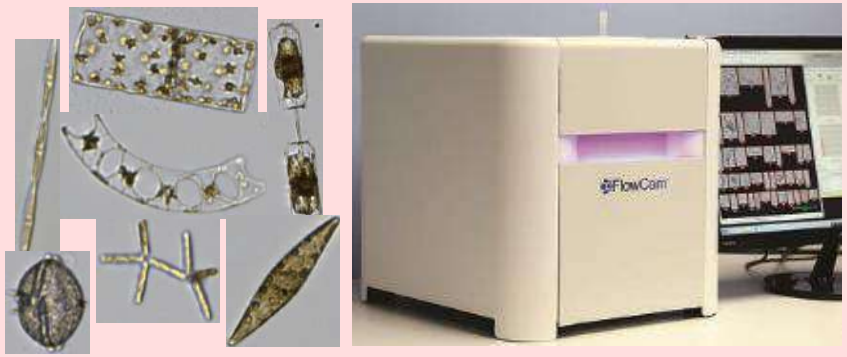
*Lefebvre A., Devreker D. (2023). *Earth System Science Data*, 15(3), 1077-1092

Imagerie du phytoplancton au LER-BL

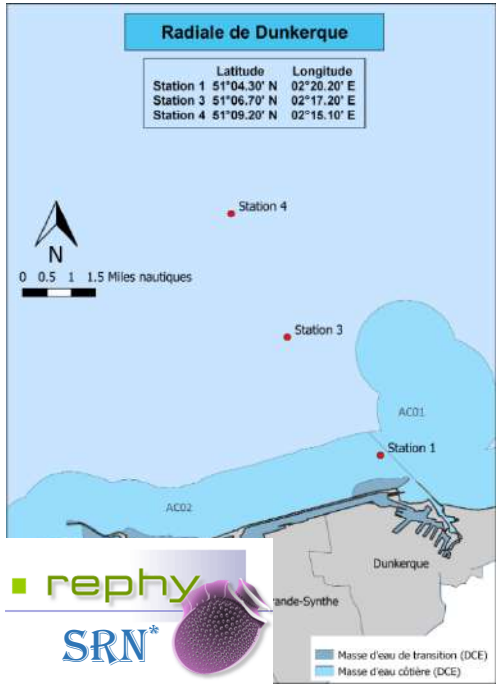
FlowCam®
Yokogawa Fluid Imaging Technologies, Inc.



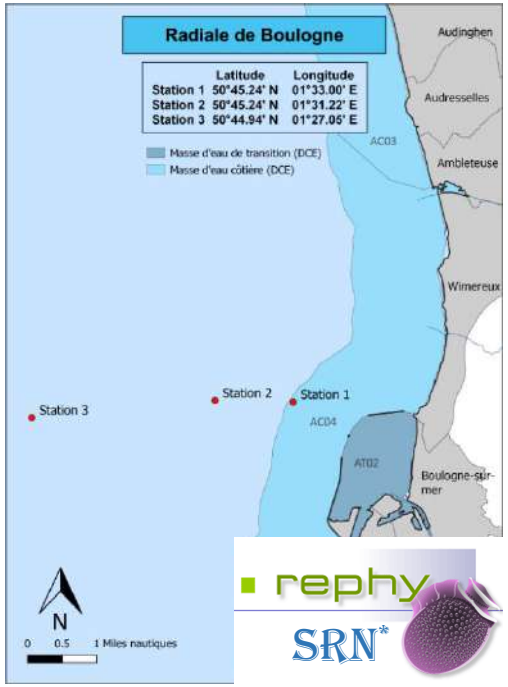
FlowCam VS Series



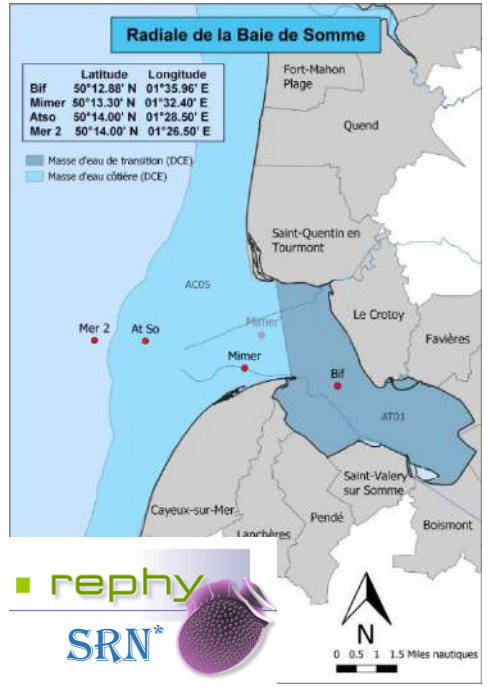
FlowCam 8100



Dunkerque
3 points
(DK1, DK3, DK4)



Boulogne-sur-Mer
3 points
(BL1, BL2, BL3)



Baie de Somme
4 points
(Bif, Mimer, AtSo, Mer2)

12 années de données → **≈ 2500 échantillons**

*Lefebvre A., Devreker D. (2023). *Earth System Science Data*, 15(3), 1077-1092

Méthode « intuitive »

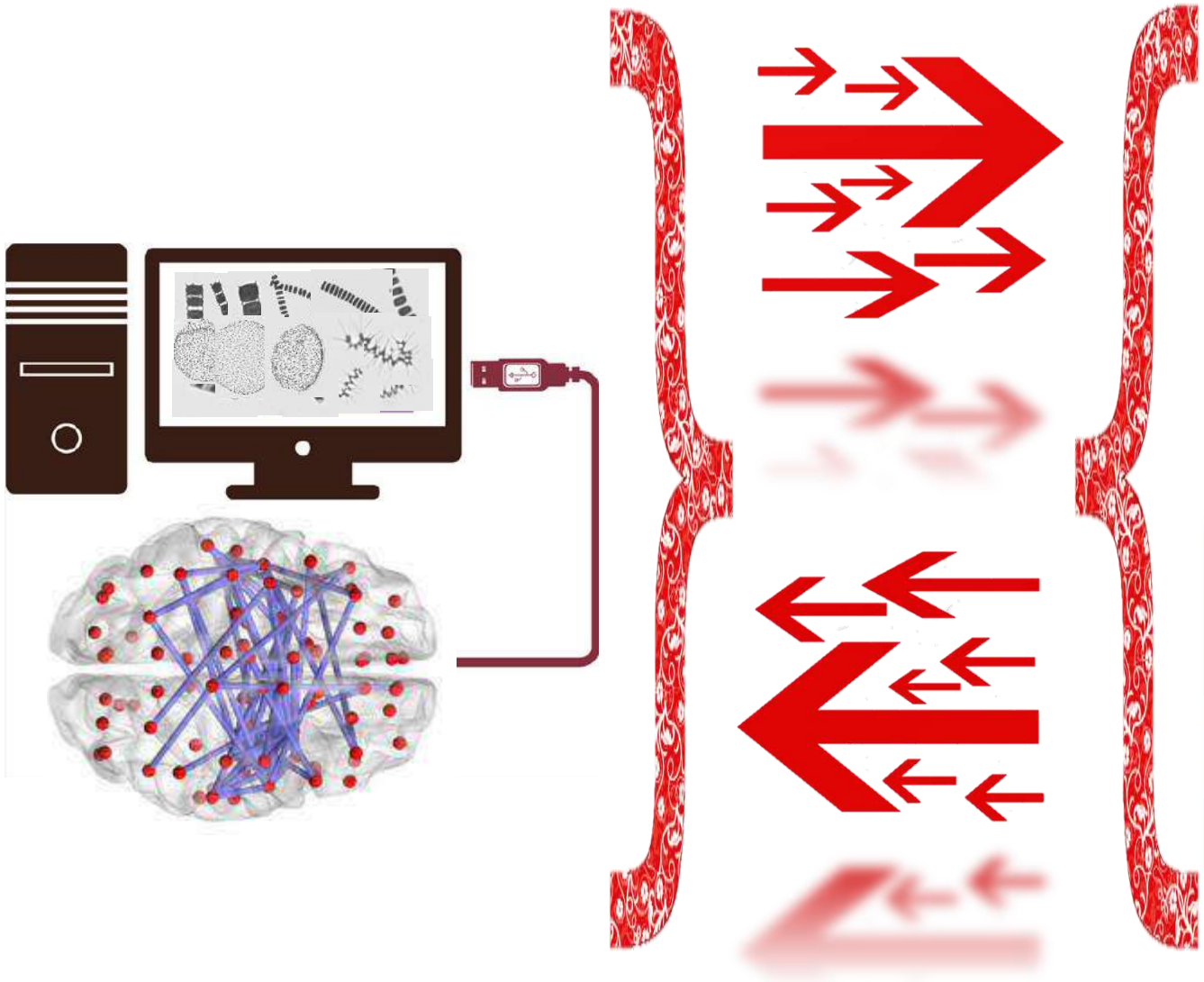
1 échantillon

=

plusieurs milliers d'images

Identification visuelle et tri manuel de chaque image

Utilisation de l'IA/ML



Objectifs principaux

- **Automatiser** les analyses
  - Quantité importante de données
  - Réduction du temps d'analyse
- **Exploiter** les données disponibles, **apprendre** et **prédire**
  - Ensembles d'apprentissage/test
  - Prédiction/forecasting

Objectifs secondaires

- **Adapter** les outils à nos besoins
  - Apprentissage actif
- Permettre une **interaction** de l'utilisateur
  - Intégration d'informations "expertes" *a priori*
  - Validation partielle des prédictions
- **Extraire** des informations complémentaires
  - Dénombrement des cellules en colonie
  - Traits fonctionnels

Développement de deux packages R

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

ZOOIMAGE

(2012-2015)

Java™

Ifremer

ECO NUM

UMONS
University of Mons

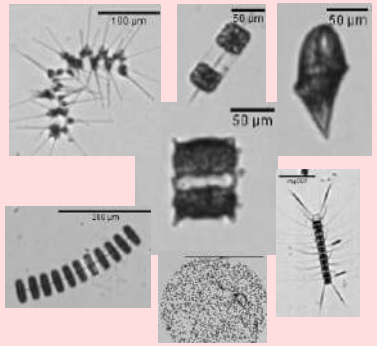
LOG

ONEMA
Office national de l'eau et des milieux aquatiques

ZooImage

ONEMA Office national de l'eau et des milieux aquatiques — Ifremer — LOG — ECO num — UMONS University of Mons

FlowCam®
Yokogawa Fluid Imaging Technologies, Inc.

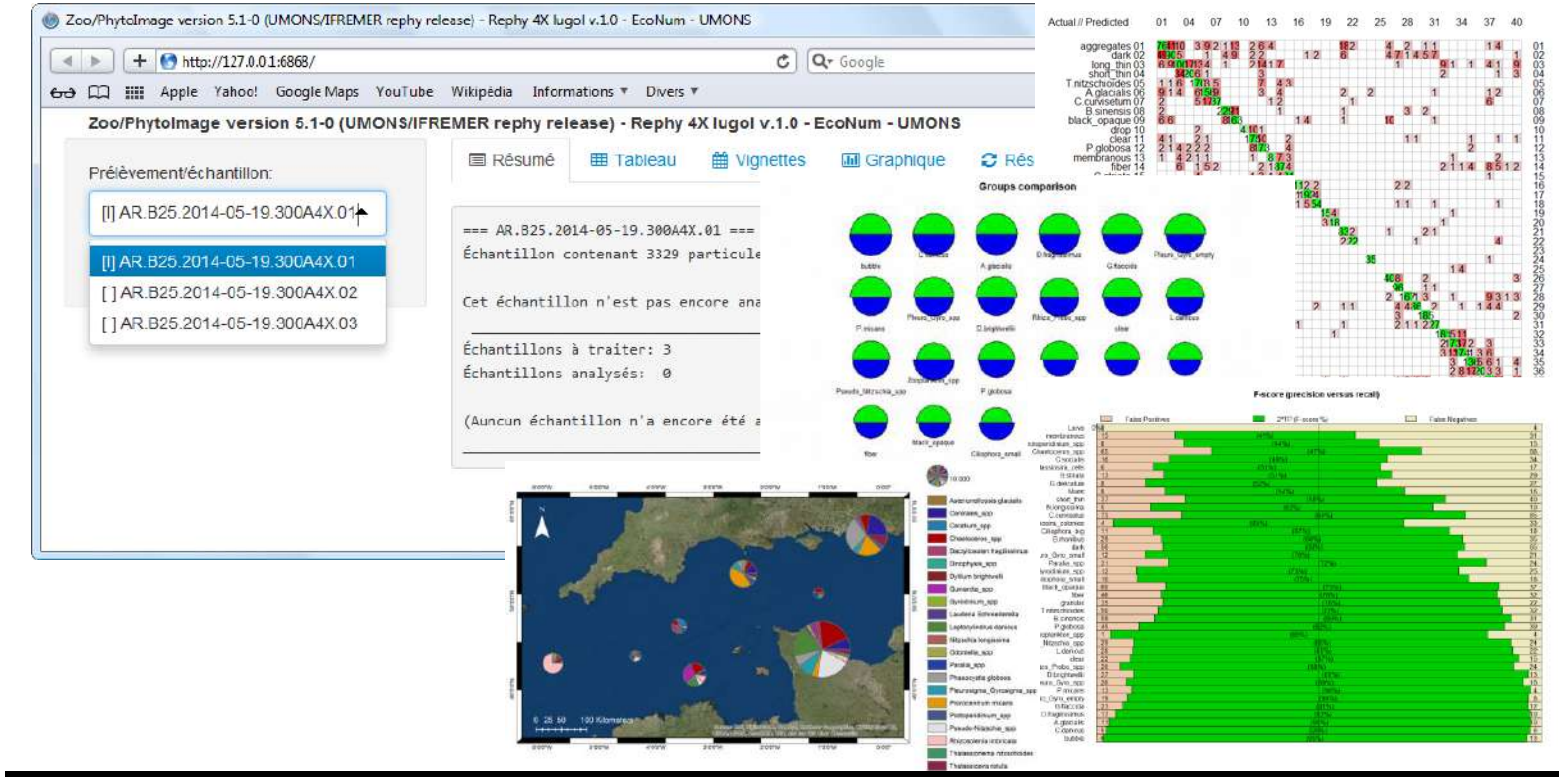


FlowCam VS Series

FlowCam 8100

Projet FlowCam/ZooImage

- Période : 2012-2015
- Convention IFREMER/UMONS/ONEMA

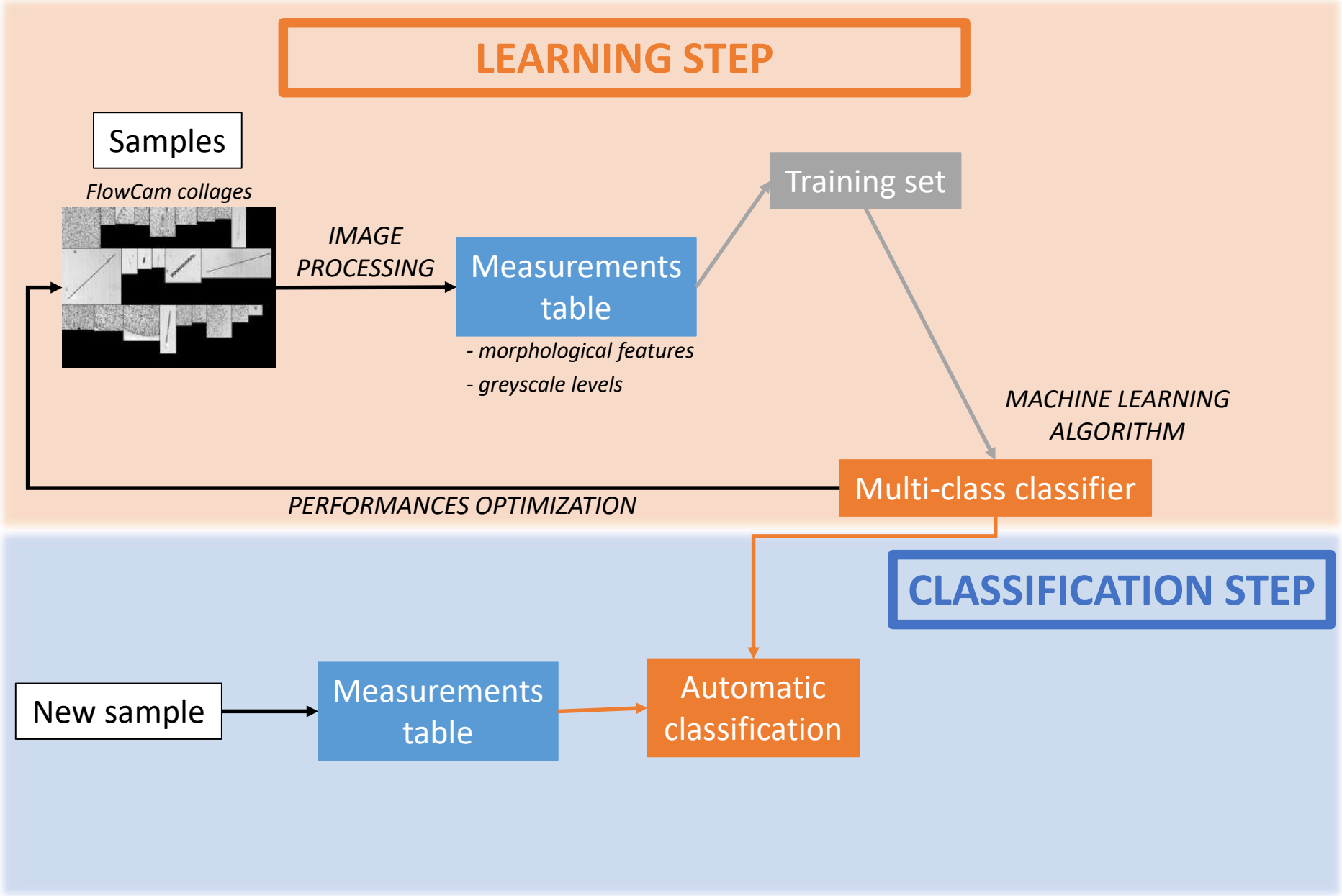


OBJECTIF

Classification automatisée avec interaction « experte »

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

Schéma de classification « traditionnelle »



ZooImage - Classification

LEARNING STEP

Samples

FlowCam collages

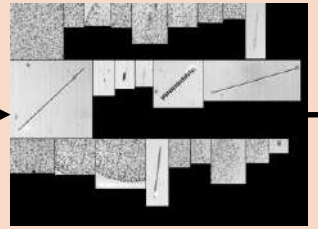


IMAGE PROCESSING

Measurements table

- *morphological features*
- *greyscale levels*

Training set

MACHINE LEARNING ALGORITHM

Multi-class classifier

PERFORMANCES OPTIMIZATION

INTEGRATION IN CLASSIFICATION STEP

New sample

Measurements table

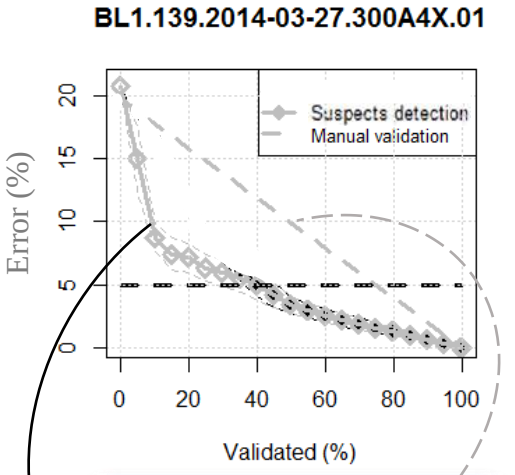
Automatic classification

Partial validation

User-interaction

Validation partielle

1. **Détection des « suspects »**
   - ➢ proba de bonne classif
   - ➢ critères bio-écologiques
2. **Validation des suspects**
   - ➢ Validation de ~30-40%



Random validation
→ décroissance **linéaire**

Suspects detection
→ décroissance **rapide**

ZooImage - Classification

INTEGRATION IN LEARNING STEP

Samples

FlowCam collages

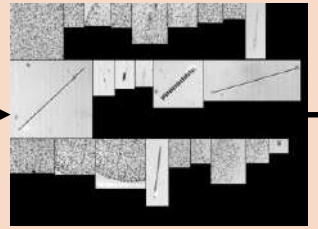


IMAGE PROCESSING

Measurements table

- *morphological features*
- *greyscale levels*

Training set

User-interaction

Contextual sample 1

Contextual sample 2

...

Contextual sample i

Adapted training set

MACHINE LEARNING ALGORITHM

Multi-class classifier

PERFORMANCES OPTIMIZATION

INTEGRATION IN CLASSIFICATION STEP

New sample

Measurements table

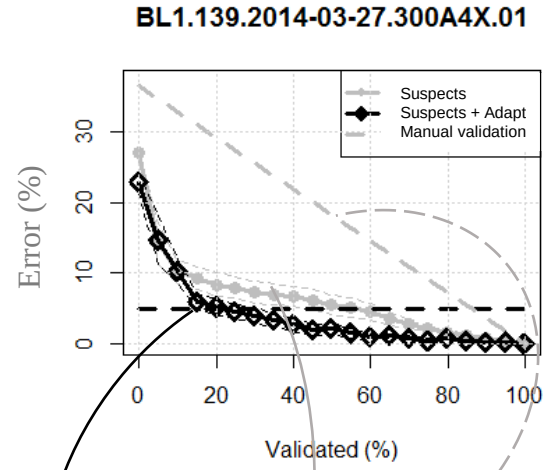
Automatic classification

Partial validation

User-interaction

Apprentissage adaptatif

1. **Données « contextuelles »**
   - Déjà classifiées
   - Même période, zone, …
2. **Validation partielle**



Manual validation
→ décroissance **linéaire**

Suspects
→ décroissance **rapide**

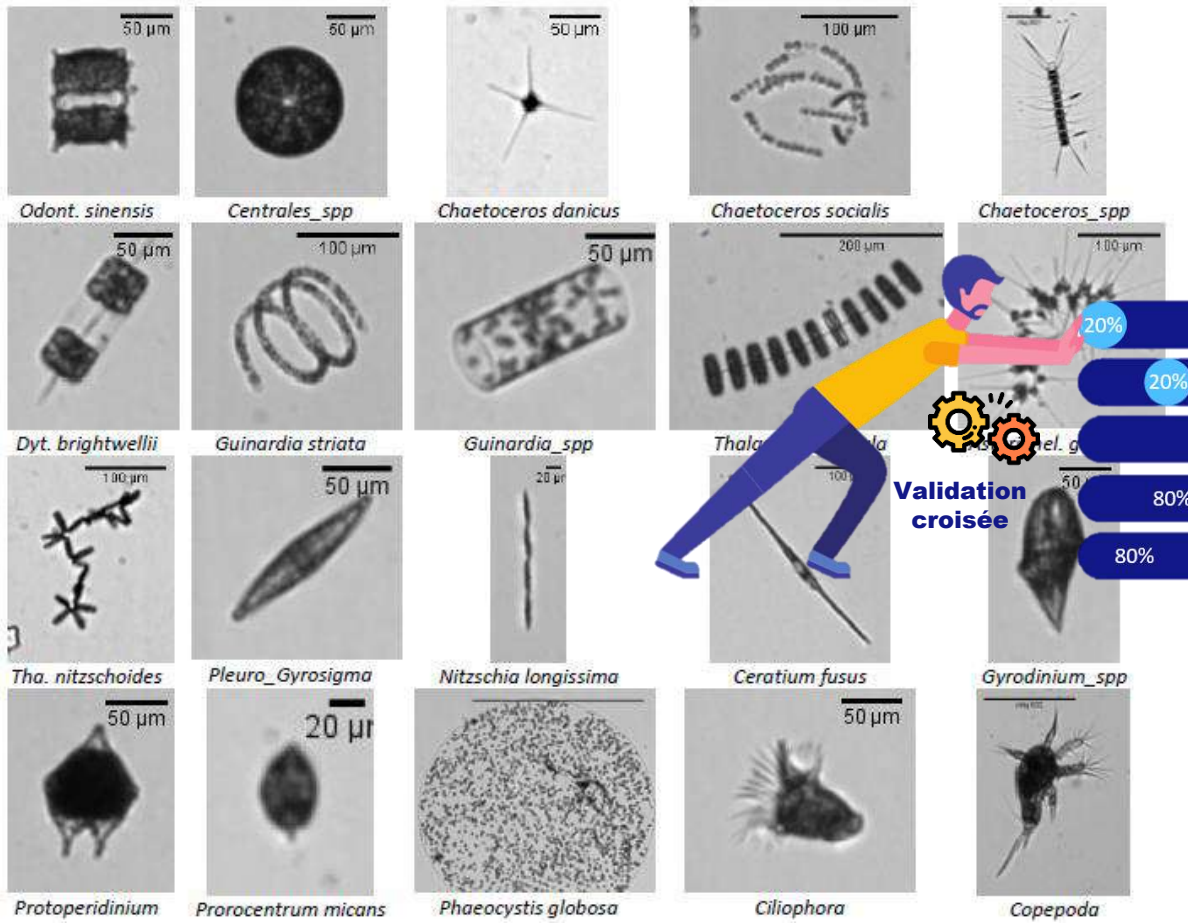
Suspects + Adapt
→ décroissance **plus rapide**

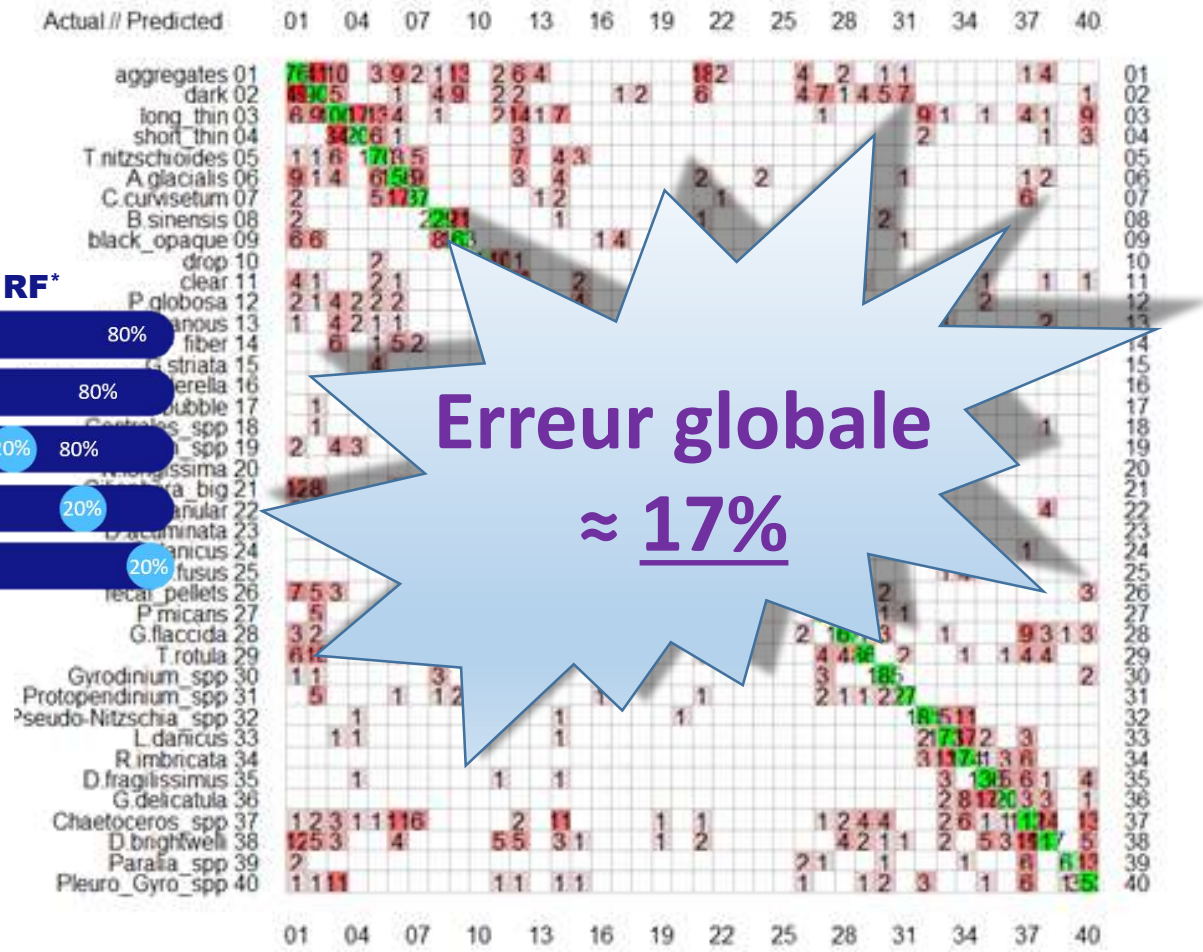
ZooImage - Training

Appareil : FlowCAM (VS Series) **Zone :** Manche-Mer du Nord

Caméra : N&B – 1280x960 pixels **Optique :** 4X-300µm

Nb groupes : 27 phytoplancton + 1 zooplancton + 12 débris





*RF : Random Forest

ZooImage - Résultats

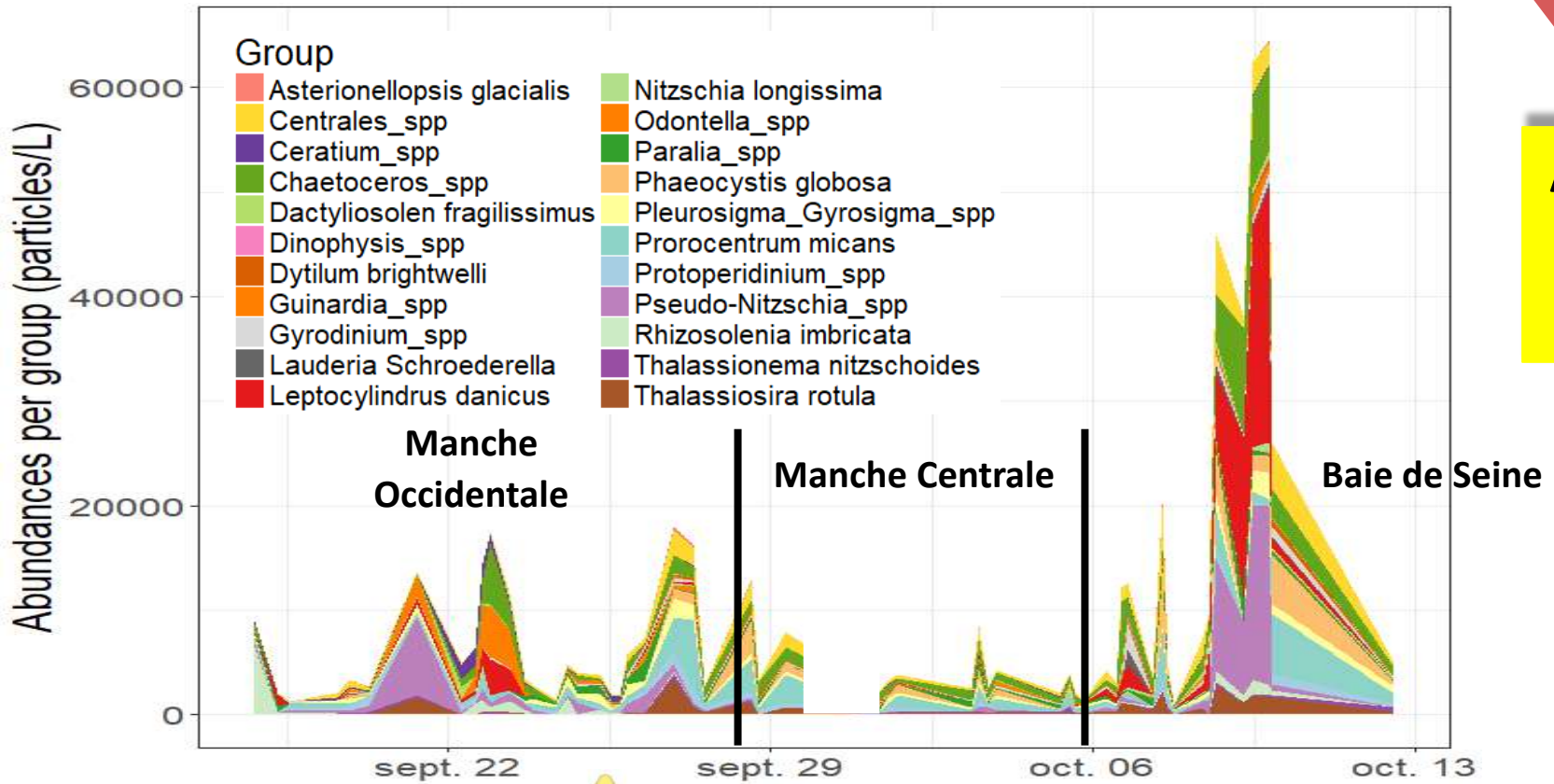
Campagne CAMANOC

Campagne multidisciplinaire

Période : 16 Septembre-12 Octobre 2014

Zone : Manche

Navire : N/O « Thalassa II » - IFREMER



Echantillonnage discret…

185 échantillons

Nb total images
=
130 000

Apprentissage adaptatif

Validation partielle

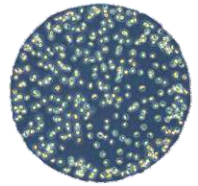
Nb images validées
≈
25 000 *(~20%)*

CHRONOPHAGE !!!

https://github.com/IFREMER-LERBL/EcoTransLearn/

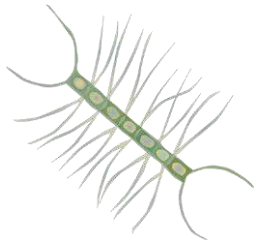


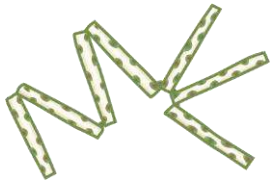
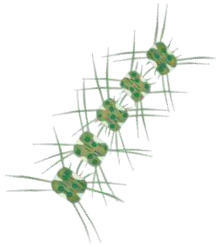
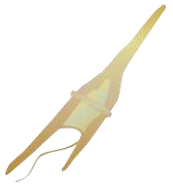


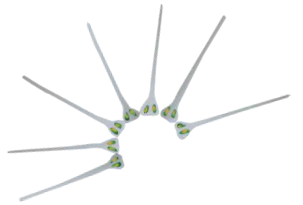



ECOTRANSLEARN

(2022)

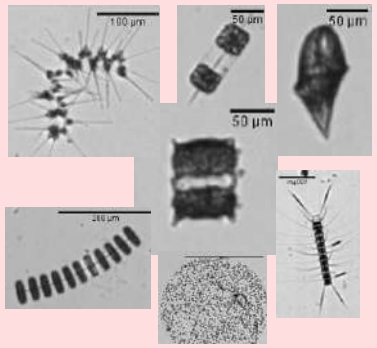




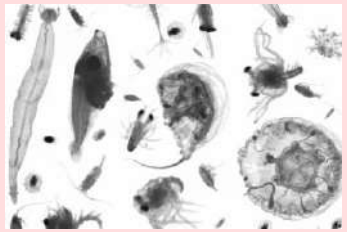



EcoTransLearn

FlowCam®
Yokogawa Fluid Imaging Technologies, Inc.

FlowCam VS Series



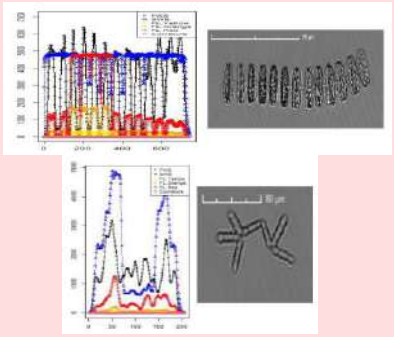


ZooScan

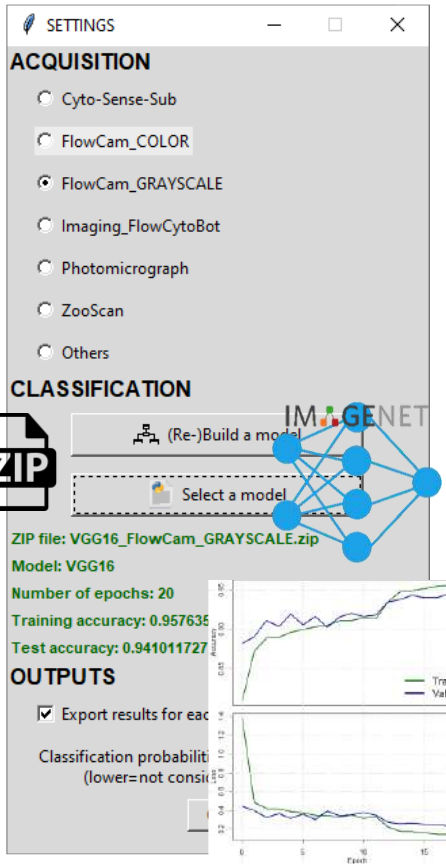
HYDROPTIC

CytoBuoy

CytoSense



Ifremer



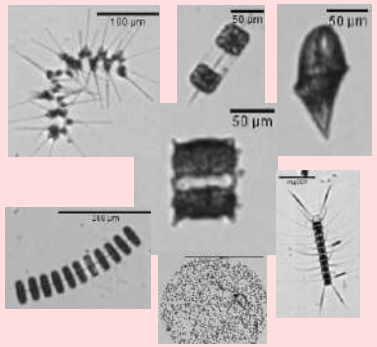
https://github.com/IFREMER-LERBL/EcoTransLearn/

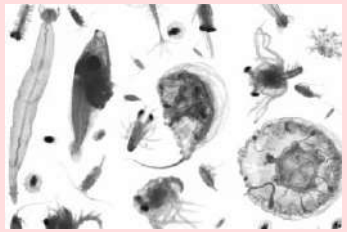
EcoTransLearn

Ifremer

FlowCam®
Yokogawa Fluid Imaging Technologies, Inc.

FlowCam VS Series



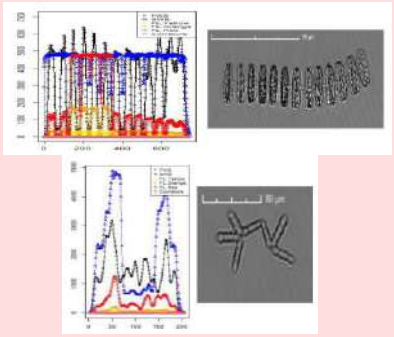


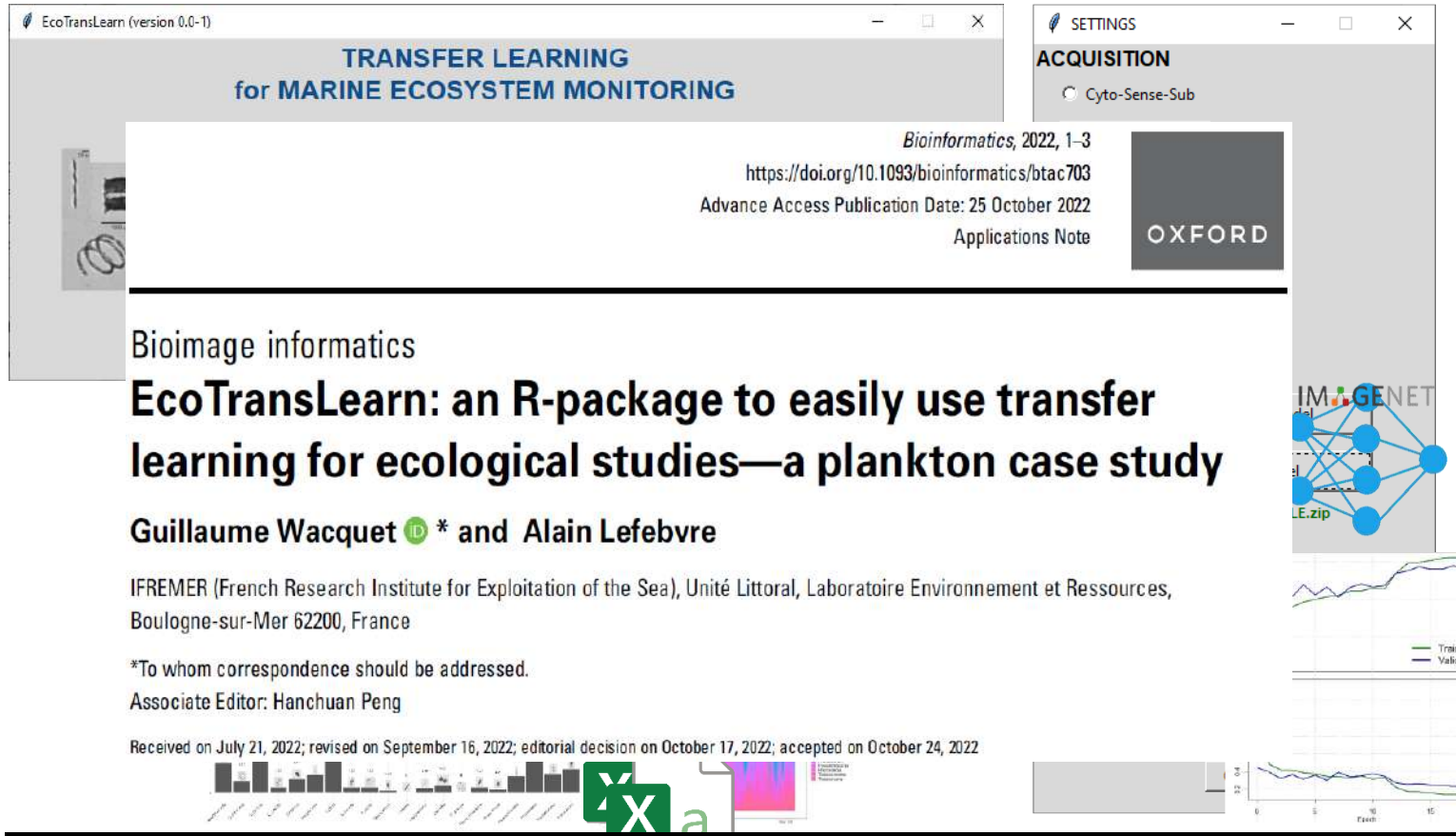
HYDROPTIC

ZooScan

CytoBuoy

CytoSense





OBJECTIF

S'affranchir de la phase de validation manuelle

Améliorer les performances de reconnaissance

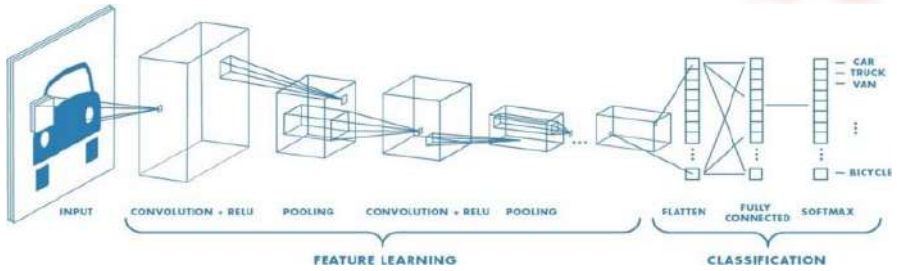
https://github.com/IFREMER-LERBL/EcoTransLearn/

EcoTransLearn

Apprentissage profond
(Deep Learning)

✖ **Construction/entraînement d'un nouveau modèle → très long**

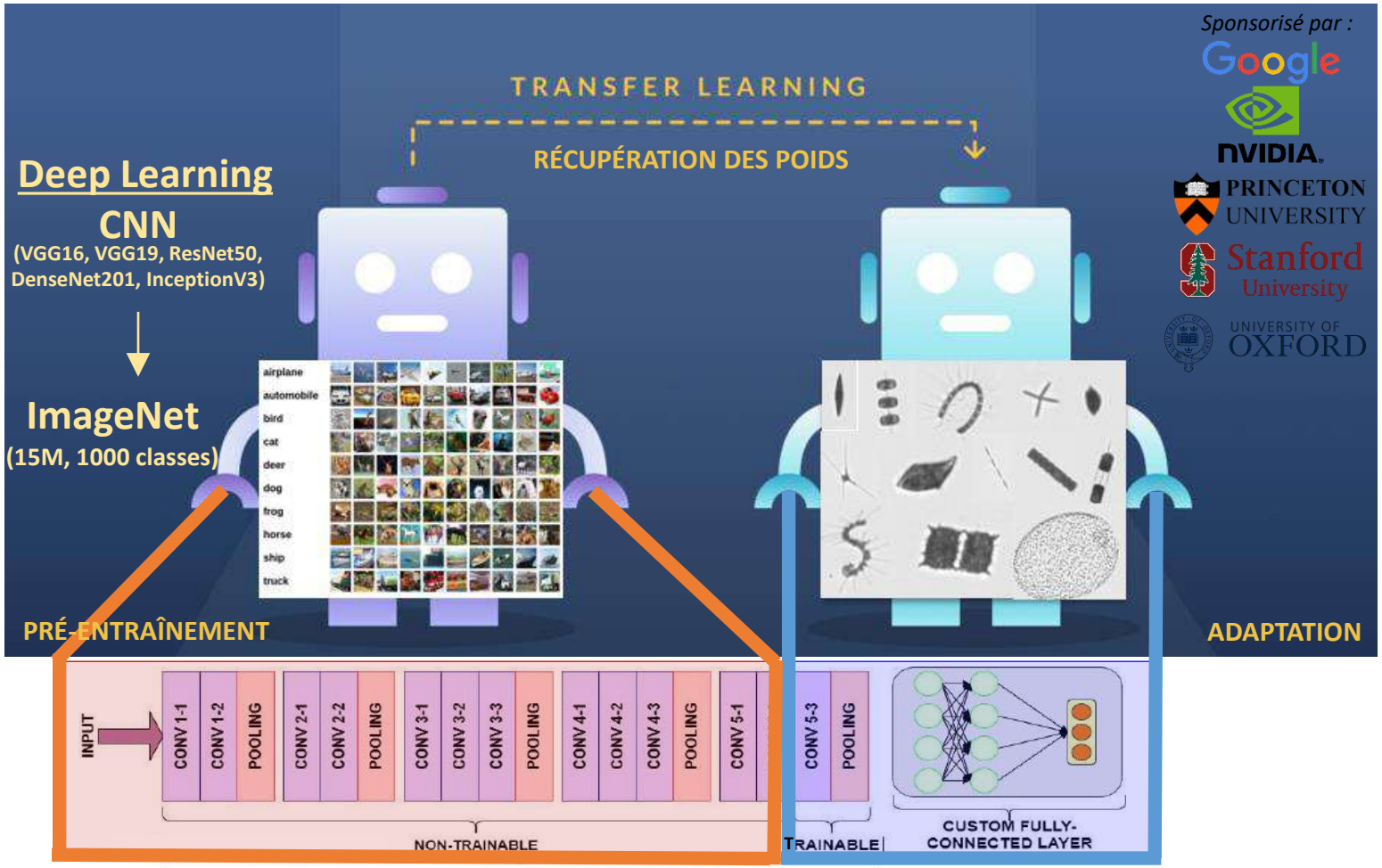
✖ **Obtention de résultats de haute précision → nécessite beaucoup d'images**



CNN : Convolutional Neural Network

Apprentissage par transfert
(Transfer Learning)

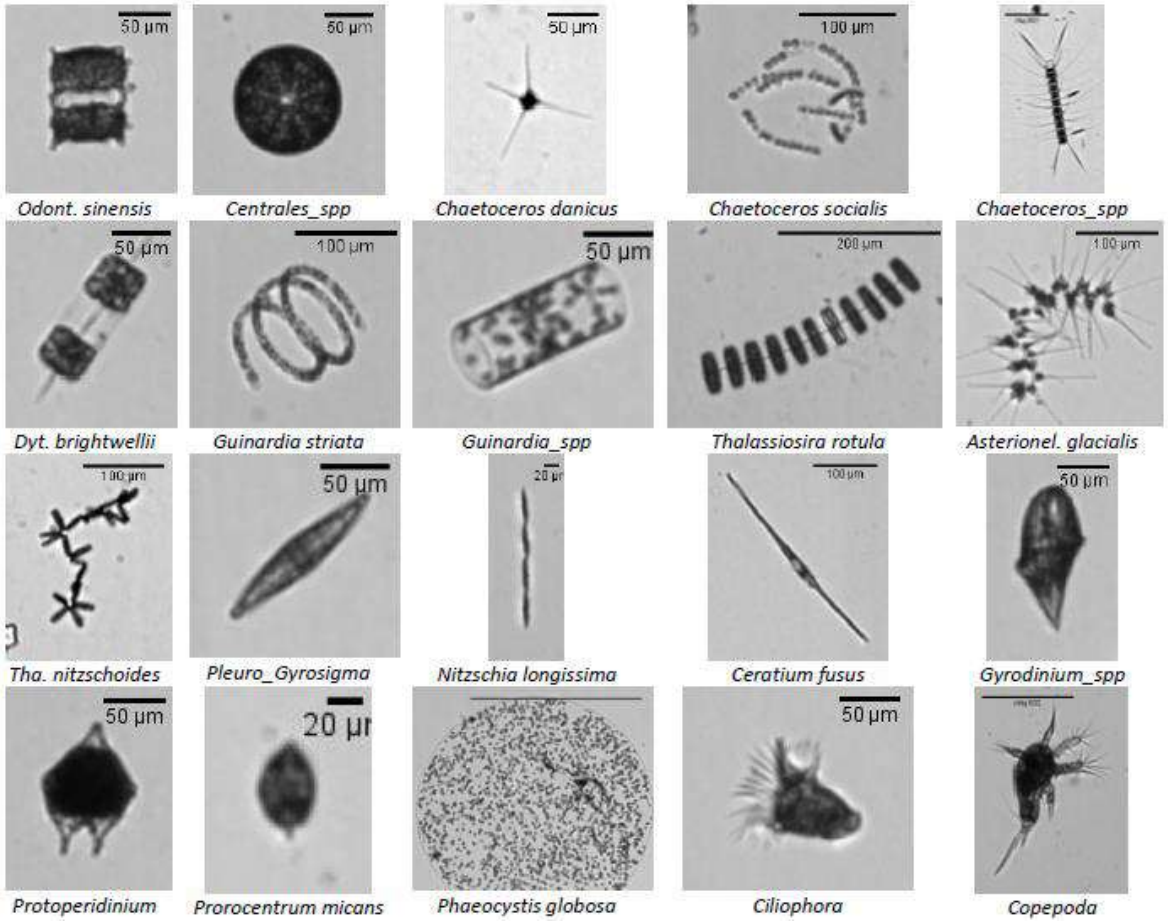
« Il est plus facile d'apprendre à jouer du piano quand on sait déjà jouer de la guitare! »



Base de données ImageNET — **Jeu de données spécifique**

EcoTransLearn - Training

Appareil : FlowCAM (VS Series)
Caméra : N&B – 1280x960 pixels
Zone : Manche-Mer du Nord
Optique : 4X-300µm

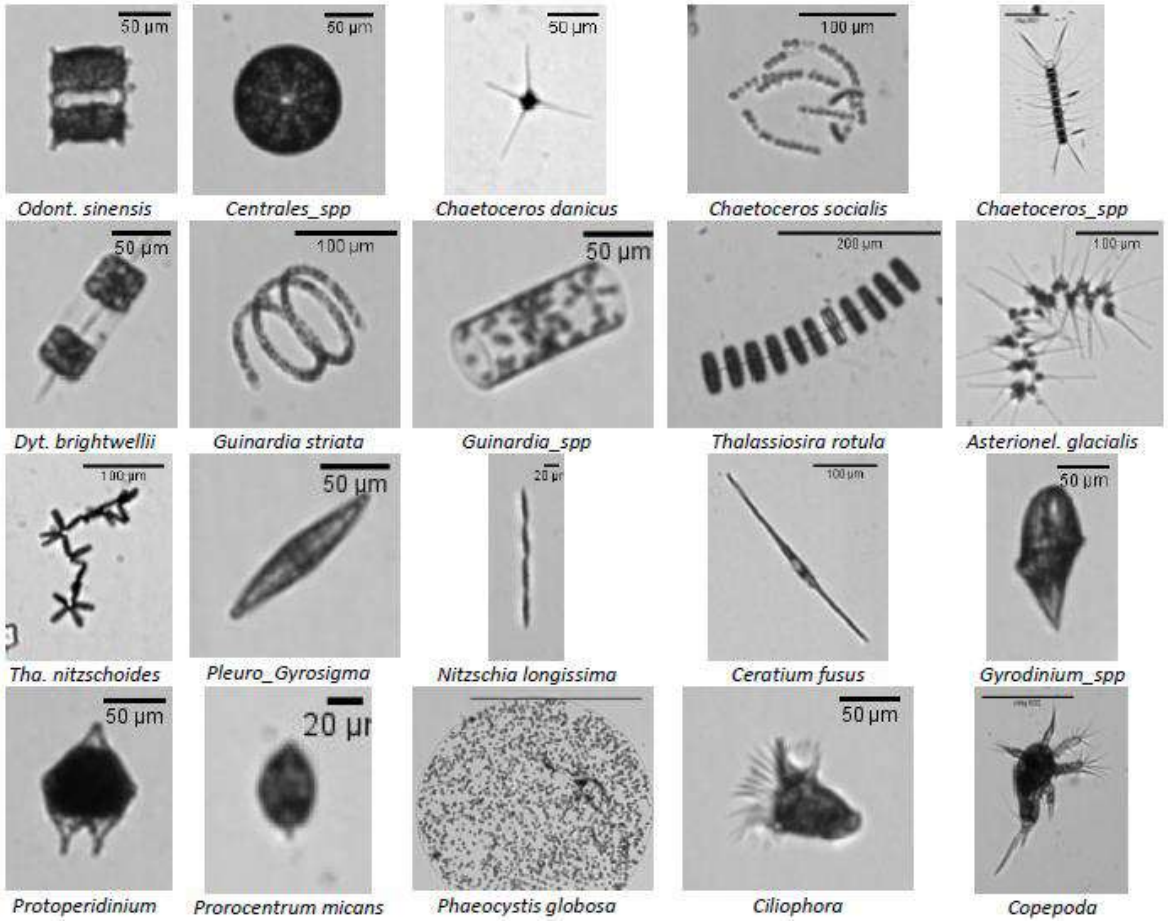


Taxa name	Nb Training set (~60%)	Nb Validation set (~40%)
Asterionellopsis_glacialis	1321	840
Chaetoceros curvisetus	377	252
Chaetoceros danicus	1295	849
Chaetoceros socialis	1012	600
Ciliophora	80	53
Dactyliosolen_spp	1254	778
Ditylum brightwellii	1490	962
Guinardia flaccida	1237	794
Guinardia striata	1251	808
Gymnodinium	114	76
Lauderia	97	64
Leptocylindrus danicus	1359	837
Odontella_mobiliensis	1284	801
Phaeocystis_globosa	1549	976
Pleuro_Gyrosigma	256	170
Prorocentrum	83	56
Pseudo_Nitzschia_spp	1267	774
Rhizosolenia_spp	1307	850
Thalassionema_nitzschioides	302	202
Thalassiosira_rotula	197	131
aggregates	1315	781
bubbles	286	191
clear_detritic	1268	809
dark_detritic	1288	744
fiber	398	265
membranous	111	74
Total	**19497**	**12203**

Train **Validation**

EcoTransLearn - Training

Appareil : FlowCAM (VS Series)
Caméra : N&B – 1280x960 pixels
Zone : Manche-Mer du Nord
Optique : 4X-300µm



Taxa name	Nb Training set (~60%)	Nb Validation set (~40%)
Asterionellopsis_glacialis	1321	840
Chaetoceros curvisetus	377	252
Chaetoceros danicus	1295	849
Chaetoceros socialis	1012	600
dark_detritic	1288	744
fiber	398	265
membranous	111	74
Total	**19497**	**12203**

DATA AUGMENTATION

Horizontal/vertical flip

Cropping

Original image

Rotation

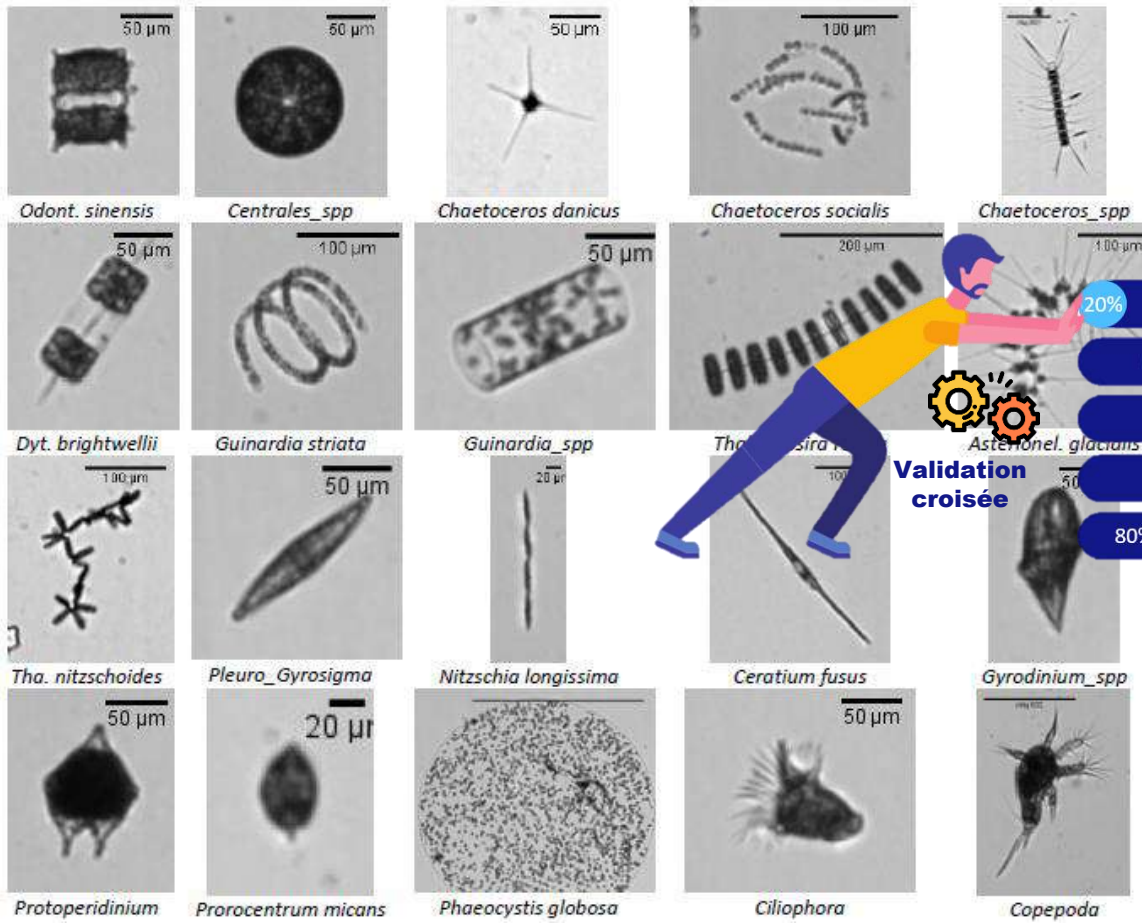
Light/contrast transformation

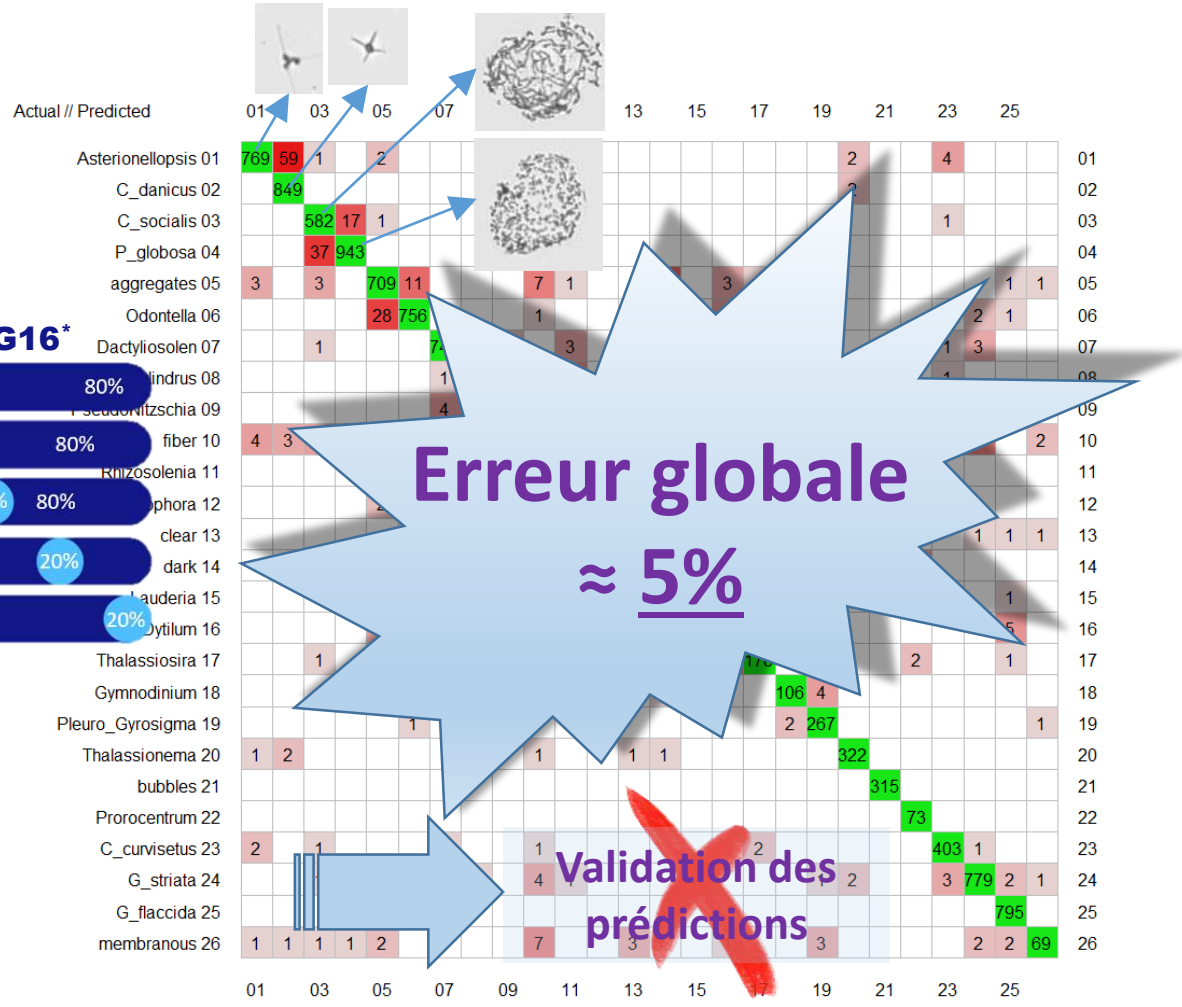
Other transformations ...

Train **Validation**

EcoTransLearn - Training

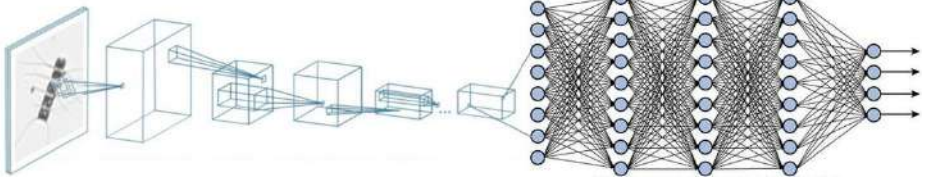
Appareil : FlowCAM (VS Series)
Caméra : N&B – 1280x960 pixels
Zone : Manche-Mer du Nord
Optique : 4X-300µm





*VGG16 : Visual Geometry Group (16 couches)

EcoTransLearn - Comparatif



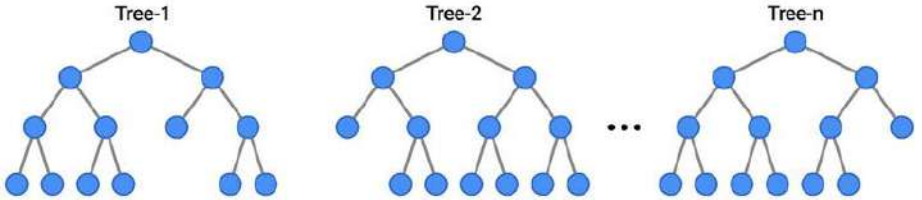
Deep Learning

(Convolutional Neural Network)

Méthodologie **EcoTransLearn**

Apprentissage (validation croisée)
Accuracy = **0.9883**

Test
Accuracy = **0.9442**



Machine Learning « classique »

(Extraction de caractéristiques + Random Forest)

Méthodologie **ZooImage**

Apprentissage (validation croisée)
Accuracy = **0.8178**

Test
Accuracy = **0.8227**

Dénombrement des cellules

ZooImage
EcoTransLearn

Campagne CAMANOC

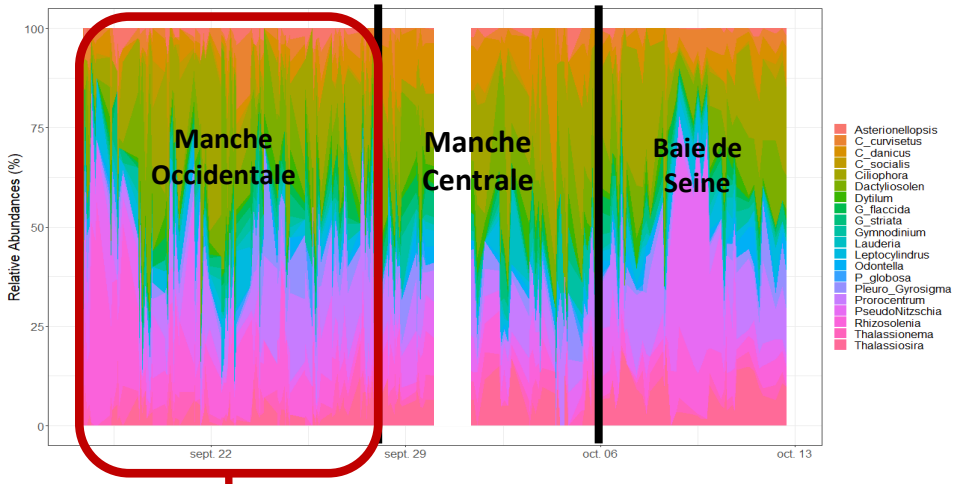
Campagne multidisciplinaire

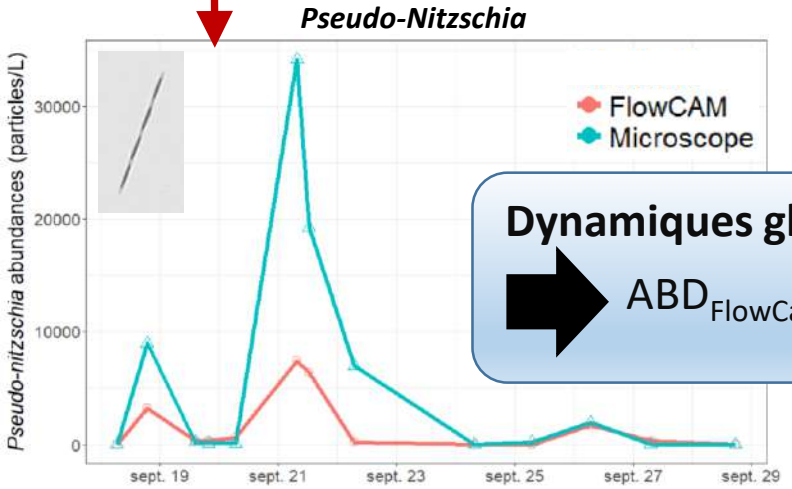
Période : 16 Septembre-12 Octobre 2014

Zone : Manche

Navire : N/O « Thalassa II » - IFREMER

Estimateurs **biomasse/biovolume** basés sur **nombre de cellules par unité de volume**





Aujourd'hui :

1 image

=

1 particule

=

1 cellule

=

1 colonie

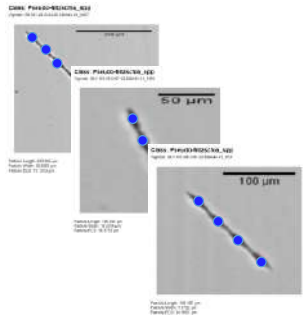
Dynamiques globales <u>proches</u>

$ABD_{FlowCam} < ABD_{Microscope}$

Dénombrement des cellules

ZooImage
EcoTransLearn

Calibration de **modèles prédictifs**
pour l'estimation du **nombre de cellules en colonies**



Labélisation

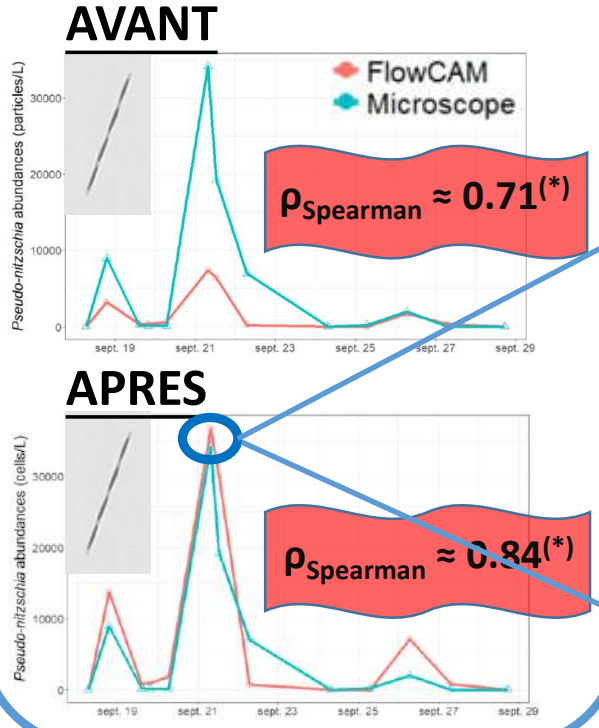
semi-automatique des images du training set
(traitement d'images : détection de contours, maxima locaux, …
+ validation manuelle)

Modèle prédictif

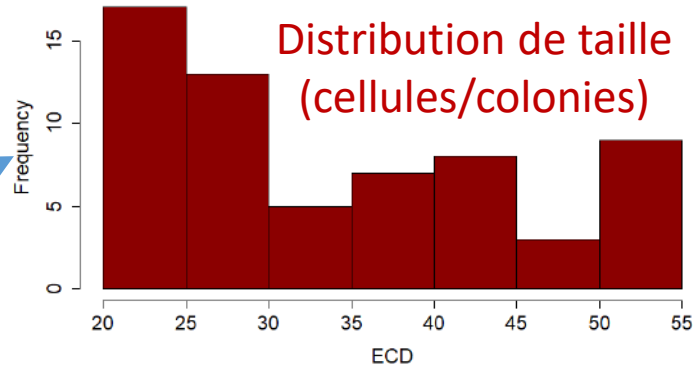
Machine/Transfer Learning
(Random Forest, ResNet, EfficientNet, LOCA, …)

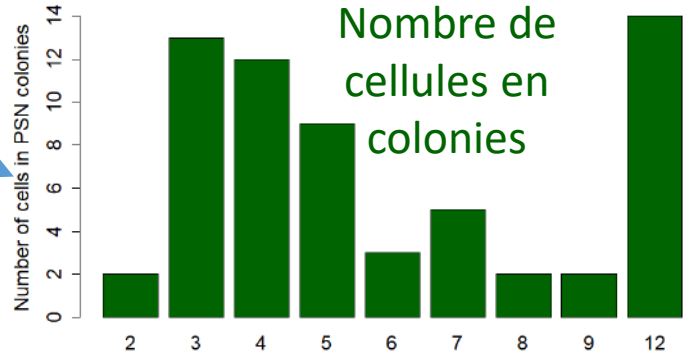
Atos

Prédiction



Traits fonctionnels

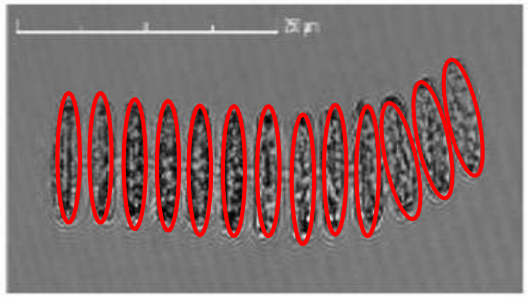




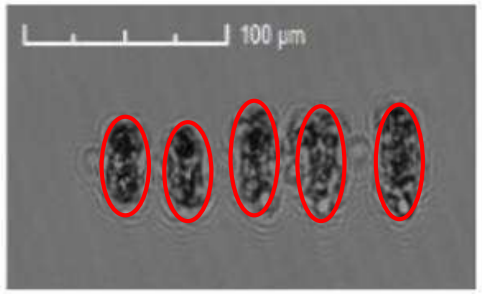
Dénombrement des cellules

Quelques résultats sur différents taxons…

Thalassiosira rotula

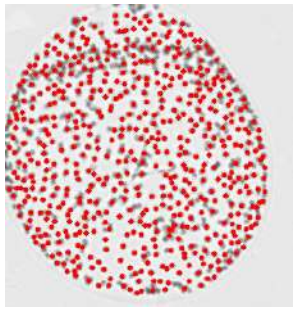


Manual: 13
Random Forest: 12

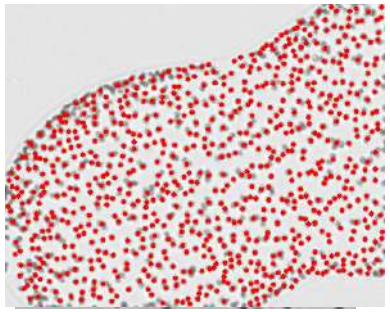


Manual: 5
Random Forest: 4

Phaeocystis globosa

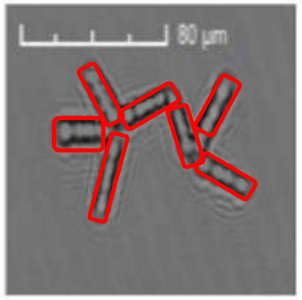


ResNet: 580

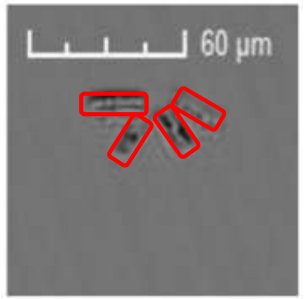


ResNet: 786

Thalassionema nitzschioides

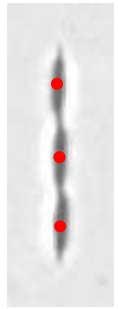


Manual: 7
Random Forest: 6

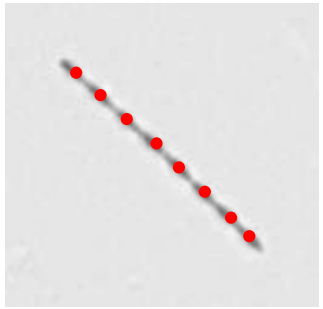


Manual: 4
Random Forest: 4

Pseudo-Nitzschia



Manual: 3
ResNet: 3



Manual: 8
ResNet: 8

Merci à tous les partenaires/collaborateurs….

Luis Felipe ARTIGAS (HDR)
CNRS-ULCO-Univ. Lille
Laboratoire d'Océanologie et Géosciences
Wimereux, France

Christophe LIET
Paul ZANONCELLI
et les étudiants de l'ENSICAEN
ATOS Inno'Lab Nord-Ouest
Colombelles, France

Philippe GROSJEAN (PR)
Université de Mons
Laboratoire d'Ecologie Numérique
Mons, Belgique

Denis HAMAD (PR)
Université du Littoral Côte d'Opale
Laboratoire d'Informatique Signal et Image de la Côte d'Opale
Calais, France

Et merci de votre attention!