

2015 - **Domaine** Outils pour la surveillance environnementale

Action 9 – FlowCAM / ZooPhytoImage

Optimisation de l'identification et du dénombrement du microphytoplancton avec le système couplé de numérisation et d'analyse d'images FlowCAM / ZooPhytoImage (système innovant)

Action 9 – Livrable 2. Mise en œuvre opérationnelle de l'outil FlowCAM-Zoo/PhytoImage dans le cadre de la surveillance REPHY. Résultats sur 12-18 mois.

Rapport final, février 2016

Auteurs principaux

Nadine NEAUD-MASSON
Alain LEFEBVRE
Catherine BELIN
Emilie GAUTHIER
Antoine HUGUET
Antoine SZIDON
Arnaud LEFEBVRE
Rémi CUVELLIEZ

AUTEURS

Pour la participation aux différentes actions, la coordination du livrable, la synthèse opérationnelle et le résumé

Catherine BELIN, coordinatrice nationale du REPHY (Ifremer Nantes), catherine.belin@ifremer.fr

Pour la mise en œuvre opérationnelle de l'outil et les actions de formation

Nadine NEAUD-MASSON, assistante à la coordination du REPHY (Ifremer Nantes), nadine.masson@ifremer.fr

Pour l'étude bancarisation

Emilie GAUTHIER, responsable de l'action Quadrige3 (Ifremer Nantes), emilie.gauthier@ifremer.fr

Catherine BELIN, coordinatrice nationale du REPHY (Ifremer Nantes), catherine.belin@ifremer.fr

Antoine HUGUET, chargé d'études (Ifremer Nantes), antoine.huguet@ifremer.fr

Nadine NEAUD-MASSON, assistante à la coordination du REPHY (Ifremer Nantes), nadine.masson@ifremer.fr

Antoine SZIDON (Umanis)

Arnaud LEFEBVRE (Umanis)

Pour le rapport de MASTER 2

Rémi CUVELLIEZ, stagiaire M2 à Ifremer Boulogne

Alain LEFEBVRE, pilote scientifique de l'action FlowCAM-Zoo/PhytoImage (Ifremer Boulogne), alain.lefebvre@ifremer.fr

Pour l'animation du Comité de pilotage de FlowCAM-Zoo/PhytoImage et les CR de réunions

Alain LEFEBVRE, pilote scientifique de l'action FlowCAM-Zoo/PhytoImage (Ifremer Boulogne), alain.lefebvre@ifremer.fr

Catherine BELIN, coordinatrice nationale du REPHY (Ifremer Nantes), catherine.belin@ifremer.fr

CORRESPONDANTS

Onema : Marie Claude XIMENES (Onema), marie-claude.ximenes@onema.fr

Ifremer : Catherine BELIN (Ifremer), catherine.belin@ifremer.fr

AUTRES CONTRIBUTEURS

Pour la mise en œuvre opérationnelle de l'outil

Alain LEFEBVRE (Ifremer Boulogne), alain.lefebvre@ifremer.fr

Danièle MAURER (Ifremer Arcachon), daniele.maurer@ifremer.fr

Mélanie BRUN (Ifremer Nantes), melanie.brun@ifremer.fr

Catherine BELIN (Ifremer Nantes), catherine.belin@ifremer.fr

Les personnels des Laboratoires Environnement Ressources de Boulogne, Port en Bessin, Dinard, Concarneau, Nantes, Arcachon

Pour l'étude bancarisation

Stéphane BOCANDE (Ifremer Brest), stephane.bocande@ifremer.fr

Dominique SOUDANT (Ifremer Nantes), dominique.soudant@ifremer.fr

Alice LAMOUREUX (Ifremer Nantes), alice.lamoureux@ifremer.fr

Gaétane DURAND (Ifremer Nantes), gaetane.durand@ifremer.fr

Catherine BORREMANS (Ifremer Brest), catherine.borremans@ifremer.fr

Philippe GROSJEAN (Université de MONS, Belgique), philippe.grosjean@umons.ac.be

Guillaume WACQUET (Université de MONS, Belgique), guillaume.wacquet@umons.ac.be

Pour le rapport de MASTER 2

Guillaume WACQUET (Université de MONS, Belgique), guillaume.wacquet@umons.ac.be

Pour la participation au Comité de pilotage de FlowCAM-Zoo/PhytoImage

Elvire ANTAJAN (Ifremer Boulogne), elvire.antajan@ifremer.fr

Luis Felipe ARTIGAS (ULCO), felipe.artigas@univ-littoral.fr

Florent COLAS (Ifremer Brest), florent.colas@ifremer.fr

Philippe GROSJEAN (Université de MONS, Belgique), philippe.grosjean@umons.ac.be

Denis HAMAD (ULCO), denis.hamad@lisic.univ-littoral.fr

Luis LAMPERT (Ifremer Brest), luis.lampert@ifremer.fr

Danièle MAURER (Ifremer Arcachon), daniele.maurer@ifremer.fr

Nadine NEAUD-MASSON, assistante à la coordination du REPHY (Ifremer Nantes), nadine.masson@ifremer.fr

Guillaume WACQUET (Université de MONS, Belgique), guillaume.wacquet@umons.ac.be

Droits d'usage : **libre accès**

Niveau géographique : **national**

Couverture géographique : **nationale**

Niveau de lecture : **experts**

RESUME

Ce livrable comprend plusieurs documents qui rendent compte de la mise en œuvre opérationnelle de l'outil FlowCAM / ZooPhytoImage dans le cadre de la surveillance REPHY. Parmi ces documents, trois rapports font un bilan sur l'année 2015 entière, ce qui n'est pas le cas des autres documents (rapport de Master 2 rendu au printemps 2015, et compte-rendus de réunions réparties sur l'année). Le présent résumé se bornera aux conclusions des trois rapports donnant une vision actualisée de cette mise en œuvre de l'outil.

Travaux réalisés en 2015 pour la mise en œuvre opérationnelle pour le REPHY

Les nombreuses numérisations réalisées en 2015 avec les trois FlowCAM présents à Ifremer (Boulogne-sur-Mer, Nantes et Arcachon) ont permis, d'une part d'alimenter les sets d'apprentissage de l'outil couplé FlowCAM-Zoo/PhytoImage, d'autre part de tester son utilisation en routine. Le temps nécessaire pour la numérisation de tous les échantillons d'eau collectés dans le cadre du RePHY Observation (37 points échantillonnés toutes les quinze semaines) a été évalué à 4,9 mois, il reste à chiffrer le temps de traitement des numérisations avec le logiciel Zoo/PhytoImage.

Le travail de particularisation du set global vers des sets spécifiques par façades maritimes distinctes Manche et Atlantique a été poursuivi en 2015 afin d'améliorer l'outil de reconnaissance. Le set d'apprentissage optimisé pour l'Atlantique comprend désormais 7 388 images réparties dans 62 groupes avec un taux d'erreur globale de 26,3 %. Des perfectionnements sont encore possibles : en effet, les outils d'évaluation des performances de reconnaissance proposés par le logiciel Zoo/PhytoImage permettent d'identifier les actions à mener pour parfaire l'optimisation d'un set. L'exploitation de ces résultats d'analyses de performances a été utilisée plus particulièrement pour contrôler le groupe des *Dinophysis* : dans ce set optimisé pour l'Atlantique, ce groupe présente un taux d'erreur de 15,5 % et obtient 80 % de vrais positifs, ce qui est très prometteur et peut encore être amélioré.

Une comparaison des résultats des abondances de *Dinophysis* obtenus au microscope et par l'outil optimisé a également été menée. Les résultats de l'échantillon testé montrent que les abondances obtenues avec l'outil FC/ZI sont sous estimées. Toutefois le test non paramétrique de Wilcoxon ne permet pas de mettre en évidence une différence significative de dénombrement de *Dinophysis* entre les deux méthodes lorsque l'on poursuit les corrections d'erreur jusqu'à validation manuelle de toutes les particules erronées jusqu'à environ 5 %.

La livraison de la dernière version du logiciel Zoo/PhytoImage v5.4 fin 2015, conduit désormais à disposer d'un outil de reconnaissance opérationnel pour l'analyse en routine des échantillons collectés sur la façade Atlantique dans le cadre du REPHY.

Un catalogue illustré des images d'un set d'apprentissage, destiné aux utilisateurs de Zoo/PhytoImage, présente un échantillon d'images de chaque classe qui constitue le set d'apprentissage Manche-Atlantique 4X, et aussi les classes dont les abondances sont insuffisantes pour être intégrés au set et pour lesquelles la collecte de nouvelles images est attendue.

Perspectives et conclusions d'une étude sur la bancarisation des données provenant de ZooPhytoImage

Le traitement des images de phytoplancton via le logiciel ZooPhytoImage, produit de très nombreuses données (identification, abondance, biovolume, et autres métriques), sur les particules contenues dans les échantillons. Une prestation d'expertise a été menée pour proposer une architecture technique de l'ensemble de la chaîne d'acquisition et définir des spécifications d'intégration des données dans la base Quadrige.

En termes d'architecture, un disque réseau partagé est actuellement alloué à ces données : l'étude a montré la pertinence de cette architecture pour les besoins actuels. Toutefois, l'espace de stockage actuel sera rapidement insuffisant. Le projet d'équipement mutualisé de calcul (supercalculateur) et de gestion de données de gros volume nommé DATARMOR, en cours de spécification au sein de l'Ifremer, a été définie comme la cible visée pour la sauvegarde des données FlowCAM et ZooPhytoImage à l'horizon de quelques années.

En termes de bancarisation dans Quadrigé, il a été décidé de ne bancariser que des informations calculées à l'échelle du taxon, et non pas à l'échelle de la particule. Ainsi, les informations relatives aux particules n'appartenant pas aux taxons phytoplanctoniques, ne seront pas bancarisées dans Quadrigé, mais resteront stockées dans le serveur, pour d'éventuelles futures utilisations ou re-jeux. De même pour des métriques n'offrant pas un intérêt immédiat sur les particules phytoplanctoniques : la liste précise des métriques utiles reste à définir, et les paramètres statistiques à bancariser doivent encore être validés (*a minima* médiane et valeurs minimale et maximale, pour chaque métrique décrivant l'ensemble des particules associées à un même taxon).

Pratiquement, les fichiers de résultats produits par ZooPhytoImage (.RData) seront transformés par un script R en fichiers .csv, puis, via un job Talend, en fichiers QuadriLabo, directement intégrables dans Quadrigé. Le référentiel Quadrigé doit être complété pour intégrer les paramètres correspondant aux nouvelles métriques à intégrer, les méthodes d'analyse, etc. Tout ceci alimentera le cahier des charges Quadrigé3.

MOTS CLES (THEMATIQUE ET GEOGRAPHIQUE)

Phytoplancton, FlowCAM, Zoo/PhytoImage, numérisation, analyse d'images, set d'apprentissage, outil de reconnaissance, architecture de données, bancarisation, Quadrigé

TITLE

Operational implementation of FlowCAM-Zoo / Phytolmage tool in the framework of the REPHY surveillance

ABSTRACT

This deliverable includes several documents that reflect the operational implementation of the tool FlowCAM / ZooPhytolmage in the framework of REPHY. Among these documents, three reports make an assessment of the entire year 2015, which is not the case with other documents (Master 2 report issued in spring 2015, and minutes of meetings over the year). This summary will be limited to the findings of the three reports providing an updated view of the implementation of the tool.

Work completed in 2015 for the operational implementation

The many scans performed in 2015 with the three FlowCAM present at Ifremer (Boulogne-sur-Mer, Arcachon and Nantes) allowed, one hand to feed the training sets of the tool FlowCAM-Zoo/Phytolmage, the other test its routine use. The time required for scanning all water samples collected in the RePHY Observation (37 points sampled every two weeks) was evaluated at 4.9 months, it remains to evaluate the processing time for scans with the software Zoo / Phytolmage.

Specific training sets for distinct coastlines Channel and the Atlantic were performed in 2015 to improve recognition tool. The training set optimized for the Atlantic now includes 7388 pictures in 62 groups with an overall error rate of 26.3%. Improvements are still possible: indeed, recognition performance assessment tools offered by the Zoo/Phytolmage software can identify actions to improve the optimization of a set. The exploitation of results of performance analysis has been particularly used to control the *Dinophysis* group: in this set optimized for the Atlantic, this group has a 15.5% error rate and gets 80% of true positives, which is very promising and could be further improved.

A comparison of *Dinophysis* abundances results obtained with the microscope and with the optimized tool was also conducted. The results of the test sample shows that the abundances obtained with the tool are underestimated. However, the nonparametric Wilcoxon test does not demonstrate a significant difference in count *Dinophysis* between the two methods, when pursuing the error corrections to manual validation of all the erroneous particles up to about 5 %.

Delivery of the latest version of Zoo/Phytolmage v5.4, late 2015, now led to have an operational reconnaissance tool for routine analysis of samples collected on the Atlantic coast in the framework of REPHY.

An illustrated catalog of images of a training set, for users of Zoo/Phytolmage, presents a sample of images of each class representing the training set Channel-Atlantic 4X, and also the classes whose abundances insufficient to be integrated into the set, for which the collection of new images is expected.

Outlook and conclusions of a study on the banking data from Zoo/Phytolmage

The treatment of phytoplankton images via Zoo/Phytolmage software, products extensive data (identification, abundance, biovolume, and other metrics) on the particles of the samples. An expertise study was conducted to propose a technical architecture of the entire acquisition chain, and to define data integration requirements in Quadriga database.

In terms of architecture, a shared network drive is currently allocated to these data: the study showed the relevance of this architecture for current needs. However, the current storage will be

soon insufficient. The shared computing equipment project (supercomputer) and large volume data management appointed DATARMOR, being specified within Ifremer, was defined as the target for backup data horizon of a few years.

In terms of banking in Quadrige, it was decided to only bank information on the scale of phytoplankton taxa, not on the scale of the particle. Thus, the information on particles not belonging to the phytoplankton taxa will not be banked in Quadrige but remain stored in the server for possible future use or re-sets. It will be the same for metrics not offering immediate interest on phytoplankton particles: the precise list of useful metric to be defined, and statistical parameters to open bank accounts still need to be validated (at least median and minimum and maximum values, for each metric describing the set of particles associated with the same taxon).

Practically, the results files produced by Zoo/PhytoImage (.RData) will be processed by an R script in .csv files, and then, via a Talend job, in Quadrilabo files directly integrated into Quadrige. The Quadrige repository must be completed to integrate the parameters corresponding to the new metrics, the methods of analysis, etc. All this will feed into the Quadrige3 specifications.

KEY WORDS (THEMATIC AND GEOGRAPHICAL AREA)

Phytoplankton, FlowCAM, Zoo/PhytoImage, scanning, image analysis, training set, recognition tool, data architecture, banking, Quadrige

SYNTHESE POUR L'ACTION OPERATIONNELLE

Ce livrable comprend plusieurs documents qui rendent compte de la mise en œuvre opérationnelle de l'outil FlowCAM / ZooPhytoImage dans le cadre de la surveillance REPHY. Ils sont présentés dans l'ordre suivant, sachant que : les trois premiers sont des rapports faisant un bilan sur l'année 2015 entière, ce qui n'est pas le cas du quatrième qui est un rapport de Master 2 rendu au printemps 2015, et des quatre autres documents qui sont des compte-rendus rendant compte des avancements, discussions, difficultés et décisions prises tout au long de l'année 2015 :

- **Projet FlowCAM / ZooPhytoImage. Travaux réalisés en 2015 pour la mise en œuvre opérationnelle pour le REPHY. N. NEAUD-MASSON, février 2016. Rapport ODE / VIGIES/16-02. 36 pages . Pages 17 à 52 du document pdf**
- **Catalogue illustré des images d'un set d'apprentissage réalisé à l'aide de l'outil FlowCAM – ZooPhytoImage utilisé dans le cadre du REPHY. Images issues de numérisations au grossissement 4X, d'échantillons d'eau, pour le set d'apprentissage global Manche-Atlantique. N. Neaud-Masson, R.INT.ODE/DYNECO/VIGIES/15-14, novembre 2015. 68 pages. Pages 53 à 121 du document pdf**
- **Perspectives et conclusions d'une étude sur la bancarisation des données provenant de ZooPhytoImage 73 pages. Pages 122 à 194 du document pdf**
- **Contribution à l'optimisation d'un réseau d'observation et de surveillance du phytoplancton par analyses d'images. Rémi CUVELLIEZ. Rapport de MASTER 2 ECOLOGIE, spécialité Fonctionnement et Gestion des Ecosystèmes Marins, Université des Sciences et Technologies de Lille (Lille 1) et Université du Littoral Côte d'Opale (ULCO), année 2014-2015. Encadrement Alain LEFEBVRE, Ifremer. 50 pages. Pages 195 à 244 du document pdf**
- **Compte-rendu d'une mission de formation à FlowCAM / ZooPhytoImage v5. Nadine Neaud-Masson, Arcachon, 31 mars au 3 avril 2015. 6 pages. Pages 245 à 250 du document pdf**
- **Compte-rendu de l'audioconférence du 28/05/2015. FlowCAM / ZooPhytoImage. Test de la V5. Ifremer Nantes & Boulogne, Université de Mons. 6 pages. Pages 251 à 256 du document pdf**
- **Lettre d'information n°1, avril 2015. Alain Lefebvre. 2 pages. Pages 257 à 258 du document pdf**
- **Compte rendu du COPIL du 4 novembre 2015 en visio conf. 8 pages. Pages 259 à 266 du document pdf**

Projet FlowCAM / ZooPhytoImage. Travaux réalisés en 2015 pour la mise en œuvre opérationnelle pour le REPHY. N. NEAUD-MASSON, février 2016. Rapport ODE / VIGIES/16-02

Trois laboratoires de l'Ifremer sont équipés d'un FlowCAM dédié au REPHY : Boulogne-sur-Mer, Nantes et Arcachon. En 2015, pour l'ensemble de ces trois laboratoires, 451 numérisations au FlowCAM ont été réalisées à l'objectif 4X pour un volume imagé de 10 ml de chaque échantillon. Ce travail a permis d'une part, d'alimenter les sets d'apprentissage dénommés "trainRephy_4Xlugol" de l'outil couplé FlowCAM-Zoo/PhytoImage (FC/ZI) et d'autre part, de tester son utilisation en routine. Il a aussi permis de chiffrer les temps nécessaires à cette activité.

L'objectif du REPHY est, à terme, d'analyser avec l'outil FC/ZI tous les échantillons de son réseau d'observation qui comprend 37 points prélevés une fois par quinzaine. Compte tenu des temps calculés à partir de la mise en œuvre en 2015, on évalue à 4,9 mois soit un peu plus d'un 1/2 temps

plein agent, le temps nécessaire pour la seule partie numérisation au FlowCAM de tous les échantillons d'eau collectés dans le cadre du Rephy Observation. Il reste à chiffrer le temps de traitement des numérisations avec le logiciel Zoo/PhytoImage.

Le rapport sur l'évaluation d'un outil de reconnaissance commun Manche-Atlantique 4X (Neaud-Masson N, Lefebvre A., février 2015) indiquait la nécessité d'une démarche de particularisation du set global vers des sets spécifiques par façades maritimes distinctes Manche et Atlantique. Ce travail a été poursuivi en 2015 afin d'améliorer l'outil de reconnaissance pour l'analyse d'échantillons provenant de l'Atlantique.

Le set d'apprentissage optimisé pour l'Atlantique présenté et évalué dans le rapport de février 2015 comprenait 7 103 images réparties dans 54 groupes et son taux d'erreur globale était de 27,5 %. A l'issue du travail réalisé tout au long de l'année 2015, ce set comprend maintenant 7 388 images réparties dans 62 groupes avec un taux d'erreur globale de 26,3 %. Il s'est enrichi de nouveaux groupes et a été remanié tout en faisant baisser le taux d'erreur globale de l'outil de reconnaissance correspondant, ce qui est très encourageant. Toutefois des perfectionnements sont encore possibles.

Les outils d'évaluation des performances de reconnaissance proposés par le logiciel Zoo/PhytoImage permettent d'identifier les actions à mener pour parfaire l'optimisation d'un set. L'exploitation de ces résultats d'analyses de performances a été utilisée plus particulièrement pour contrôler le groupe des *Dinophysis* qui, par ailleurs, a fait l'objet d'une comparaison des résultats de dénombrements obtenus par microscopie avec ceux obtenus avec l'outil FC/ZI présenté dans ce rapport. Dans ce set optimisé pour l'Atlantique, le groupe des *Dinophysis* comporte 174 vignettes, présente un taux d'erreur de 15,5 % et obtient 80 % de vrais positifs. Ce score de reconnaissance est très prometteur et peut encore être amélioré. Il a été possible de déterminer que la plus grande confusion qui s'opère avec le groupe des *Dinophysis*, concerne le groupe des *Prorocentrum*. C'est donc dans ce dernier que les vignettes sources de confusion seront en priorité recherchées et écartées du set.

Une comparaison des résultats des abondances de *Dinophysis* obtenus par lecture au microscope inversé et par l'outil couplé FC/ZI v5.4 avec l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique a été menée. Les résultats de l'échantillon testé montrent que les abondances obtenues avec l'outil FC/ZI sont sous estimées. Toutefois le test non paramétrique de Wilcoxon réalisé sur les résultats ne permet pas de mettre en évidence une différence significative de dénombrement de *Dinophysis* avec la méthode d'analyse au microscope lorsque l'on poursuit les corrections d'erreur jusqu'à validation manuelle de toutes les particules erronées jusqu'à environ 5 %. Néanmoins, la confirmation ou l'infirmité du résultat de comparaison des méthodes FC/ZI (SD) et microscopie nécessiterait un nombre de répliques plus conséquent.

Un contrôle du classement final des vignettes d'un des répliques a permis de comptabiliser et retrancher le nombre de faux positifs (vignettes classées à tort dans le groupe *Dinophysis*) et d'ajouter le nombre de faux négatifs. Ainsi, pour l'un des répliques testés, l'abondance de *Dinophysis* obtenue est égale à la médiane des abondances obtenues par microscopie. La sous-estimation ne semble donc pas imputable à la technique de numérisation de l'échantillon au FlowCAM, mais serait plutôt due au taux d'erreur de l'outil de reconnaissance utilisé ici. Enfin, un examen des vignettes classées dans les catégories de débris a mis en évidence un bon nombre de faux négatifs qui concerne d'autres groupes. Ceci induit la nécessité de revoir les vignettes des groupes de débris.

Les éléments présentés dans ce rapport constituent une base pour poursuivre l'optimisation du set "trainRephy_4Xlugol" adapté pour l'Atlantique ainsi que l'amélioration des performances de l'outil de reconnaissance correspondant. Grâce au logiciel Zoo/PhytoImage v5.4 dont nous disposons, ceci sera aisément mis en œuvre et permettra de disposer rapidement d'un outil de reconnaissance opérationnel pour l'analyse en routine des échantillons collectés sur la façade Atlantique.

Catalogue illustré des images d'un set d'apprentissage réalisé à l'aide de l'outil FlowCAM – ZooPhytoImage utilisé dans le cadre du REPHY. Images issues de numérisations au grossissement 4X, d'échantillons d'eau, pour le set d'apprentissage global Manche-Atlantique. N. Neaud-Masson, R.INT.ODE/DYNECO/VIGIES/15-14, novembre 2015.

Un set d'apprentissage "global" Manche-Atlantique 4X composé d'images provenant des banques d'images créées sur les trois sites d'acquisition a été constitué. Les images sont issues de 113 échantillons d'eau de mer d'Atlantique de 2011 à 2014 et 127 de Manche de 2013. Ce set est une base de travail pour les utilisateurs. Il doit être complété notamment pour inclure les classes peu abondantes ou encore non alimentées. Ce set doit aussi être optimisé par les utilisateurs pour créer un outil de reconnaissance performant et adapté.

Les utilisateurs du logiciel ZooPhytoImage, plus particulièrement lors des étapes de corrections d'erreurs de l'analyse d'échantillon, doivent avoir une connaissance visuelle de la représentation des particules dans chaque classe. D'autant plus s'ils n'ont pas participé à la constitution de la bibliothèque d'image. Il est apparu donc nécessaire de créer un catalogue de ces images, car visualiser les images directement dans l'arborescence numérique s'avère très fastidieux.

Ce document présente un échantillon d'images de chaque classe qui constitue ce set d'apprentissage Manche-Atlantique 4X, et aussi les classes dont les abondances sont insuffisantes pour être intégrés au set et pour lesquelles la collecte de nouvelles images est attendue.

Perspectives et conclusions d'une étude sur la bancarisation des données provenant de ZooPhytoImage

Cette partie comprend plusieurs documents, dont un rapport d'étude final, des documents techniques et des compte-rendus de réunions. Une synthèse de l'ensemble est décrite dans le premier document. Ces documents sont présentés dans l'ordre suivant :

- Synthèse de l'étude sur la bancarisation des données ZooPhytoImage. E. Gauthier, février 2016. 6 pages
- Rapport d'étude final. Etude bancarisation des données ZooPhytoImage. Rapport Umanis. Référence F005236CO2. Version 3.1. 02/02/2016. 26 pages
- Arborescence de travail pour l'analyse d'un échantillon par ZooPhytoImage. Document technique. N. Neaud-Masson, décembre 2015. 2 pages
- Fichier Quadrilabo : éléments pour remplir le fichier Quadrilabo, servant d'intermédiaire entre les fichiers csv en provenance de ZooPhytoImage, et Quadrige. Document technique. E. Gauthier, décembre 2015. 3 pages
- Compte-rendu réunion : Bancarisation des données provenant du traitement par ZooPhytoImage. Nantes, 29 septembre 2014 + Réflexions ultérieures, décembre 2014 et janvier 2015. 6 pages
- Compte-rendu de la réunion de lancement de la prestation par Umanis. Nantes, 1^{er} décembre 2015. 15 pages
- Compte-rendu de la réunion d'avancement de la prestation par Umanis. Nantes, 7 janvier 2016. 6 pages
- Compte-rendu de la réunion de restitution de la prestation par Umanis. Nantes, 2 février 2016. 8 pages

En résumé :

Architecture

L'analyse du phytoplancton via la numérisation par FlowCAM, suivie du traitement des résultats via le logiciel ZooPhytoImage, produit des données d'identification, d'abondance, de biovolume, et d'autres métriques quantitatives sur les particules contenues dans les échantillons. Nombre de ces données sont nouvelles par rapport aux résultats de la surveillance du phytoplancton effectuée dans le cadre du réseau REPHY et obtenus par analyse des échantillons au microscope optique. Ce nouveau protocole d'analyse permet de conserver à la fois une image des particules contenues dans les échantillons, et l'ensemble des mesures effectuées sur toutes ces particules. Une prestation d'expertise, menée par la société Umanis a été menée pour proposer une architecture technique de l'ensemble de la chaîne d'acquisition et définir des spécifications d'intégration des données dans la base Quadrige.

Actuellement, un disque réseau partagé a été alloué à ces données. L'étude d'Umanis a montré la pertinence de cette architecture pour les besoins actuels, toutefois, l'espace disque nécessaire va rapidement augmenter, et l'espace de stockage actuel sera rapidement insuffisant. Or, un projet d'équipement mutualisé de calcul (supercalculateur) et de gestion de données de gros volume nommé DATARMOR est en cours de spécification au sein de Ifremer. Cette architecture est donc la cible visée pour la sauvegarde des données FlowCAM et ZooPhytoImage à l'horizon de quelques années.

Bancarisation

En termes de procédure d'intégration des données : les fichiers de résultats produits par ZooPhytoImage sont des fichiers .RData. Un script R transforme ces fichiers .RData en fichiers .csv. La procédure retenue pour l'intégration de ces fichiers dans Quadrige est de les convertir au format Quadrilabo via un job Talend. Enfin, les résultats d'analyse sont intégrés sur un échantillon préalablement créé dans Quadrige. Le référentiel Quadrige doit être complété pour intégrer les paramètres correspondant aux nouvelles métriques à intégrer, les méthodes d'analyse, etc. Les programmes et stratégies REPHY dans Quadrige devront ensuite être mis à jour avec ces nouveaux référentiels.

Pour ce qui concerne les données à intégrer dans Quadrige (sachant que tout le reste sera sauvegardé sous forme de fichiers) : le nombre de résultats générés par les analyses ZooPhytoImage étant très important, il est décidé de ne bancariser que des métriques calculées à l'échelle du taxon, et non pas à l'échelle de la particule. Par ailleurs, les informations relatives aux particules n'appartenant pas aux taxons phytoplanctoniques, ne seront pas bancarisées dans Quadrige. La liste précise des métriques utiles reste à définir, et les paramètres statistiques à bancariser doivent encore être validés (*a priori* médiane et valeurs minimale et maximale).

En termes de perspectives et d'actions à poursuivre : les besoins pour la sauvegarde des données sur DATARMOR ont été remontés, et intégreront le cahier des charges en cours d'élaboration. En 2016, de nombreuses actions ont été listées : mise à jour de la procédure de numérisation, identification des métriques pertinentes à intégrer dans Quadrige², mise à jour du référentiel Quadrige² et des programmes et stratégies, rédaction de consignes de saisie des métadonnées. Par ailleurs, les spécifications de conversion des fichiers ZooPhytoImage au format Quadrilabo, seront détaillées, et le script correspondant devra être développé. Enfin, les résultats de l'ensemble de ce travail alimenteront le cahier des charges Quadrige³.

Contribution à l'optimisation d'un réseau d'observation et de surveillance du phytoplancton par analyses d'images. Rémi CUVELLIEZ. Rapport de MASTER 2 ECOLOGIE, spécialité Fonctionnement et Gestion des Ecosystèmes Marins, Université des Sciences et Technologies de Lille (Lille 1) et Université du Littoral Côte d'Opale (ULCO), année 2014-2015. Encadrement Alain LEFEBVRE, Ifremer.

Le système couplé FlowCAM/ZooPhytoImage semble être un outil très utile pour l'étude de nombreux échantillons. Il nécessite cependant quelques améliorations. Tout d'abord, il est actuellement nécessaire d'utiliser un set d'apprentissage spécifique à la région d'utilisation afin de diminuer au mieux les erreurs de classification. Pour les espèces coloniales, il serait nécessaire de disposer d'un module de dénombrement des colonies (actuellement en développement). Il est difficile actuellement de corriger les erreurs de l'outil de reconnaissance, c'est pourquoi un module d'apprentissage actif va être ajouté au logiciel. Enfin, le logiciel est toujours en version bêta et nécessite encore un travail afin de corriger les bugs et d'améliorer son fonctionnement pour le rendre simple d'utilisation. Il semble également nécessaire de développer un système de numérisation permettant une meilleure qualité (zoom x10) tout en gardant un minimum de temps de passage de l'échantillon (10 minutes par numérisation). Ces objectifs ouvrent donc la porte au développement d'un nouvel appareil appelé FastCAM. Pour conclure, ce système semble être un bon moyen de substitution à la microscopie.

N.B. les remarques relatives aux améliorations à prévoir décrites dans ce rapport de MASTER 2 finalisé au printemps 2015, ont toutes trouvé des solutions pour ce qui concerne le logiciel Zoo/PhytoImage, avec l'ajout d'un module de dénombrement des colonies, d'un apprentissage actif, et l'optimisation de la procédure de validation (cf. livrable 1). Quant à l'amélioration du matériel de numérisation, se reporter au livrable 3. Ce rapport est donc mis ici à titre d'information, mais ses conclusions ne doivent être plus considérées comme valides au jour d'aujourd'hui.

Compte-rendu d'une mission de formation à FlowCAM / ZooPhytoImage v5. Nadine Neaud-Masson, Arcachon, 31 mars au 3 avril 2015.

Les objectifs de cette formation étaient d'apporter une mise à niveau aux personnels Ifremer Arcachon (trois stagiaires) pour l'utilisation de l'outil FlowCAM couplé au logiciel ZooPhytoImage version 5, et d'établir un plan d'action 2015. La formation a inclus des présentations théoriques et des travaux pratiques sur le FlowCAM et le logiciel. Les principales conclusions à ce stade de l'année ont été :

- numérisation : en 2015 Arcachon s'engage à numériser quatre échantillons flore totale par mois, avec trois répliqués
- traitement par ZooPhytoImage : même si l'interface utilisateur a considérablement amélioré l'importation et la mise en forme des données, le module d'analyse, plus particulièrement de correction d'erreur, ne répond pas complètement aux attentes car il est extrêmement chronophage ; pour ces raisons, Arcachon propose de ne pas commencer les traitements avant que ces problèmes ne soient résolus

Compte-rendu de l'audioconférence du 28/05/2015. FlowCAM / ZooPhytoImage. Test de la V5. Ifremer Nantes & Boulogne, Université de Mons.

L'objectif de cette audio-conférence était de faire un debriefing suite à l'échange de mails au sujet de ZooPhytoImage V5 et aux difficultés rencontrées par les utilisateurs, et sur les travaux en cours et à venir. Outre un certain nombre de décisions techniques sur les améliorations à prévoir pour ZooPhytoImage, les points suivants ont été discutés :

- quelle information peut-on attendre d'un tel outil en termes de précision taxonomique ? en effet, l'augmentation du nombre d'échantillons analysés va de pair avec une diminution de la résolution des identifications. Il est cependant rappelé que la plupart du temps, les données REPHY sont agrégées avant d'être exploitées
- il est souligné l'importance de l'échange des données acquises (numérisées et traitées) entre les différents partenaires, pour le développement et l'optimisation des méthodes de traitement
- l'importance du signalement des bugs doit être faite le plus précisément possible avec un rapport comprenant un maximum de détails (avec copie ou film d'écran). Il est crucial que Ifremer ait à demeure une personne ressource maîtrisant ZooPhytoImage

Lettre d'information n°1, avril 2015. Alain Lefebvre

Cette lettre d'information était destinée à informer le comité de pilotage des travaux en cours ou à venir.

Compte rendu du COPIL du 4 novembre 2015 en visio conf.

Ce COPIL a abordé plusieurs sujets, dont un état d'avancement sur :

- le FastCAM, qui permet de numériser un échantillon avec un grandissement X10 à une vitesse quinze fois supérieure au FlowCAM. Si le FastCAM s'avère un meilleur outil que le FlowCAM, le problème reste la construction à l'échelle industrielle : la difficulté est de trouver un industriel qui ne craindrait pas un procès avec FluidImaging. L'autre solution serait une construction et une maintenance en interne Ifremer. Il est aussi possible de démarcher FluidImaging pour éviter la surenchère de leur part s'il se sent concurrencé : la direction de la valorisation d'Ifremer va être démarchée pour voir si cette option est envisageable.
- le logiciel ZooPhytoImage : l'université de Mons a diffusé le 3 novembre 2015 la version 5.4.4. du logiciel accompagné d'un manuel utilisateur rédigé en français. Les principales nouveautés de cette version sont : la refonte du code de ZooPhytoImage pour être compatible avec la dernière version de R (3.2.2), des améliorations dans l'interface utilisateur, l'incorporation du comptage de cellules par colonies, l'optimisation de la correction d'erreur qui permet une diminution très significative du nombre de vignettes à valider, l'incorporation au processus de l'apprentissage actif, un manuel d'utilisation en français et fortement orienté par le traitement IFREMER dans le cadre du REPHY est ajouté au manuel en anglais plus généraliste, la correction d'une série de bugs.
- la bancarisation des données issues de ZooPhytoImage : un Appel d'Offre a été lancé par Ifremer Nantes afin de permettre d'engager le processus. La société Umanis a été retenue. A ce stade, il semble que l'utilisation de niveaux de bancarisation sera nécessaire : un niveau d'information bancarisé dans Quadriges2 (paramètres intégrés de numérisation et des

résultats de la classification) et un niveau d'information détaillé bancarisé dans un système différent (images, protocoles de numérisation et de classification, etc.).

Les projets en cours ou à venir ont été recensés : projets et études en lien direct ou indirect avec l'analyse d'images en flux afin de faciliter notamment les prises de contacts, l'optimisation des montages de projets et des demandes de moyens humain et financier.

Le fonctionnement du Comité de Pilotage a été revu.

Communication et Valorisation

Communications grand public

Lefebvre A., Belin C., (2015). Article pour le document « Les temps forts » de l'ONEMA. Suivi du phytoplancton côtier : vers des analyses automatisées. Parution 2016.

Lefebvre A., 2015. L'observation et la surveillance de l'environnement marin en Manche orientale : de la basse fréquence vers la haute fréquence. Comité de bassin de l'Agence de l'Eau Artois Picardie, CAB, 24/04/2015, Boulogne-sur-Mer.

Lefebvre A., 2015. Jerico-next, la traque des microalgues en temps réel. Le Marin, 22/05/2015.

Communications scientifiques dans des groupes de travail, colloques

Artigas L.F., Bonato S., Creach V., Hébert P.A., Lefebvre A., Lizon F., Poisson-Caillault E., Rijkeboer M., Thyssen M., Veen A., 2015. On the combination of semi-automated approaches and tools for measuring phytoplankton dynamics in coastal waters: implications for monitoring networks. Colloque ASLO, Aquatic Science Meeting 2015, 22-27 février 2015, Granada, Espagne.

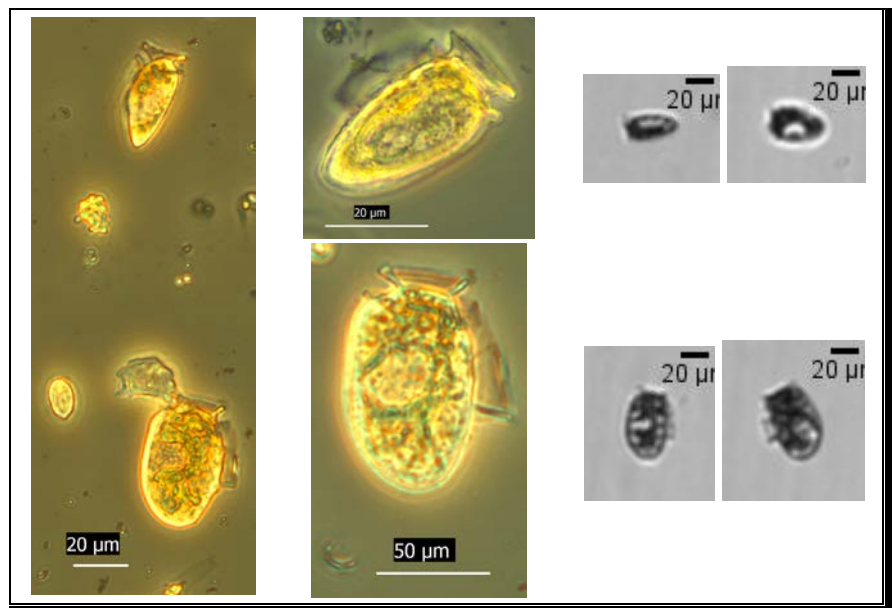
Grosjean P., Denis K., Wacquet G., Ali N., Parent J.Y., Lancelot C., Hamad D., Artigas F., Lefebvre A., Belin C., 2015. Towards a better classification of plankton digital images: suspect detection, error correction and real-time. Colloque ASLO, Aquatic Science Meeting 2015, 22-27 février 2015, Granada, Espagne.

Artigas L.F., Bonato S., Créach V., Didry M., Gomez F., Guiselin N., Hamad D., Hébert P.A., Houliez E., Lefebvre A., Lampert L., Lizon F., Poisson-Caillault E., Prévost E., Rijkeboer M., Thyssen M., Veen A., Rutten T., Wacquet G., 2015. Monitoring of phytoplankton and Harmful Algal Blooms in coastal waters by combining innovative semi-automated tools (scanning flow cytometry & spectral fluorometry). Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Lefebvre A., Neau-Masson N., Maurer D., Wacquet G., Grosjean P., Colas F., Tardivel M., Artigas L.F., Belin C., 2015. Optimization of the monitoring strategy for the French National Phytoplankton and Phycotoxins Network (REPHY) using semi-automated digital images analysis. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Wacquet G., Grosjean P., Hamad D., Lefebvre A., Neaud-Masson N., Colas F., Maurer D., Artigas L.F., 2015. Zoo/PhytoImage : current advances in the semi-automated classification of plankton digital images. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Projet FlowCAM – Zoo/PhytoImage. Travaux réalisés en 2015 pour la mise en œuvre opérationnelle pour le REPHY



RESUME

1	INTRODUCTION	4
2	Nombre de numérisations et traitements Zoo/PhytoImage réalisés en 2015	5
2.1	Laboratoire de NANTES	5
2.2	Laboratoire de BOULOGNE sur MER	8
2.3	Laboratoire d'ARCACHON.....	10
2.4	Bilan et perspectives.....	12
3	Optimisation du set d'apprentissage "trainRephy_4Xlugol" pour l'Atlantique	13
3.1	Description du set "trainRephy_4Xlugol" optimisé et adapté pour l'Atlantique	13
3.2	Performances du set "trainRephy_4Xlugol" optimisé pour l'Atlantique	16
3.3	Conclusion et perspectives.....	23
4	Comparaison des dénombrements de <i>Dinophysis</i> (Microscope vs FlowCAM-Zoo/PhytoImage)	24
4.1	Matériel et méthode	24
4.2	Résultats	26
4.3	Contrôle du classement des vignettes.....	30
4.4	Discussion et suites à donner	32

Annexes

ANNEXE I : Bibliographie	33
ANNEXE II : Liste des tableaux, figures, images et cartes	34

Contributeurs et relecteurs :

Alain Lefebvre (LER/BL)
Danièle Maurer (LER/AR)
Mélanie Brun (ODE/VIGIES)
Catherine Belin (ODE/VIGIES)

Ont participé à la collecte des données :

Le personnel mettant en œuvre le REPHY dans les laboratoires de Boulogne sur Mer, Port en Bessin, Dinard, Concarneau, Nantes et Arcachon

RESUME

Trois sites de l'Unité Littoral du département Océanographie et Dynamique des Ecosystèmes (ODE) de l'Ifremer sont équipés d'un FlowCAM dédié au REPHY. Ces sites sont les Laboratoires Environnement Ressources de : Boulogne-sur-Mer (LER/BL), Nantes (LER/MPL/NT) et Arcachon (LER/AR). En 2015, pour l'ensemble de ces trois, 451 numérisations au FlowCAM ont été réalisées à l'objectif 4X pour un volume imagé de 10 ml de chaque échantillons. Ce travail a permis d'une part, d'alimenter les sets d'apprentissage dénommés "trainRephy_4Xlugol" de l'outil couplé FlowCAM-Zoo/phytolmage (FC/ZI) et d'autre part, de tester son utilisation en routine. Il a aussi permis de chiffrer les temps nécessaires pour cette activité.

L'objectif du REPHY est, à terme, d'analyser avec l'outil FC/ZI tous les échantillons de son réseau d'observation qui comprend 37 points prélevés une fois par quinzaine. Si, par exemple, on répartit équitablement les échantillons vers les trois sites équipés d'un FlowCAM dont dispose actuellement le Rephy, chaque laboratoire aurait 296 numérisations à réaliser (888 par an au total). **Compte tenu des temps calculés à partir de la mise en œuvre en 2015, on évalue à 4,9 mois soit un peu plus d'un 1/2 temps plein agent, le temps nécessaire pour la seule partie numérisation au FlowCAM de tous les échantillons d'eau collectés dans le cadre du Rephy Observation.**

Ces données ainsi que les celles qui restent à chiffrer concernant le temps de traitement des numérisations avec le logiciel Zoo/Phytolmage permettront de mener une réflexion sur la faisabilité et, le cas échéant, s'orienter vers des choix de revue à la baisse, par exemple en sélectionnant des lieux de prélèvement d'intérêt particulier permettant ainsi de limiter la charge analytique des laboratoires.

Le rapport sur l'évaluation d'un outil de reconnaissance commun Manche-Atlantique 4X (Neaud-Masson N, Lefebvre A., février 2015) indiquait la nécessité d'une démarche de particularisation du set global vers des sets spécifiques par façades maritimes distinctes Manche et Atlantique. Ce travail a été poursuivi en 2015 afin d'améliorer l'outil de reconnaissance pour l'analyse d'échantillons provenant de l'Atlantique.

Le **set d'apprentissage optimisé pour l'Atlantique** présenté et évalué dans le rapport de février 2015 comprenait 7 103 images réparties dans 54 groupes et son taux d'erreur globale était de 27,5 %. A l'issue du travail réalisé tout au long de l'année 2015, ce set **comprend maintenant 7 388 images réparties dans 62 groupes avec un taux d'erreur globale de 26,3 %**. Il s'est enrichi de nouveaux groupes et a été remanié tout en faisant baisser le taux d'erreur globale de l'outil de reconnaissance correspondant, ce qui est très encourageant. Toutefois des perfectionnements sont encore possibles.

Les outils d'évaluation des performances de reconnaissance proposés par le logiciel Zoo/Phytolmage permettent d'identifier les actions à mener pour parfaire l'optimisation d'un set. L'exploitation de ces résultats d'analyses de performances a été utilisée plus particulièrement pour contrôler le groupe des *Dinophysis* qui, par ailleurs, a fait l'objet d'une comparaison des résultats de dénombrements obtenus par microscopie avec ceux obtenus avec l'outil FC/ZI présenté dans ce rapport. *Dinophysis* étant un taxon cible du REPHY, il est important que les performances de l'outil permette sa détection et son dénombrement avec le moins d'incertitude possible.

Dans ce set optimisé pour l'Atlantique, le groupe des *Dinophysis* comporte 174 vignettes, présente un taux d'erreur de 15,5 % et obtient 80 % de vrais positifs. Ce score de reconnaissance est très prometteur et peut encore être amélioré. La matrice de confusion générée par Zoo/Phytolmage a permis de déterminer que la plus grande confusion qui s'opère avec le groupe des *Dinophysis*, concerne le groupe des *Prorocentrum*. C'est donc dans ce dernier que les vignettes sources de confusion seront en priorité recherchées et écartées du set. Les autres groupes sources de confusion subiront tout de même ce contrôle.

Les évaluations des performances pour les autres groupes seront réalisées ultérieurement dans l'optique de poursuivre l'optimisation du set. De nouveaux tests seront réalisés avec les acquisitions déjà disponibles d'échantillons variés d'Atlantique ainsi qu'avec de nouvelles acquisitions.

Outre ces outils permettant de cibler les problèmes affectant les performances de l'outil de reconnaissance, les compétences et **l'expérience des analystes du phytoplancton s'avèrent toutefois indispensable pour identifier d'autres sources potentielles de problèmes.**

Une comparaison des résultats des abondances de *Dinophysis* obtenus par lecture au microscope inversé et par l'outil couplé FC/ZI v5.4 avec l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique a été menée. Les résultats de l'échantillon testé montrent que les abondances obtenues avec l'outil FC/ZI sont sous estimées. Toutefois le test non paramétrique de Wilcoxon réalisés sur les résultats ne permet pas de mettre en évidence une différence significative de dénombrement de *Dinophysis* avec la méthode d'analyse au microscope lorsque l'on poursuit les corrections d'erreur jusqu'à validation manuelle de toutes les particules erronées jusqu'à environ 5 %. Néanmoins, la confirmation ou l'infirmité du résultat de comparaison des méthodes FC/ZI (SD) et microscopie nécessiterait un nombre de réplicas plus conséquent.

Un contrôle du classement final des vignettes d'un des réplicas a permis de comptabiliser et retrancher le nombre de faux positifs (vignettes classées à tort dans le groupe *Dinophysis*) et d'ajouter le nombre de faux négatifs. Ainsi, pour l'un des réplicas testés, l'abondance de *Dinophysis* obtenue est égale à la médiane des abondances obtenues par microscopie. **La sous estimation ne semble donc pas imputable à la technique de numérisation de l'échantillon au FlowCAM, mais serait plutôt due au taux d'erreur de l'outil de reconnaissance utilisé ici.**

Ce contrôle du classement des vignettes de *Dinophysis* a permis aussi de confirmer le **fort taux de confusion avec le groupe des *Proocentrum*** détecté lors de l'évaluation des performances de l'outil de reconnaissance optimisé pour l'Atlantique. Enfin, un examen des vignettes classées dans les catégories de débris a mis en évidence un bon nombre de faux négatifs qui concerne d'autres groupes. Ceci induit la nécessité de revoir les vignettes des groupes de débris.

Les éléments présentés dans ce rapport constituent une base pour poursuivre l'optimisation du set "trainRephy_4Xlugol" adapté pour l'Atlantique ainsi que l'amélioration des performances de l'outil de reconnaissance correspondant. Grâce au logiciel Zoo/PhytoImage v5.4 dont nous disposons, ceci sera aisément mis en œuvre et permettra de disposer rapidement d'un outil de reconnaissance opérationnel pour l'analyse en routine des échantillons collectés sur la façade Atlantique dans le cadre du REPHY.

1 INTRODUCTION

Le logiciel Zoo/PhytoImage dans sa version 5.4, couplé au FlowCAM, devient un outil complètement abouti pour une utilisation opérationnelle de l'analyse d'image et de la classification supervisée appliquée au monitoring du phytoplancton.

Totalement adapté aux observations du phytoplancton réalisées dans le cadre du réseau d'observation REPHY, il permet de mieux répondre aux sollicitations présentes et futures concernant l'amélioration des connaissances sur le compartiment phytoplancton (biodiversité, dynamique, phénologie,...) et l'évaluation de la qualité des eaux littorales et marines dans le cadre des exigences européennes, telles que la DCE et la DCSMM, ou de la convention OSPAR (contribution à l'évaluation de l'état d'eutrophisation des masses d'eau).

Les nouvelles techniques de correction d'erreur introduites dans la version 5 du logiciel utilisant ont permis une diminution drastique du taux d'erreur. L'information liée à la validation manuelle des vignettes par l'expert est introduite afin d'optimiser l'approche de type apprentissage actif. Dans ce contexte d'apprentissage actif, la classification ne se fait plus uniquement sur base d'un set d'apprentissage initial et d'un « classifieur » relativement figés. Zoo/PhytoImage identifie de manière probabiliste les particules qui sont sans doute mal classées (les particules suspectes). Lors de la phase de validation par l'expert, l'identification de ces particules suspectes, et les corrections éventuelles sont prises en compte localement au niveau de l'échantillon.

Le rapport sur l'évaluation d'un outil de reconnaissance global Manche-Atlantique 4X (Neaud-Masson N, Lefebvre A., février 2015) indiquait la nécessité d'une démarche de particularisation de ce set global vers des sets spécifiques par façades maritimes distinctes Manche et Atlantique déjà initiée début 2015. La démarche consiste, à partir du set d'apprentissage global, à ajouter des vignettes validées localement, à éliminer progressivement des vignettes au fur et à mesure que des données locales viennent compléter le set. Ceci afin d'améliorer l'outil de reconnaissance pour l'analyse d'échantillons spécifiques de l'Atlantique.

Ce document décrit la poursuite de cette démarche en 2015. Les résultats obtenus par l'analyse d'échantillons avec l'outil FlowCAM-Zoo/PhytoImage sont ensuite comparés avec ceux des lectures réalisées au microscope en ciblant plus particulièrement le groupe des *Dinophysis* taxon cible du REPHY.

En préambule, un inventaire des numérisations d'échantillons aux FlowCAM des trois sites équipés est présenté. Ce travail permet d'alimenter l'évolution des sets spécifiques par façade et de fournir de la matière pour l'évaluation de l'outil Zoo/PhytoImage. Il a aussi permis de quantifier les temps nécessaires pour la partie numérisation d'échantillons au FlowCAM.

2 Nombre de numérisations et traitements Zoo/Phytolmage réalisés en 2015

2.1 Laboratoire de NANTES

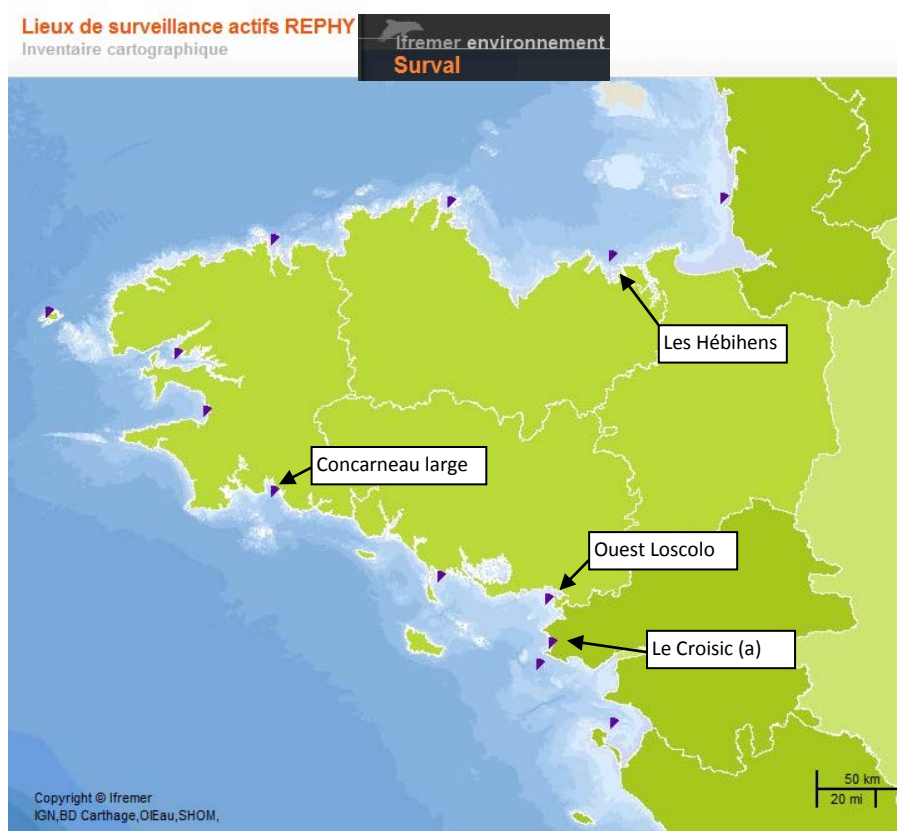
Les prélèvements d'eau de quatre lieux du réseau national REPHY (Réseau d'Observation et de Surveillance du Phytoplancton et de l'Hydrologie), situés en Bretagne – Pays de Loire, faisant l'objet d'un suivi régulier des flores phytoplanctoniques totales et faisant partie du réseau DCE ont été choisis pour être systématiquement numérisés au FlowCAM de Nantes selon la Procédure de numérisation d'échantillon d'eau lugolée du LER/MPL de Nantes, version initiale Tunin-Ley A., Maurer D., (2011) adaptée par Wacquet G., (2014) révisée par Neaud-Masson N., (2014).

Au total, 52 échantillons ont été numérisés (plusieurs répliques). Les numérisations qui ont été commencées à être traitées avec le logiciel Zoo/Phytolmage version 5.4 sont utilisées pour évaluer l'outil de reconnaissance. Certaines vignettes issues de ces échantillons ont pu alimenter le set d'apprentissage "trainRephy_4Xlugol" et enrichir celui-ci avec la création de nouvelles classes.

Les quatre lieux de surveillance retenus sont les suivants :

Lieu de surveillance (libellé)	Latitude	Longitude	Masse d'eau DCE
022-P-018 - les Hébihens	48.643036	-2.207224	FRGC03 - Rance - Fresnaye
047-P-016 - Concarneau large	47.833405	-3.949922	FRGC28 - Concarneau (large)
063-P-002 - Ouest Loscolo	47.457596	-2.538268	FRGC44 - Baie de Vilaine (cote)
068-P-001 - Le Croisic (a)	47.299995	-2.5141633	FRGC45 - Baie de Vilaine (large)

Tableau 1 : Liste des points retenus pour une numérisation systématique des échantillons au laboratoire de Nantes.



Carte 1 : Carte des points retenus pour une numérisation systématique des échantillons au laboratoire de Nantes.

Par ailleurs, des échantillons provenant de quatre lieux de surveillance de la côte Normande ont été sélectionnés en raison des abondances de *Dinophysis* dénombrés dans la perspective d'une comparaison des résultats obtenus avec l'outil FlowCAM-Zoo/Phytolmage (FC/ZI) et ceux des dénombrements au microscope (cf. point 4).

Les points concernés sont les suivants :

Lieu de surveillance (libellé)	Latitude	Longitude	Masse d'eau DCE
010-P-002 - Antifer ponton pêche	49.650417	0.15015987	FRHC16 - Le Havre - Antifer
010-P-109 - Cabourg	49.298405	-0.117002614	FRHC15 - Cote Fleurie
010-P-102 - Ouistreham 1 mille	49.310905	-0.24283881	FRHC14 - Baie de Caen
010-P-187 – Luc 1 mille	49.325904	-0.32917428	FRHC14 - Baie de Caen

Tableau 2 : Liste des points retenus pour une numérisation des échantillons au laboratoire de Nantes en vue d'une comparaison des dénombrements de *Dinophysis*



Carte 2 : Carte des points retenus pour une numérisation des échantillons au laboratoire de Nantes en vue d'une comparaison des dénombrements de *Dinophysis*.

Au total 217 numérisations ont été effectuées à Nantes par **Nadine Neaud-Masson (ODE-VIGIES)**, répartis sur 24 dates, elles sont stockées dans l'arborescence suivante :

<ul style="list-style-type: none"> Acquisition nantes 2015 4X lugol 15-03-13_NN4XQUANTI <ul style="list-style-type: none"> BN.043.2015.03.10.300A4X.01 BN.043.2015.03.10.300A4X.02 BN.043.2015.03-10.300A4X.03 grey 15-03-20_NN4XQUANTI <ul style="list-style-type: none"> NT.021.2015-03-16.300A4X.01 NT.021.2015-03-16.300A4X.02 NT.021.2015-03-16.300A4X.03 15-04-21_NN4XQUANTI <ul style="list-style-type: none"> grey NT.027.2015-04-13.300A4X.01 NT.027.2015-04-13.300A4X.02 NT.027.2015-04-13.300A4X.03 15-04-24_NN4XQUANTI <ul style="list-style-type: none"> BN.072.2015-04-14.300A4X.01 BN.072.2015-04-14.300A4X.02 BN.072.2015-04-14.300A4X.03 grey 15-05-06_NN4XQUANTI <ul style="list-style-type: none"> BN.083.2015-04-28.300A4X.01 BN.083.2015-04-28.300A4X.02 BN.083.2015-04-28.300A4X.03 BO.139.2015-04-28.300A4X.01 BO.139.2015-04-28.300A4X.02 BO.139.2015-04-28.300A4X.03 grey 15-05-11_NN4XQUANTI <ul style="list-style-type: none"> grey NT.031.2015-04-27.300A4X.01 NT.031.2015-04-27.300A4X.02 NT.031.2015-04-27.300A4X.03 NT.039.2015-05-11.300A4X.01 NT.039.2015-05-11.300A4X.02 NT.039.2015-05-11.300A4X.03 TM.074.2015-04-28.300A4X.01 TM.074.2015-04-28.300A4X.02 TM.074.2015-04-28.300A4X.03 15-05-26_NN4XQUANTI <ul style="list-style-type: none"> BN.091.2015-04-28.300A4X.01 BN.091.2015-04-28.300A4X.02 BN.091.2015-04-28.300A4X.03 BO.151.2015-05-12.300A4X.01 BO.151.2015-05-12.300A4X.02 BO.151.2015-05-12.300A4X.03 grey NT.046.2015-05-26.300A4X.01 NT.046.2015-05-26.300A4X.02 NT.046.2015-05-26.300A4X.03 TM.086.2015-05-11.300A4X.01 TM.086.2015-05-11.300A4X.02 TM.086.2015-05-11.300A4X.03 TM.086.2015-05-11.300A4X.04 	<ul style="list-style-type: none"> 15-06-24_NN4XQUANTI <ul style="list-style-type: none"> grey NT.062.2015-06-22.300A4X.01 NT.062.2015-06-22.300A4X.02 NT.062.2015-06-22.300A4X.03 TM.135.2015-06-22.300A4X.01 TM.135.2015-06-22.300A4X.02 TM.135.2015-06-22.300A4X.04 15-07-16_NN4XQUANTI <ul style="list-style-type: none"> BN.113.2015-06-23.300A4X.01 BN.113.2015-06-23.300A4X.02 BN.113.2015-06-23.300A4X.03 BN.125.2015-07-07.300A4X.01 BN.125.2015-07-07.300A4X.02 BN.125.2015-07-07.300A4X.03 grey NT.075.2015-07-06.300A4X.01 NT.075.2015-07-06.300A4X.02 NT.075.2015-07-06.300A4X.03 NT.076.2015-07-07.300A4X.01 NT.076.2015-07-07.300A4X.02 NT.076.2015-07-07.300A4X.03 TM.154.2015-07-07.300A4X.01 TM.154.2015-07-07.300A4X.02 TM.154.2015-07-07.300A4X.03 	<ul style="list-style-type: none"> 15-08-11_NN4XQUANTI <ul style="list-style-type: none"> BN.129.2015-07-21.300A4X.01 BN.129.2015-07-21.300A4X.02 BN.129.2015-07-21.300A4X.03 grey NT.083.2015-07-20.300A4X.01 NT.083.2015-07-20.300A4X.02 NT.083.2015-07-20.300A4X.03 NT.087.2015-08-03.300A4X.01 NT.087.2015-08-03.300A4X.02 NT.087.2015-08-03.300A4X.03 	<ul style="list-style-type: none"> 15-08-12_NN4XQUANTI <ul style="list-style-type: none"> grey TM.159.2015-07-21.300A4X.01 TM.159.2015-07-21.300A4X.02 TM.159.2015-07-21.300A4X.03 TM.159.2015-07-21.300A4X.04 TM.169.2015-08-10.300A4X.01 TM.169.2015-08-10.300A4X.02 TM.169.2015-08-10.300A4X.03 TM.169.2015-08-10.300A4X.04 15-08-22_NN4XQUANTI <ul style="list-style-type: none"> BO.219.2015-06-23.300A4X.01 BO.219.2015-06-23.300A4X.02 BO.219.2015-06-23.300A4X.03 BO.219.2015-06-23.300A4X.04 BO.306.2015-08-04.300A4X.01 BO.306.2015-08-04.300A4X.02 BO.306.2015-08-04.300A4X.03 BO.322.2015-08-17.300A4X.01 BO.322.2015-08-17.300A4X.02 grey TM.183.2015-08-27.300A4X.01 	<ul style="list-style-type: none"> 15-09-03_NN4XQUANTI <ul style="list-style-type: none"> grey N.315.2015-08-12.300A4X.01 N.315.2015-08-12.300A4X.02 N.315.2015-08-12.300A4X.03 N.315.2015-08-12.300A4X.04 N.315.2015-08-12.300A4X.05 N.315.2015-08-12.300A4X.06 N.315.2015-08-12.300A4X.07 N.315.2015-08-12.300A4X.08 N.315.2015-08-12.300A4X.09 N.315.2015-08-12.300A4X.10 N.316.2015-08-12.300A4X.01 N.316.2015-08-12.300A4X.02 N.316.2015-08-12.300A4X.03 N.316.2015-08-12.300A4X.04 15-09-11_NN4XQUANTI <ul style="list-style-type: none"> grey N.316.2015-08-12.300A4X.05 N.316.2015-08-12.300A4X.06 N.316.2015-08-12.300A4X.07 N.316.2015-08-12.300A4X.08 N.316.2015-08-12.300A4X.09 N.316.2015-08-12.300A4X.10 N.317.2015-08-12.300A4X.01 N.317.2015-08-12.300A4X.02 N.317.2015-08-12.300A4X.03 N.317.2015-08-12.300A4X.04 N.317.2015-08-12.300A4X.05 N.317.2015-08-12.300A4X.06 N.317.2015-08-12.300A4X.07 N.317.2015-08-12.300A4X.08 N.317.2015-08-12.300A4X.09 N.317.2015-08-12.300A4X.10 NT.098.2015-09-01.300A4X.01 NT.098.2015-09-01.300A4X.02 NT.098.2015-09-01.300A4X.03 15-10-02_NN4XQUANTI <ul style="list-style-type: none"> BN.142.2015-08-10.300A4X.01 BN.142.2015-08-10.300A4X.02 BN.142.2015-08-10.300A4X.03 grey 15-10-09_NN4XQUANTI <ul style="list-style-type: none"> BN.175.2015-09-07.300A4X.01 BN.175.2015-09-07.300A4X.02 BN.175.2015-09-07.300A4X.03 BO.335.2015-09-01.300A4X.01 BO.335.2015-09-01.300A4X.02 BO.335.2015-09-01.300A4X.03 BO.335.2015-09-01.300A4X.04 grey 15-10-16_NN4XQUANTI <ul style="list-style-type: none"> BO.357.2015-09-15.300A4X.01 BO.357.2015-09-15.300A4X.02 BO.357.2015-09-15.300A4X.03 grey NT.109.2015-09-28.300A4X.01 NT.109.2015-09-28.300A4X.02 NT.109.2015-09-28.300A4X.03 TM.198.2015-09-09.300A4X.01 TM.198.2015-09-09.300A4X.02 TM.198.2015-09-09.300A4X.03 TM.198.2015-09-09.300A4X.04 TM.198.2015-09-09.300A4X.05 TM.204.2015-09-21.300A4X.01 TM.204.2015-09-21.300A4X.02 TM.204.2015-09-21.300A4X.03 	<ul style="list-style-type: none"> 15-10-26_NN4XQUANTI <ul style="list-style-type: none"> BN.181.2015-09-21.300A4X.01 BN.181.2015-09-21.300A4X.02 BN.181.2015-09-21.300A4X.03 BN.181.2015-09-21.300A4X.04 BN.181.2015-09-21.300A4X.05 grey 15-11-13_NN4XQUANTI <ul style="list-style-type: none"> BN.199.2015-10-06.300A4X.01 BN.199.2015-10-06.300A4X.02 BN.199.2015-10-06.300A4X.03 grey NT.135.2015-10-26.300A4X.01 TM.231.2015-10-20.300A4X.01 TM.231.2015-10-20.300A4X.02 TM.231.2015-10-20.300A4X.03 TM.231.2015-10-20.300A4X.04 TM.239.2015-11-03.300A4X.01 TM.239.2015-11-03.300A4X.02 TM.239.2015-11-03.300A4X.03 TM.239.2015-11-03.300A4X.04
---	---	--	--	--	---

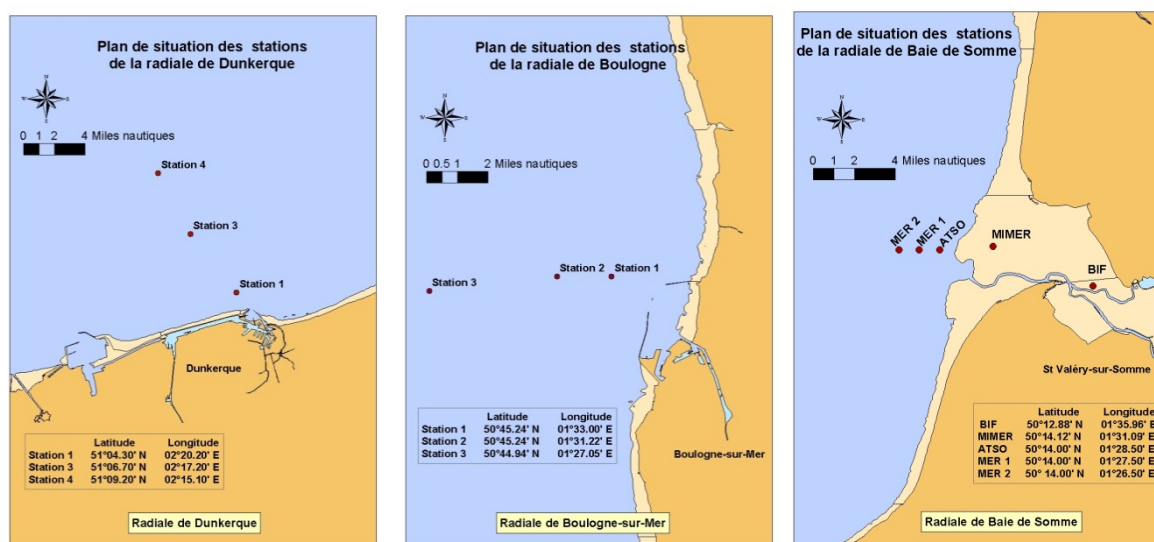
Image 1 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire de Nantes.

2.2 Laboratoire de BOULOGNE sur MER

Le laboratoire de Boulogne-sur-Mer, en collaboration avec l'Agence de l'Eau Artois Picardie, a mis en place en 1992 un réseau de Suivi Régional des Nutriments (S.R.N.) sur le littoral du Nord, du Pas-de-Calais et de la Picardie. Les objectifs de ce suivi sont d'évaluer l'influence des apports continentaux (par exemple nitrates et phosphates) sur le milieu marin et leurs conséquences sur d'éventuels processus d'eutrophisation. L'accent est mis sur le rôle des apports continentaux et marins en nutriments sur les apparitions de *Phaeocystis globosa*, qui est une algue phytoplanctonique responsable de la formation d'écumes nauséabondes sur le littoral et dont la présence en masse peut s'avérer néfaste pour l'écosystème. Il a pour but également d'estimer l'efficacité des stations d'épuration dans l'élimination de tels rejets. L'acquisition régulière des données permet l'établissement d'un suivi à long terme de l'évolution de la qualité des eaux littorales.

Les campagnes de mesures ont lieu mensuellement sur onze points de janvier à décembre, sauf entre mars et juin où l'échantillonnage devient bimensuel du fait de la présence de *Phaeocystis*. Les prélèvements sont effectués au niveau de trois radiales situées dans les eaux côtières de Dunkerque, de Boulogne-sur-Mer et en Baie de Somme. Les données sont utilisées par l'Ifremer et les partenaires locaux (Universités, Agence de l'Eau, ...) en soutien aux programmes de recherches nationaux, internationaux et également dans le contexte des réflexions menées au sein des groupes de travail de la Directive Cadre sur l'Eau et de la convention. Tous les prélèvements d'eau ont été numérisés au FlowCAM.

L'analyse des résultats des numérisations acquises à Boulogne en 2015 grâce au logiciel Zoo/Phytolmage v5.4 est en cours de programmation et devrait être intégrée à un projet de thèse en lien avec l'étude de la dynamique des blooms phytoplanctoniques en Manche orientale - baie sud de la Mer du Nord, dans le contexte des efflorescences de la prymnésiophytécée *Phaeocystis globosa*.



Carte 3 : Cartes des points retenus pour une numérisation des échantillons au laboratoire de Boulogne sur Mer.

Au total 120 numérisations ont été effectuées par **Camille Blondel et Pascale Hébert (LER-BL)**, répartis sur 26 dates, elles sont stockées dans l'arborescence suivante :

<ul style="list-style-type: none"> ▲ 2015-02-20 <ul style="list-style-type: none"> Atso.181.2015-01-19.300A4X.01 Bif.181.2015-01-19.300A4X.01 grey Me2.181.2015-01-19.300A4X.01 Mim.181.2015-01-19.300A4X.01 ▲ 2015-02-25 <ul style="list-style-type: none"> Bil.182.2015-01-23.300A4X.01 Biz.182.2015-01-23.300A4X.01 Bis.182.2015-01-23.300A4X.01 Dk1.182.2015-01-20.300A4X.01 Dk3.182.2015-01-20.300A4X.01 Dk4.182.2015-01-20.300A4X.01 grey ▲ 2015-02-26 <ul style="list-style-type: none"> Atso.183.2015-02-18.300A4X.01 Bif.183.2015-02-18.300A4X.01 grey Me2.183.2015-02-18.300A4X.01 Mim.183.2015-02-18.300A4X.01 ▲ 2015-03-23 <ul style="list-style-type: none"> Bil.184.2015-03-05.300A4X.01 Biz.184.2015-03-05.300A4X.01 Bis.184.2015-03-05.300A4X.01 grey ▲ 2015-03-25 <ul style="list-style-type: none"> Dk1.185.2015-03-06.300A4X.01 Dk3.185.2015-03-06.300A4X.01 Dk4.185.2015-03-06.300A4X.01 grey ▲ 2015-04-08 <ul style="list-style-type: none"> Atso.186.2015-03-10.300A4X.01 Bif.186.2015-03-10.300A4X.01 grey Me2.186.2015-03-10.300A4X.01 Mim.186.2015-03-10.300A4X.01 	<ul style="list-style-type: none"> ▲ 2015-04-10 <ul style="list-style-type: none"> Atso.187.2015-03-20.300A4X.01 Bif.187.2015-03-20.300A4X.01 Bil.187.2015-03-24.300A4X.01 Biz.187.2015-03-24.300A4X.01 Bis.187.2015-03-24.300A4X.01 grey Me2.187.2015-03-20.300A4X.01 Mim.187.2015-03-20.300A4X.01 ▲ 2015-04-17 <ul style="list-style-type: none"> Bil.188.2015-04-08.300A4X.01 Biz.188.2015-04-08.300A4X.01 Bis.188.2015-04-08.300A4X.01 Dk1.188.2015-04-09.300A4X.01 Dk3.188.2015-04-09.300A4X.01 Dk4.188.2015-04-09.300A4X.01 grey ▲ 2015-04-22 <ul style="list-style-type: none"> Atso.189.2015-04-16.300A4X.01 Bif.189.2015-04-16.300A4X.01 grey Me2.189.2015-04-16.300A4X.01 Mim.189.2015-04-16.300A4X.01 ▲ 2015-05-11 <ul style="list-style-type: none"> Bl1.190.2015-04-21.300A4X.01 Bl2.190.2015-04-21.300A4X.01 Bl3.190.2015-04-21.300A4X.01 grey ▲ 2015-05-12 <ul style="list-style-type: none"> Dk1.191.2015.24-04.300A4x.001 Dk3.191.2015.24-04.300A4x.01 Dk4.191.2015.24-04.300A4x.01 grey 	<ul style="list-style-type: none"> ▲ 2015-06-01 <ul style="list-style-type: none"> Atso.192.2015-05-11.300A4X Bif.192.2015-05-11.300A4X Bil.192.2015-05-12.300A4X Biz.192.2015-05-12.300A4X Bis.192.2015-05-12.300A4X grey Me2.192.2015-05-11.300A4X Mim.192.2015-05-11.300A4X ▲ 2015-07-02 <ul style="list-style-type: none"> Atso.193.2015-06-04.300A4X.01 Bif.193.2015-06-04.300A4X.01 Dk1.193.2015-05-27.300A4X.01 Dk1.193.2015-06-04.300A4X.01 Dk3.193.2015-05-27.300A4X.01 Dk4.193.2015-05-27.300A4X.01 grey Me2.193.2015-06-04.300A4X.01 Mim.193.2015-06-04.300A4X.01 ▲ 2015-07-03 <ul style="list-style-type: none"> Atso.194.2015-06-17.300A4X.01 Bif.194.2015-06-17.300A4X.01 Bil.194.2015-06-15.300A4X.01 Biz.194.2015-06-15.300A4X.01 Bis.194.2015-06-15.300A4X.01 Dk1.194.2015-06-16.300A4X.01 Dk3.194.2015-06-16.300A4X.01 Dk4.194.2015-06-16.300A4X.01 grey Me2.194.2015-06-17.300A4X.01 Mim.194.2015-06-17.300A4X.01 ▲ 2015-07-07 <ul style="list-style-type: none"> Bil.195.2015-06-26.300A4X.01 Biz.195.2015-06-26.300A4X.01 Bis.195.2015-06-26.300A4X.01 grey 	<ul style="list-style-type: none"> ▲ 2015-07-08 <ul style="list-style-type: none"> Dk1.196.2015-06-26.300A4X.01 Dk3.196.2015-06-26.300A4X.01 Dk4.196.2015-06-26.300A4X.01 grey ▲ 2015-07-22 <ul style="list-style-type: none"> Atso.197.2015-07-16.300A4X.01 Bif.197.2015-07-16.300A4X.01 Bil.197.2015-07-16.300A4X.01 Biz.197.2015-07-16.300A4X.01 Bis.197.2015-07-16.300A4X.01 grey Me2.197.2015-07-16.300A4X.01 Mim.197.2015-07-16.300A4X.01 ▲ 2015-08-25 <ul style="list-style-type: none"> Atso.199.2015-08-11.300A4X.01 Bif.199.2015-08-11.300A4X.01 Bil.199.2015-08-12.300A4X.01 Biz.199.2015-08-12.300A4X.01 Bis.199.2015-08-12.300A4X.01 Dk1.199.2015-08-19.300A4X.01 Dk3.199.2015-08-19.300A4X.01 Dk4.199.2015-08-19.300A4X.01 grey Me2.199.2015-08-11.300A4X.01 Mim.199.2015-08-11.300A4X.01 ▲ 2015-10-12 <ul style="list-style-type: none"> Bil.200.2015-08-31.300A4X.01 Bil.200.2015-09-25.300A4X.01 Biz.200.2015-09-25.300A4X.01 Bis.200.2015-09-25.300A4X.01 	<ul style="list-style-type: none"> ▲ 2015-10-13 <ul style="list-style-type: none"> Atso.201.2015-09-28.300A4X.01 Bif.201.2015-09-28.300A4X.01 grey Mer2.201.2015-09-28.300A4X Mimer.201.2015-09-28.300A4X.01 ▲ 2015-10-28 <ul style="list-style-type: none"> Atso.202.2015-10-19.300A4X.01 Bif.202.2015-10-19.300A4X.01 grey Me2.202.2015-10-19.300A4X.01 Mim.202.2015-10-19.300A4X.01 ▲ 2015-11-04 <ul style="list-style-type: none"> Bl1.203.2015-10-26.300A4X.01 Bl2.203.2015-10-26.300A4X.01 Bl3.203.2015-10-26.300A4X.01 grey ▲ 2015-11-12 <ul style="list-style-type: none"> Dk1.204.2015-10-23.300A4X.01 Dk3.204.2015-10-23.300A4X.01 Dk4.204.2015-10-23.300A4X.01 grey ▲ 2015-12-23 <ul style="list-style-type: none"> Bl1.205.2015-11-12.A4X.01 Dk1.205.2015-12-15.300A4X.01 Dk3.205.2015-12-15.300A4X.01 Dk4.205.2015-12-15.300A4X.01 grey ▲ 2015-12-28 <ul style="list-style-type: none"> Atso.206.2015-12-14.300A4X.01 Bif.206.2015-12-14.300A4X.01 grey Me2.206.2015-12-14.300A4X.01 Mim.206.2015-12-14.300A4X.01
---	---	--	--	--

Image 2 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire de Boulogne sur Mer.

2.3 Laboratoire d'ARCACHON

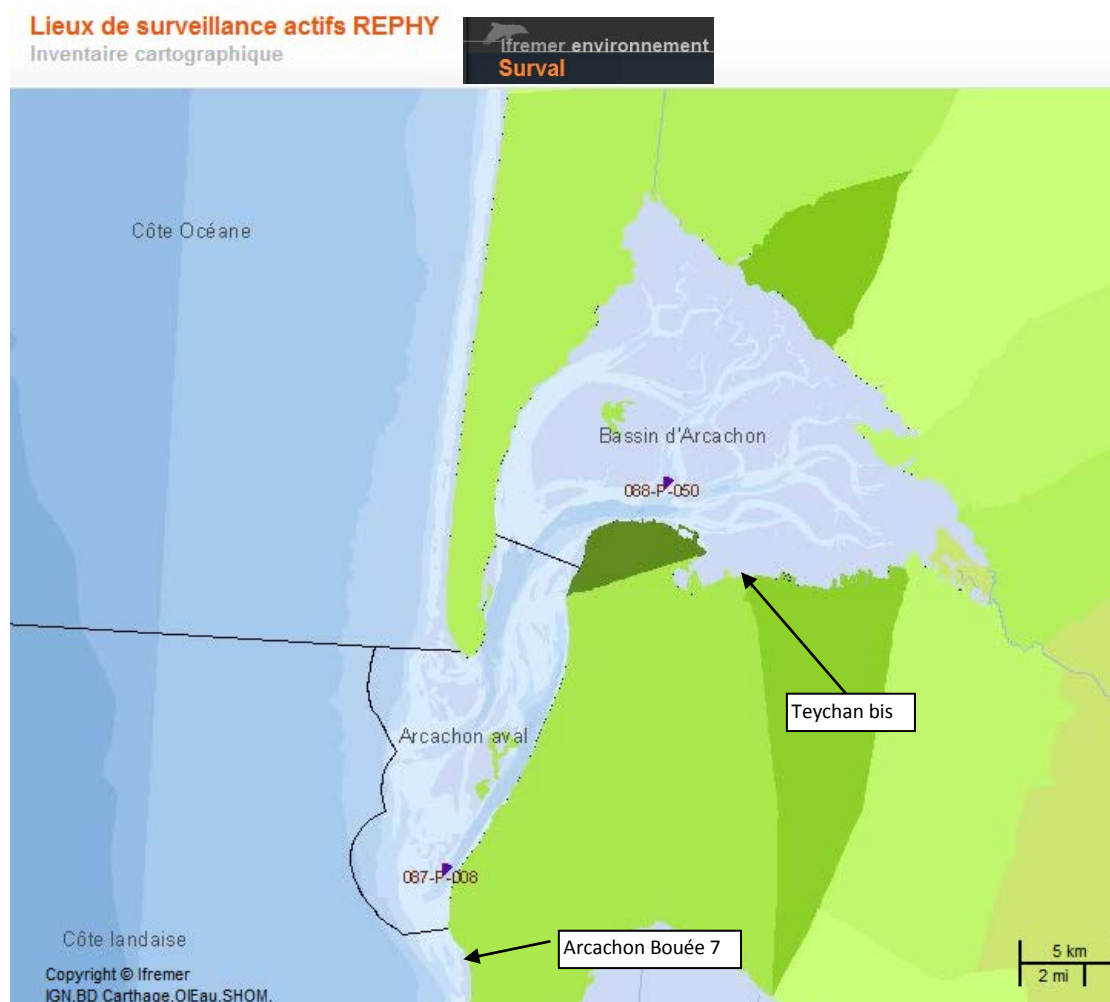
Depuis avril 2015, tous les échantillons collectés sur les points "Arcachon bouée7" et "Teychan bis" dans le bassin d'Arcachon, ont été numérisés au FlowCAM. Ces points ont été choisis car ils font partie du réseau de surveillance DCE et que dans le cadre du REPHY, les flores microphytoplanctoniques totales sont analysées au microscope sur deux échantillons par mois sur chacun de ces points.

L'analyse de ces échantillons avec le logiciel Zoo/Phytolmage v5.4 et la comparaison avec les résultats obtenus par microscopie est programmée en 2016.

Les points concernés sont les suivants :

Lieu de surveillance (libellé)	Latitude	Longitude	Masse d'eau DCE
088-P-050 - Teychan bis	44.673615	-1.1585778	FRFC06 - Arcachon amont
087-P-008 - Arcachon - Bouée 7	44.54228	-1.2637435	FRFC07 - Arcachon aval

Tableau 3 : Liste des points retenus pour une numérisation systématique des échantillons au laboratoire d'Arcachon.



Carte 4 : Carte des points retenus pour une numérisation des échantillons au laboratoire d'Arcachon en vue d'une comparaison des dénombrements de *Dinophysis*.

Au total 109 numérisations ont été effectuées par **Claire Meteigner et Myriam Perrière (LER-AR)**, réparties sur 16 dates, elles sont stockées dans l'arborescence suivante :

<ul style="list-style-type: none"> Acquisition_Arcachon-2015 <ul style="list-style-type: none"> 15-06-29_CM4XQUANTI <ul style="list-style-type: none"> AR.B20.2015-04-13.300A4X.01 AR.B20.2015-04-13.300A4X.02 AR.B20.2015-04-13.300A4X.03 grey 15-06-30_CM4XQUANTI <ul style="list-style-type: none"> AR.B23.2015-04-27.300A4X.01 AR.B23.2015-04-27.300A4X.02 AR.B23.2015-04-27.300A4X.03 AR.B27.2015-05-11.300A4X.01 AR.B27.2015-05-11.300A4X.02 AR.B27.2015-05-11.300A4X.03 AR.T26.2015-04-13.300A4X.01 AR.T26.2015-04-13.300A4X.02 AR.T26.2015-04-13.300A4X.03 AR.T29.2015-04-27.300A4X.01 AR.T29.2015-04-27.300A4X.02 AR.T29.2015-04-27.300A4X.03 grey 15-07-01_CM4XQUANTI <ul style="list-style-type: none"> AR.B31.2015-05-26.300A4X.01 AR.B31.2015-05-26.300A4X.02 AR.B31.2015-05-26.300A4X.03 AR.T33.2015-05-11.300A4X.01 AR.T33.2015-05-11.300A4X.02 AR.T33.2015-05-11.300A4X.03 AR.T37.2015-05-26.300A4X.01 AR.T37.2015-05-26.300A4X.02 AR.T37.2015-05-26.300A4X.03 grey 	<ul style="list-style-type: none"> 15-07-02_CM4XQUANTI <ul style="list-style-type: none"> AR.B35.2015-06-09.300A4X.01 AR.B35.2015-06-09.300A4X.02 AR.B35.2015-06-09.300A4X.03 grey 15-07-03_CM4XQUANTI <ul style="list-style-type: none"> AR.B41.2015-06-25.300A4X.01 AR.B41.2015-06-25.300A4X.02 AR.B41.2015-06-25.300A4X.03 AR.T41.2015-06-09.300A4X.01 AR.T41.2015-06-09.300A4X.02 AR.T41.2015-06-09.300A4X.03 AR.T47.2015-06-25.300A4X.01 AR.T47.2015-06-25.300A4X.02 AR.T47.2015-06-25.300A4X.03 grey 15-07-09_CM4XQUANTI <ul style="list-style-type: none"> AR.B43.2015-07-09.300A4X.01 AR.B43.2015-07-09.300A4X.02 AR.B43.2015-07-09.300A4X.03 AR.T53.2015-07-09.300A4X.01 AR.T53.2015-07-09.300A4X.02 AR.T53.2015-07-09.300A4X.03 grey 	<ul style="list-style-type: none"> 15-07-27_CM4XQUANTI <ul style="list-style-type: none"> AR.B45.2015-07-24.300A4X.01 AR.B45.2015-07-24.300A4X.02 AR.B45.2015-07-24.300A4X.03 AR.T57.2015-07-24.300A4X.01 AR.T57.2015-07-24.300A4X.02 AR.T57.2015-07-24.300A4X.03 grey 15-08-11_CM4XQUANTI <ul style="list-style-type: none"> AR.B47.2015-08-10.300A4X.01 AR.B47.2015-08-10.300A4X.02 AR.B47.2015-08-10.300A4X.03 AR.T59.2015-08-10.300A4X.01 AR.T59.2015-08-10.300A4X.02 AR.T59.2015-08-10.300A4X.03 grey 15-09-08_CM4XQUANTI <ul style="list-style-type: none"> AR.B50.2015-08-25.300A4X.01 AR.B50.2015-08-25.300A4X.02 AR.B50.2015-08-25.300A4X.03 AR.B52.2015-09-07.300A4X.01 AR.B52.2015-09-07.300A4X.02 AR.B52.2015-09-07.300A4X.03 AR.T63.2015-08-25.300A4X.01 AR.T63.2015-08-25.300A4X.02 AR.T63.2015-08-25.300A4X.03 AR.T64.2015-09-07.300A4X.01 AR.T64.2015-09-07.300A4X.02 AR.T64.2015-09-07.300A4X.03 grey 	<ul style="list-style-type: none"> 15-09-29_CM4XQUANTI <ul style="list-style-type: none"> AR.B54.2015-09-21.300A4X.01 AR.B54.2015-09-21.300A4X.02 AR.B54.2015-09-21.300A4X.03 AR.T67.2015-09-21.300A4X.01 AR.T67.2015-09-21.300A4X.02 AR.T67.2015-09-21.300A4X.03 grey 15-11-05_MPR4XQUANTI <ul style="list-style-type: none"> AR.B55.2015-10-08.300A4X.01 AR.B55.2015-10-08.300A4X.02 AR.B55.2015-10-08.300A4X.03 AR.T68.2015-10-08.300A4X.01 AR.T68.2015-10-08.300A4X.02 AR.T68.2015-10-08.300A4X.03 AR.T71.2015-10-20.300A4X.01 AR.T71.2015-10-20.300A4X.02 AR.T71.2015-10-20.300A4X.03 Grey 15-11-06_MPR4XQUANTI <ul style="list-style-type: none"> AR.B58.2015-10-20.300A4X_01 AR.B58.2015-10-20.300A4X_02 AR.B58.2015-10-20.300A4X_03 AR.B60.2015-11-05.300A4X_01 AR.B60.2015-11-05.300A4X_02 AR.B60.2015-11-05.300A4X_03 AR.T73.2015-11-05.300A4X_01 AR.T73.2015-11-05.300A4X_02 AR.T73.2015-11-05.300A4X_03 Grey 	<ul style="list-style-type: none"> 15-11-23_MPR4XQUANTI <ul style="list-style-type: none"> AR.B62.2015-11-19.300A4X_01 AR.B62.2015-11-19.300A4X_02 AR.B62.2015-11-19.300A4X_03 Grey 15-11-24_MPR4XQUANTI <ul style="list-style-type: none"> AR.T75.2015-11-19.300A4X_01 AR.T75.2015-11-19.300A4X_02 AR.T75.2015-11-19.300A4X_03 Grey 15-12-22_MPR4XQUANTI <ul style="list-style-type: none"> AR.B64.2015-12-03.300A4X_01 AR.B64.2015-12-03.300A4X_02 AR.B64.2015-12-03.300A4X_03 AR.B66.2015-12-18.300A4X_01 AR.B66.2015-12-18.300A4X_02 AR.B66.2015-12-18.300A4X_03 AR.T77.2015-12-03.300A4X_01 AR.T77.2015-12-03.300A4X_02 AR.T77.2015-12-03.300A4X_03 Grey 16-01-04_MPR4XQUANTI <ul style="list-style-type: none"> AR.T79.2015-12-18.300A4X_01 AR.T79.2015-12-18.300A4X_02 AR.T79.2015-12-18.300A4X_03 GREY
--	---	---	---	--

Image 3 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire d'Arcachon

2.4 Bilan et perspectives

En 2015, pour l'ensemble des trois laboratoires équipés d'un FlowCAM, 451 numérisations réparties sur 66 dates ont été réalisées à l'objectif 4X pour un volume imagé de 10 mL. Les laboratoires d'Arcachon et de Nantes ont réalisé systématiquement au moins 3 répliques de chaque échantillon, alors que le laboratoire de Boulogne n'a numérisé chaque échantillon qu'une seule fois. **Afin de s'assurer l'obtention d'une numérisation exploitable, il faut vérifier sur les images acquises qu'il n'y a pas eu colmatage ou obstruction par une bulle provoquant ainsi la prise d'images multiples d'une même particule. Si c'est le cas, une nouvelle numérisation de l'échantillon doit être réalisée.**

La durée d'une numérisation d'un volume de 10 mL imagé à l'objectif 4X est de 11 minutes et demie en moyenne avec les appareils de Boulogne et de Nantes et de 21 minutes avec l'appareil d'Arcachon. Cette différence est due au système de pompage de l'échantillon : pompe péristaltique à Arcachon et pompe seringue à Nantes et Boulogne sur mer. Cette différence technique implique aussi une différence de contexte de numérisation. A ces temps s'ajoutent les temps de mise en route de l'appareil, des rinçages entre chaque échantillon, de nettoyage de l'appareil en clôture de session d'acquisition ainsi que la gestion documentaire environnante telle que l'enregistrement des échantillons par exemple. Tenant compte de cela, l'estimation du temps consacré au passage d'un échantillon au FlowCAM est fixée à 15 minutes avec les appareils de Boulogne et de Nantes et à 25 minutes à Arcachon.

L'objectif à terme pour le Rephy est d'analyser avec l'outil FC/ZI tous les échantillons de de sa composante Observation qui comprend 37 points prélevés une fois par quinzaine. Si, par exemple, on répartit équitablement les échantillons vers les trois FlowCAM dont dispose actuellement le Rephy, chaque laboratoire aurait 296 numérisations à réaliser (888 par an au total). Pour les laboratoires de Boulogne et de Nantes, cela représenterai 1,3 mois, et pour le laboratoire d'Arcachon 2,2 mois, soit un total de 4,9 mois ce qui équivaut à un peu plus d'1/2 plein temps agent.

Le Rephy doit utiliser ces données ainsi que les estimations qui restent à chiffrer et qui concerne le traitement des numérisation avec le logiciel Zoo/PhytoImage, afin de mener une réflexion sur la faisabilité et, le cas échéant, s'orienter vers des choix à la baisse, par exemple, en sélectionnant des lieux de prélèvement d'intérêt particulier et limiter ainsi la charge analytique des laboratoires.

3 Optimisation du set d'apprentissage "trainRephy_4Xlugol" pour l'Atlantique

Pour rappel, d'un set d'apprentissage est constitué à partir d'une banque d'images issues de l'appareil de numérisation (ici, le FlowCAM) et représentatives des particules rencontrées dans les échantillons à analyser ultérieurement. Ces images sont identifiées visuellement par l'observateur puis classées dans autant de dossiers ou sous-dossiers que nécessaires pour représenter les niveaux d'identification souhaités.

Un outil de reconnaissance automatique, ou semi-automatique, du phytoplancton utilise le set d'apprentissage précédemment créé pour « entraîner » un algorithme de classification supervisée à reconnaître la nature des particules sur la base des mesures obtenues sur les images numériques. Cet outil, nommé classier en anglais, peut être assimilé à une boîte noire capable de déterminer le groupe d'une image uniquement sur la base des paramètres mesurés sur celle-ci.

Cette partie décrit la poursuite, en 2015, du travail d'optimisation du set d'apprentissage global Manche-Atlantique 4X afin d'améliorer ses performances pour l'analyse d'échantillons typiques d'Atlantique, ceci en suivant les recommandations de Tunin-Ley et Maurer (2011), et de Wacquet et Lefebvre (2014).

L'optimisation de l'outil de reconnaissance passe par un examen des vignettes de chaque groupe du set d'apprentissage, d'ajout de vignettes validées localement et d'élimination des vignettes jugées sources de confusions ou inopportunes dans l'analyse des échantillons locaux.

Les performances de l'outil de reconnaissance issu du set d'apprentissage obtenu sont ensuite évaluées grâce aux différentes représentations graphiques générées par les outils du logiciel Zoo/PhytoImage.

L'étude des résultats de l'évaluation des performances de l'outil de reconnaissance permet enfin de cibler les actions supplémentaires à mettre en œuvre sur le set d'apprentissage pour parfaire l'optimisation de l'outil de reconnaissance correspondant.

3.1 Description du set "trainRephy_4Xlugol" optimisé et adapté pour l'Atlantique

L'enrichissement du set "trainRephy_4Xlugol" a été permis par l'acquisition de nouvelles vignettes lors des numérisations effectuées tout au long de l'année 2015.

L'examen des vignettes a abouti à l'élimination de celles qui suggèrent des risques d'erreur dans certains groupes et à l'écartement du groupe des *Phaeocystis* car l'occurrence de ce taxon est rare dans les échantillons d'Atlantique et, de plus, génère de nombreuses confusions avec d'autres groupes comme celui des *Chaetoceros socialis* par exemple.

Par ailleurs, un remaniement des groupes des particules de débris a été opéré dans le but d'homogénéiser les images.

A l'issue de ce travail, les débris, avec un exemple de vignette représentative, sont répartis dans les groupes suivants:

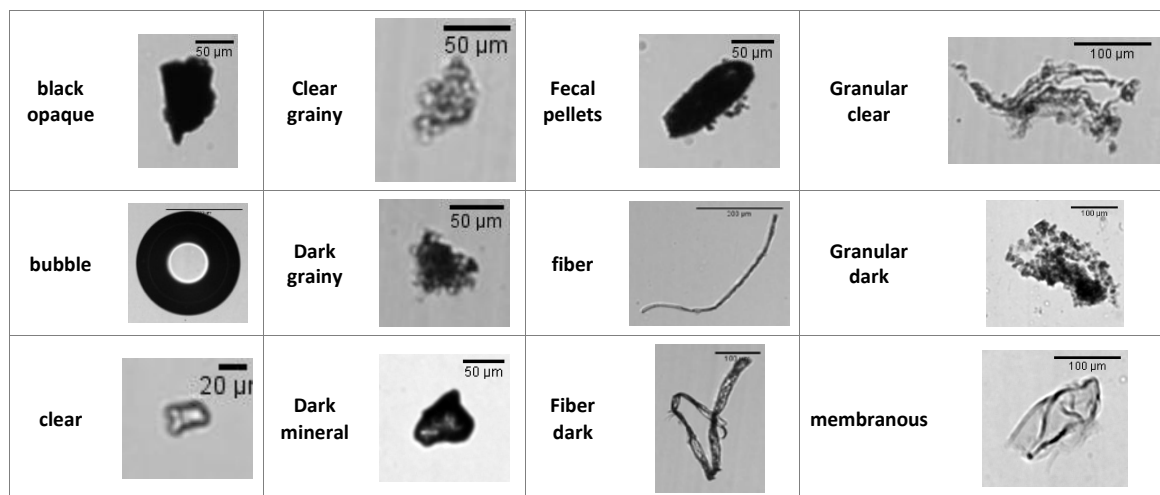


Image 4 : Groupes illustrés des particules "débris" du set "trainRephy_4Xlugol" optimisé pour l'Atlantique.

Trois groupes de phytoplancton ont pu être intégrés au set d'apprentissage grâce à l'obtention d'images supplémentaires de ces taxa.

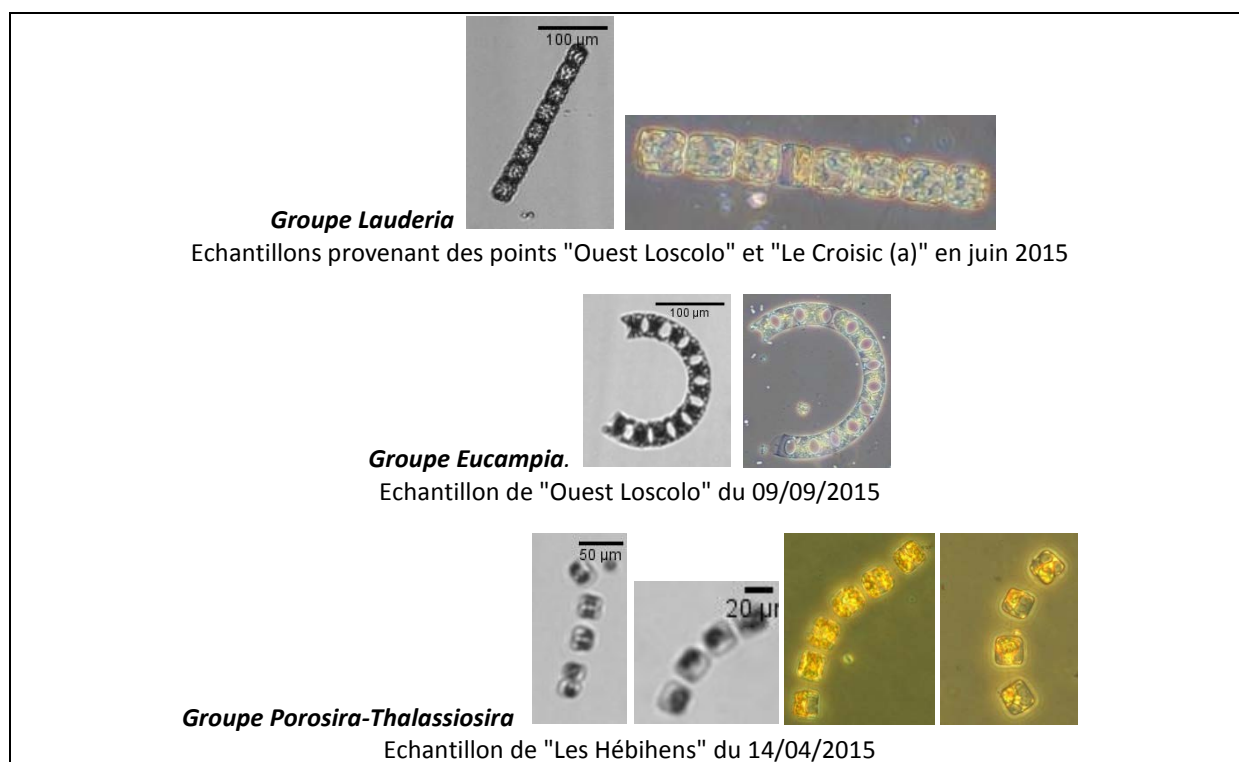


Image 5 : Groupes illustrés des taxons intégrés en 2015 au set "trainRephy_4Xlugol" optimisé pour l'Atlantique (images du FlowCAM et photos prises au microscope).

Les utilisateurs de cet outil n'ayant pas forcément participé à l'élaboration du set d'apprentissage ont besoin de visualiser les vignettes du set afin de s'y référer pour les étapes de correction d'erreur. A cette fin, un catalogue d'images du set a été réalisé¹.

¹ Neaud-Masson N (2015). Catalogue illustré des images d'un set d'apprentissage réalisé à l'aide de l'outil FlowCAM – ZooPhytoImage utilisé dans le cadre du REPHY - Images issues de numérisations au grossissement 4X, d'échantillons d'eau, pour le set d'apprentissage global Manche-Atlantique. R.INT.ODE/DYNECO/VIGIES/15-14

Ce set comprend 7 388 images classées dans 62 groupes dont les effectifs sont présentés dans le Tableau 4.

- 28 groupes phytoplanctoniques appartenant aux Bacillariophyceae
- 15 groupes phytoplanctoniques appartenant aux Dinophyceae
- 4 groupes phytoplanctoniques autres
- 3 groupes zooplanctoniques
- 12 groupes détritiques

Bacillariophyceae		Nb	Dinophyceae		Nb
<i>Asterionellopsis glacialis</i>		179	<i>Alexandrium</i>		187
<i>Bacteriastrum</i>		176	<i>Alexandrium affine</i>		93
<i>Cerataulina</i>		123	<i>Ceratium</i>		52
<i>Chaetoceros</i>		203	Dinophysis		174
<i>Chaetoceros curvisetus</i>		194	<i>Dinophysis tripos</i>		35
<i>Chaetoceros danicus</i>		62	<i>Dissodinium Pyrocystis</i>		41
<i>Chaetoceros socialis</i>		60	<i>Gonyaulax</i>		48
<i>Cylindrotheca Hantzschia Nitzschia</i>		68	<i>Gymnodinium Gyrodinium Torodinium</i>		124
<i>Dactyliosolen fragilissimus</i>		175	<i>Lepidodinium mucous</i>		170
<i>Ditylum brightwellii</i>		184	<i>Neoceratium furca lineatum</i>		151
<i>Eucampia</i>		107	<i>Neoceratium fusus</i>		183
<i>Guinardia delicatula</i>		31	<i>Peridinium Protoperidinium</i>		177
<i>Guinardia flaccida</i>		202	<i>Polykrikos</i>		22
<i>Guinardia striata</i>		78	<i>Prorocentrum</i>		258
<i>Gyrosigma Pleurosigma</i>		221	<i>Scrippsiella Ensiculifera Pentapharsodinium Bysmatrum</i>		149
<i>Lauderia</i>		43			
<i>Leptocylindrus</i>		197	zooplancton		
<i>Odontella</i>		148	<i>Ciliophora_cils</i>		94
<i>Paralia</i>		84	<i>Ciliophora_cone</i>		94
<i>Porosira-Thalassiosira</i>		127	Zooplankton spp		34
<i>Proboscia Rhizosolenia</i>		161			
<i>Pseudo-nitzschia</i>		185	Débris		
<i>Rhizosolenia setigera</i>		41	black opaque		160
<i>Skeletonema</i>		153	bubble		200
<i>Thalassionema</i>		130	clear		163
<i>Thalassiosira cells</i>		38	Clear grainy		57
<i>Thalassiosira big chaines</i>		107	Dark grainy		41
<i>Thalassiosira small chaines</i>		97	Dark mineral		200
			Fecal pellets		20
Autres			fiber		120
<i>Phaeocystis</i>		264	Fiber dark		59
<i>Dictyocha</i>		94	Granular clear		74
<i>Euglenoidea</i>		53	Granular dark		100
<i>Noctiluca</i>		43	membranous		80

Tableau 4 : Distribution des nombres de particules par groupe du set "trainRephy_4Xlugol" optimisé pour l'Atlantique.

Le classifieur généré avec ce set estime que 5 207 vignettes sont de vrais positifs et que le pourcentage global d'erreur est de 26,28 %.

La répartition dans l'ordre croissant des pourcentages d'erreur par classe est la suivante :

groupes	Error (%)	groupes	Error (%)
bubble	5,00	clear	28,83
<i>Euglenoidea</i>	5,66	fiber	29,17
<i>Ceratium</i>	5,77	<i>Zooplankton spp</i>	29,41
<i>Prorocentrum</i>	6,59	<i>Cerataulina</i>	32,52
<i>Neoceratium furca lineatum</i>	6,62	<i>Eucampia</i>	32,71
<i>Alexandrium</i>	6,95	granular dark	33,00
<i>Noctiluca</i>	6,98	<i>Rhizosolenia setigera</i>	36,59
<i>Guinardia flaccida</i>	8,42	<i>Chaetoceros socialis</i>	36,67
<i>Neoceratium fusus</i>	8,74	<i>Cylindrotheca Hantzschia Nitzschia Ceratoneis</i>	36,77
<i>Chaetoceros danicus</i>	11,29	<i>Thalassiosira cells</i>	36,84
<i>Asterionellopsis glacialis</i>	12,29	<i>Lauderia</i>	37,21
<i>Gyrosigma Pleurosigma</i>	13,12	<i>Gonyaulax</i>	37,50
<i>Pseudo-nitzschia</i>	13,51	<i>Gymnodinium Gyrodinium Torodinium</i>	39,52
<i>Lepidodinium mucous</i>	14,12	<i>Polykrikos</i>	40,91
<i>Dactyliosolen fragilissimus</i>	15,43	<i>Ciliophora cils</i>	42,55
<i>Dinophysis</i>	15,52	<i>Thalassiosira big chaines</i>	43,93
<i>Skeletonema</i>	15,69	membranous	45,00
<i>Scripsiella Ensiculifera Pentapharsodinium Bysmatrum</i>	17,45	<i>Thalassiosira small chaines</i>	46,39
<i>Dictyocha</i>	18,09	<i>Peridinium Protoperidinium</i>	47,46
<i>Alexandrium affine</i>	18,28	<i>Guinardia striata</i>	50,00
<i>Porosira-Thalassiosira</i>	19,69	<i>Dissodinium Pyrocystis</i>	51,22
<i>Leptocylindrus</i>	19,80	<i>Chaetoceros</i>	56,16
<i>Paralia</i>	20,24	fiber dark	59,32
<i>Bacteriastrium</i>	22,16	fecal pellets	60,00
black opaque	22,50	dark grainy	63,06
<i>Ditylum brightwellii</i>	23,37	dark mineral	63,42
<i>Odontella</i>	23,65	granular clear	66,22
<i>Thalassionema</i>	24,62	clear grainy	73,68
<i>Proboscia Rhizosolenia</i>	24,85	<i>Ciliophora cone</i>	74,47
<i>Chaetoceros curvisetus pseudocurvisetus</i>	25,77	<i>Guinardia delicatula</i>	100,00
<i>Dinophysis tripos</i>	28,57		

Tableau 5 : Estimation des erreurs par classe du set "trainRephy_4Xlugol".

3.2 Performances du set "trainRephy_4Xlugol" optimisé pour l'Atlantique

A cette étape, le logiciel Zoo/PhytoImage propose des outils d'analyse des performances de l'outil de reconnaissance ou "Classifieur", en montrant les différences entre la classification manuelle des vignettes et celle obtenue par le logiciel. Les représentations matricielles et graphiques permettent à l'utilisateur de visualiser les résultats.

Les figures qui suivent montrent les analyses des performances du classifieur généré par le set "trainRephy_4Xlugol" optimisé pour l'Atlantique.

Nous utilisons ici l'exploitation de ces résultats d'analyses de performances pour cibler plus particulièrement le groupe des ***Dinophysis*** qui, par ailleurs, fait l'objet d'une comparaison des résultats de dénombrements obtenus par microscopie et avec l'outil FC/ZI présentée au point 4 de ce rapport.

Les évaluations des performances pour les autres groupes seront réalisées ultérieurement dans l'optique de poursuivre l'optimisation du set. De nouveaux tests seront alors réalisés avec des acquisitions d'échantillons variés d'Atlantique déjà disponibles et listés au point 1.

Dans cet outil de reconnaissance, le groupe des *Dinophysis* comporte 174 vignettes et présente un taux d'erreur de 15,5 %.

F-score (precision versus recall)

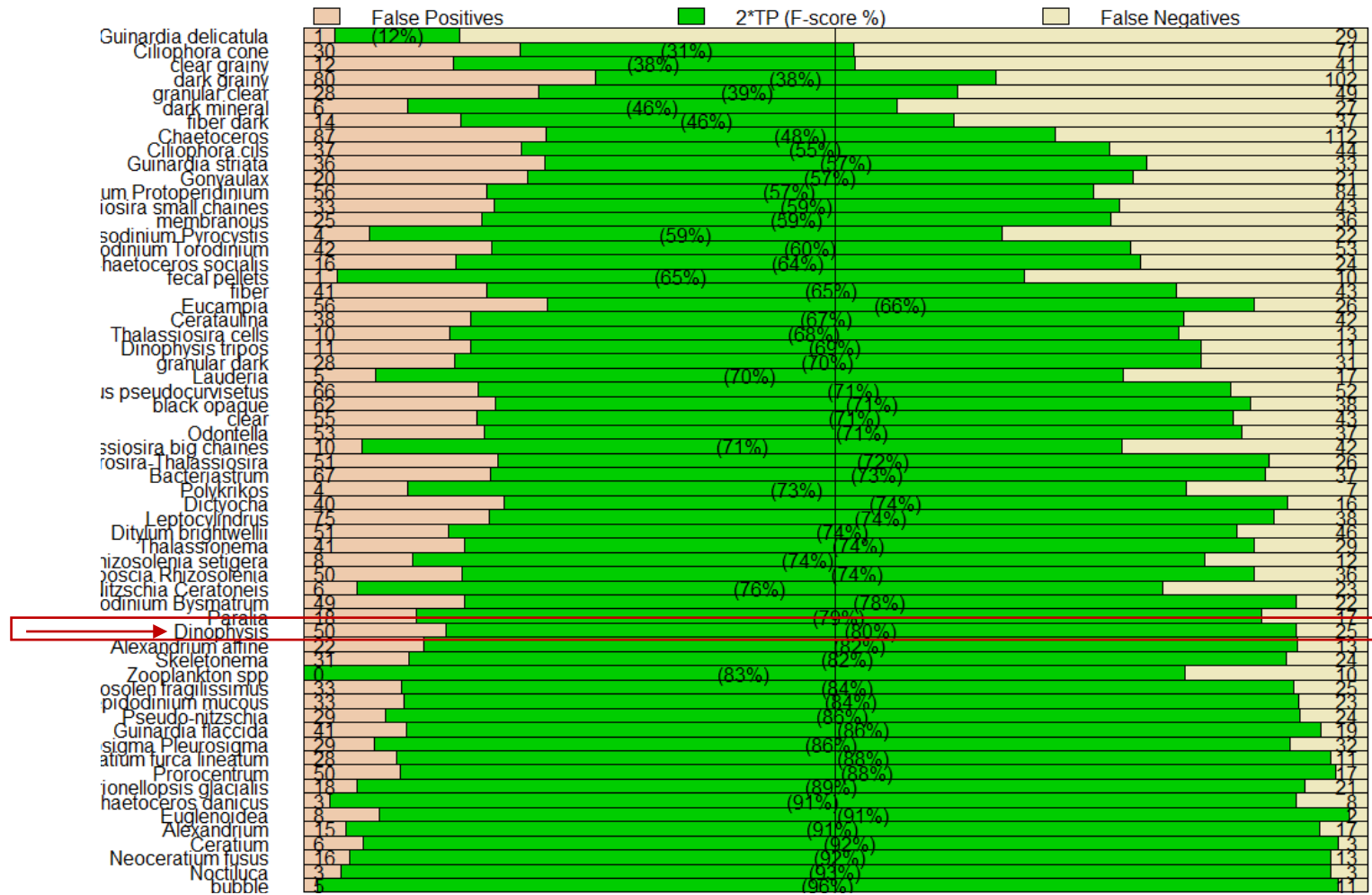


Figure 1 : Représentation graphique 1 des taux de "Précision" et de "Recall" calculés par groupe pour l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique et généré par l'algorithme Random Forest.

Groups comparison

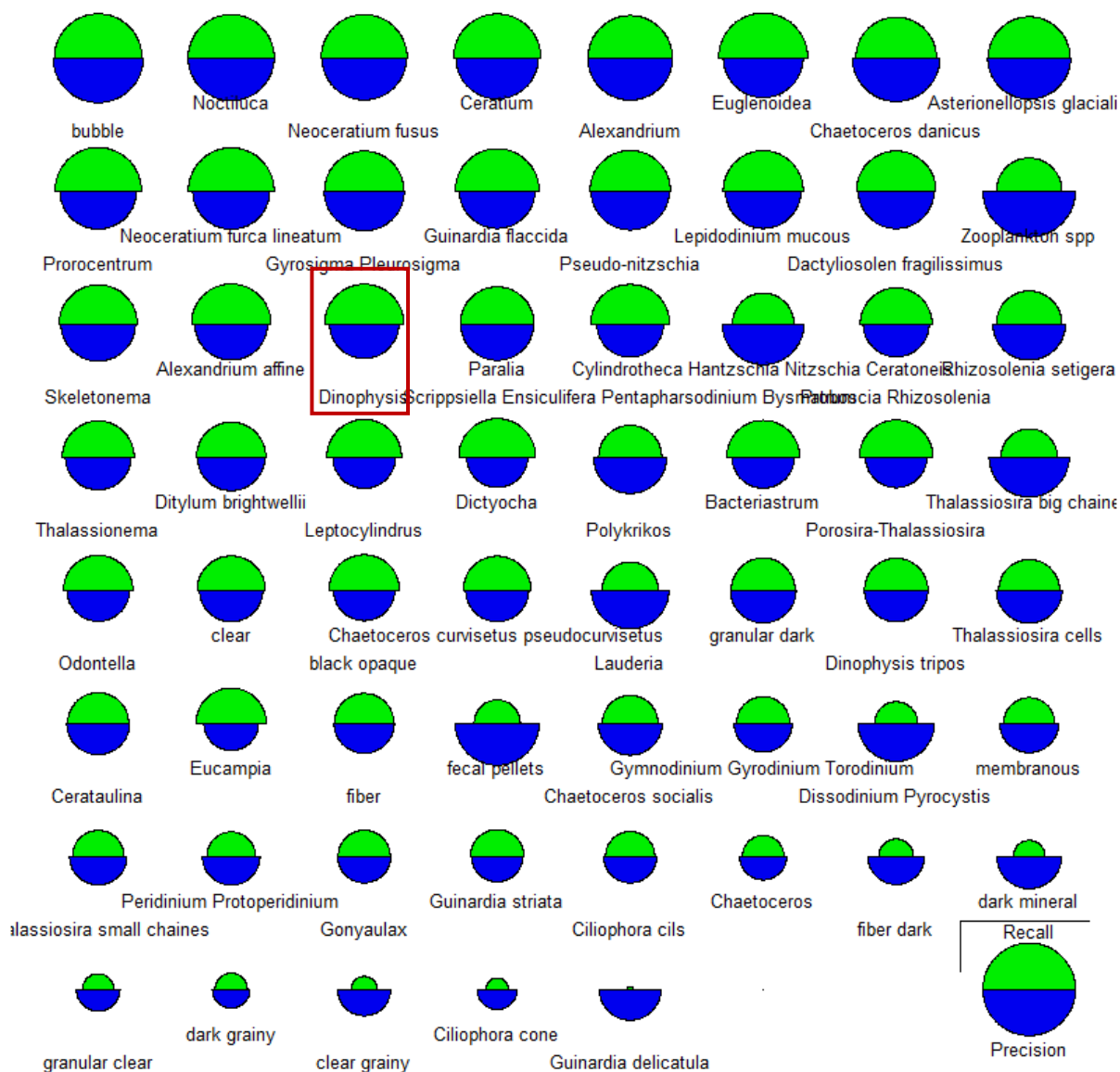


Figure 2 : Représentation graphique 2 des taux de "Précision" et de "Recall" calculés par groupe pour l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique et généré par l'algorithme Random Forest.

La Figure 1 permet de visualiser les taux de "Précision" et de "Recall", les faux positifs et les faux négatifs de chaque groupe. La Figure 2 permet de visualiser simplement et plus rapidement les taux de "Précision" et de "Recall".

Sur la Figure 1, le groupe des *Dinophysis* obtient 80 % de vrais positifs. Ce score de reconnaissance est très encourageant. La barre verte est décalée vers la droite ce qui induit un risque de surestimation pour ce groupe avec un biais égal à +25.

La Figure 2 n'apporte pas plus d'information pour le groupe des *Dinophysis*. Toutefois, cette figure sera utilisée ultérieurement pour l'évaluation prévue des autres groupes.

Afin de connaître le nom des groupes comportant des particules identifiées à tort comme étant des *Dinophysis* et les groupes vers lesquels des particules de *Dinophysis* ont été classés à tort, la visualisation de la matrice de confusion est nécessaire. Ceci permettra de pointer les groupes à revoir afin d'améliorer les discriminations.

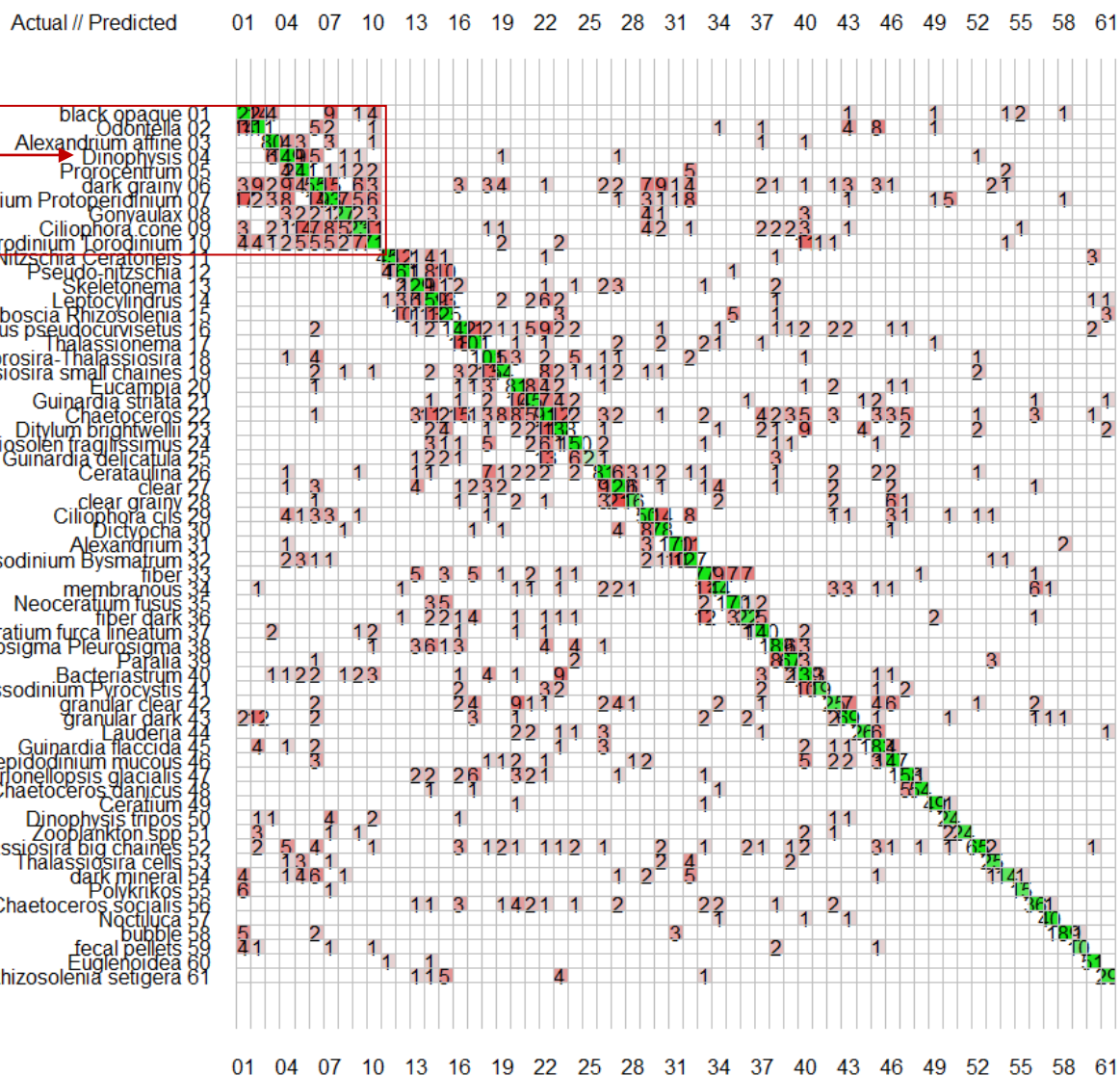


Figure 3 : Matrice de confusion numérique relative à l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique calculée par validation croisée avec Random Forest

La Figure 3 permet de mettre en évidence les sources de confusion entre les groupes en indiquant le nombre de vignettes correctement identifiées par l'algorithme (diagonale en vert) et les identifications erronées. Les chiffres (cases rouges) en dehors de la diagonale correspondent aux erreurs d'identification de l'outil de reconnaissance, avec en ligne les faux négatifs et en colonne les faux positifs.

Cette matrice de confusion remanie l'ordre des groupes afin de maximiser les confusions à proximité de la diagonale, et faciliter la visualisation des confusions induites par l'outil de reconnaissance.

En raison du grand nombre de groupes et de la longueur de certains intitulés, cette figure est difficilement lisible. Néanmoins, autour de *Dinophysis* on peut identifier les groupes qui génèrent des confusions.

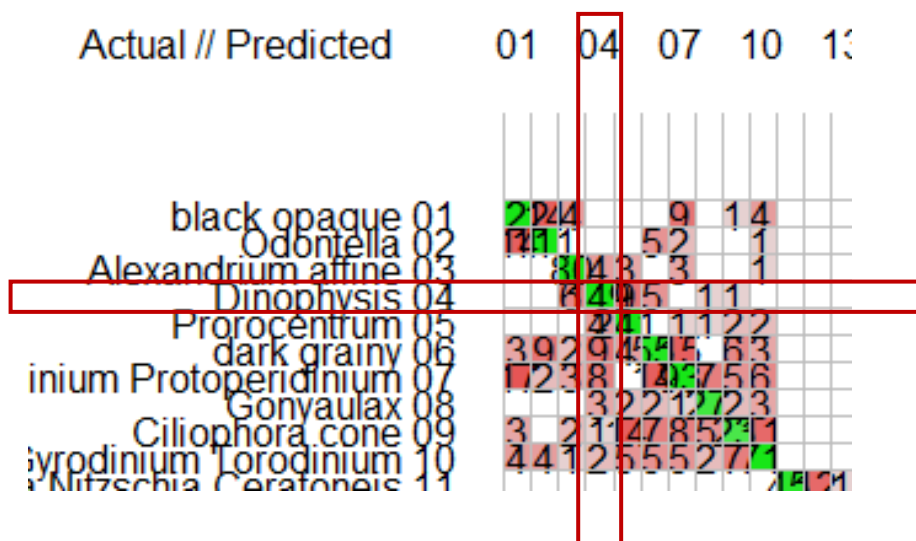


Figure 4 : matrice de confusion numérique Zoom de la Figure 3 autour du groupe *Dinophysis*.

La Figure 5 éclaire sur les groupes qui présentent les plus grands nombres de confusions avec le groupe des *Dinophysis*.

Aussi, sur la Figure 3, en colonne on identifie les faux positifs, c'est à dire les images des groupes incorrectement classées dans les *Dinophysis*, et, en ligne les faux négatifs, c'est à dire les images de *Dinophysis* incorrectement classées dans d'autres groupes, et qui sont listés dans le tableau suivant :

<i>Dinophysis</i>	Faux positifs	Faux négatifs
<i>Alexandrium affine</i> ,	4	6
<i>Prorocentrum</i> ,	4	9
dark grainy,	9	5
<i>Peridinium Protoperidinium</i> ,	8	
<i>Gonyaulax</i> ,	3	1
<i>Ciliophora cone</i>	1	1
<i>Gymnodinium Gyrodinium Torodinium</i>	2	
	1	
<i>Thalassiosira small chaines</i>		1
<i>Cerataulina</i>	1	
Clear	1	1
<i>Ciliophora cils</i>	1	
<i>Alexandrium</i>	4	
<i>Scripsiella Ensiculifera Pentapharsodinium Bysmatrum</i>	1	
<i>Bacteriastrum</i>	2	
<i>Guinardia flaccida</i>	1	
<i>Thalassiosira big chaines</i>	5	1
<i>Thalassiosira cells</i>	1	
Dark mineral	1	

Tableau 6 : Listes des groupes et nombre de vignettes concernées par des confusions avec le groupe *Dinophysis*.

Ces confusions résultent de similitudes de forme de particules constituant ces groupes. Afin d'optimiser la reconnaissance, un nouvel examen des vignettes de ces groupes devra être réalisé.

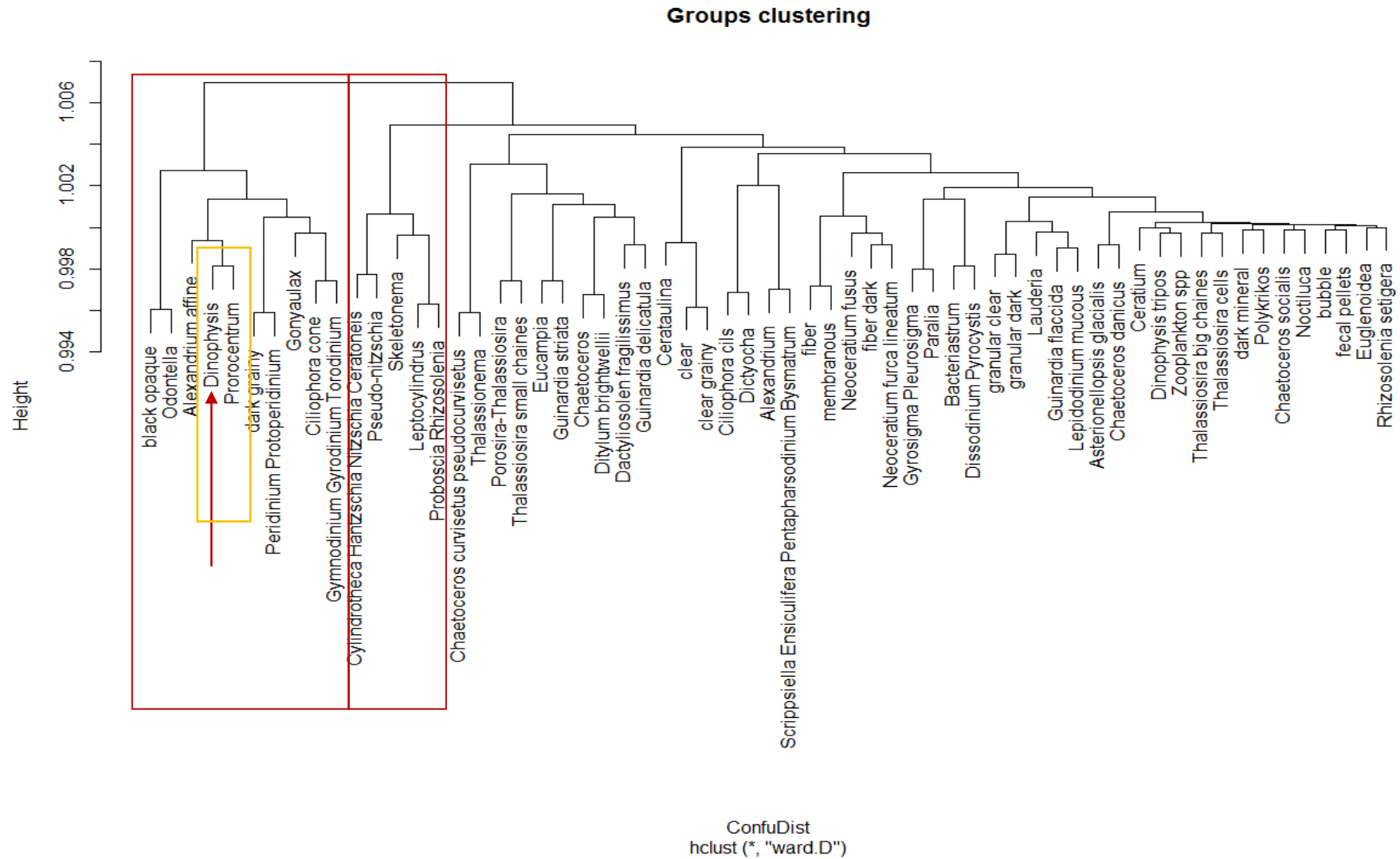


Figure 5 : Dendrogramme relatif à l'outil de reconnaissance "trainRephy_4Xlugol", calculé par validation croisée avec Random Forest.

Avec le groupe de débris "dark grainy", le groupe des *Prorocentrum* semble opérer la plus grande confusion avec les *Dinophysis*. Les images suivantes illustrent cette confusion :

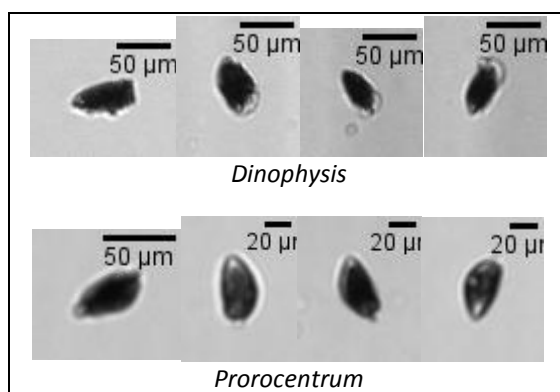


Image 6 : Echantillon d'images illustrant la confusion entre les groupes *Dinophysis* et *Prorocentrum*.

L'analyse des figures 3 et 5 permet de déterminer les ensembles de groupes générant le plus de confusion. Le deuxième ensemble de confusion concerne les groupes : *Cylindrotheca*, *Hantzschia*, *Ceratoneis*, *Pseudonitzschia*, *Skeletonema*, *Leptocylindrus* et *Proboscia*, *Rhizosolenia*. Ainsi de suite les ensembles de groupes de fortes confusions sont identifiés et, un nouvel examen des vignettes doit être réalisé en tenant compte de cette information. Cela consistera d'une part à éliminer les vignettes d'un groupe dont la morphologie évoque celle d'un autre, et, d'autre part, enrichir chacun groupe de faible effectif avec des vignettes typiques de ce groupe. Ce travail fait parti de la poursuite de l'optimisation du set.

Outre ces outils permettant de cibler les problèmes affectant les performances de l'outil de reconnaissance, les compétences et l'expérience des analystes du phytoplancton restent indispensables pour pointer d'autres sources d'affaiblissement. Ainsi, pour le taxon *Chaetoceros danicus*, le pourcentage d'erreur est estimé à 11,3 %, avec de faibles taux de faux positifs et de faux négatifs, et très peu de confusions possibles avec d'autres groupes. Mais ce groupe est constitué d'images essentiellement issues d'échantillons en provenance de la Manche et l'observateur averti constate que la morphologie des cellules de ce taxon est très différente lorsqu'elles proviennent de l'Atlantique. Les images suivantes illustrent cet exemple.

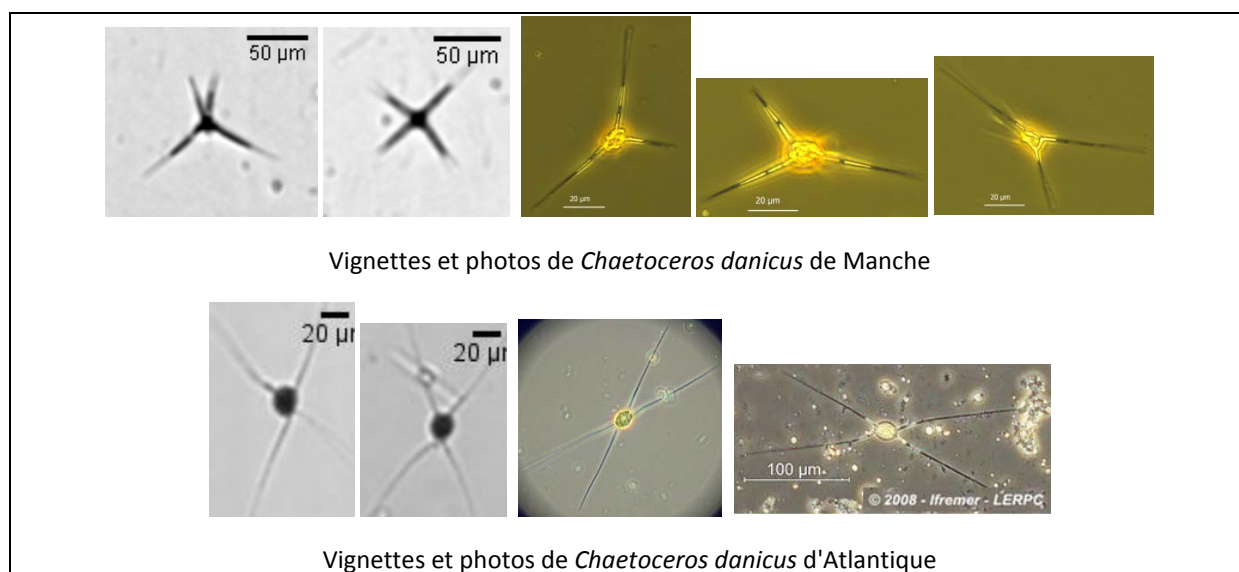


Image 7 : vignettes et photos illustrant les différences morphologiques entre les cellules de *Chaetoceros danicus* de Manche et d'Atlantique.

3.3 Conclusion et perspectives

La démarche de particularisation du set global Manche-Atlantique vers des sets spécifiques par façades maritimes distinctes préconisée dans le rapport sur l'évaluation d'un outil de reconnaissance commun Manche-Atlantique 4X (Neaud-Masson N, Lefebvre A., février 2015) été poursuivie en 2015.

Le **set d'apprentissage optimisé pour l'Atlantique** présenté et évalué dans le rapport de février 2015 comprenait 7 103 images réparties dans 54 groupes et son taux d'erreur globale était de 27,5 %. A l'issue du travail réalisé tout au long de l'année 2015, ce set **comprend maintenant 7 388 images réparties dans 62 groupes avec un taux d'erreur globale de 26,3 %**. Ce set s'est enrichi de nouveaux groupes et a été remanié tout en faisant baisser le taux d'erreur global de l'outil de reconnaissance correspondant, ce qui est très satisfaisant. Toutefois **des perfectionnements sont encore possibles**.

Grâce aux outils proposés par le logiciel Zoo/Phytolmage permettant d'évaluation des performances du classier, les actions à mener pour parfaire l'optimisation d'un set sont ciblées. Les groupes à revoir sont ceux dont le taux d'erreur individuel est élevé, et/ou qui présentent de forts taux de "recall" et, ceci, en s'appuyant sur la matrice de confusion afin d'identifier les groupes générant les confusions.

L'exploitation des résultats d'analyses de performances a été utilisée plus particulièrement pour contrôler le groupe des *Dinophysis*. Dans ce set optimisé pour l'Atlantique, le groupe des *Dinophysis* comporte 174 vignettes, présente un taux d'erreur de 15,5 % et obtient 80% de vrais positifs. Ce score de reconnaissance est très encourageant. Afin d'améliorer les discriminations, les groupes comportant des particules identifiées à tort comme étant des *Dinophysis* et les groupes vers lesquels des particules de *Dinophysis* ont été classés à tort, ont été visualisés sur la matrice de confusion. Ainsi, il a été déterminé que **l'une des plus grandes confusions qui s'opère avec le groupe des *Dinophysis*, concerne le groupe des *Proocentrum***.

Outre ces outils permettant de cibler les problèmes affectant les performances de l'outil de reconnaissance, les compétences et **l'expérience des analystes s'avèrent indispensable pour identifier d'autres sources potentielles de problèmes**. Par exemple, pour le cas où la morphologie des cellules d'un taxon s'avère être très différente lorsqu'elles proviennent de façades maritimes différentes.

Les éléments présentés dans ce rapport constituent une base pour poursuivre l'optimisation du set RePHY-4X adapté pour l'Atlantique ainsi que l'amélioration des performances de l'outil de reconnaissance correspondant. Cela consistera d'une part à éliminer les images d'un groupe pour lesquelles la morphologie de la particule évoque celle typique d'un autre groupe et, d'autre part, enrichir chaque groupe de faible effectif avec des vignettes typiques de ce groupe. **Grâce à ces éléments, ces actions seront aisément mises en œuvre et permettront de disposer rapidement d'un outil de reconnaissance opérationnel pour l'analyse en routine des échantillons collectés sur la façade Atlantique dans le cadre du REPHY.**

4 Comparaison des dénombrements de *Dinophysis* (Microscope vs FlowCAM-Zoo/PhytoImage)

Dans cette section, nous comparons les résultats des abondances de *Dinophysis* obtenues par lecture au microscope inversé selon la méthode appliquée dans le cadre du RePHY², et par l'outil couplé FlowCAM-Zoo/PhytoImage v5.4 (FC/ZI), avec numérisation selon la procédure appliquée à Nantes³ et utilisation de l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique présenté au point 3.

4.1 Matériel et méthode

Les tests ont été réalisés sur 4 échantillons sélectionnés en raison des abondances différentes en *Dinophysis* qu'ils présentent. *Dinophysis* est un taxon cible du REPHY et il est important que les performances de l'outil permette sa détection et son dénombrement avec le moins d'incertitude possible.

Les échantillons sélectionnés pour ce test sont les suivants :

Lieu de prélèvement	Date de prélèvement
010-P-002 - Antifer ponton pêche	10/08/2015
010-P-109 – Cabourg	13/08/2015
010-P-102 - Ouireham 1 mille	13/08/2015
010-P-187 – Luc 1 mille	13/08/2015

Tableau 7 : Echantillons sélectionnés pour le test.

Les échantillons ont été transmis par le laboratoire Ifremer de Port en Bessin par colis au laboratoire de Nantes la dernière semaine d'août 2015.

Les numérisations au FlowCAM de Nantes ont été réalisées les 2,3 et 11 septembre.

Dix numérisations de 10 ml à l'objectif 4X selon la procédure de référence⁴ ont été réalisées sur chaque échantillon avec le FlowCAM du laboratoire de Nantes soit 40 numérisations au total.

Seules les données obtenues avec l'échantillon du point "Antifer ponton pêche" ont été traitées avec le logiciel Zoo/PhytoImage version 5.4 et sont présentées ici. L'exploitation des données des autres échantillons n'est pas achevée et est programmée pour 2016.

Le schéma suivant décrit le déroulement de l'acquisition des données au laboratoire de Nantes.

² Neaud-Masson N (2015). **Observation et dénombrement du phytoplancton marin par microscopie optique photonique - Spécifications techniques et méthodologiques appliquées au REPHY. Document de méthode.**

NF EN 15204 (2006), **Qualité de l'eau – Norme guide pour le dénombrement du phytoplancton par microscopie inversée (méthode d'Utermöhl)**. Décembre 2006.

³ Neaud-Masson N (2014). **Projet FlowCAM - ZooPhytoImage. Procédure de numérisation d'échantillon d'eau lugolée au FlowCAM de Nantes.** Version révisée adaptée à l'usage du LER/MPL de Nantes – Version initiale Tunin-Ley A., Maurer D., (2011) adaptée par Wacquet G. (2014)

⁴ Procédure de numérisation d'échantillons d'eau lugolée au FlowCAM de Nantes – version de juin 2014.

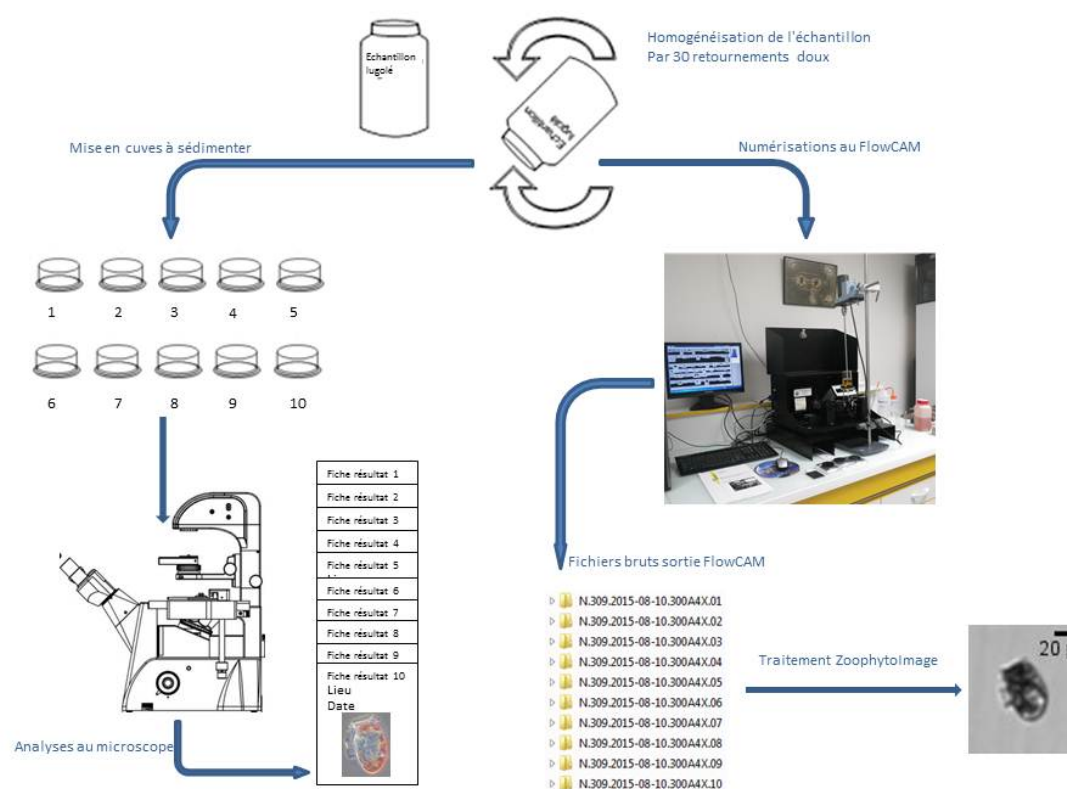


Figure 6 : Schéma d'expérience pour comparaison des dénombrements des *Dinophysis* par microscopie vs FlowCAM-Zoo/Phytolmage.

Dix répliques du même échantillon ont été réparties en cuves de sédimentation de 10 ml le 10 septembre 2015 et les dénombrements des *Dinophysis* ont été effectués au laboratoire de Nantes, au microscope inversé le 11 septembre. Les dénombrements au microscope ont été réalisés selon les prescriptions du document de méthode Neaud-Masson Nadine (2015) conforme à la norme EN 15204 (2006).

4.2 Résultats

Le Tableau 8 suivant présente quelques données (nombre total de particules dans 10 ml, dates et durées des numérisations) générées, lors des numérisations, par le logiciel d'acquisition VisualSpreadsheet conçu pour le FlowCAM (fichiers *run_summary.txt).

Code Fichier Echantillon	Nombre de particules acquises	Dates-Heures et durées des numérisations
N.309.2015-08-10.300A4X.01	Particle Count: 7 282	Start: 2015-09-02 08:05:00 End: 2015-09-02 08:18:33 Sampling Time: 00:12:57
N.309.2015-08-10.300A4X.02	Particle Count: 6 666	Start: 2015-09-02 08:22:39 End: 2015-09-02 08:36:57 Sampling Time: 00:13:43
N.309.2015-08-10.300A4X.03	Particle Count: 6 589	Start: 2015-09-02 08:41:33 End: 2015-09-02 08:54:52 Sampling Time: 00:12:44
N.309.2015-08-10.300A4X.04	Particle Count: 7 838	Start: 2015-09-02 08:57:09 End: 2015-09-02 09:10:49 Sampling Time: 00:13:02
N.309.2015-08-10.300A4X.05	Particle Count: 7 087	Start: 2015-09-02 09:13:42 End: 2015-09-02 09:27:07 Sampling Time: 00:12:50
N.309.2015-08-10.300A4X.06	Particle Count: 6 412	Start: 2015-09-02 09:30:23 End: 2015-09-02 09:43:40 Sampling Time: 00:12:41
N.309.2015-08-10.300A4X.07	Particle Count: 7 703	Start: 2015-09-02 09:46:08 End: 2015-09-02 09:59:37 Sampling Time: 00:12:55
N.309.2015-08-10.300A4X.08	Particle Count: 7 266	Start: 2015-09-02 10:01:04 End: 2015-09-02 10:14:45 Sampling Time: 00:13:05
N.309.2015-08-10.300A4X.09	Particle Count: 6 517	Start: 2015-09-02 10:19:26 End: 2015-09-02 10:33:15 Sampling Time: 00:13:11
N.309.2015-08-10.300A4X.10	Particle Count: 8 367	Start: 2015-09-02 10:53:00 End: 2015-09-02 11:07:33 Sampling Time: 00:13:57

Tableau 8 : Dates, heures et durées des numérisations des dix répliques de l'échantillon d'eau du point "Antifer ponton pêche" prélevé le 10/08/2015.

Le contexte utilisé dans cette procédure de numérisation, permet une durée d'acquisition d'environ 13 minutes pour un volume imagé de 10 mL d'échantillon. Les dix acquisitions ont été réalisées dans la foulée, ce qui a pris 3 heures au total. Ces performances sont tout à fait satisfaisantes dans la perspective de l'utilisation en routine de cet outil pour le REPHY.

Pour l'échantillon concerné, c'est à dire prélevé au point "Antifer ponton pêche" le 11 août 2015, le nombre total de particules dénombrées dans 10 ml .est d'environ 7 000 en moyenne.

Concernant les résultats obtenus par dénombrement au microscope, si l'on considère que les populations comptées respectent une distribution de Poisson dans la chambre de sédimentation, on peut évaluer la précision des comptages. L'intervalle de confiance (pour un niveau de signification de 95 %) est calculé grâce au pourcentage de confiance suivant :

$$\% \text{ confiance} = \pm 200\% / \sqrt{n}$$

où **n** est le nombre de cellules comptées.

Les résultats ainsi obtenus sont les suivants :

Numéro réplique	Nb de cellules comptées (1/2 cuve 10 mL)	% confiance	Min (cel./L)	microscope Nb <i>Dinophysis</i> /L	Max (cell./L)
1	141	16.84	23 450	28 200	32 950
2	119	18.33	19 437	23 800	28 163
3	95	20.52	15 101	19 000	22 899
4	118	18.41	19 255	23 600	27 945
5	109	19.16	17 624	21 800	25 976
6	151	16.28	25 285	30 200	35 115
7	132	17.41	21 804	26 400	30 996
8	122	18.11	19 982	24 400	28 818
9	150	16.33	25 101	30 000	34 899
10	127	17.75	20 892	25 400	29 908

Tableau 9 : Résultats des analyses au microscope des dix répliques de l'échantillon d'eau du point "Antifer ponton pêche" prélevé le 10/08/2015

Code Fichier Echantillon-réplique	FlowCAM-ZI (EC) Nb <i>Dinophysis</i> /L	FlowCAM-ZI (SD) Nb <i>Dinophysis</i> /L
N.309.2015-08-10.300A4X.01	14 900	22 400
N.309.2015-08-10.300A4X.02	14 100	17 300
N.309.2015-08-10.300A4X.03	12 700	19 500
N.309.2015-08-10.300A4X.04	18 600	24 900
N.309.2015-08-10.300A4X.05	16 800	25 200
N.309.2015-08-10.300A4X.06	14 800	20 400
N.309.2015-08-10.300A4X.07	18 800	21 600
N.309.2015-08-10.300A4X.08	15 900	19 800
N.309.2015-08-10.300A4X.09	15 000	22 700
N.309.2015-08-10.300A4X.10	20 400	21 900

Tableau 10 : Résultats des abondances de *Dinophysis* obtenus avec l'outil couplé FlowCAM-Zoo/Phytolmage avec deux stratégies de correction d'erreur.

Parmi les évolutions du logiciel, lors des étapes de correction d'erreur de l'analyse d'un échantillon, la version 5.4 apporte un nouvel indicateur permettant de choisir la fin des corrections manuelles d'erreur lorsque le seuil de dissimilarité différentielle atteint 5 %, c'est à dire que les abondances obtenues à l'étape courante sont fortement similaires à celles obtenues à la précédente étape (ligne EC sur la Figure 7) ou bien de poursuivre le processus jusqu'à correction/validation manuelle de toutes les particules suspectes jusqu'à environ 5 % (ligne SD sur la Figure 8). Le Tableau 10 indique les résultats obtenus avec les deux méthodes pour chaque réplique.

N.309.2015-08-10.300A4X.10_4/Step3

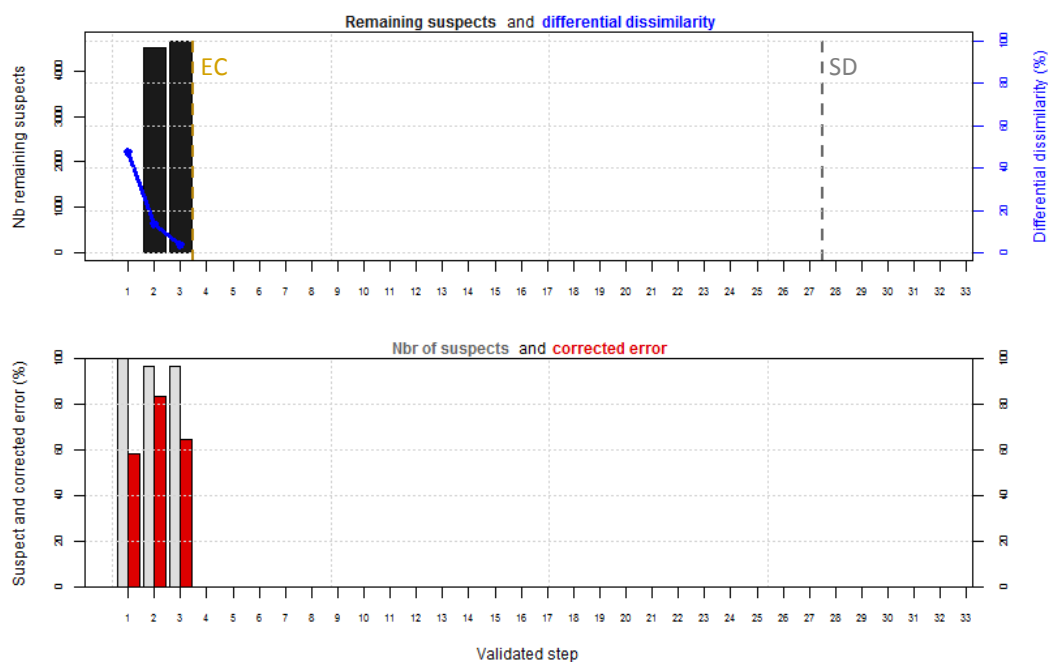


Figure 7 : Exemple de rapport du processus de validation après trois étapes de corrections d'erreurs et arrêt possible pour atteinte de l'objectif après (C)orrection d'(E)rreur (indicateur EC).

N.309.2015-08-10.300A4X.05_12/Step20

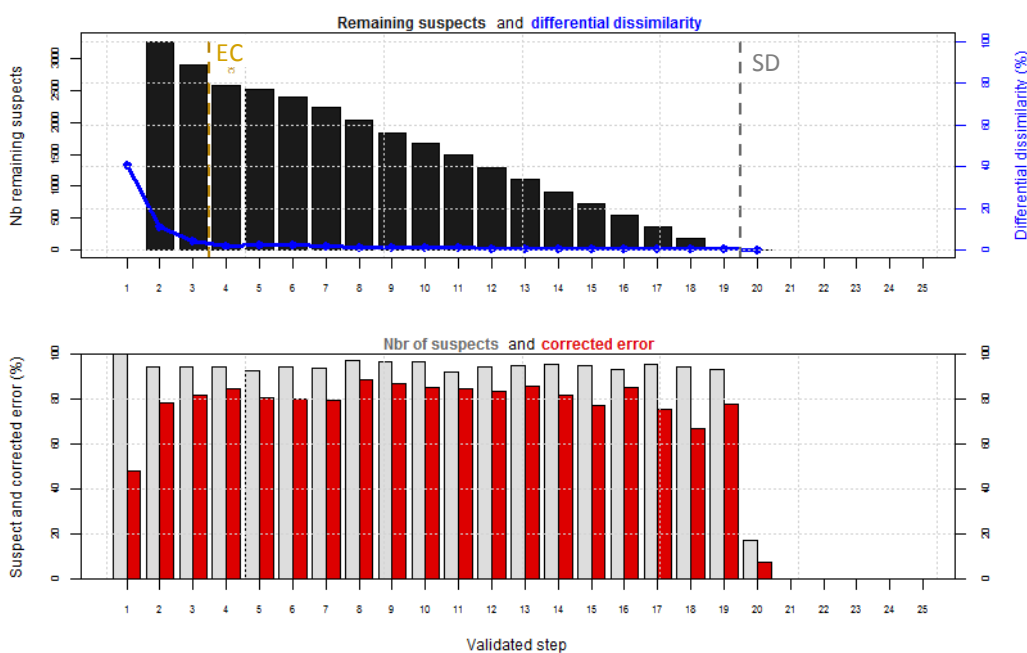


Figure 8 : Exemple de rapport du processus de validation après vingt étapes de corrections d'erreurs et atteinte de l'objectif par (D)étection (et correction) des (S)uspects (indicateur SD).

Tous les résultats sont représentés dans le graphe de la Figure 9 suivante.

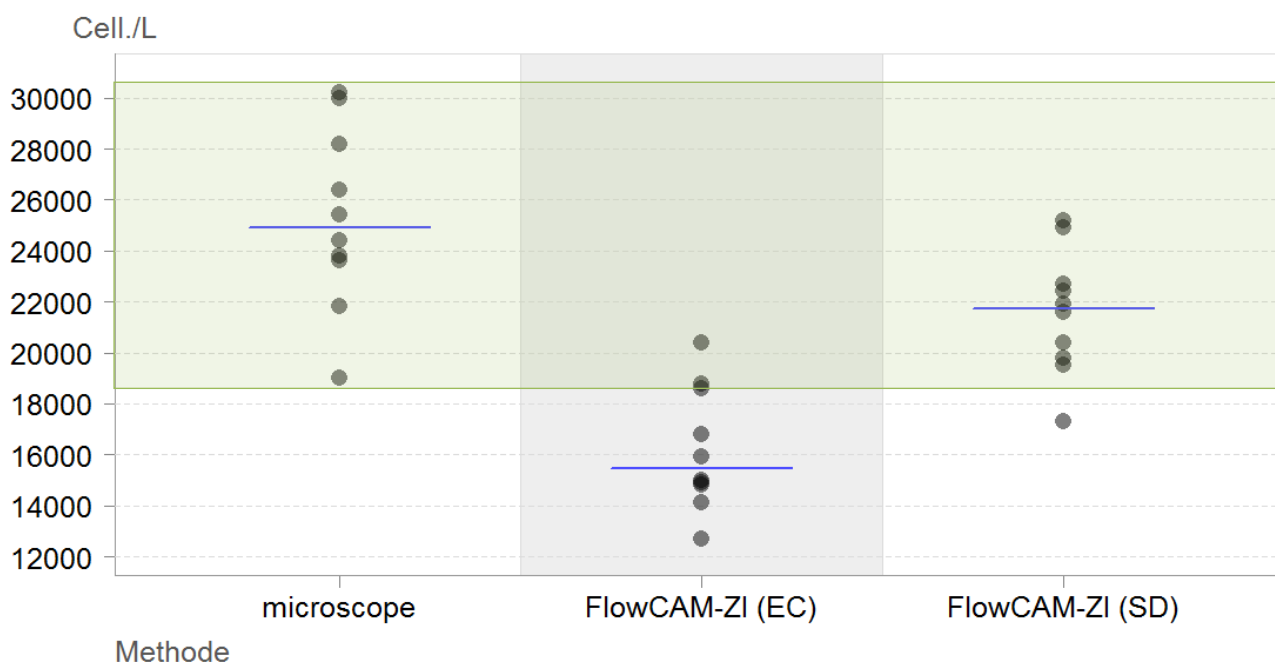


Figure 9 : Représentation graphique des abondances de *Dinophysis* obtenues par microscopie et avec l'outil FC/ZI. Les points gris représentent les valeurs obtenues pour les 10 répliques et les lignes bleues représentent la médiane pour chaque méthode.

Trois tests de comparaison par paire ont été appliqués aux données, afin de tester la différence entre les moyennes des résultats des trois méthodes prises deux à deux. Le nombre de répliques étant petit, le test non paramétrique de Wilcoxon a été utilisé, pour données appariées en ce qui concerne la comparaison FC/ZI (SD) vs FC/ZI (EC) et pour données non appariées en ce qui concerne les comparaisons FC/ZI (SD ou EC) vs microscopie. Pour chaque test, le seuil de significativité (*i.e.* risque de première espèce) a été corrigé au moyen de la méthode de Bonferroni, pour prendre en compte l'augmentation du risque de première espèce global lors de tests multiples. Ce seuil est alors de 0.017. Pour chaque paire de méthodes, l'hypothèse H_0 est la suivante : "Il n'y a pas de différence de dénombrement de *Dinophysis* entre les deux méthodes".

L'hypothèse H_0 est rejetée pour les paires de méthodes

"Microscopie vs FC/ZI (EC)" (p-value = 0.00002) et "FC/ZI (SD) vs FC/ZI (EC)" (p-value = 0.002).

L'hypothèse H_0 ne peut pas être rejetée pour la paire de méthodes

"Microscopie vs FC/ZI (SD)" (p-value = 0.023 > 0.017).

Les abondances obtenues avec l'outil FC/ZI sont très sous estimées lorsque l'on choisi d'arrêter le processus de correction d'erreur après l'atteinte de l'indicateur "EC". Lorsque l'on poursuit les corrections d'erreur jusqu'à atteindre l'indicateur "SD", la sous-estimation subsiste mais les résultats obtenus ne permettent pas de mettre en évidence une différence significative avec la méthode d'analyse au microscope. Néanmoins, la valeur de la p-value, inférieure au seuil habituel de 0.05 mais supérieure au seuil corrigé de 0.017, est à relativiser vis-à-vis du petit nombre de répliques et de la puissance plus faible des tests non paramétriques. La confirmation ou l'infirmité du résultat de comparaison des méthodes FC/ZI (SD) et microscopie nécessiterait un nombre de répliques plus conséquent. On remarque par ailleurs, que la variabilité des résultats semble plus faible avec le FC/ZI qu'en microscopie.

4.3 Contrôle du classement des vignettes

L'échantillon d'eau renfermait deux espèces de *Dinophysis* de taille et morphologies différentes dont les photos et vignettes sont présentées Image 8.

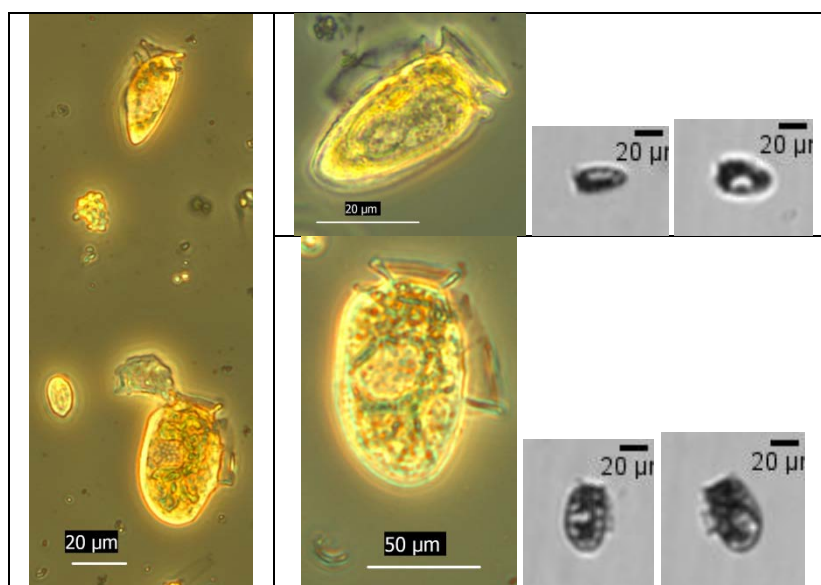


Image 8 : Photos prises au microscope et vignettes correspondantes issues du traitement FlowCAM-Zoo/PhytoImage des cellules de *Dinophysis* présentes dans l'échantillon provenant du point "Antifer ponton pêche" du 10/08/2015.

Afin d'évaluer le nombre de vignettes placées à tort dans d'autres groupes, un examen visuel des images classées à l'issue de l'analyse a été réalisé sur le réplica numéro 01.

4.3.1 CONTROLE DES FAUX POSITIFS

Pour le réplica **N.309.2015-08-10.300A4X.01**, 224 vignettes ont été classées dans le groupe des *Dinophysis*. Sur ces 224 vignettes, 20 ont été classées à tort dans ce groupe. Ce sont les vignettes suivantes.

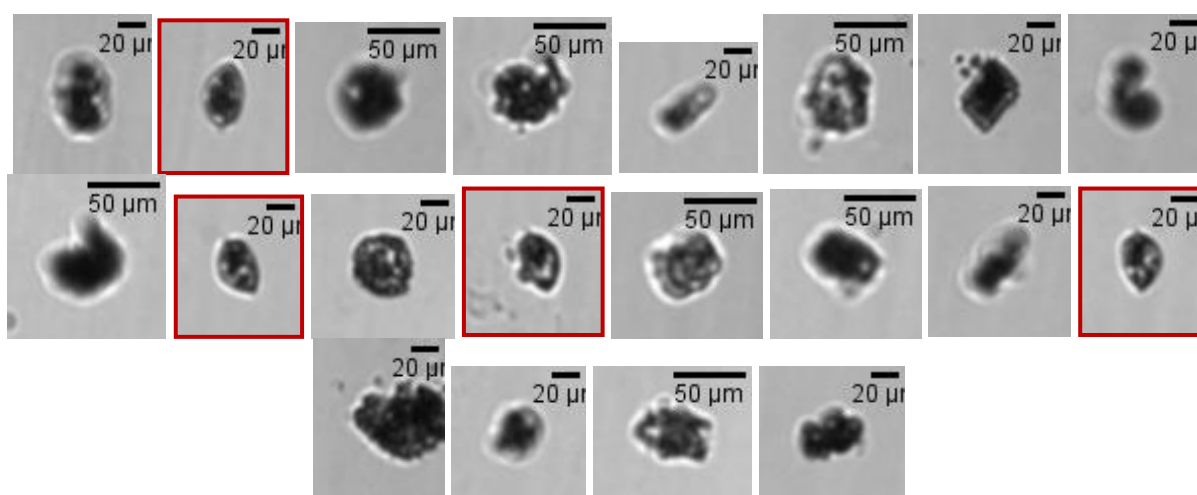


Image 9 : Faux positifs – Images du réplica N.309.2015-08-10.300A4X.01 classées à tort dans le groupe "*Dinophysis*"

Quatre des 20 images faux positifs sont des *Prorocentrum*. Ceci confirme le fort taux de confusion entre les groupes *Dinophysis* et *Prorocentrum* dans l'outil de reconnaissance utilisé ici et renforce la nécessité de revoir les images du set d'apprentissage pour ces groupes afin de réduire cette confusion.

4.3.2 CONTROLE DES FAUX NEGATIFS

Tous les groupes de débris ont été contrôlés. Ils contiennent au total 1 311 vignettes dont les 45 suivantes sont des images de *Dinophysis* classées à tort dans un de ces groupes.

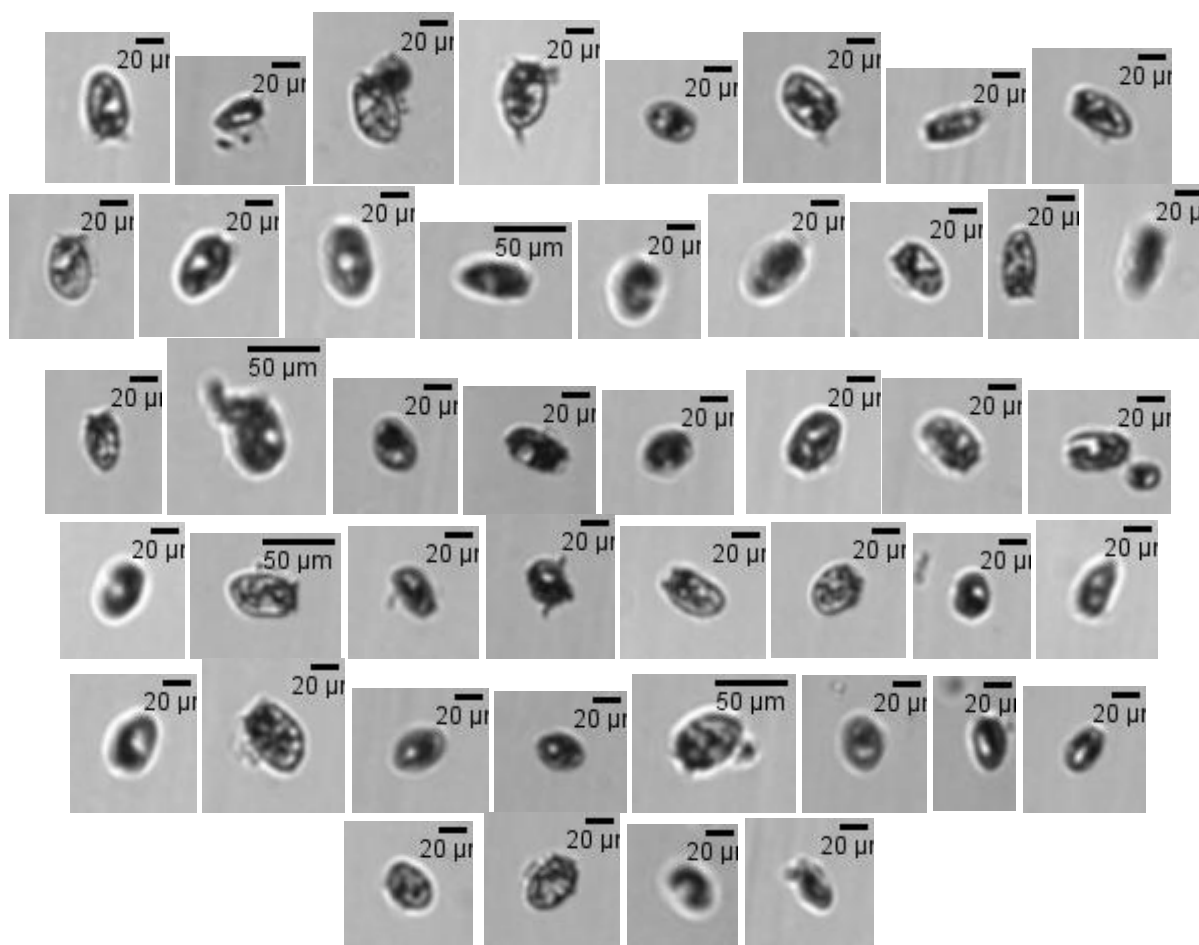


Image 10 : Faux négatifs – Images de du réplica N.309.2015-08-10.300A4X.01 classées à tort dans un groupe de débris.

Pour le réplica **N.309.2015-08-10.300A4X.01** après analyse avec l'outil FC/ZI avec correction d'erreur jusqu'à atteindre l'indicateur "SD", l'abondance de *Dinophysis* obtenue est de 22 400 cellules/L. Si on retranche les faux positifs et ajoute les vignettes faussement classées dans les groupes des débris, on obtient une abondance de 24 900 cellules/L qui est égale à la médiane des abondances obtenues par microscopie.

Puisque le nombre réel de vignettes de *Dinophysis* obtenues par la numérisation au FlowCAM est égal au nombre attendu compte tenu des résultats de dénombrements au microscope, la sous-estimation résultant de l'analyse avec Zoo/PhytoImage ne semble pas imputable à la technique de numérisation de l'échantillon au FlowCAM mais serait plutôt due au taux d'erreur de l'outil de reconnaissance utilisé. Ceci est à vérifier avec les résultats des réplicas de tous les échantillons afin de valider la méthode de numérisation pour les *Dinophysis*.

A noter aussi que de nombreuses vignettes représentant des cellules phytoplanctoniques, et particulièrement des *Prorocentrum*, sont observées dans les catégories de débris pour cet échantillon. Les catégories des débris du set d'apprentissage doivent donc faire l'objet d'un nouvel examen détaillé afin de détecter les sources de confusion et, le cas échéant, envisager un remaniement de celles-ci.

4.4 Discussion et suites à donner

Afin d'évaluer les performances de l'outil couplé FlowCAM-Zoo/phytolmage v5.4 + outil de reconnaissance issu du set Manche Atlantique optimisé pour l'Atlantique, les résultats d'abondance de *Dinophysis* obtenus avec cet outil ont été comparés à ceux obtenus sur le même échantillon d'eau par analyse au microscope selon la méthode appliquée au RePHY. Faute de temps, les résultats présentés dans ce rapport ne concernent qu'un échantillon sur les quatre qui avaient été sélectionnés pour mener ce test. Les numérisations au FlowCAM obtenues avec les trois autres échantillons sont disponibles mais n'ont pas encore été analysés avec le logiciel Zoo/Phytolmage.

Dinophysis a été choisi pour mener ces tests car il est un taxon cible du REPHY, il est donc important que les performances de l'outil permette sa détection lorsque sa concentration est faible et son dénombrement avec le moins d'incertitude possible.

Pour réaliser ce test, 10 numérisations de 10 ml à l'objectif 4X selon la procédure de référence⁵ ont été réalisées avec le FlowCAM du laboratoire de Nantes puis traitées avec le logiciel Zoo/Phytolmage version 5.4. Parallèlement, 10 répliques du même échantillon ont été réparties en cuves de sédimentation de 10 ml sur lesquelles les dénombrements des *Dinophysis* ont été effectués au microscope inversé selon la méthode Utermöhl décrite dans la norme AFNOR⁶.

Parmi les évolutions du logiciel, lors des étapes de correction d'erreur de l'analyse d'un échantillon, la version 5.4 apporte un nouvel indicateur permettant de choisir la fin des corrections manuelles d'erreur lorsque le seuil de dissimilarité différentielle atteint 5 %, c'est à dire que les abondances obtenues à l'étape courante sont fortement similaires à celles obtenues à la précédente étape (indicateur "EC"), ou bien de poursuivre le processus jusqu'à correction/validation manuelle de toutes les particules suspectes jusqu'à environ 5 % (indicateur "SD").

Les résultats de l'échantillon testé montrent que les abondances obtenues avec l'outil FC/ZI sont très sous estimées lorsque l'on choisi d'arrêter le processus de correction d'erreur après l'atteinte de l'indicateur "EC". Lorsque l'on poursuit les corrections d'erreur jusqu'à atteindre l'indicateur "SD", la sous-estimation subsiste mais les résultats obtenus ne permettent pas de mettre en évidence une différence significative avec la méthode d'analyse au microscope. On remarque par ailleurs que la variabilité des résultats semble plus faible avec le FC/ZI qu'en microscopie.

Un contrôle du classement final des vignettes d'un des répliques a permis de comptabiliser les faux positifs et les faux négatifs. Ainsi, pour le réplique contrôlé, lorsqu'on retranche les faux positifs, c'est à dire les vignettes qui ne sont pas de *Dinophysis* mais qui sont classées par erreur dans ce groupe, et qu'on ajoute les vignettes de *Dinophysis* faussement classées dans les groupes des débris, on obtient une abondance de *Dinophysis* par litre qui est égale à la médiane des abondances obtenues par microscopie.

Ce résultat permet de penser que **la sous estimation ne semble pas imputable à la technique de numérisation de l'échantillon au FlowCAM, mais qu'elle est plutôt due au taux d'erreur de l'outil de reconnaissance** utilisé ici. La même analyse des résultats des répliques de tous les échantillons doit être réalisée afin de confirmer ou d'infirmer cette hypothèse

Ce contrôle du classement des vignettes de *Dinophysis* a permis aussi de confirmer le fort taux de confusion avec le groupe des *Prorocentrum* détecté lors de l'évaluation des performances de l'outil de reconnaissance optimisé pour l'Atlantique.

Enfin, un examen des vignettes classées dans les catégories de débris a mis en évidence un bon nombre de faux négatifs qui concerne d'autres groupes. Ceci induit la nécessité de **revoir les vignettes des groupes de débris et de cibler les confusions résultant des similitudes de formes avec les particules d'intérêt.**

⁵ Procédure de numérisation d'échantillons d'eau lugolée au FlowCAM de Nantes – version de juin 2014

⁶ NF EN 15204 (2006), Qualité de l'eau – Norme guide pour le dénombrement du phytoplancton par microscopie inversée (méthode d'Utermöhl). Décembre 2006.

ANNEXE I : Bibliographie

Tunin-Ley A., Maurer D., (2011). **Mise en œuvre opérationnelle d'un système couplé de numérisation (FlowCAM) et de traitement d'images (ZooPhytoImage), pour l'analyse automatisée, ou semi-automatisée, de la composition phytoplanctonique d'échantillons d'eau de mer.** *Rapport RST/LER/AR/11/002*

Grosjean P., Wacquet G. (2014). ZooPhytoImage Version 5.4-0 – **Analyse d'images de plancton assistée par ordinateur - Manuel Utilisateur.** UMONS. *ZooPhytoImage, logiciel gratuit (Open Source) pour l'analyse d'images numériques de plancton*

Wacquet G., Grosjean P., (2014). **Apprentissage actif, training set adaptatif.** UMONS Juin-Août 2014.

Wacquet G., Lefebvre A., (2014). **Protocole sur les outils de reconnaissance optimisés du phytoplancton dans le cadre du REPHY (zone Manche-Atlantique). Version 1.3 Mars 2013-Février 2014** *RST/ODE/LERBL.*

Neaud-Masson N (2014). **Projet FlowCAM – ZooPhytoImage. Procédure de numérisation d'échantillon d'eau lugolée au FlowCAM de Nantes.** Version révisée adaptée à l'usage du LER/MPL de Nantes – Version initiale Tunin-Ley A., Maurer D., (2011) adaptée par Wacquet G. (2014)

Neaud-Masson N, Lefebvre A., (2015). **Rapport sur l'évaluation d'un outil de reconnaissance commun Manche-Atlantique 4X** - fusion des sets d'apprentissage construits sur trois sites (Arcachon, Boulogne sur Mer et Nantes)

Neaud-Masson N (2015). **Catalogue illustré des images d'un set d'apprentissage réalisé à l'aide de l'outil FlowCAM – ZooPhytoImage utilisé dans le cadre du REPHY - Images issues de numérisations au grossissement 4X, d'échantillons d'eau, pour le set d'apprentissage global Manche-Atlantique.** *R.INT.ODE/DYNECO/VIGIES/15-14*

NF EN 15204 (2006), **Qualité de l'eau – Norme guide pour le dénombrement du phytoplancton par microscopie inversée (méthode d'Utermöhl).** Décembre 2006.

Neaud-Masson N (2015). **Observation et dénombrement du phytoplancton marin par microscopie optique photonique - Spécifications techniques et méthodologiques appliquées au REPHY.** Document de méthode. <http://archimer.ifremer.fr/doc/00292/40293/>

ANNEXE II : Liste des tableaux, figures, images et cartes

LISTE DES TABLEAUX

Tableau 1 : Liste des points retenus pour une numérisation systématique des échantillons au laboratoire de Nantes.....	5
Tableau 2 : Liste des points retenus pour une numérisation des échantillons au laboratoire de Nantes en vue d'une comparaison des dénombrements de <i>Dinophysis</i>	6
Tableau 3 : Liste des points retenus pour une numérisation systématique des échantillons au laboratoire d'Arcachon.....	10
Tableau 4 : Distribution des nombres de particules par groupe du set "trainRephy_4Xlugol" optimisé pour l'Atlantique.....	15
Tableau 5 : Estimation des erreurs par classe du set "trainRephy_4Xlugol".....	16
Tableau 6 : Listes des groupes et nombre de vignettes concernées par des confusions avec le groupe <i>Dinophysis</i>	20
Tableau 7 : Echantillons sélectionnés pour le test.....	24
Tableau 8 : Dates, heures et durées des numérisations des dix répliques de l'échantillon d'eau du point "Antifer ponton pêche" prélevé le 10/08/2015.....	26
Tableau 9 : Résultats des analyses au microscope des dix répliques de l'échantillon d'eau du point "Antifer ponton pêche" prélevé le 10/08/2015.....	27
Tableau 10 : Résultats des abondances de <i>Dinophysis</i> obtenus avec l'outil couplé FlowCAM-Zoo/PhytoImage avec deux stratégies de correction d'erreur.....	27

LISTE DES FIGURES

Figure 1 : Représentation graphique 1 des taux de "Précision" et de "Recall" calculés par groupe pour l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique et généré par l'algorithme Random Forest..	17
Figure 2 : Représentation graphique 2 des taux de "Précision" et de "Recall" calculés par groupe pour l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique et généré par l'algorithme Random Forest. .	18
Figure 3 : Matrice de confusion numérique relative à l'outil de reconnaissance "trainRephy_4Xlugol" optimisé pour l'Atlantique calculée par validation croisée avec Random Forest.....	19
Figure 4 : matrice de confusion numérique Zoom de la Figure 3 autour du groupe <i>Dinophysis</i>	20
Figure 5 : Dendrogramme relatif à l'outil de reconnaissance "trainRephy_4Xlugol", calculé par validation croisée avec Random Forest.....	21
Figure 6 : Schéma d'expérience pour comparaison des dénombrements des <i>Dinophysis</i> par microscopie vs FlowCAM-Zoo/PhytoImage.....	25
Figure 7 : Exemple de rapport du processus de validation après trois étapes de corrections d'erreurs et arrêt possible pour atteinte de l'objectif après (C)orrection d'(E)rreur (indicateur EC).....	28
Figure 8 : Exemple de rapport du processus de validation après vingt étapes de corrections d'erreurs et atteinte de l'objectif par (D)étection (et correction) des (S)uspects (indicateur SD).	28
Figure 9 : Représentation graphique des abondances de <i>Dinophysis</i> obtenues par microscopie et avec l'outil FC/ZI. Les points gris représentent les valeurs obtenues pour les 10 répliques et les lignes bleues représentent la médiane pour chaque méthode.	29

LISTE DES IMAGES

Image 1 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire de Nantes	7
Image 2 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire de Boulogne sur Mer.....	9
Image 3 : Capture d'écran de l'arborescence des numérisations faite en 2015 au laboratoire d'Arcachon	11
Image 4 : Groupes illustrés des particules "débris" du set "trainRephy_4Xlugol" optimisé pour l'Atlantique.....	14
Image 5 : Groupes illustrés des taxons intégrés en 2015 au set "trainRephy_4Xlugol" optimisé pour l'Atlantique (images du FlowCAM et photos prises au microscope)	14
Image 6 : Echantillon d'images illustrant la confusion entre les groupes <i>Dinophysis</i> et <i>Prorocentrum</i>	22
Image 7 : vignettes et photos illustrant les différences morphologiques entre les cellules de <i>Chaetoceros danicus</i> de Manche et d'Atlantique	22
Image 8 : Photos prises au microscope et vignettes correspondantes du FlowCAM des cellules de <i>Dinophysis</i> présentes dans l'échantillon provenant du point "Antifer ponton pêche" du 10/08/2015.....	30
Image 9 : Faux positifs – Images du réplica N.309.2015-08-10.300A4X.01 classées à tort dans le groupe " <i>Dinophysis</i> "	30
Image 10 : Faux négatifs – Images de du réplica N.309.2015-08-10.300A4X.01 classées à tort dans un groupe de débris.	31

LISTE DES CARTES

Carte 1 : Carte des points retenus pour une numérisation systématique des échantillons au laboratoire de Nantes.....	5
Carte 2 : Carte des points retenus pour une numérisation des échantillons au laboratoire de Nantes en vue d'une comparaison des dénombrements de <i>Dinophysis</i>	6
Carte 3 : Cartes des points retenus pour une numérisation des échantillons au laboratoire de Boulogne sur Mer	8
Carte 4 : Carte des points retenus pour une numérisation des échantillons au laboratoire d'Arcachon en vue d'une comparaison des dénombrements de <i>Dinophysis</i>	10

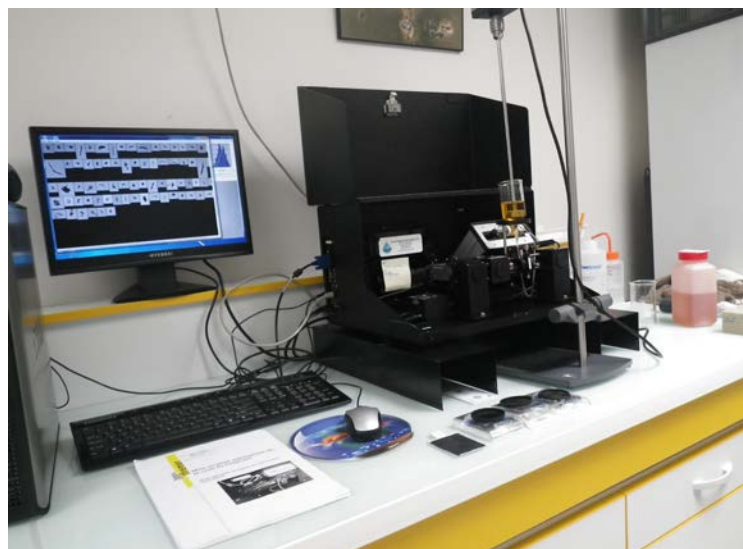
Contributeurs :

Camille Blondel (LER-BL)
Alain Lefebvre (LER-BL)
Pascale Hébert (LER-BL)
Danièle Maurer (LER-AR)
Claire Meteigner (LER-AR)
Myriam Rumèbe-Perrière (LER-AR)
Guillaume Wacquet (LER-BL)

Novembre 2015 - R.INT.ODE/DYNECO/VIGIES/15-14

Catalogue illustré des images d'un set d'apprentissage réalisé à l'aide de l'outil FlowCAM – ZooPhytoImage utilisé dans le cadre du REPHY

Images issues de numérisations au grossissement 4X, d'échantillons d'eau, pour le set d'apprentissage global Manche-Atlantique



sommaire

Catalogue illustré des images
d'un set d'apprentissage réalisé à l'aide
de l'outil FlowCAM – ZooPhytoImage
utilisé dans le cadre du REPHY

Images issues de numérisations
au grossissement 4X, d'échantillons d'eau,
pour le set d'apprentissage global
Manche-Atlantique

Sommaire

1. Introduction	6
2. Description des fiches	8
3. Classes non intégrées au set	9
3.1. <i>Asteroplanus karianus</i>	9
3.2. <i>Helicotheca</i>	9
3.3. <i>Hemiaulus</i>	9
3.4. <i>Proboscia indica</i>	9
3.5. <i>Dinophysis acuta</i>	10
3.6. <i>Diplopsalis Diplopelta Oblea</i>	10
3.7. <i>Heterocapsa</i>	10
3.8. <i>Pyramimonas</i>	10
3.9. <i>Pisces larva</i>	11
3.10. tintinnids.....	12
3.11. polen	12
4. Classes des particules intégrées au set (débris)	13
4.1. Black opaque	13
4.2. Bubble.....	13
4.3. Clear	14
4.4. Clear grainy	14
4.5. Dark grainy	15
4.6. Dark mineral	15
4.7. Fecal pellets.....	16
4.8. Fiber.....	16
4.9. Fiber dark.....	17
4.10. Granular clear.....	18
4.11. Granular dark.....	19
4.12. Membranous.....	20
5. Classes phytoplancton intégrées au set	21
5.1. Euglenoidea.....	21
5.2. <i>Phaeocystis</i>	22
5.3. <i>Ceratium</i>	23
5.4. <i>Neoceratium furca lineatum</i>	24
5.5. <i>Neoceratium fusus</i>	25
5.6. <i>Dinophysis</i>	26
5.7. <i>Dinophysis tripos</i>	27
5.8. <i>Alexandrium</i>	28
5.9. <i>Alexandrium affine</i>	28
5.10. <i>Gonyaulax</i>	29
5.11. <i>Gymnodinium Gyrodinium Torodinium</i>	30
5.12. <i>Lepidodinium mucous</i>	31
5.13. <i>Noctiluca</i>	32
5.14. <i>Peridinium Protoperidinium</i>	33
5.15. <i>Scrippsiella Ensiculifera Pentapharsodinium Bysmatrum</i>	34
5.16. <i>Polykrikos</i>	35
5.17. <i>Prorocentrum</i>	36
5.18. <i>Dissodinium Pyrocystis</i>	37
5.19. <i>Cylindrotheca Hantzschia Nitzschia</i>	38
5.20. <i>Pseudo-nitzschia</i>	39
5.21. <i>Bacteriastrum</i>	40
5.22. <i>Chaetoceros</i>	41
5.23. <i>Chaetoceros curvisetus</i>	42

sommaire

5.24. Chaetoceros danicus.....	43
5.25. Chaetoceros socialis.....	44
5.26. Asterionellopsis glacialis.....	45
5.27. Cerataulina	46
5.28. Leptocylindrus.....	47
5.29. Ditylum brightwellii.....	48
5.30. Eucampia.....	49
5.31. Lauderia.....	50
5.32. Paralia.....	51
5.33. Gyrosigma Pleurosigma	52
5.34. Dactyliosolen fragilissimus	53
5.35. Guinardia delicatula.....	54
5.36. Guinardia flaccida.....	55
5.37. Guinardia striata	56
5.38. Proboscia Rhizosolenia	57
5.39. Rhizosolenia setigera	58
5.40. Skeletonema.....	59
5.41. Thalassionema	60
5.42. Thalassiosira big chaines	61
5.43. Thalassiosira cells	62
5.44. Thalassiosira small chaines.....	63
5.45. Odontella	64
5.46. Dictyocha	65
6. Classes du zooplancton intégrées au set.....	66
6.1. Ciliophora big.....	66
6.2. Ciliophora cone.....	67
6.3. Ciliophora small	68
6.4. Zooplankton spp	69



1. Introduction

Dans la démarche de constituer un set d'apprentissage pour l'outil de reconnaissance généré grâce au logiciel ZooPhytoImage, trois laboratoires de l'Ifremer équipés d'un FlowCam dédié au REPHY ont participé à l'acquisition d'images* : Boulogne-sur-mer (LER/BL), Nantes (LER/MPL/NT) et Arcachon (LER/AR).

Un set d'apprentissage "global" Manche-Atlantique 4X composé d'images provenant des banques d'images créées sur les trois sites d'acquisition a été constitué. Les images sont issues de 113 échantillons d'eau de mer d'Atlantique de 2011 à 2014 et 127 de Manche de 2013. Ce set est une base de travail pour les utilisateurs. Il doit être complété notamment pour inclure les classes peu abondantes ou encore non alimentées. Ce set doit aussi être optimisé par les utilisateurs pour créer un outil de reconnaissance performant et adapté.

Les utilisateurs du logiciel ZooPhytoImage, plus particulièrement lors des étapes de corrections d'erreurs de l'analyse d'échantillon, doivent avoir une connaissance visuelle de la représentation des particules dans chaque classe. D'autant plus s'ils n'ont pas participé à la constitution de la bibliothèque d'image. Il est apparu donc nécessaire de créer un catalogue de ces images, car visualiser les images directement dans l'arborescence numérique s'avère très fastidieux.

Ce document présente un échantillon d'images de chaque classe qui constitue ce set d'apprentissage Manche-Atlantique 4X, et aussi les classes dont les abondances sont insuffisantes pour être intégrés au set et pour lesquelles la collecte de nouvelles images est attendue (<LowAbundances>). L'ordre de présentation des classes intégrées au set correspond à l'ordre d'apparition de ces classes dans l'outil ZooPhytoImage, qui correspond à l'ordre des classes dans l'arborescence des répertoires contenant les vignettes.

*** Contributeurs à l'acquisition d'images et constitution du set :**

LER/BL : Guillaume Wacquet, Camille Blondel, Pascale Hébert et Alain Lefebvre,

VIGIES/Nantes : Nadine Neaud-Masson,

LER/AR : Danièle Maurer, Claire Meteigner et Myriam Perrière.

Ce set comprend 7081 images classées dans 61 classes dont les effectifs sont présentés dans le tableau suivant

- 27 classes phytoplanctoniques appartenant aux *Bacillariophyceae*.
- 15 classes phytoplanctoniques appartenant aux *Dinophyceae*
- 4 classes phytoplanctoniques autres
- 3 classes zooplanctoniques
- 12 classes détritiques

Tableau 1 : Distribution des particules incluses entre les différentes classes

<i>Bacillariophyceae</i>	Nb
<i>Asterionellopsis glacialis</i>	179
<i>Bacteriastrium</i>	176
<i>Cerataulina</i>	123
<i>Chaetoceros</i>	203
<i>Chaetoceros curvisetus</i>	194
<i>Chaetoceros danicus</i>	62
<i>Chaetoceros socialis</i>	60
<i>Cylindrotheca Hantzschia Nitzschia</i>	68
<i>Dactyliosolen fragilissimus</i>	175
<i>Ditylum brightwellii</i>	184
<i>Eucampia</i>	107
<i>Guinardia delicatula</i>	31
<i>Guinardia flaccida</i>	202
<i>Guinardia striata</i>	78
<i>Gyrosigma Pleurosigma</i>	221
<i>Lauderia</i>	43
<i>Leptocylindrus</i>	197
<i>Odontella</i>	148
<i>Paralia</i>	84
<i>Proboscia Rhizosolenia</i>	161
<i>Pseudo-nitzschia</i>	185
<i>Rhizosolenia setigera</i>	41
<i>Skeletonema</i>	153
<i>Thalassionema</i>	130
<i>Thalassiosira cells</i>	38
<i>Thalassiosira big chaines</i>	107
<i>Thalassiosira small chaines</i>	97

Autres classes	
<i>Phaeocystis</i>	264
<i>Dictyocha</i>	94
<i>Euglenoidea</i>	53
<i>Noctiluca</i>	43

<i>Dinophyceae</i>	Nb
<i>Alexandrium</i>	187
<i>Alexandrium affine</i>	93
<i>Ceratium</i>	52
<i>Dinophysis</i>	174
<i>Dinophysis tripos</i>	35
<i>Dissodinium Pyrocystis</i>	41
<i>Gonyaulax</i>	48
<i>Gymnodinium Gyrodinium Torodinium</i>	124
<i>Lepidodinium mucous</i>	170
<i>Neoceratium furca lineatum</i>	151
<i>Neoceratium fusus</i>	183
<i>Peridinium Protoperidinium</i>	177
<i>Polykrikos</i>	22
<i>Prorocentrum</i>	258
<i>Scrippsiella Ensiculifera Pentapharsodinium Bysmatrum</i>	149

zooplancton

<i>Ciliophora_cils</i>	94
<i>Ciliophora_cone</i>	94
Zooplankton spp	34

Débris

black opaque	160
bubble	200
clear	163
Clear grainy	57
Dark grainy	41
Dark mineral	200
Fecal pellets	20
fiber	120
Fiber dark	59
Granular clear	74
Granular dark	100
membranous	80

2. Description des fiches

Nom de la classe

Position de la classe dans l'arborescence du set

3.19. *Pseudo-nitzschia*

Ochrophyta \ Bacillariophyceae \ Bacillariaceae \ *Pseudo-nitzschia*

Photo numérique (échelle variable) microscope optique à contraste de phase.
Auteur : Nadine Masson-Neaud @Ifremer

Images du laboratoire de Boulogne sur Mer

Images obtenues par numérisation au FlowCAM et traitement avec ZooPhytoImage

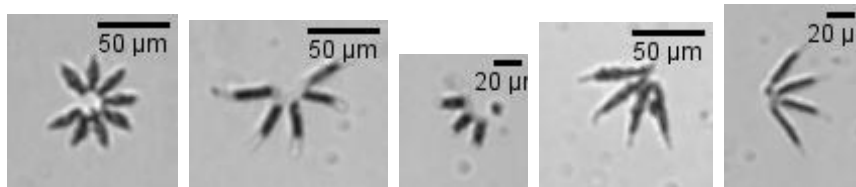
Images du laboratoire de Nantes

Images du laboratoire d'Arcachon

3. Classes non intégrées au set

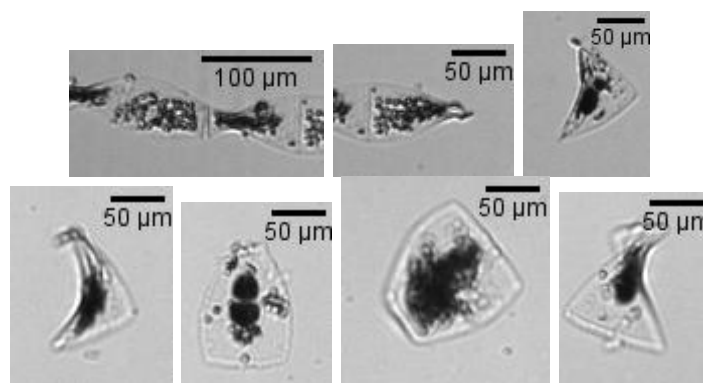
Sous répertoires de <LowAbundances> ces classes nécessitent l'acquisition de nouvelles images pour pouvoir être intégrées au set.

3.1. *Asteroplanus karianus*



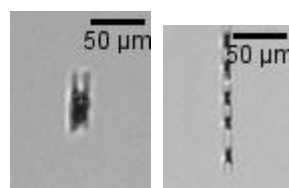
NT

3.2. *Helicotheca*



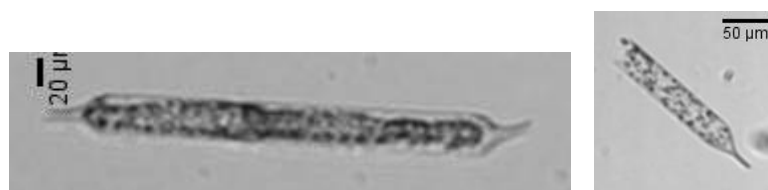
AR

3.3. *Hemiaulus*



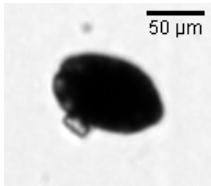
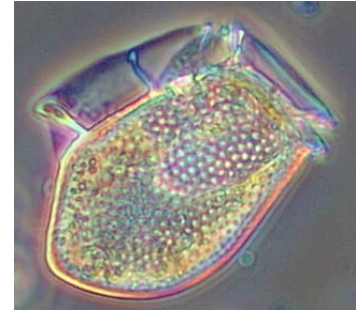
AR

3.4. *Proboscia indica*

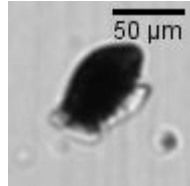
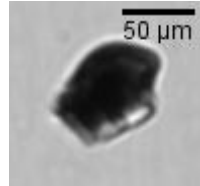


NT

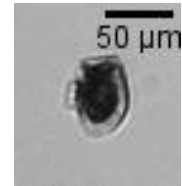
3.5. *Dinophysis acuta*



BL

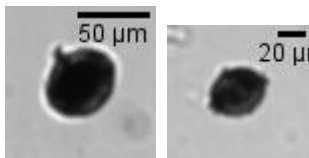


NT

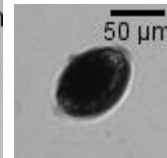
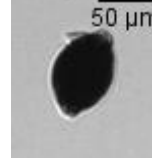
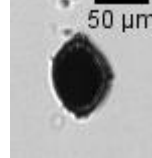


AR

3.6. *Diplopsalis Diplopelta Oblea*

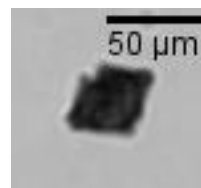
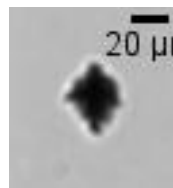


NT



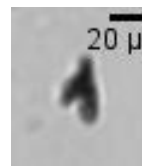
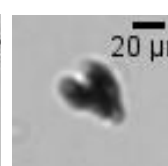
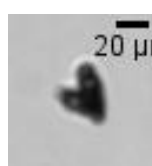
AR

3.7. *Heterocapsa*



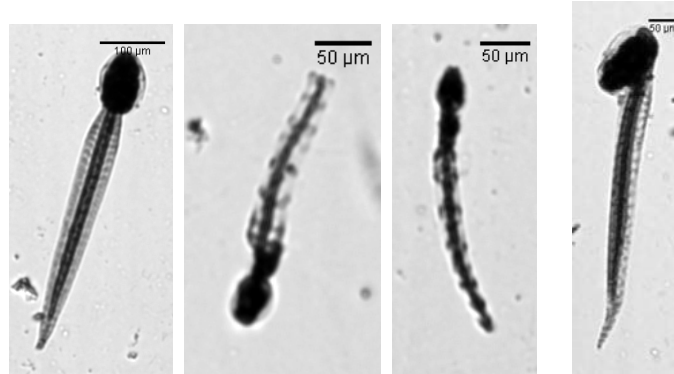
NT

3.8. *Pyramimonas*

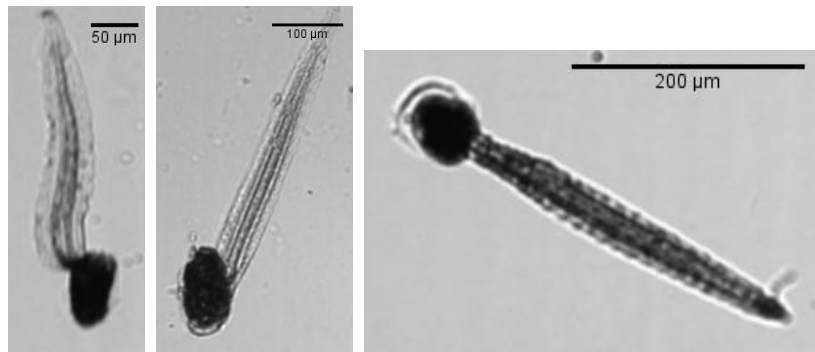


NT

3.9. Pisces larva

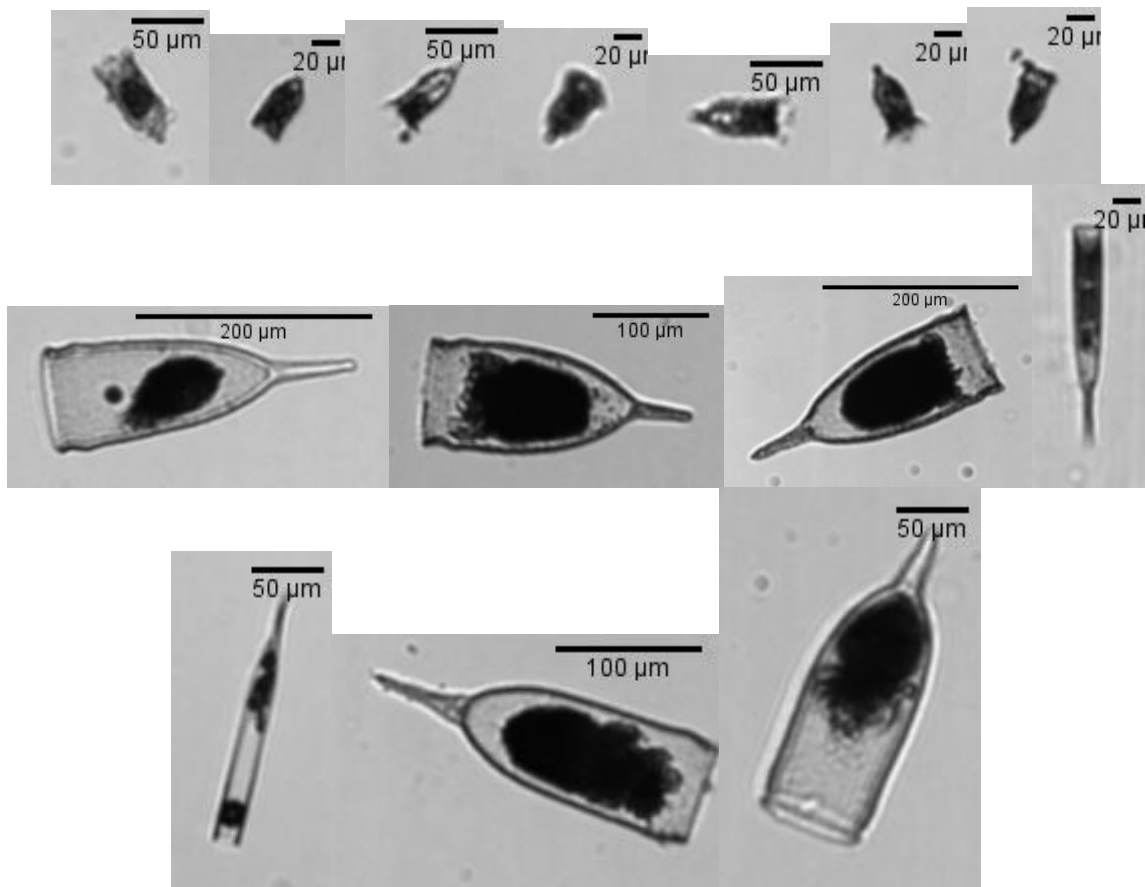


BL



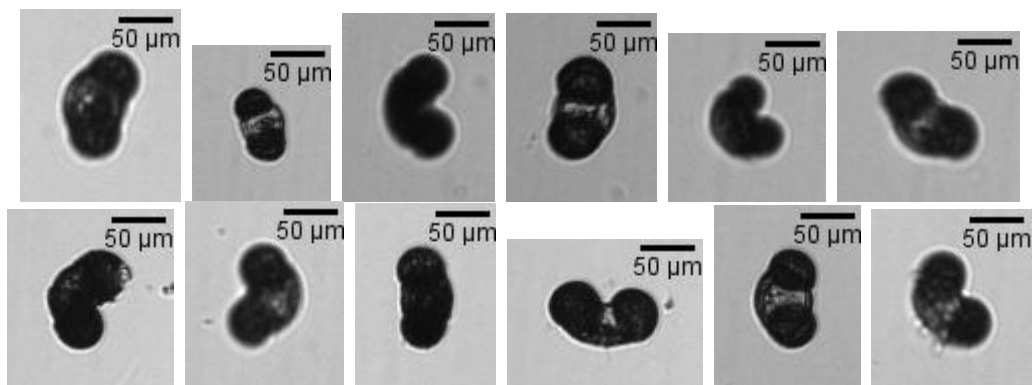
NT

3.10. tintinnids



NT

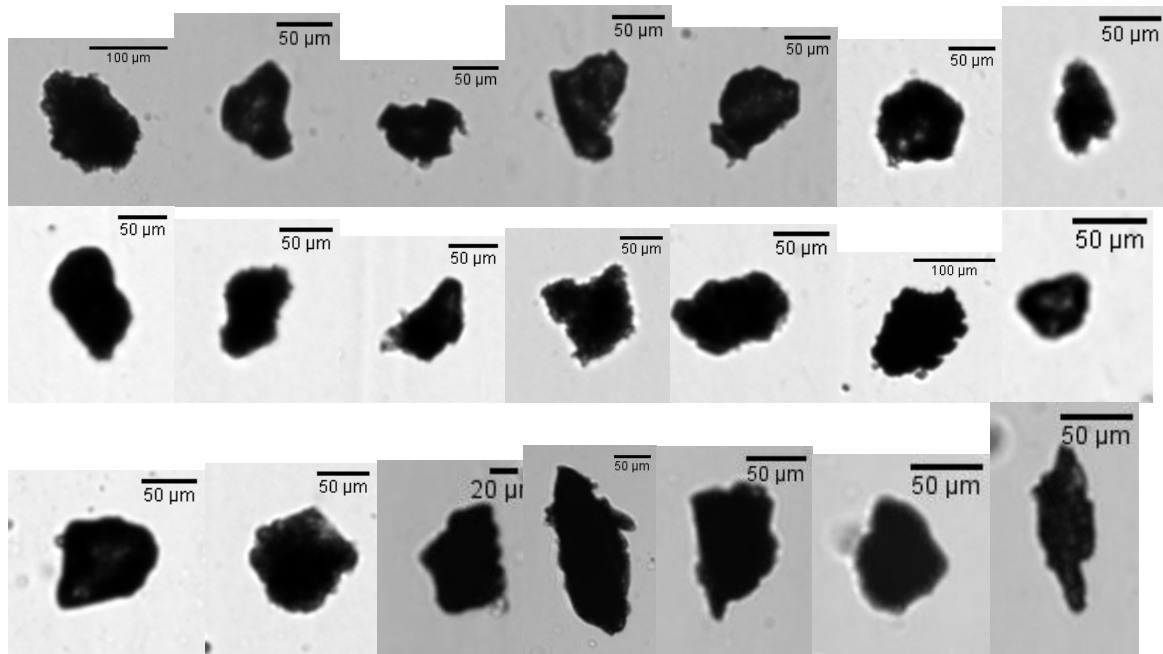
3.11. polen



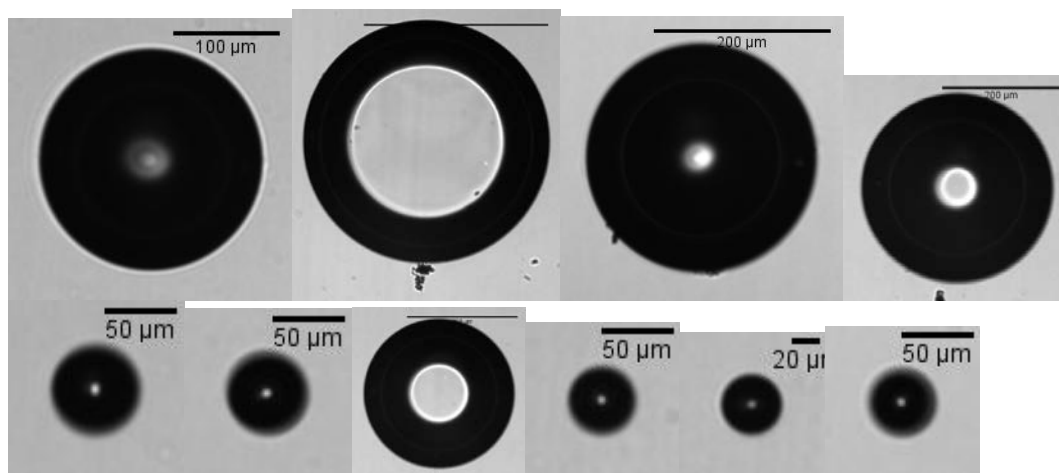
AR

4. Classes des particules intégrées au set (débris)

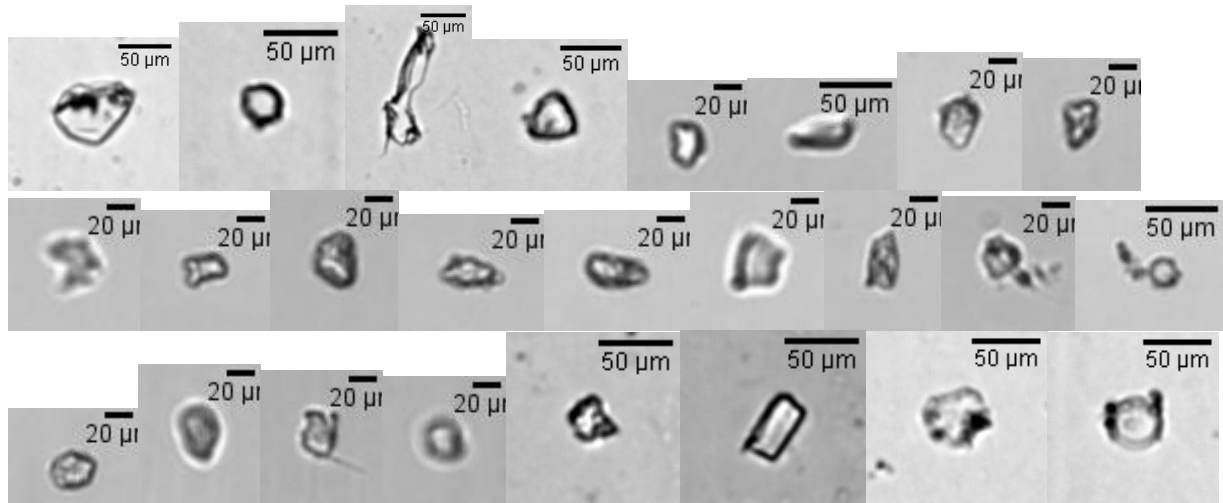
4.1. Black opaque



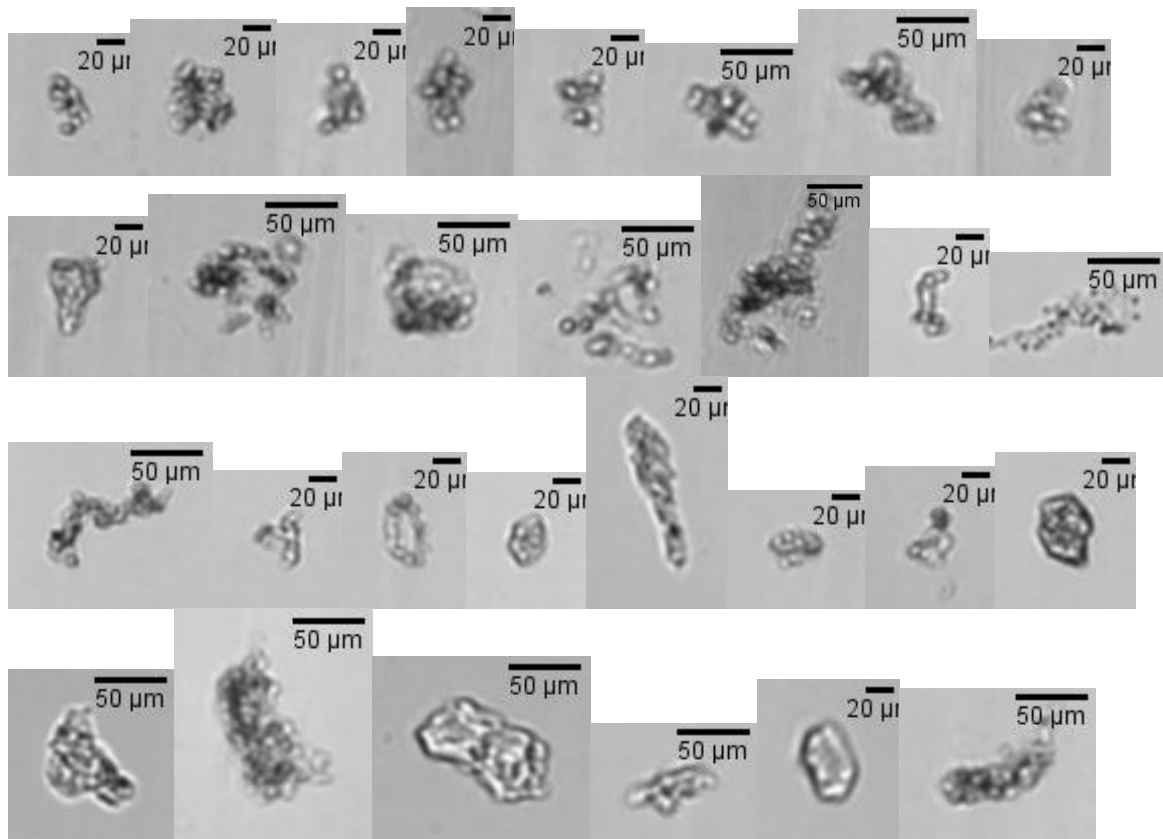
4.2. Bubble



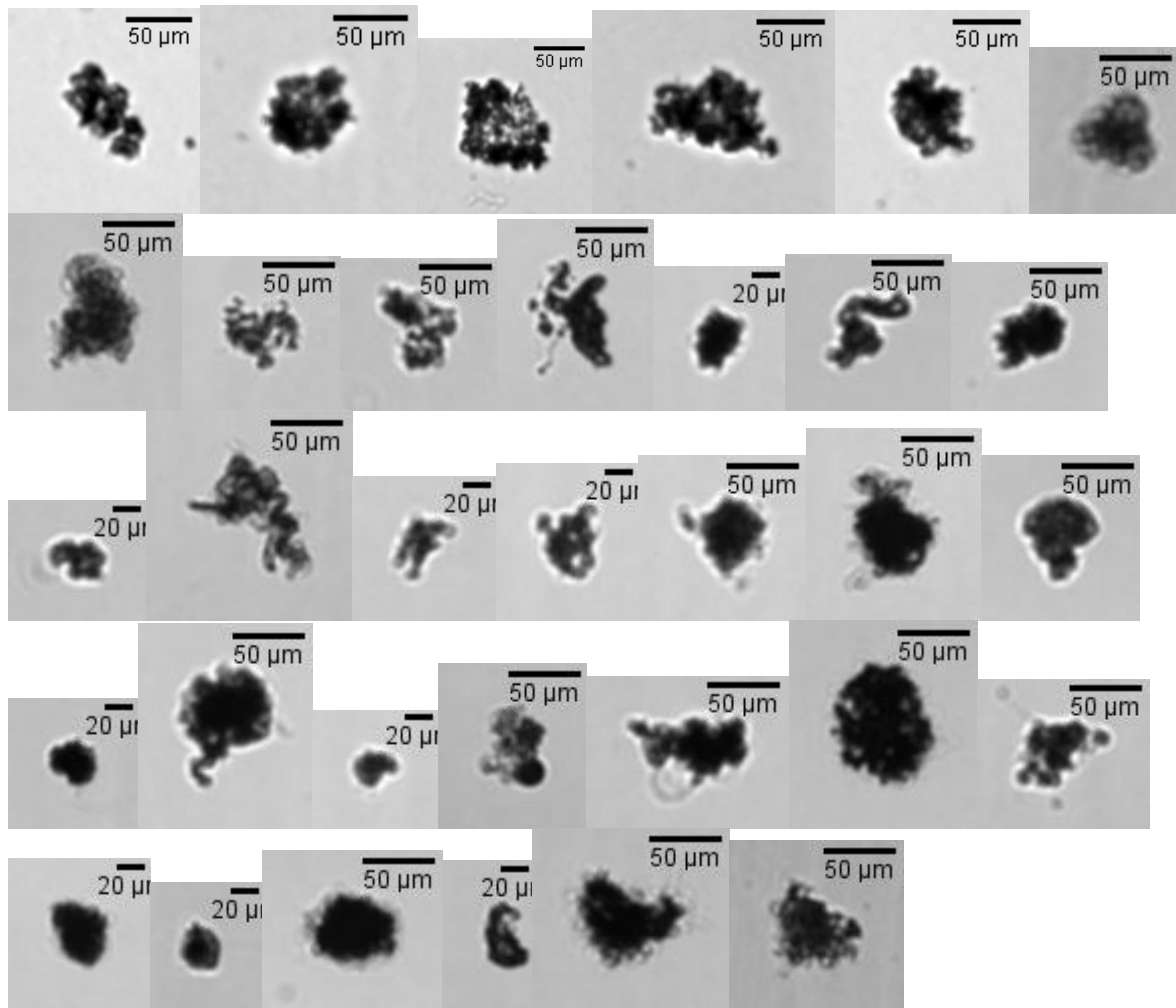
4.3. Clear



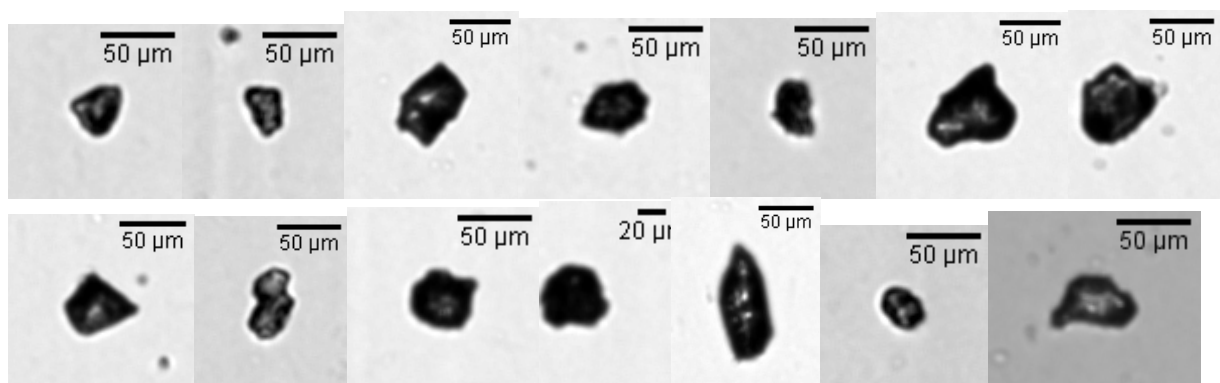
4.4. Clear grainy



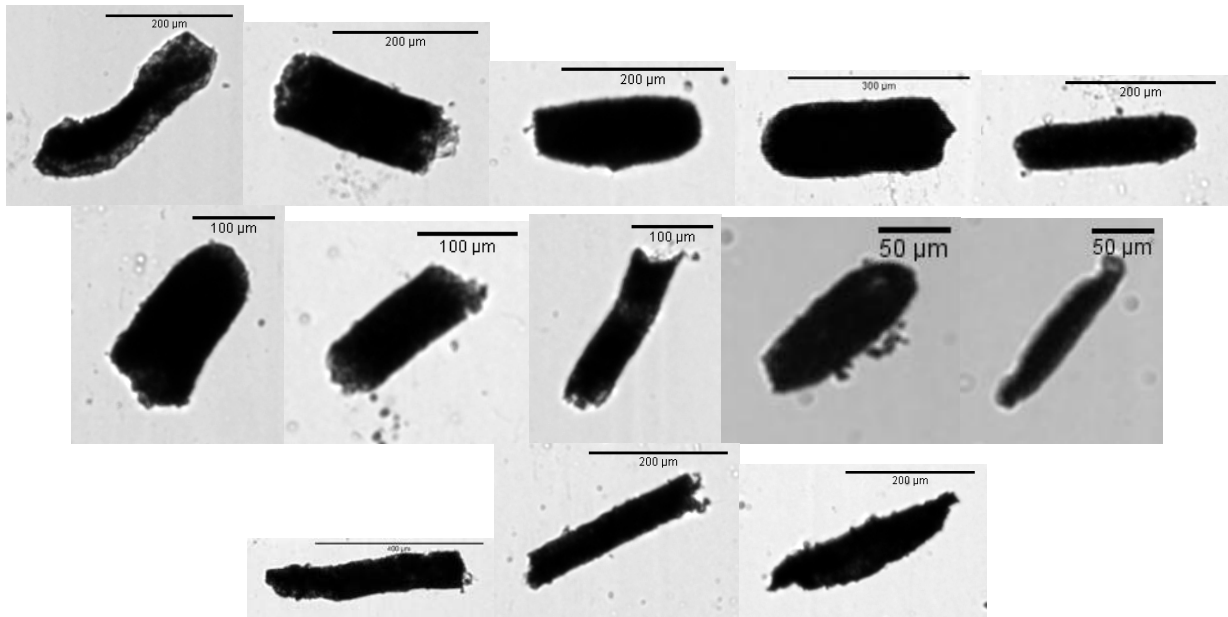
4.5. Dark grainy



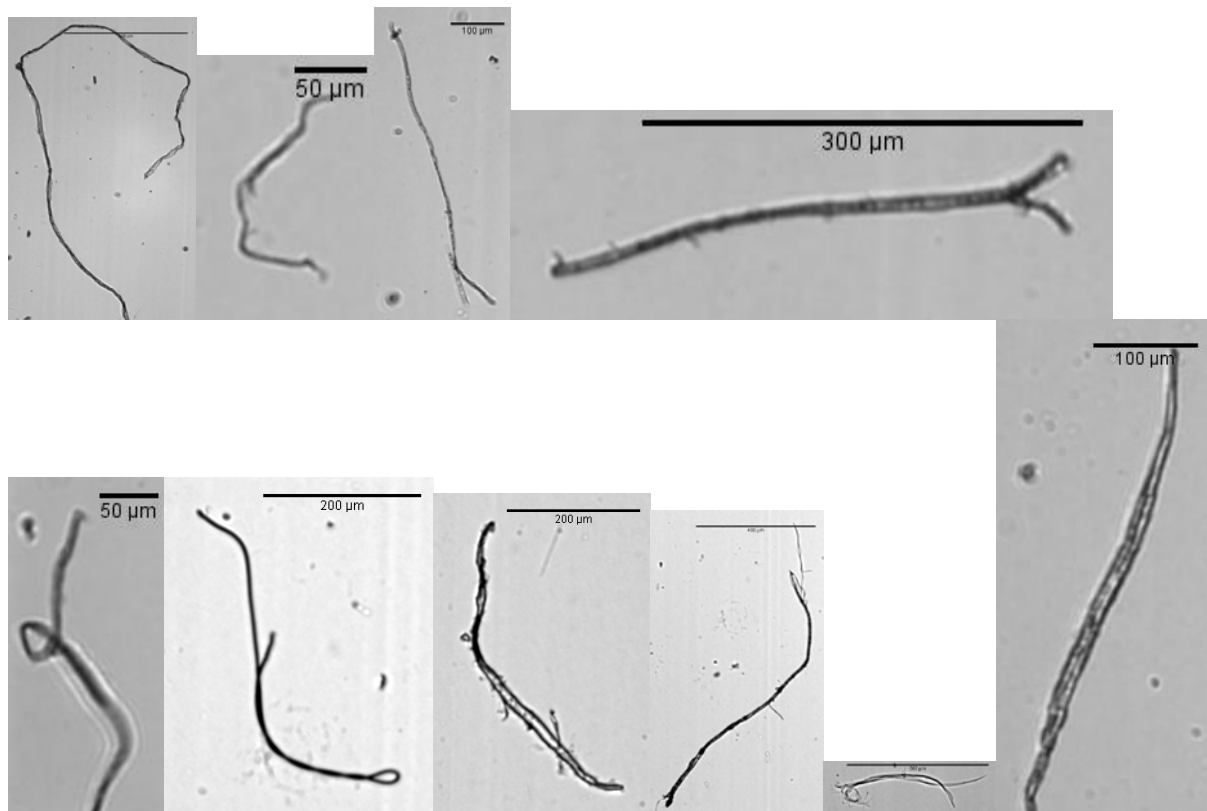
4.6. Dark mineral



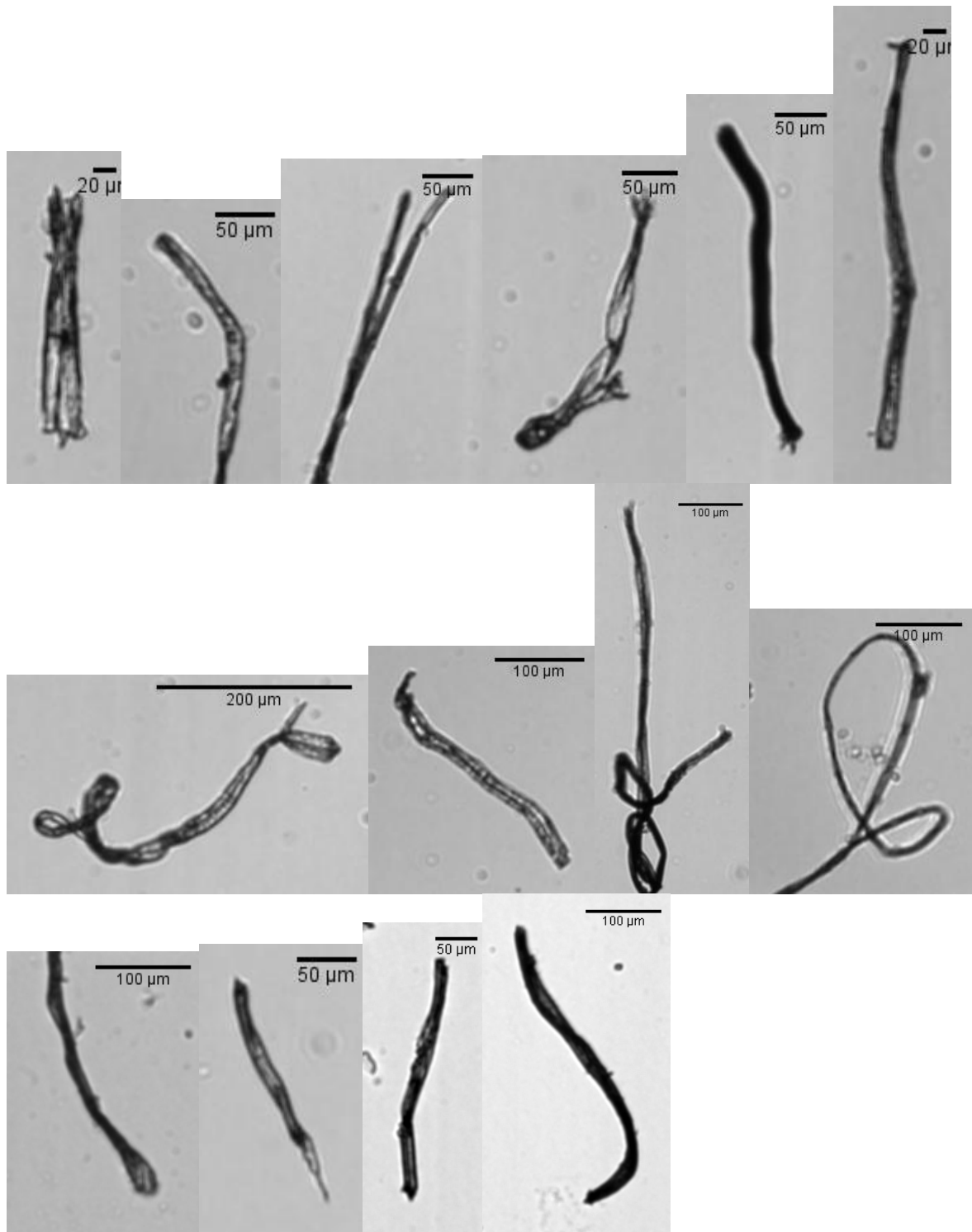
4.7. Fecal pellets



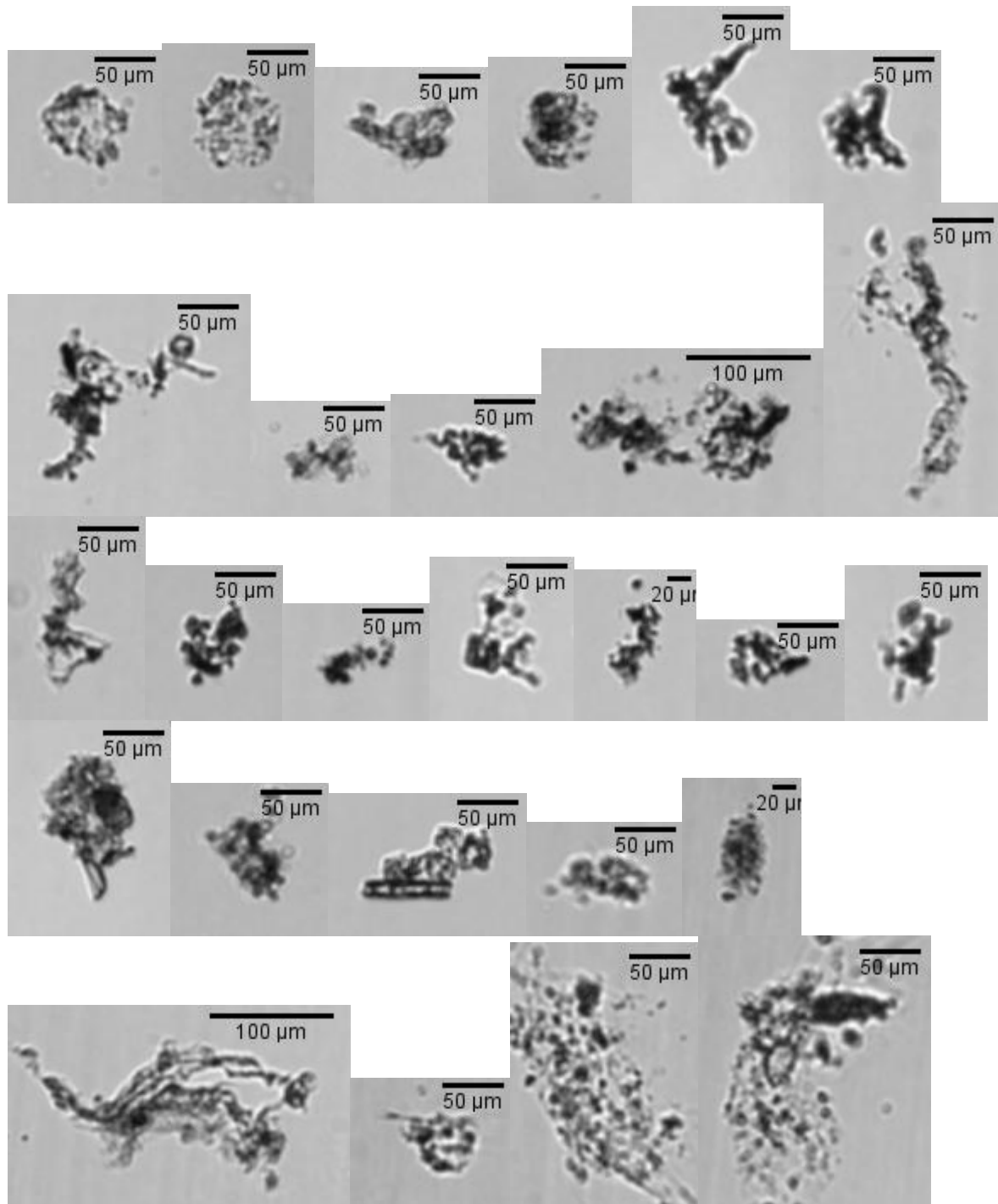
4.8. Fiber



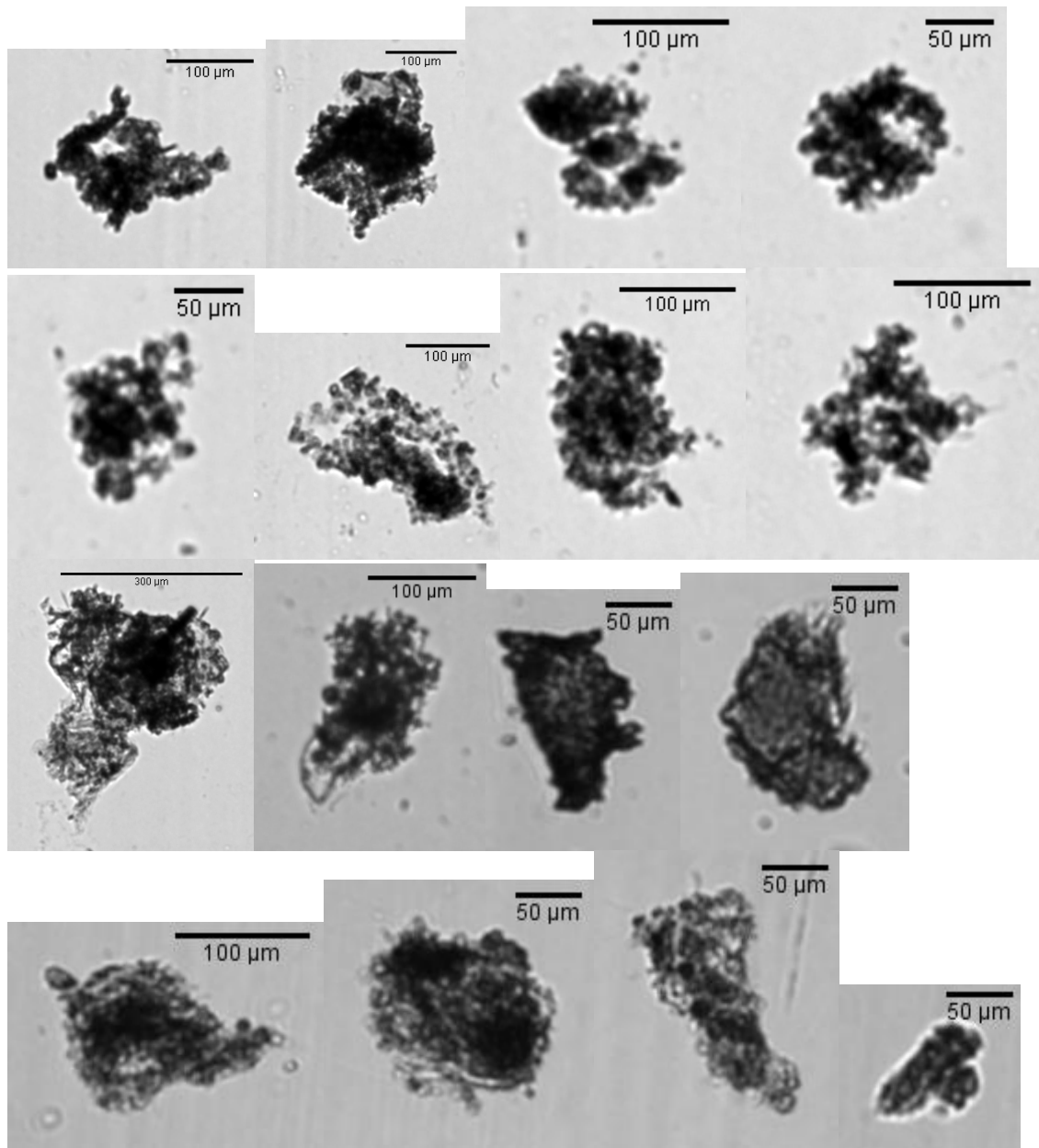
4.9. Fiber dark



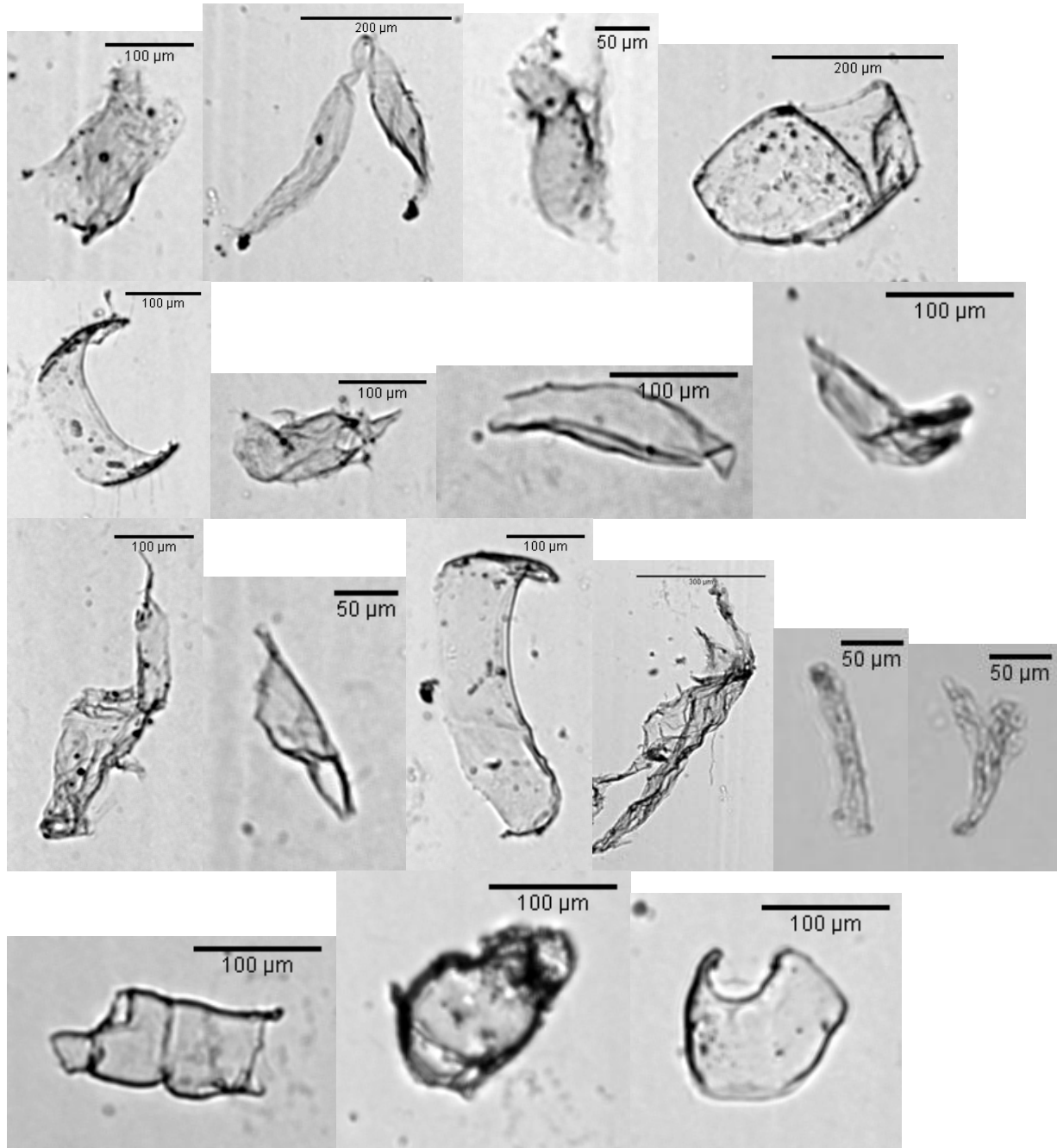
4.10. Granular clear



4.11. Granular dark



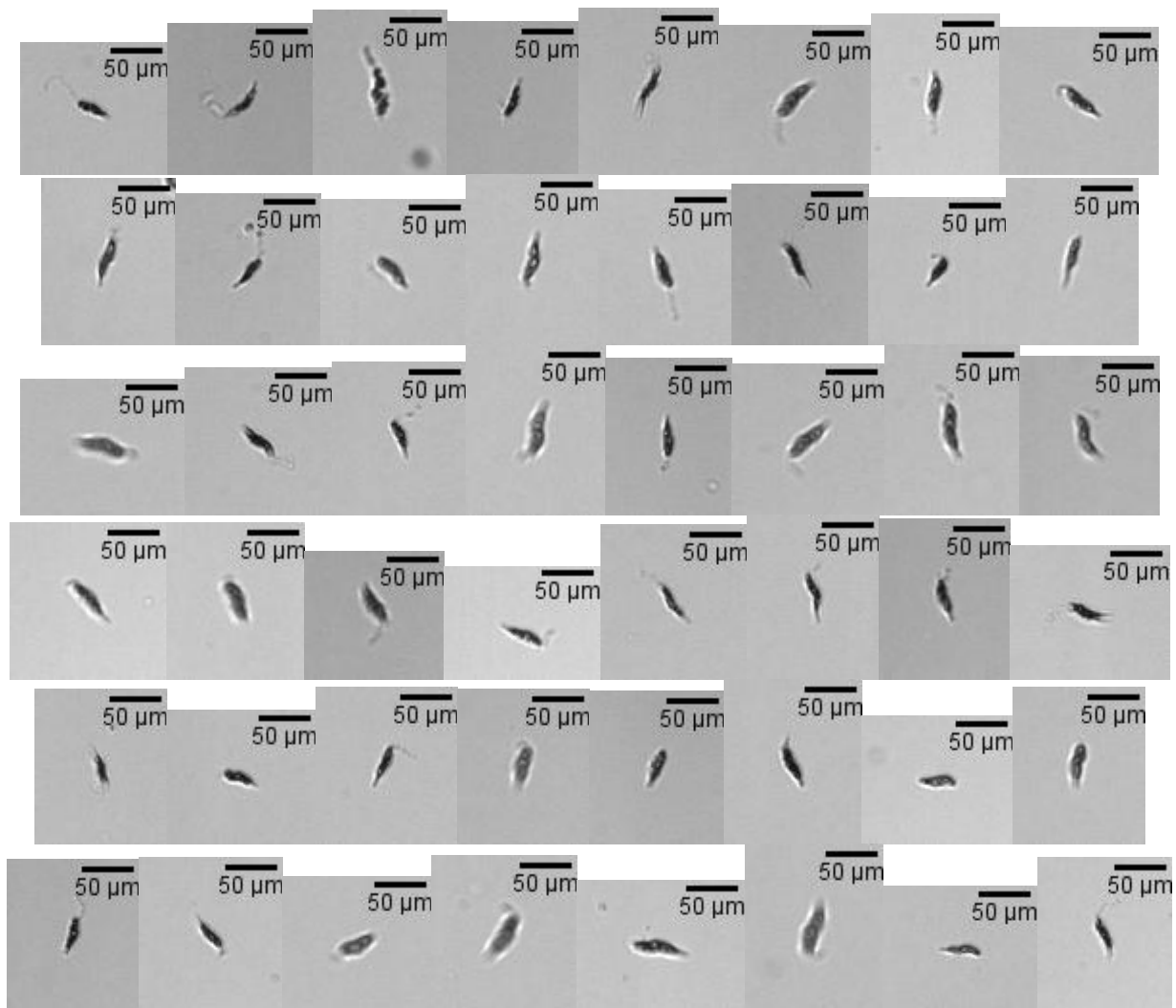
4.12. Membranous



5. Classes phytoplancton intégrées au set

5.1. *Euglenoidea*

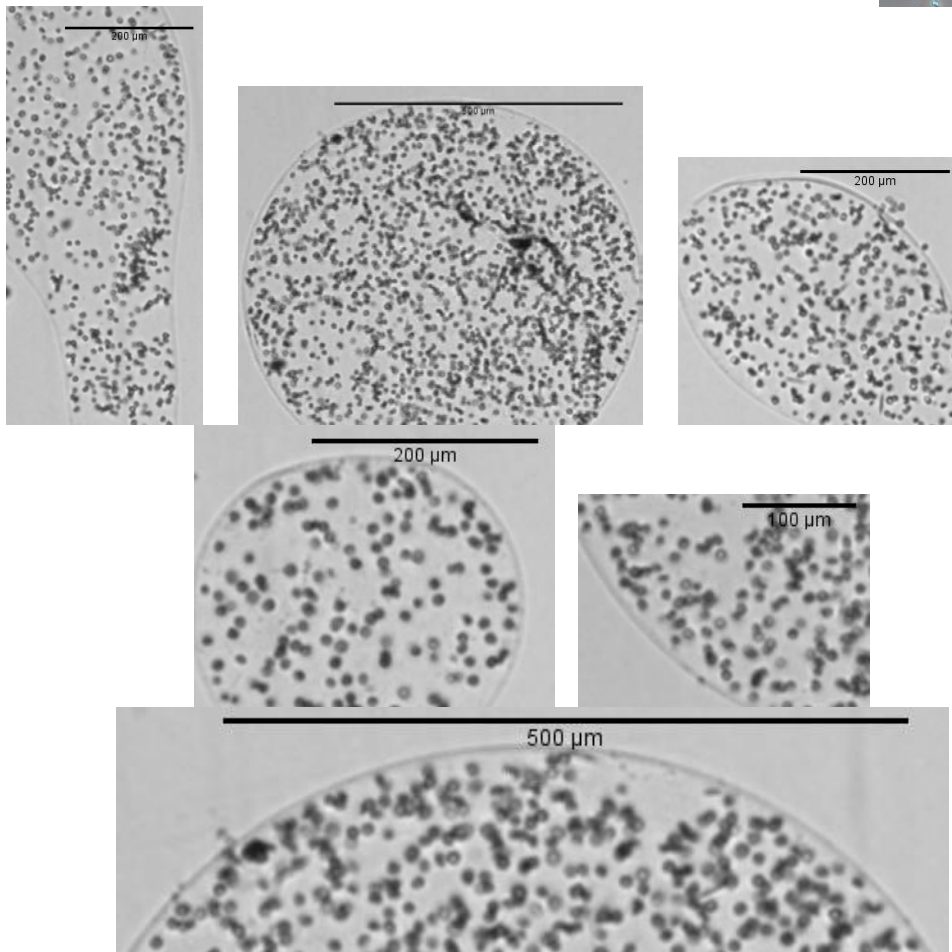
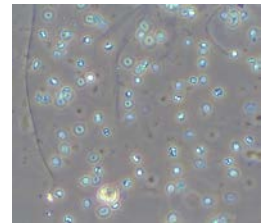
Euglenozoa\Euglenoidea



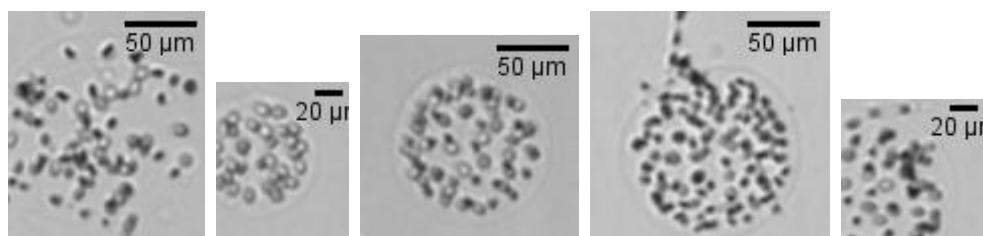
AR

5.2. *Phaeocystis*

Haptophyta\Prymnesiophyceae\Phaeocystaceae\Phaeocystis



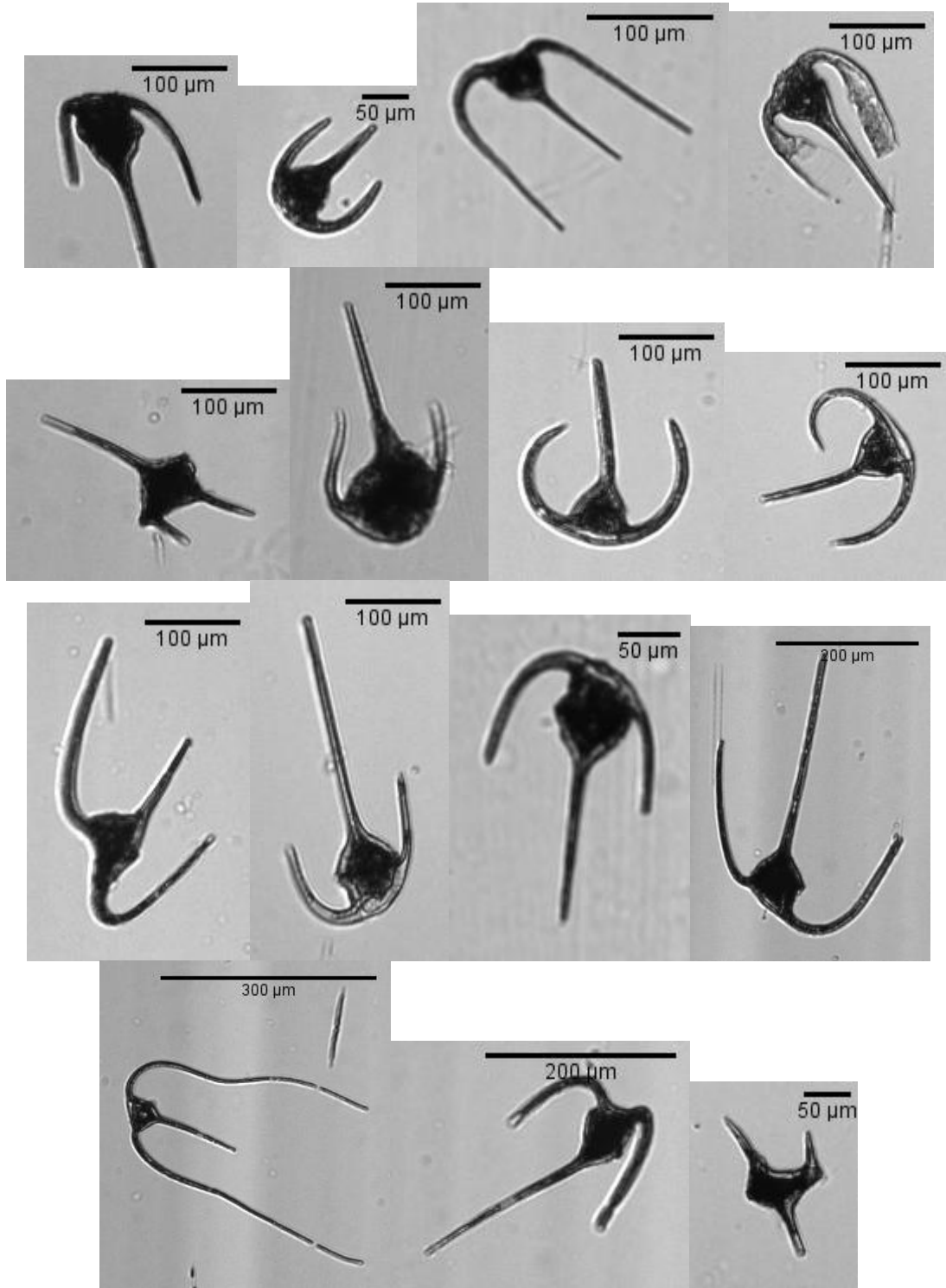
BL



NT

5.3. Ceratium

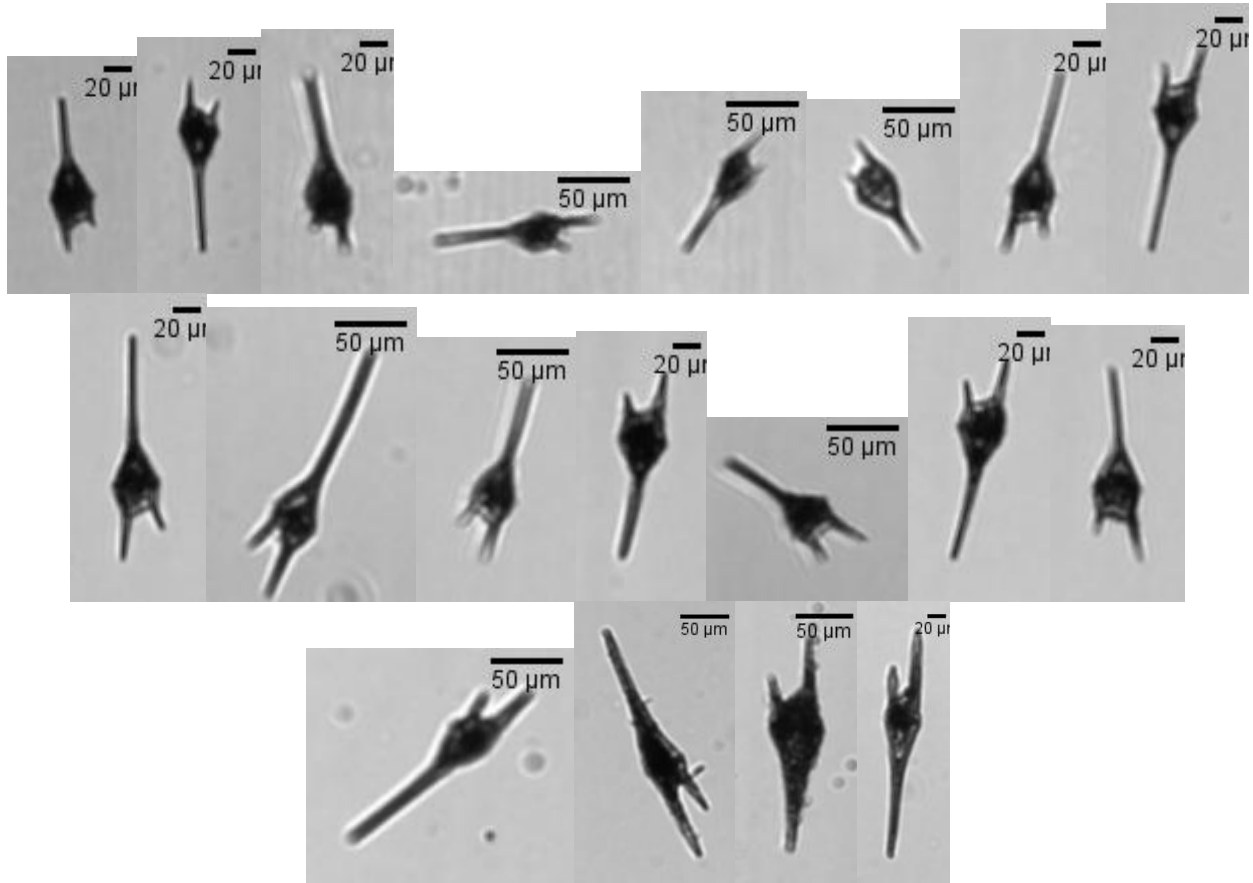
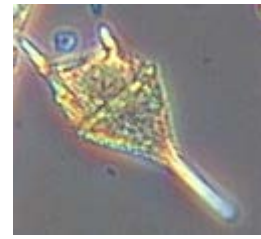
Myzozoa\Dinophyceae\Ceratiaceae\Ceratium



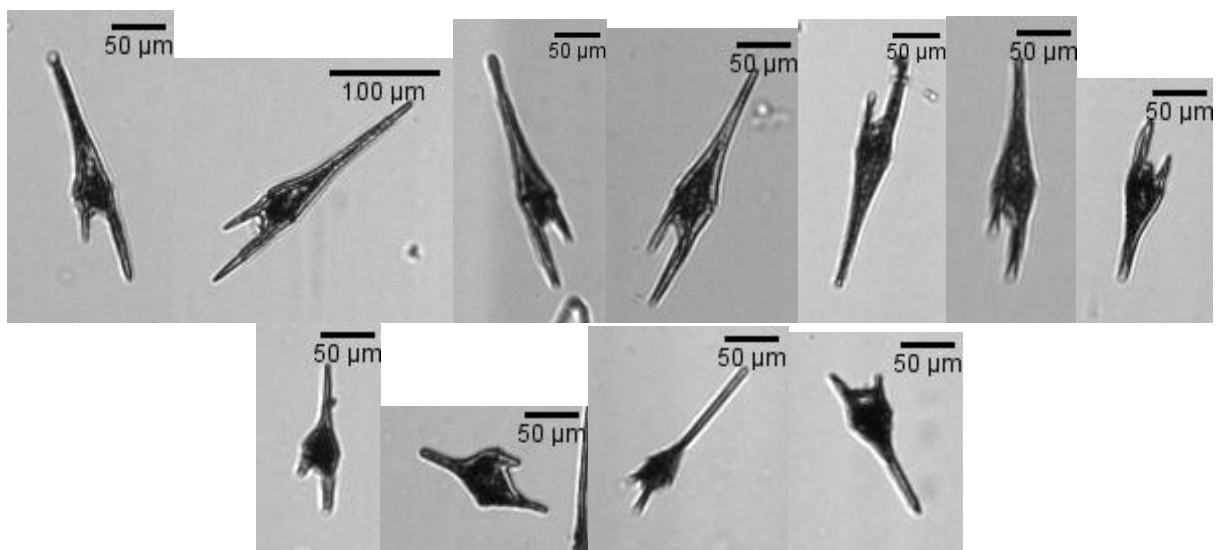
AR

5.4. *Neoceratium furca lineatum*

Myzozoa\Dinophyceae\Ceratiaceae*Neoceratium furca lineatum*



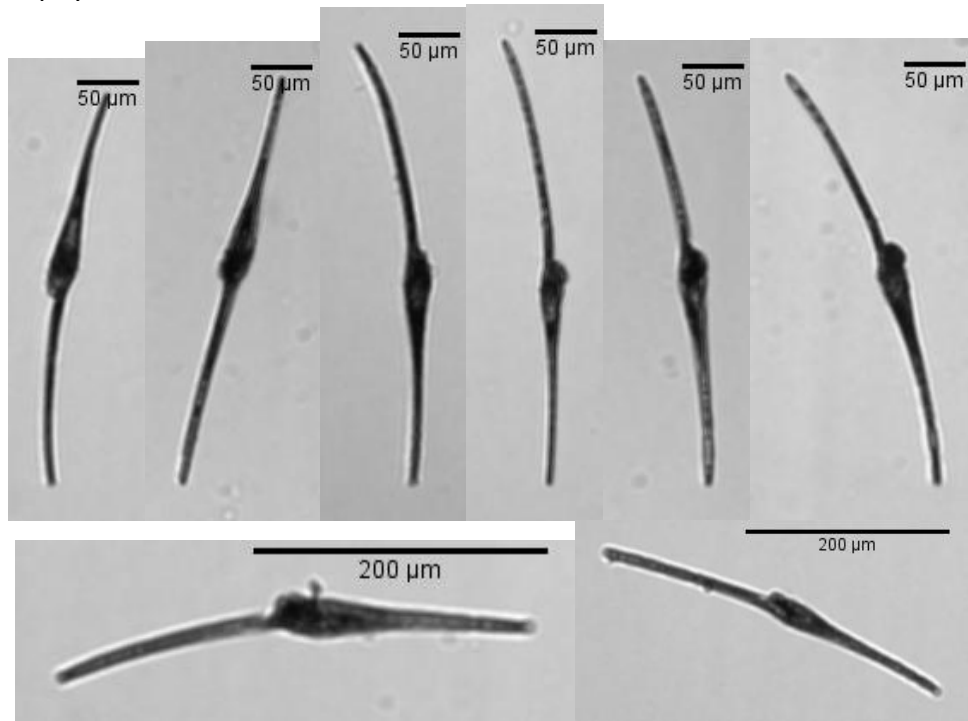
NT



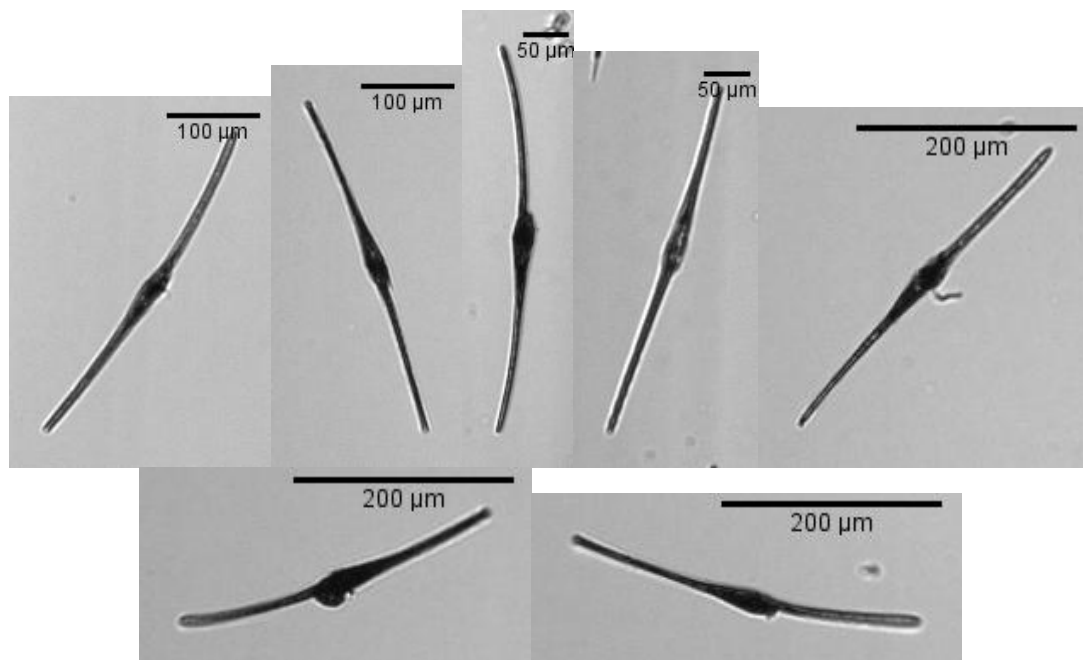
AR

5.5. *Neoceratium fusus*

Myzozoa\Dinophyceae\Ceratiaceae\Neoceratium fusus



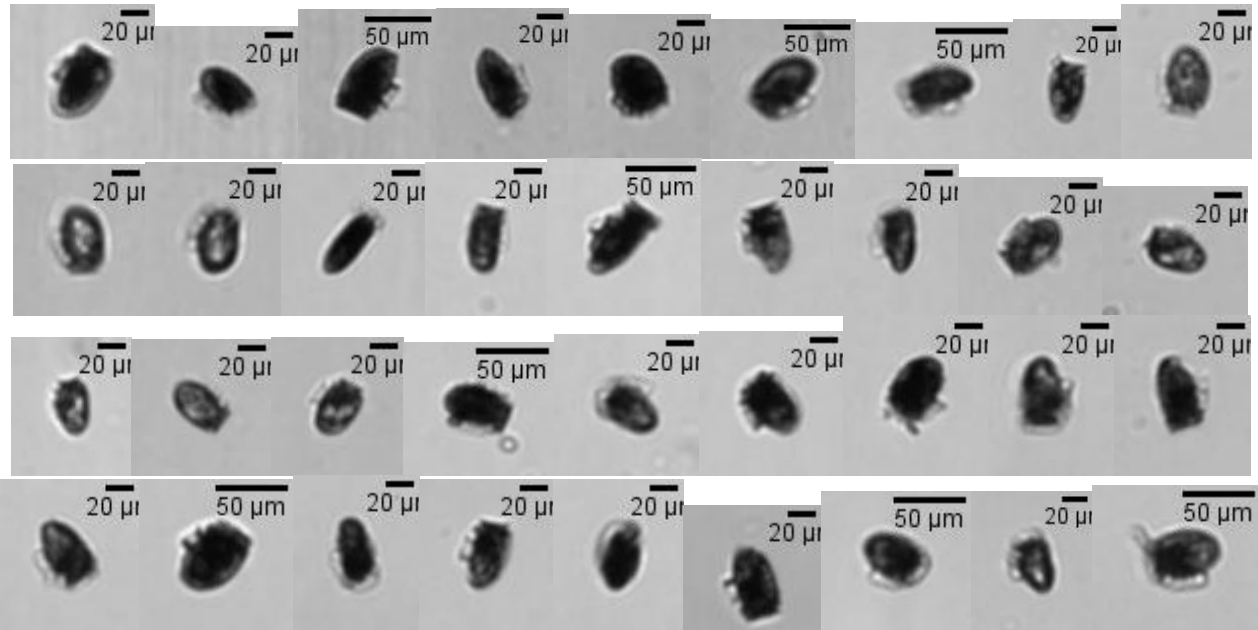
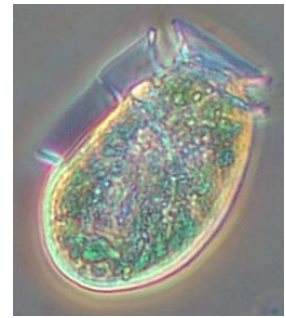
NT



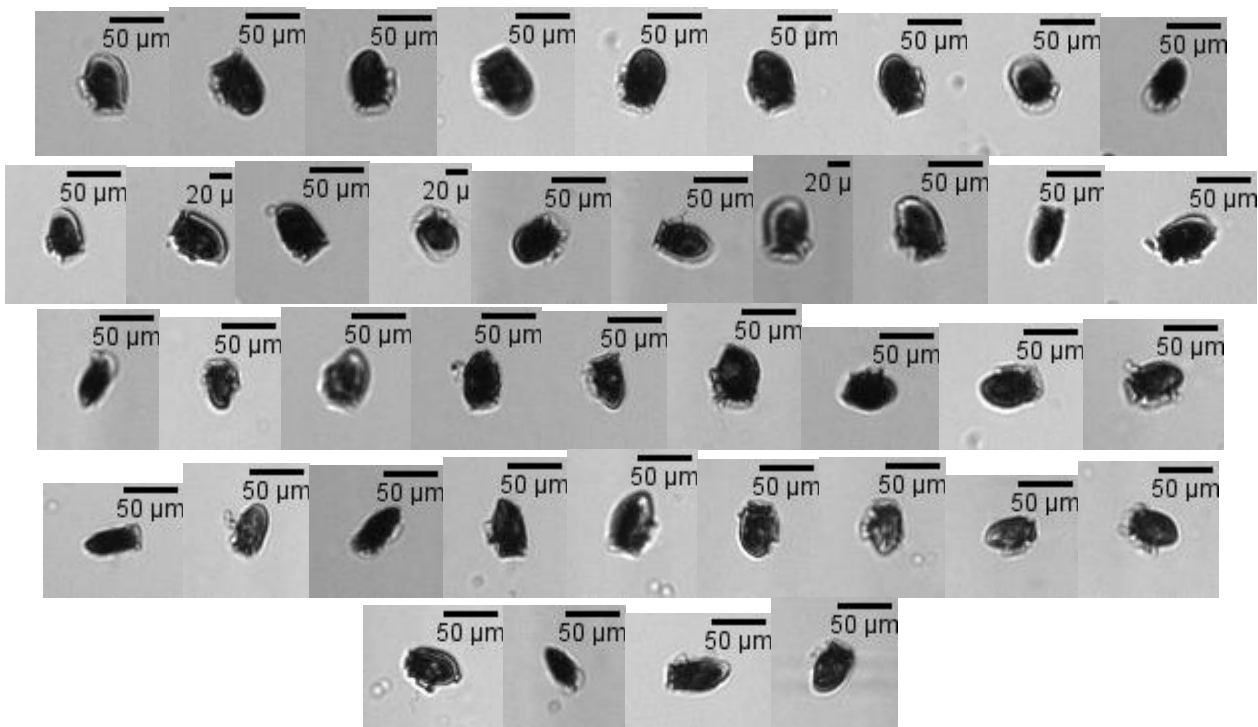
AR

5.6. Dinophysis

Myzozoa\Dinophyceae\Dinophysiaceae\Dinophysis



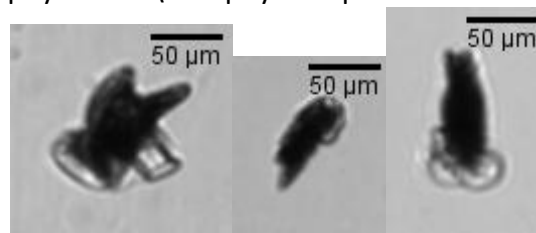
NT



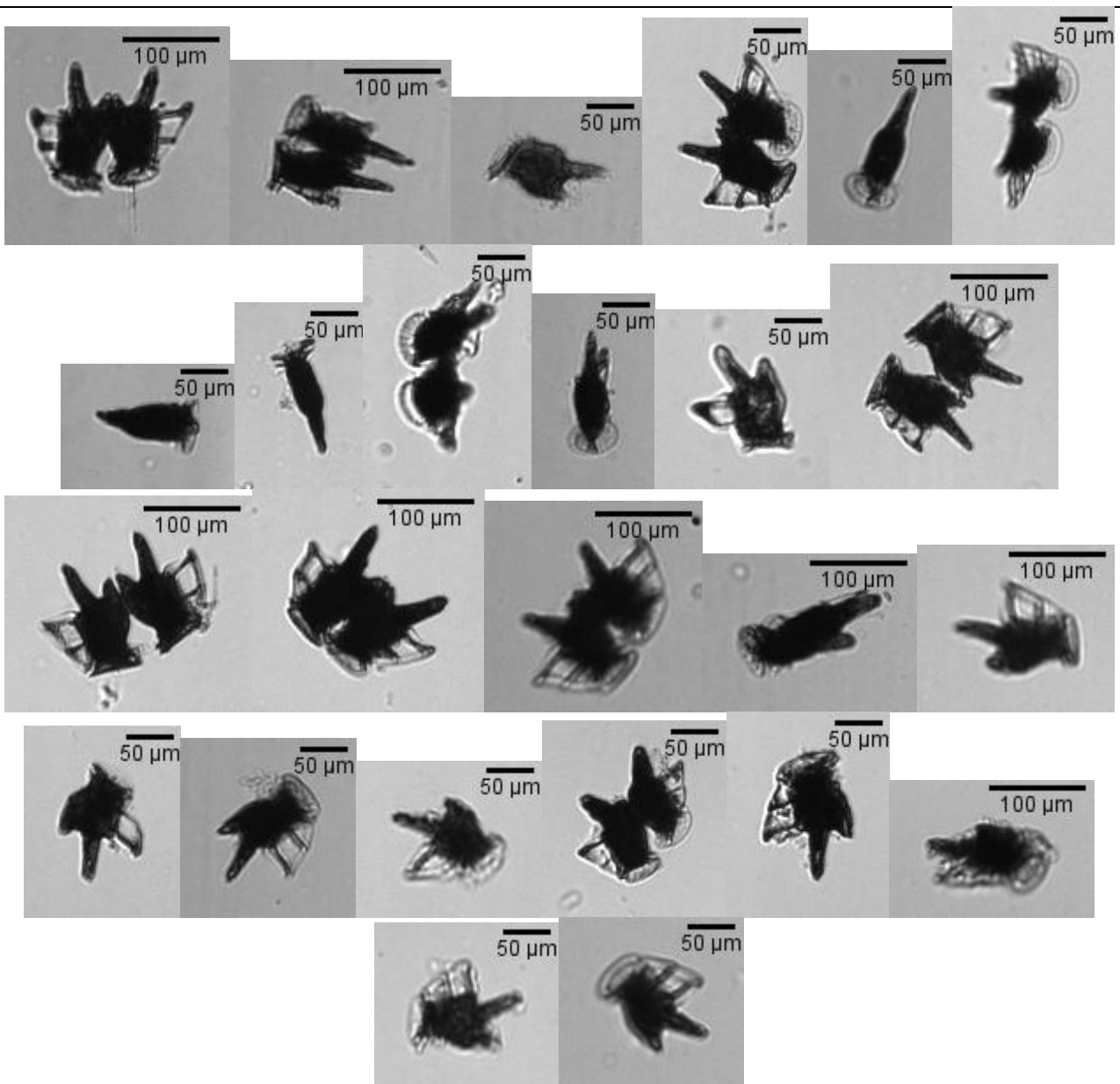
AR

5.7. *Dinophysis tripos*

Myzozoa\Dinophyceae\Dinophysiaceae\Dinophysis tripos



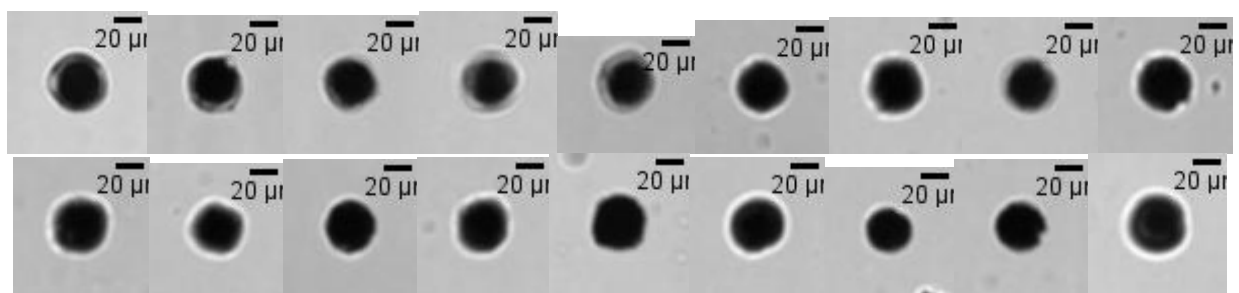
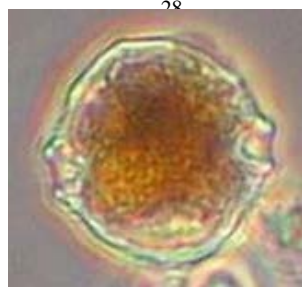
NT



AR

5.8. *Alexandrium*

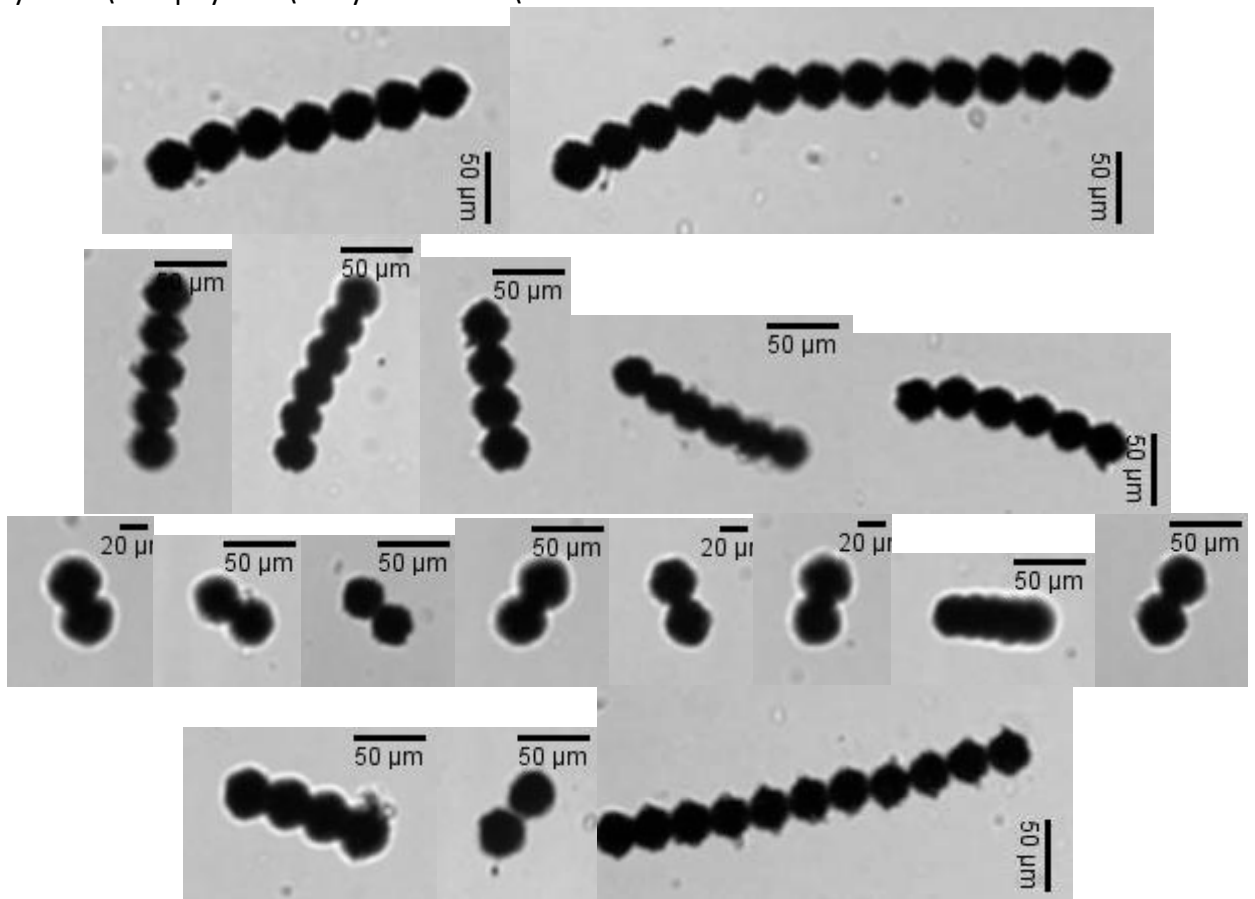
Myzozoa\Dinophyceae\Gonyaulacaceae\Alexandrium



NT

5.9. *Alexandrium affine*

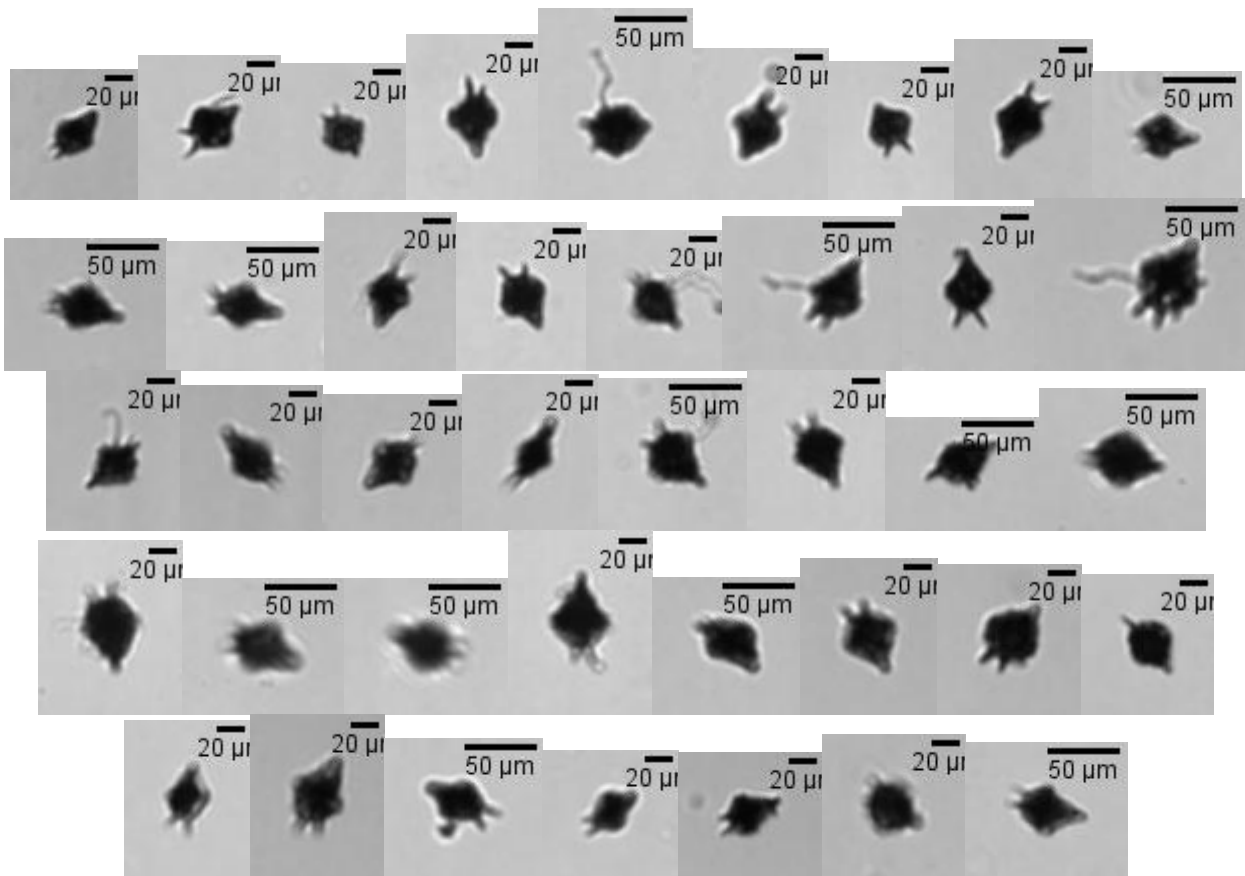
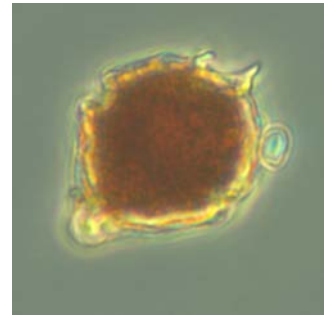
Myzozoa\Dinophyceae\Gonyaulacaceae\Alexandrium affine



NT

5.10. *Gonyaulax*

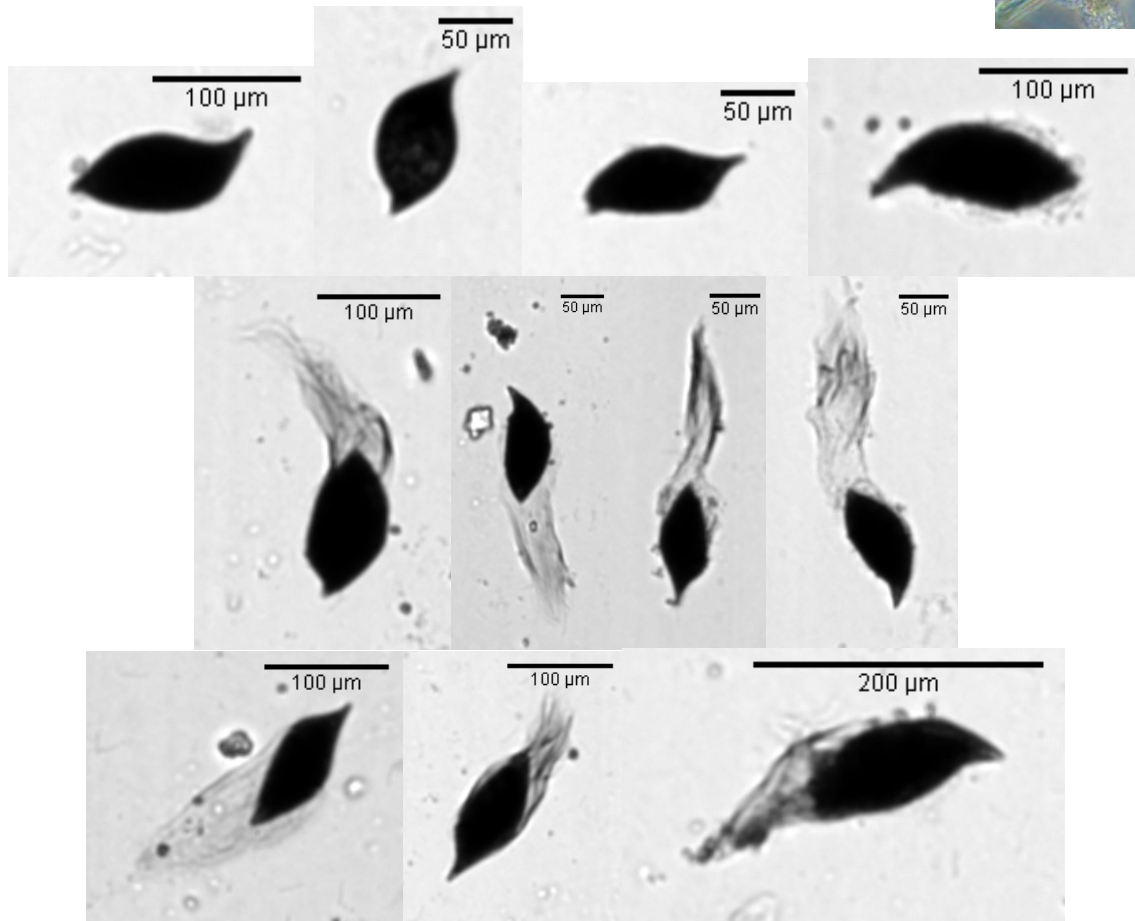
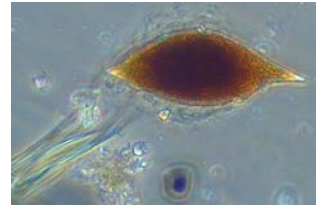
Myzozoa\Dinophyceae\Gonyaulacaceae\Gonyaulax



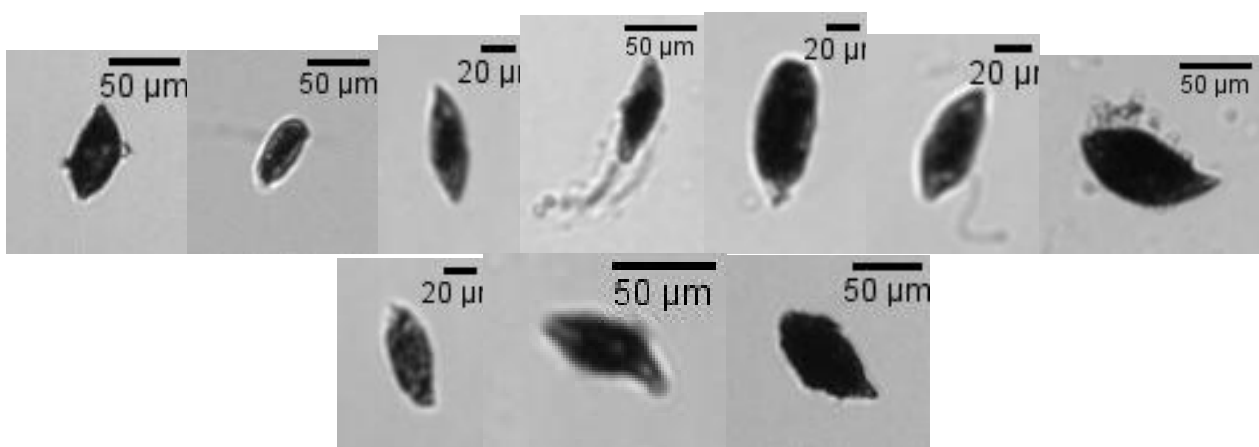
NT

5.11. *Gymnodinium Gyrodinium Torodinium*

Myzozoa\Dinophyceae\Gymnodiniaceae\Gymnodinium Gyrodinium Torodinium



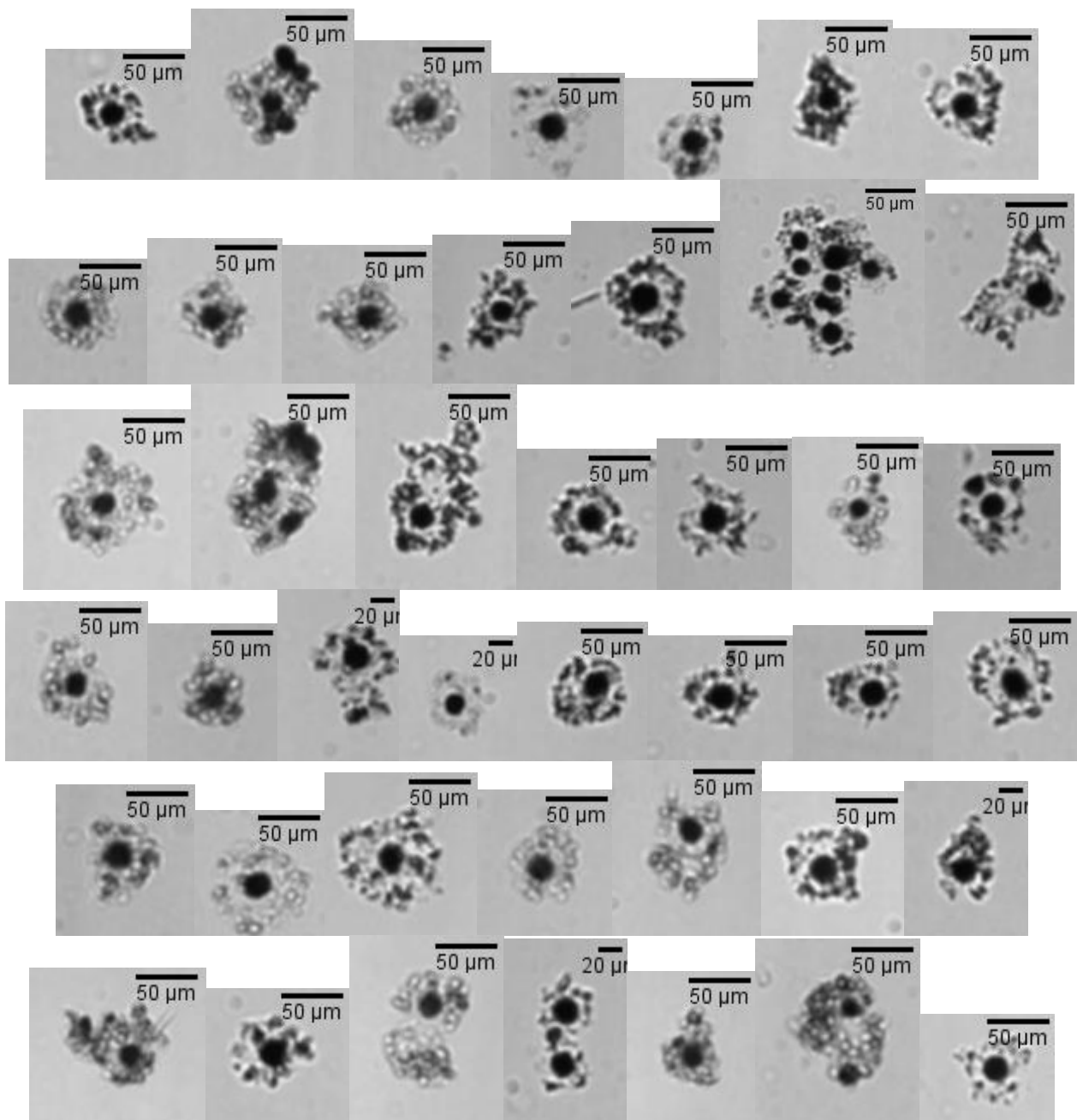
BL



NT

5.12. *Lepidodinium mucous*

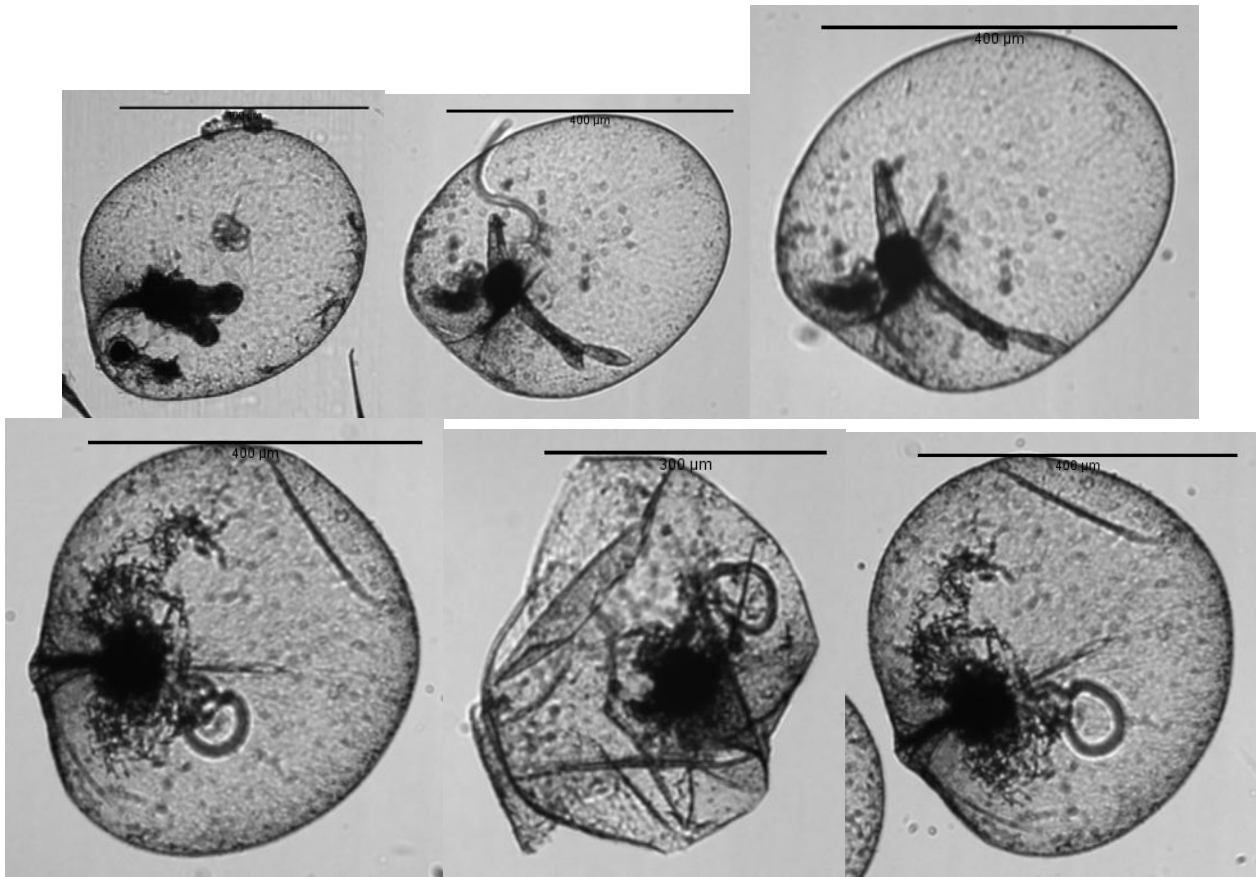
Myzozoa\Dinophyceae\Gymnodiniaceae\Lepidodinium mucous



NT

5.13. *Noctiluca*

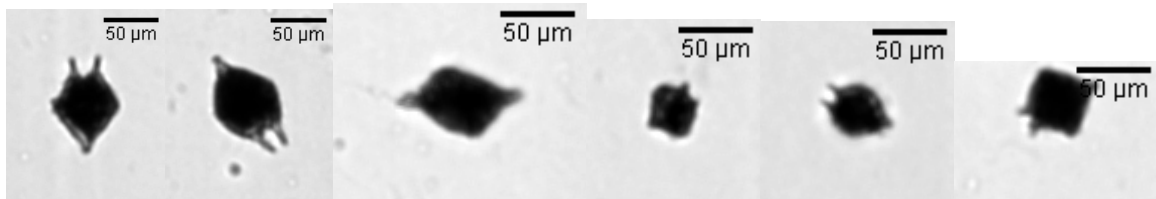
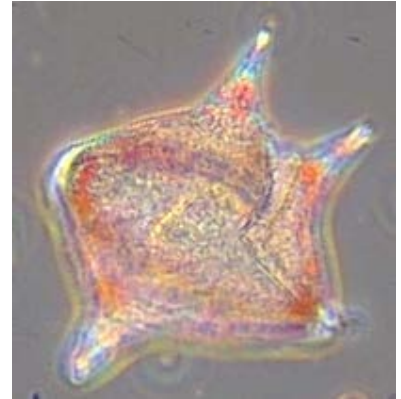
Myzozoa\Dinophyceae>Noctilucaceae>Noctiluca



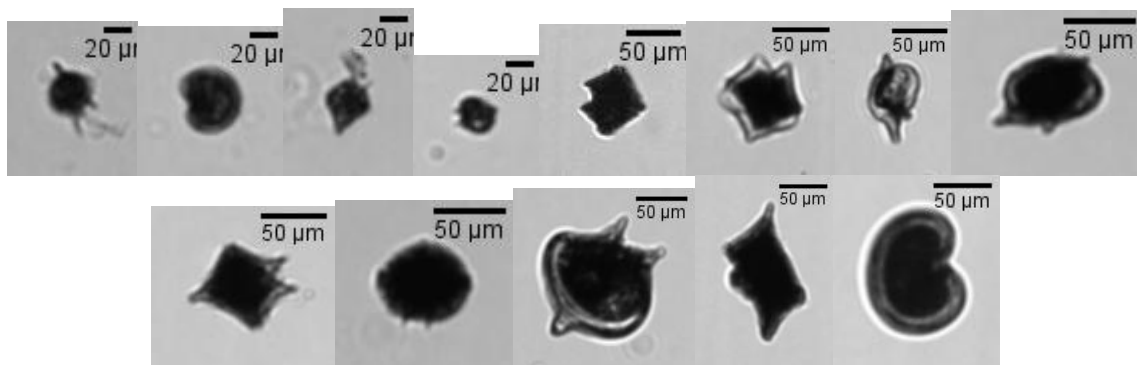
NT

5.14. *Peridinium Protoperidinium*

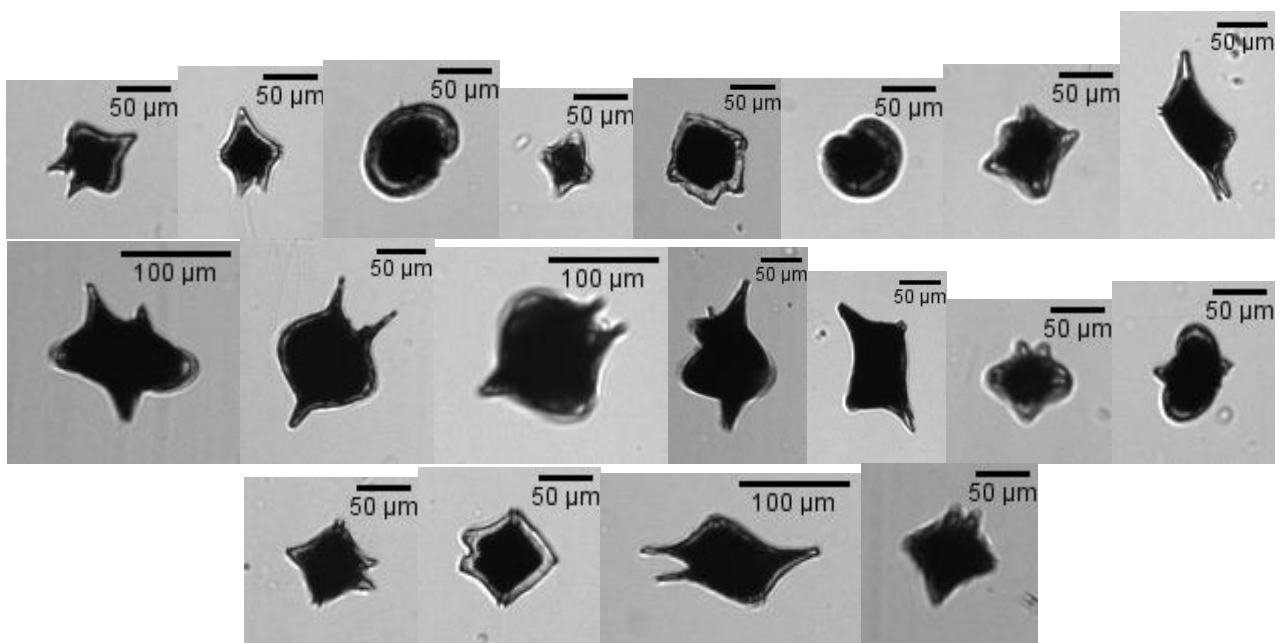
Myzozoa\Dinophyceae\Peridiniaceae Protoperidiniaceae
 \Peridinium Protoperidinium



BL



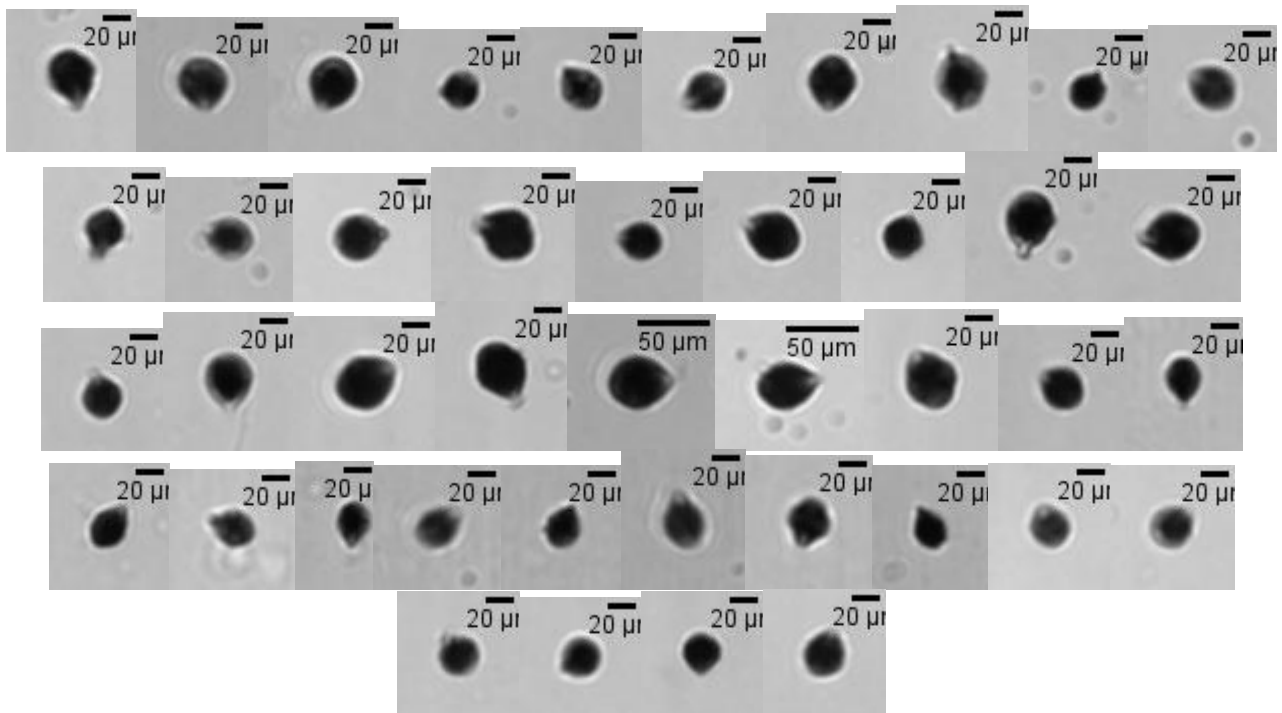
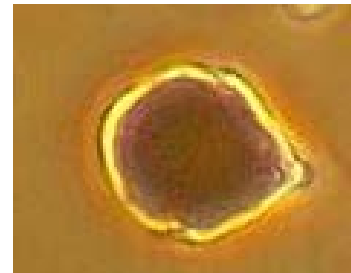
NT



AR

5.15. *Scrippsiella* *Ensiculifera* *Pentapharsodinium* *Bysmatrum*

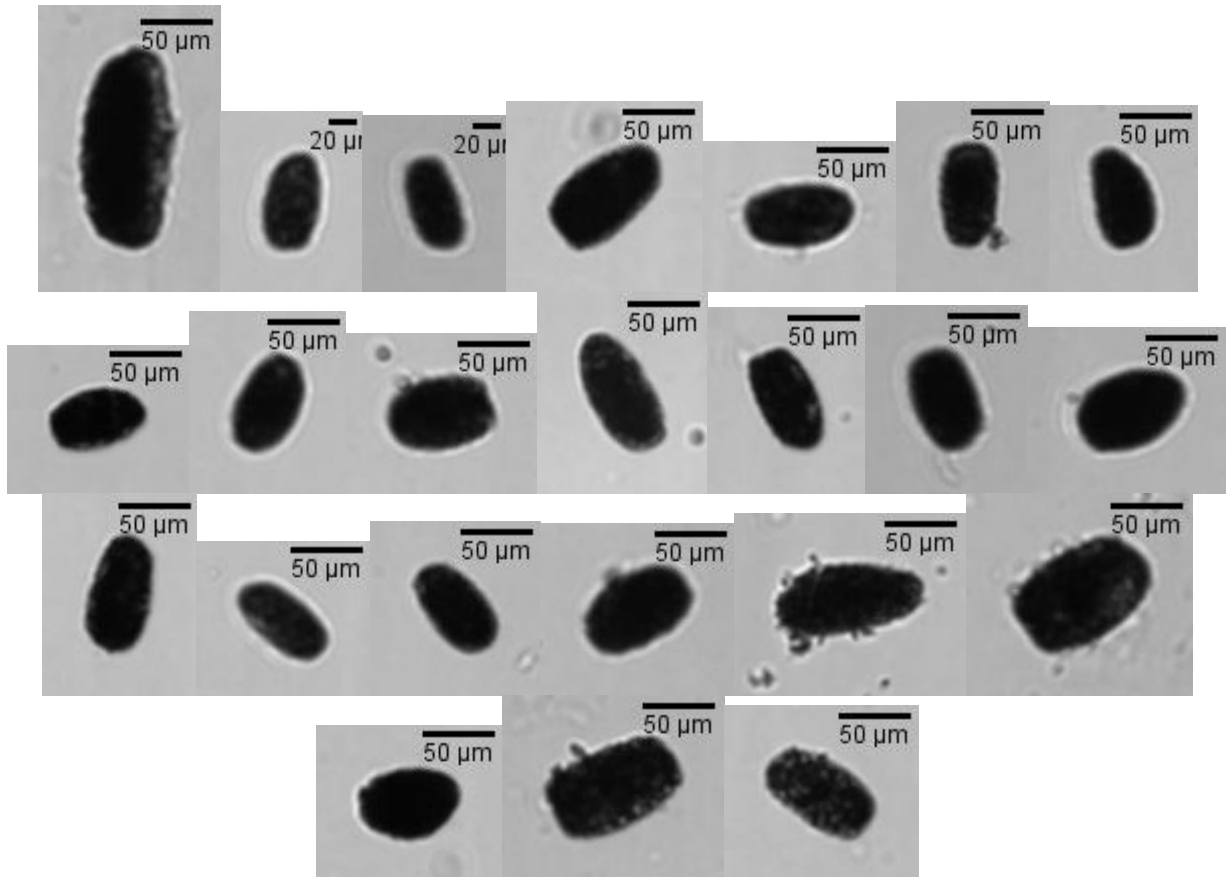
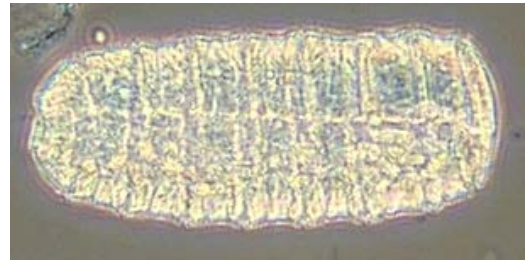
Myzozoa\Dinophyceae\Peridiniaceae Protoperidiniaceae\Scrippsiella Ensiculifera
Pentapharsodinium Bysmatrum



NT

5.16. *Polykrikos*

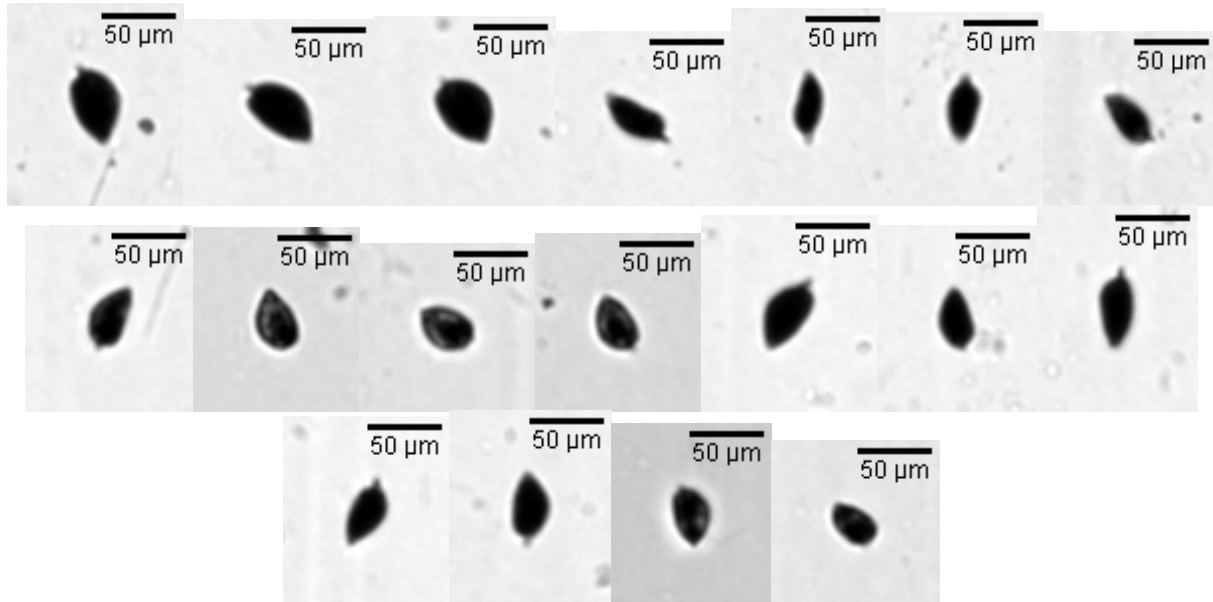
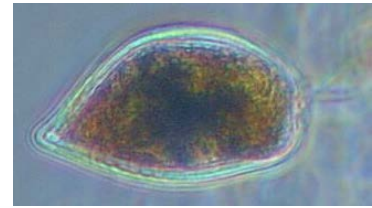
Myzozoa\Dinophyceae\Polykrikaceae\Polykrikos



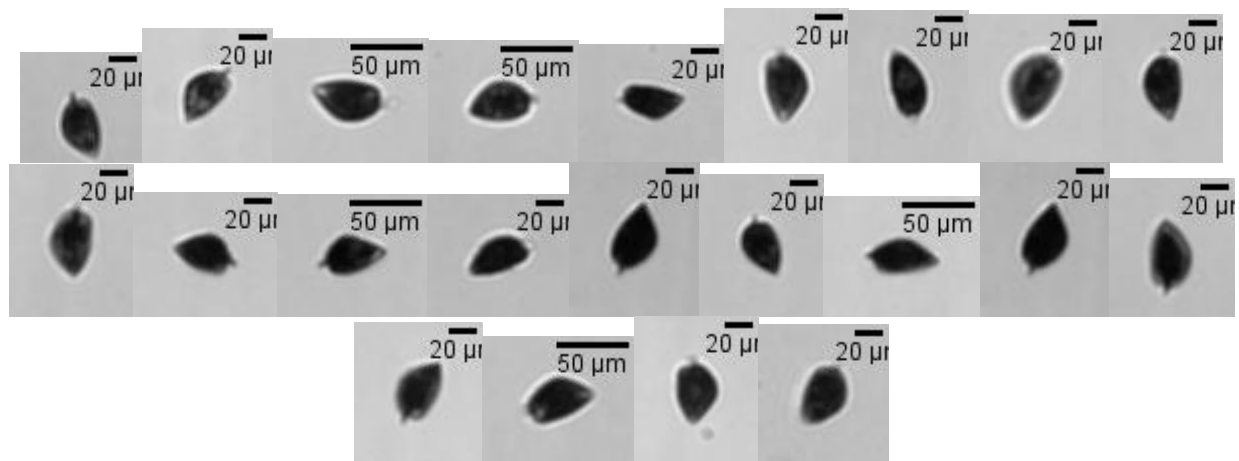
NT

5.17. Prorocentrum

Myzozoa\Dinophyceae\Prorocentraceae\Prorocentrum



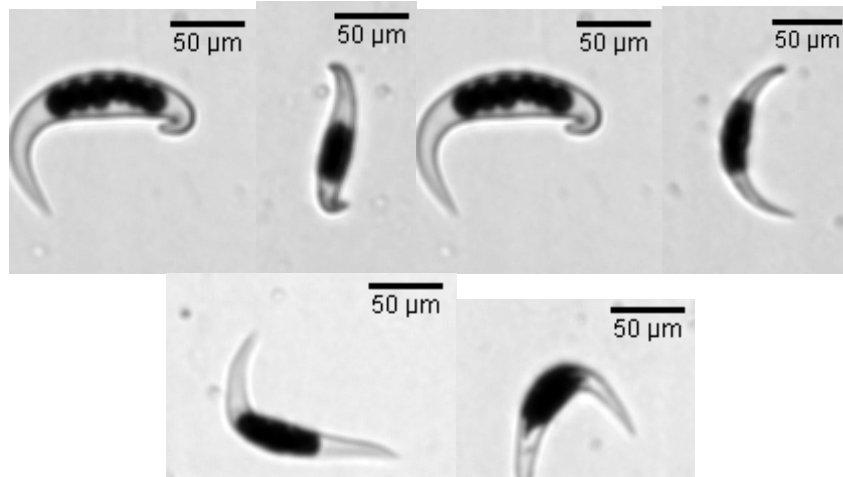
BL



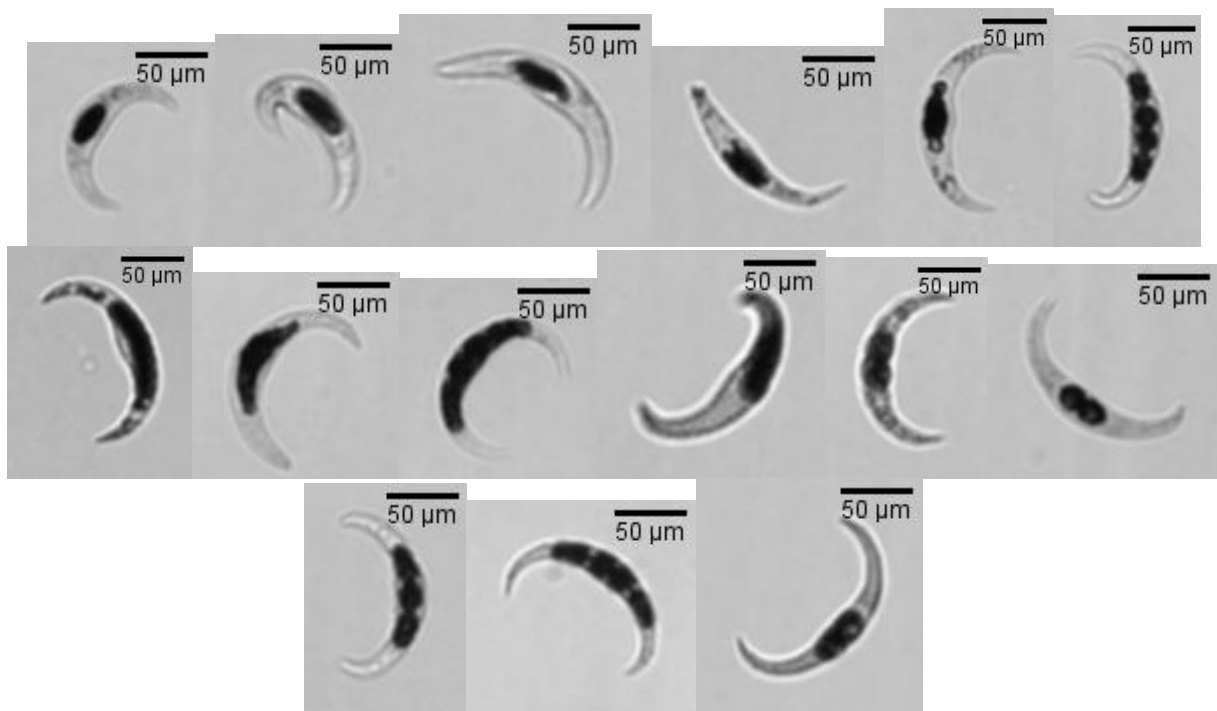
NT

5.18. *Dissodinium Pyrocystis*

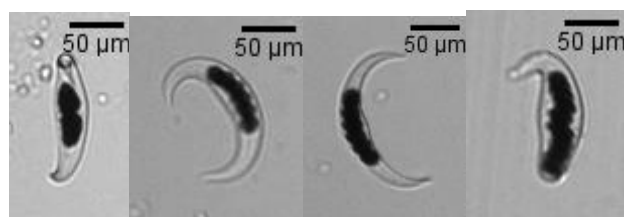
Myzozoa\Dinophyceae\Pyrocystaceae\Dissodinium Pyrocystis



BL



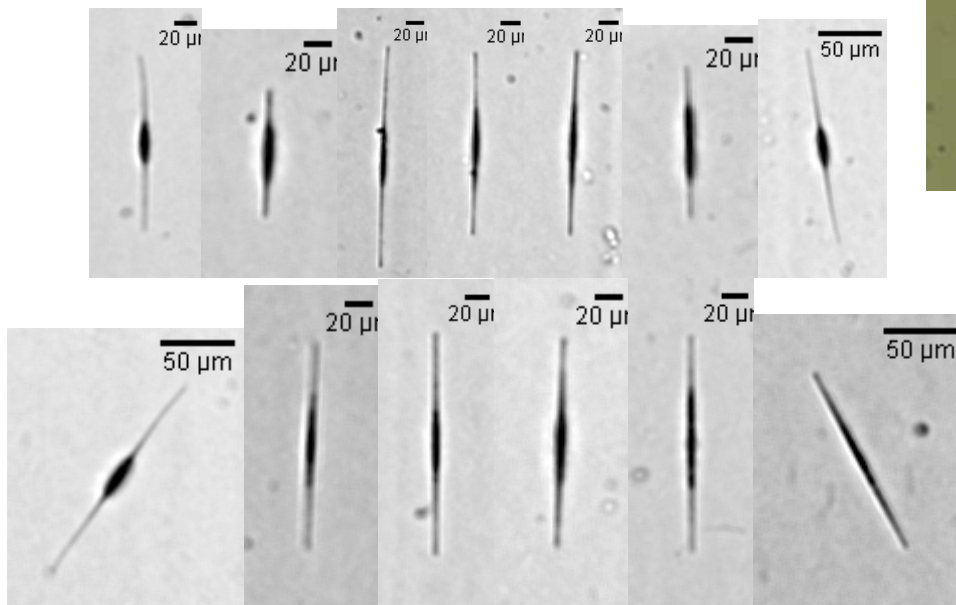
NT



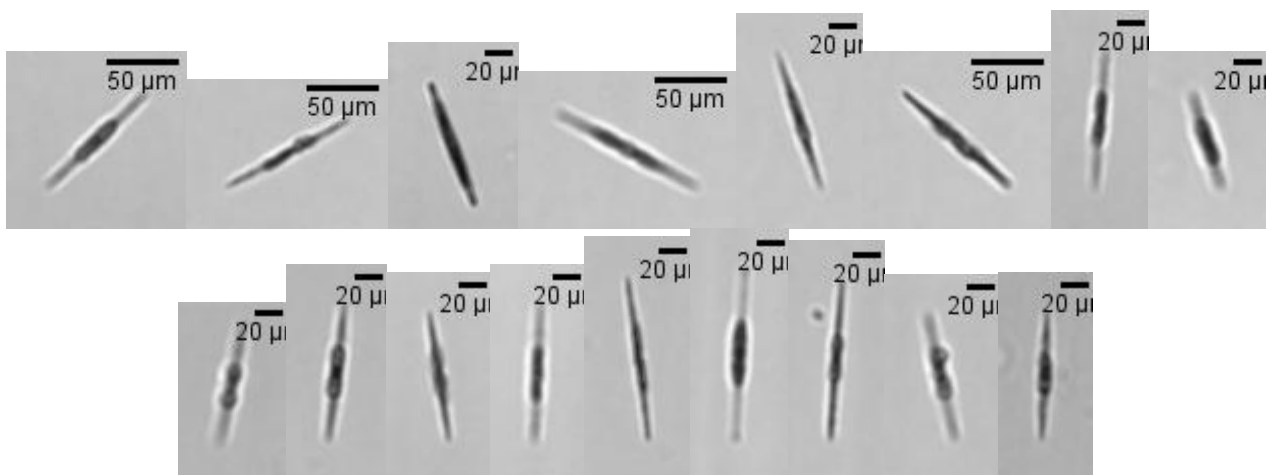
AR

5.19. *Cylindrotheca Hantzschia Nitzschia*

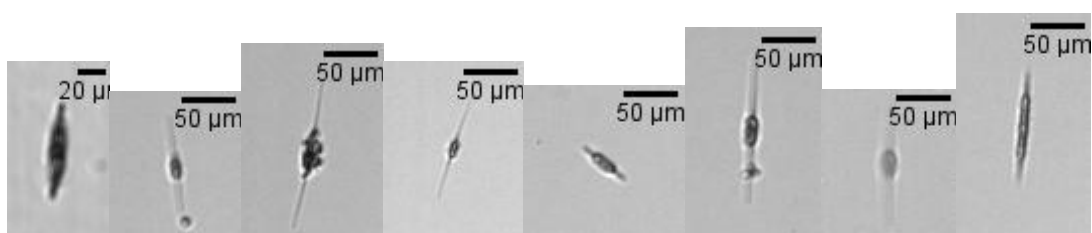
Ochrophyta\Bacillariophyceae\Bacillariaceae\Cylindrotheca Hantzschia Nitzschia



BL



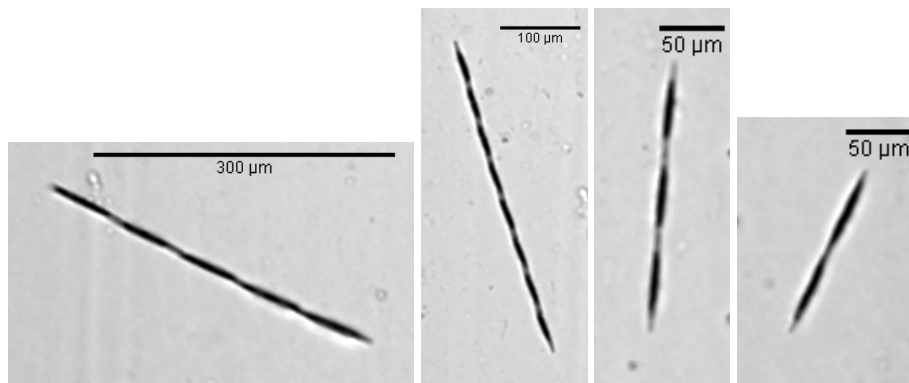
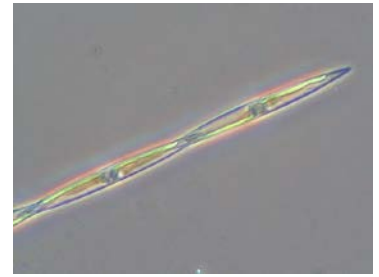
NT



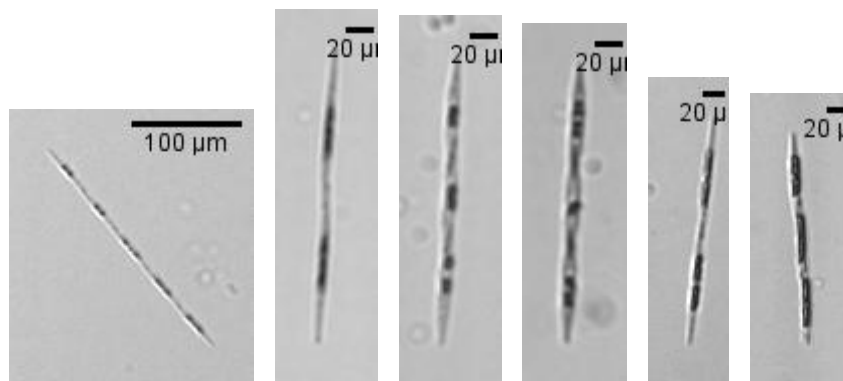
AR

5.20. *Pseudo-nitzschia*

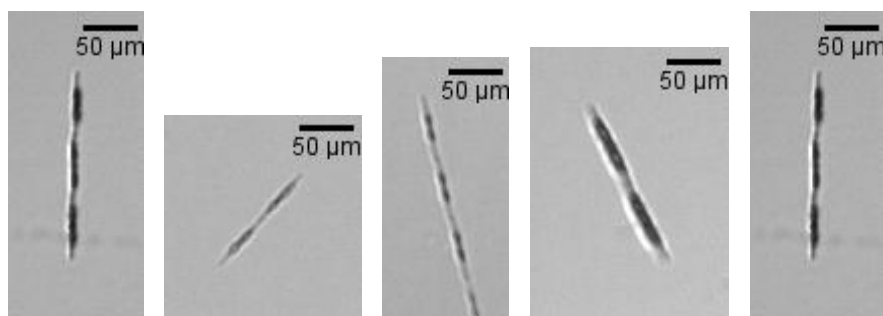
Ochrophyta\Bacillariophyceae\Bacillariaceae\Pseudo-nitzschia



BL



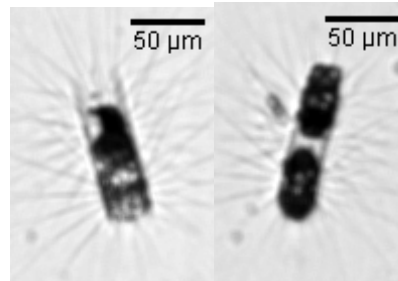
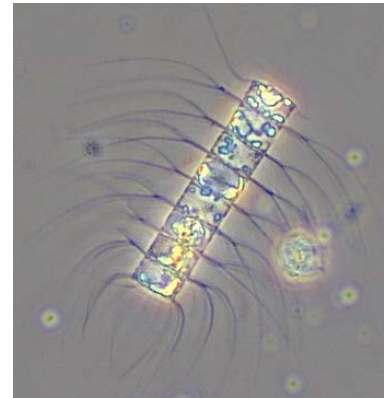
NT



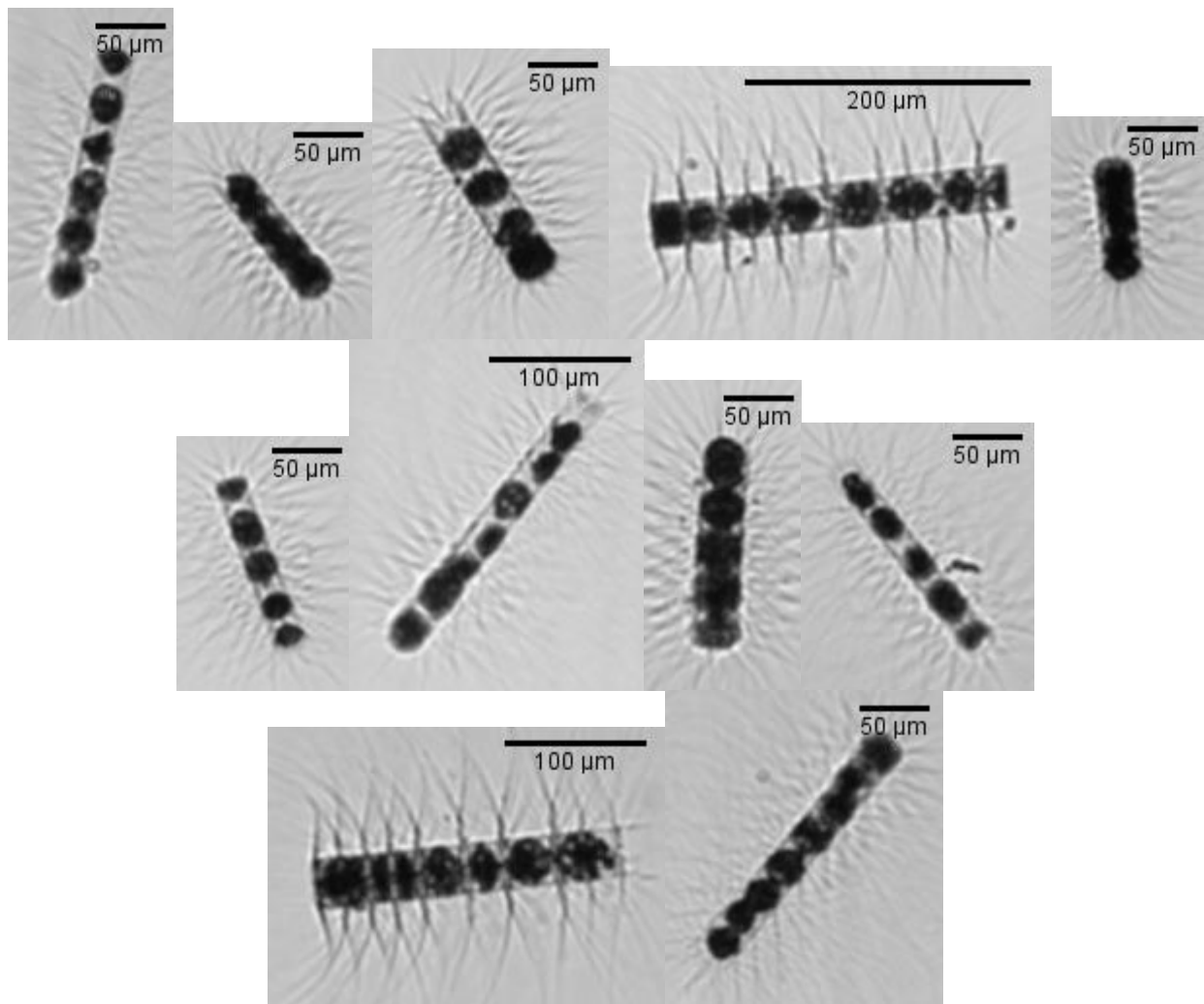
AR

5.21. *Bacteriastrium*

Ochrophyta \ Bacillariophyceae \ Chaetocerotaceae \ *Bacteriastrium*



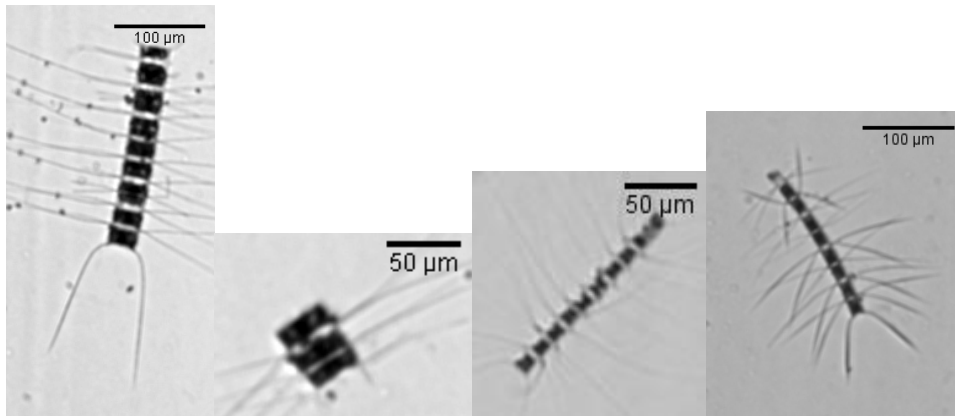
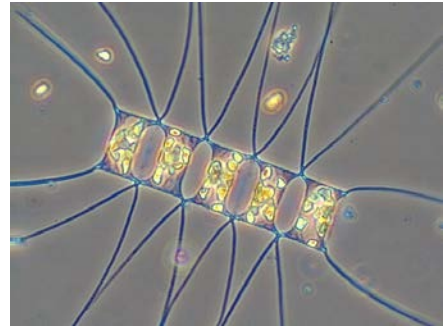
BL



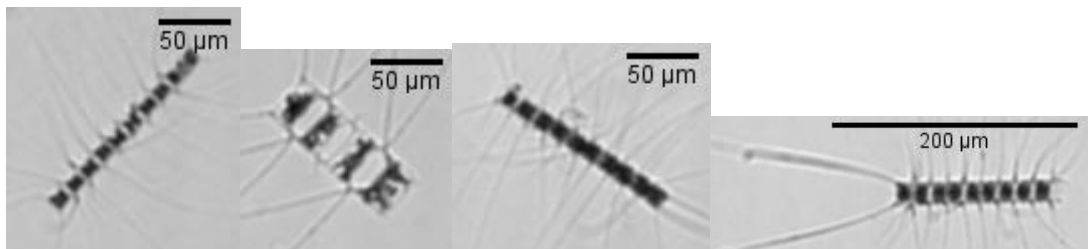
NT

5.22. Chaetoceros

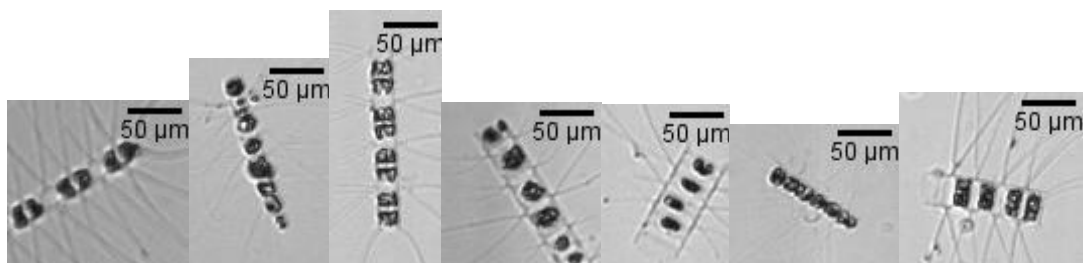
Ochrophyta\Bacillariophyceae\Chaetocerotaceae\Chaetoceros



BL



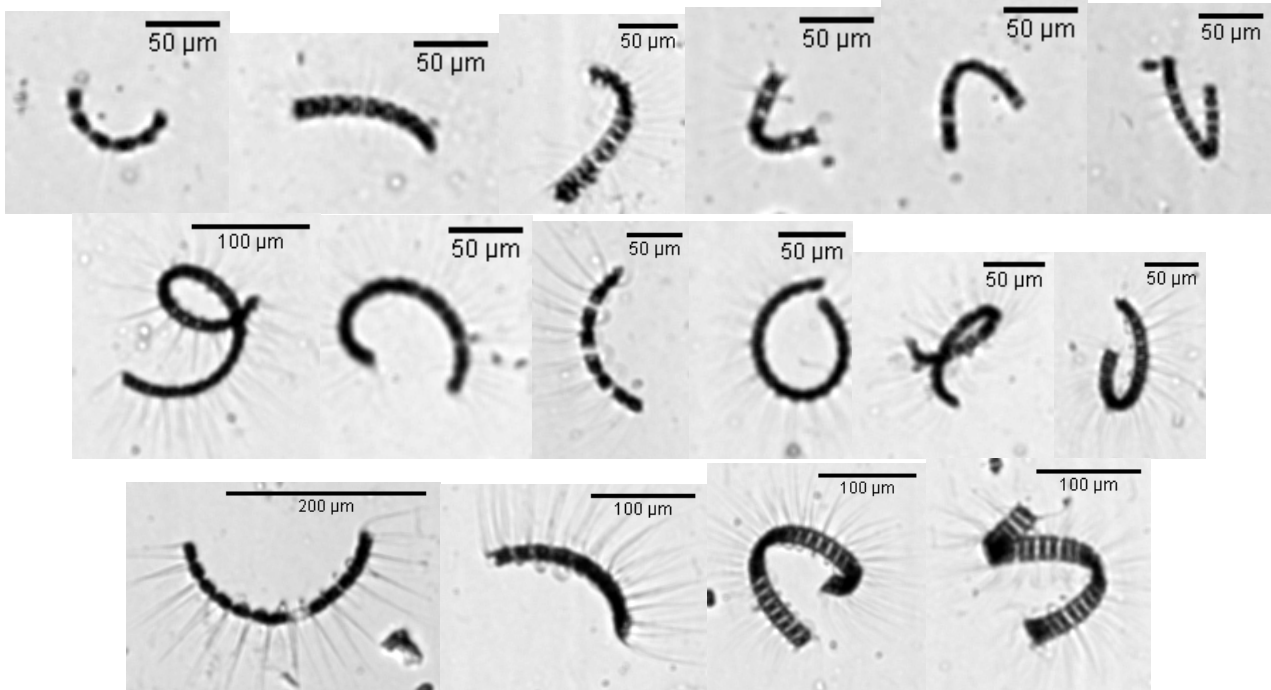
NT



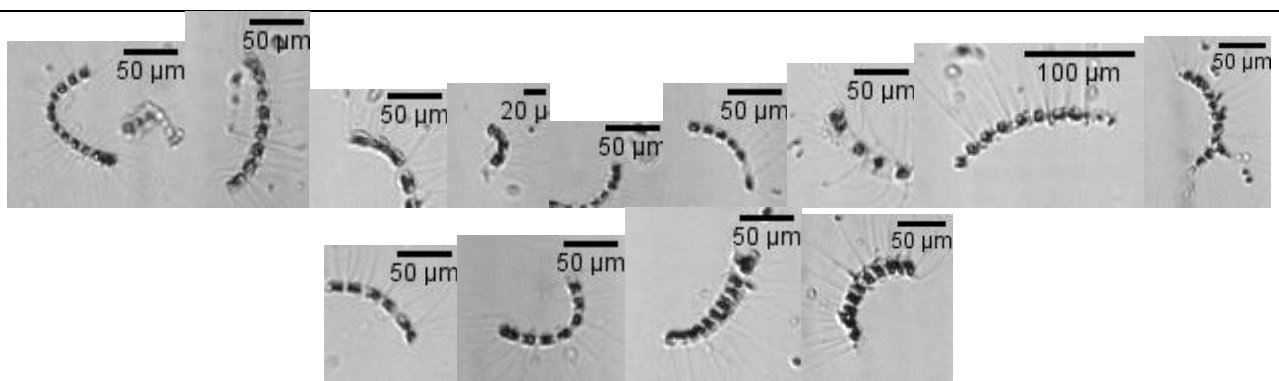
AR

5.23. *Chaetoceros curvisetus*

Ochrophyta\Bacillariophyceae\Chaetocerotaceae\Chaetoceros curvisetus



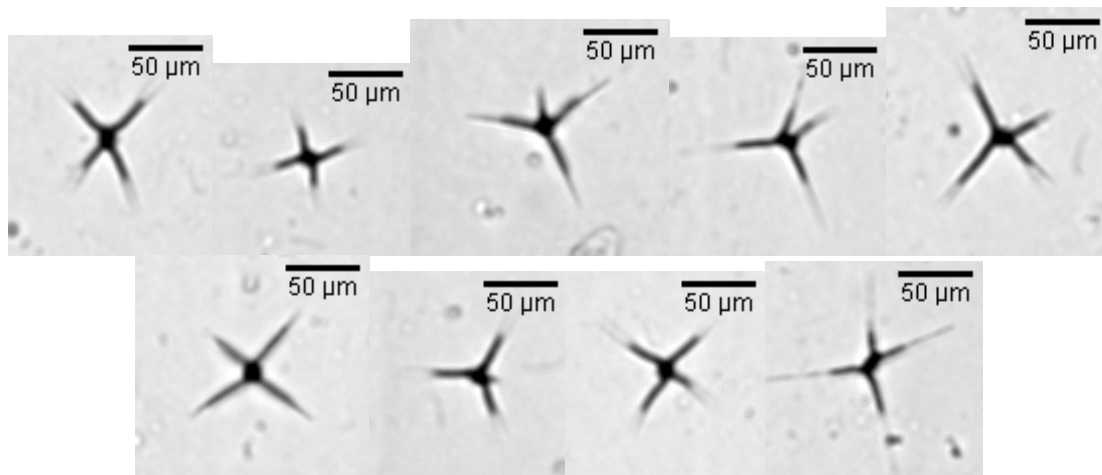
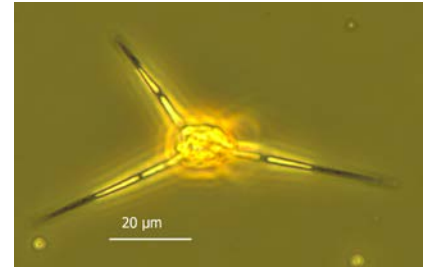
BL



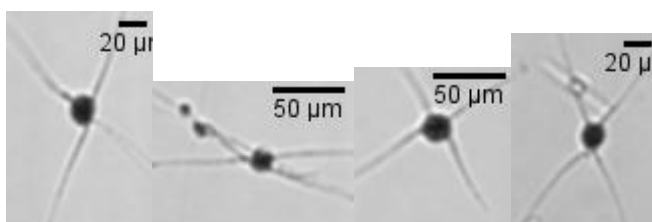
NT

5.24. *Chaetoceros danicus*

Ochrophyta\Bacillariophyceae\Chaetocerotaceae\Chaetoceros danicus



BL

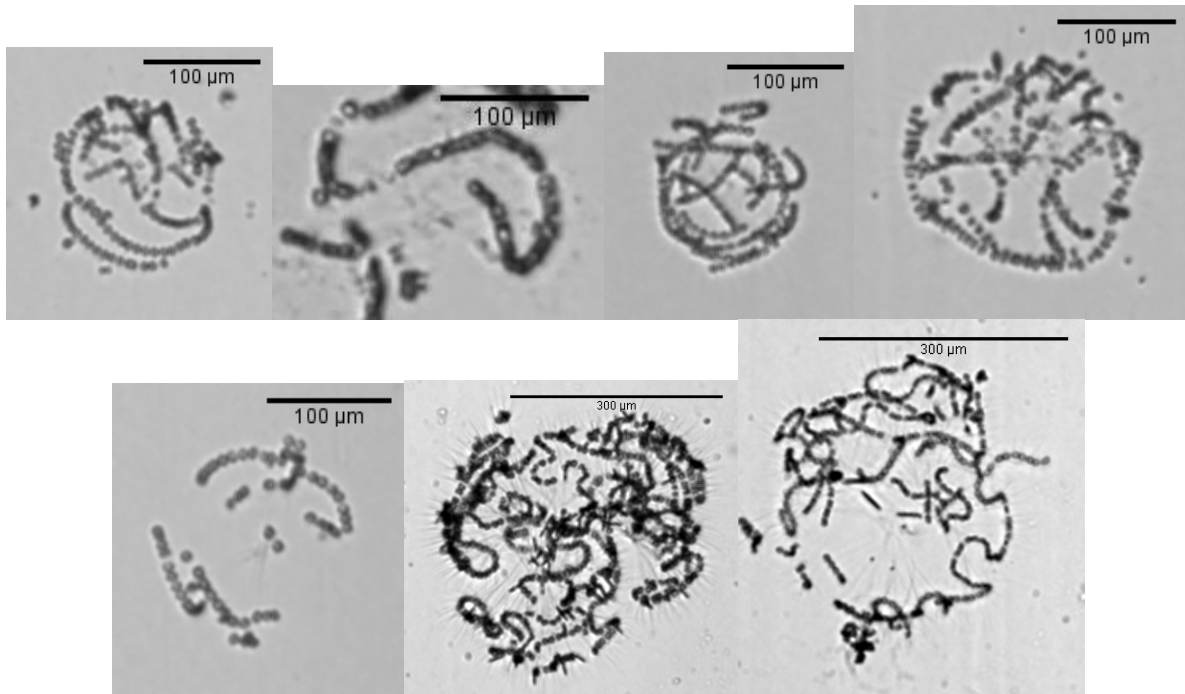


NT

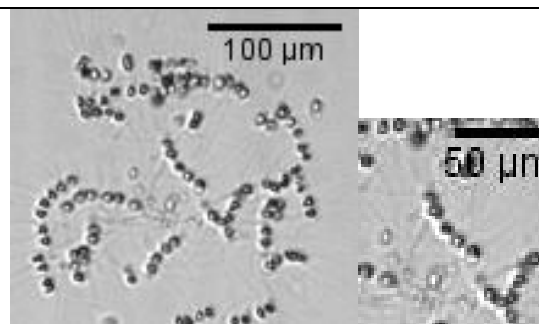


5.25. *Chaetoceros socialis*

Ochrophyta\Bacillariophyceae\Chaetocerotaceae\Chaetoceros socialis



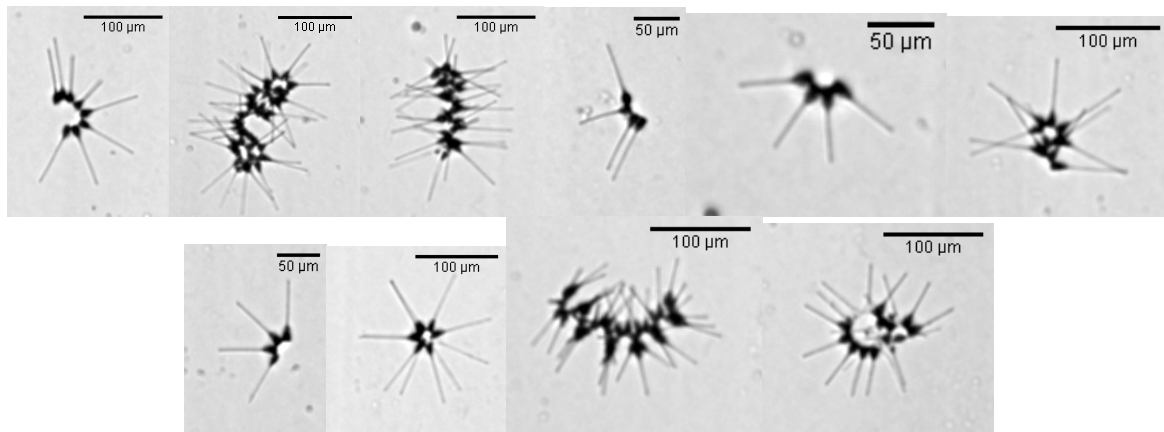
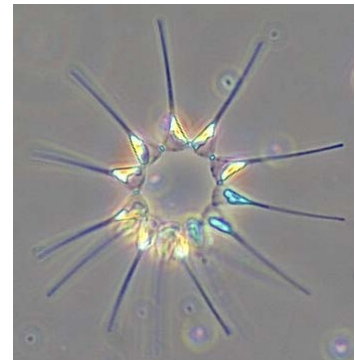
BL



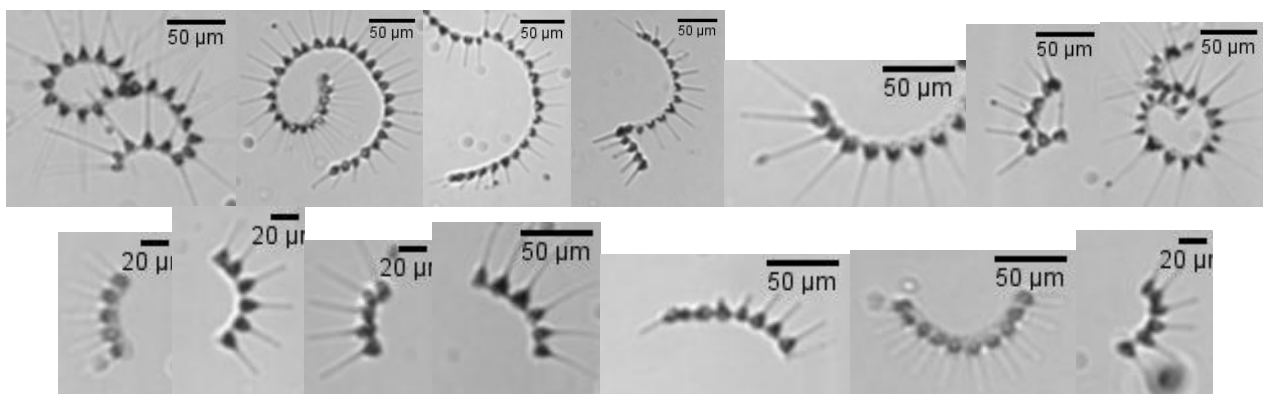
NT

5.26. *Asterionellopsis glacialis*

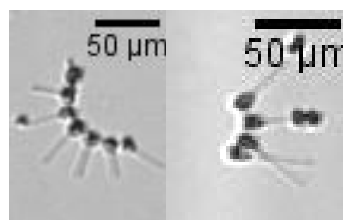
Ochrophyta\Bacillariophyceae\Fragilariaceae\Asterionellopsis glacialis



BL



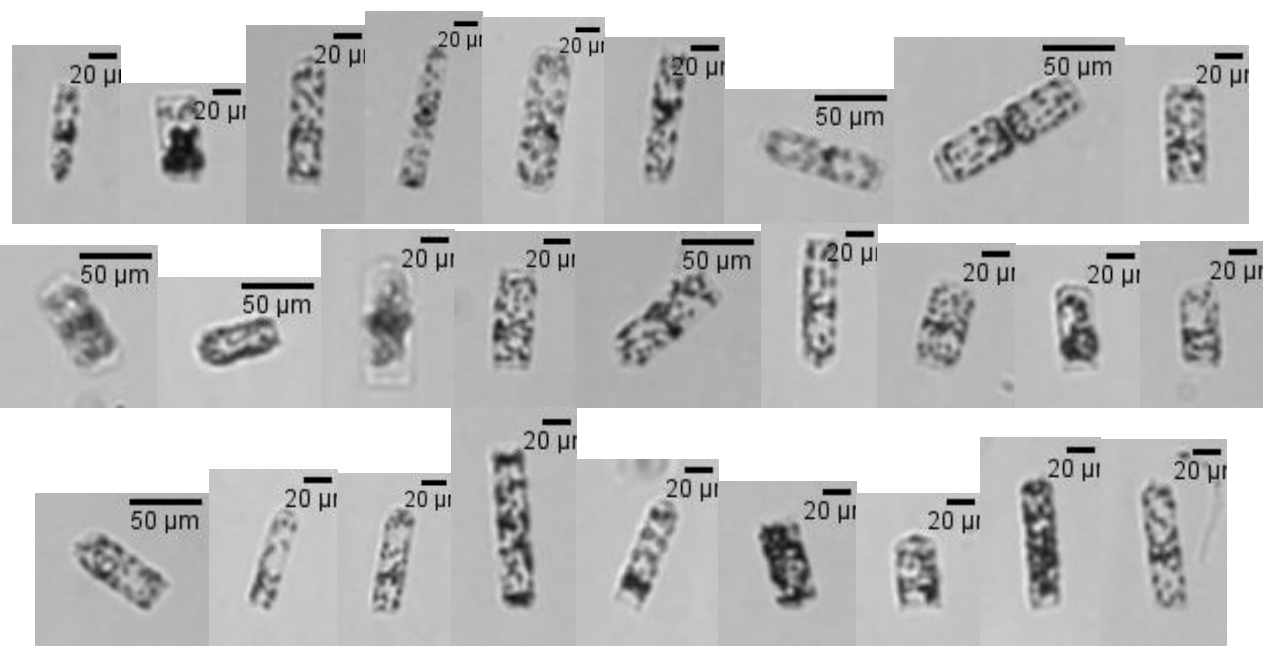
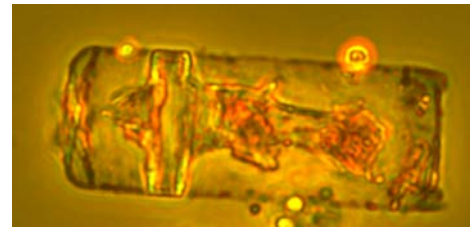
NT



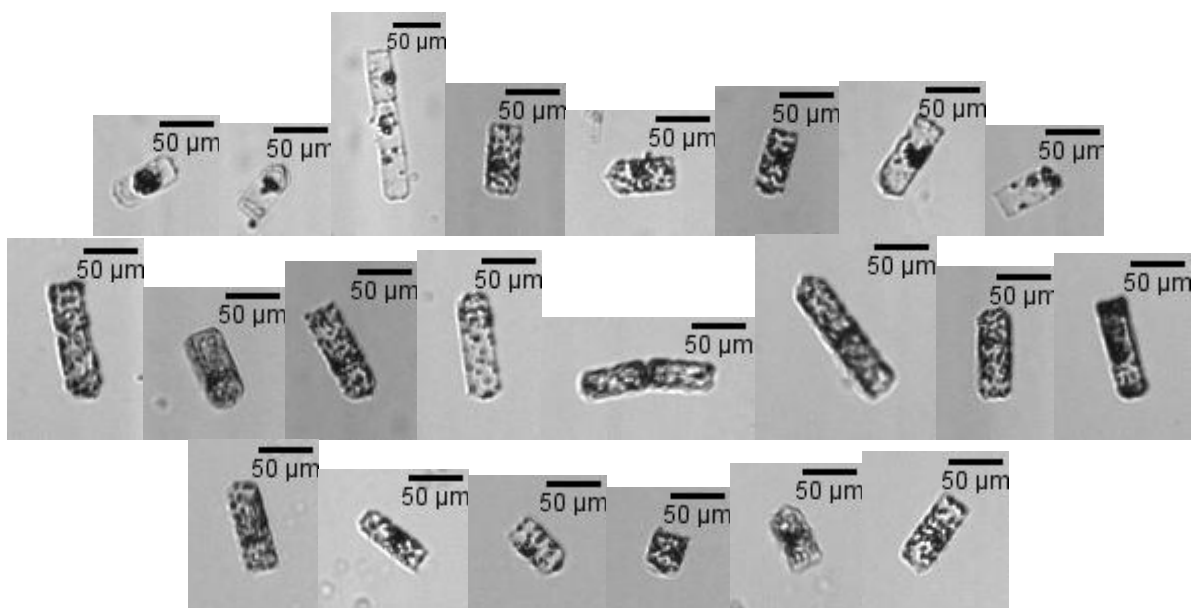
AR

5.27. *Cerataulina*

Ochrophyta\Bacillariophyceae\Hemiaulaceae\Cerataulina



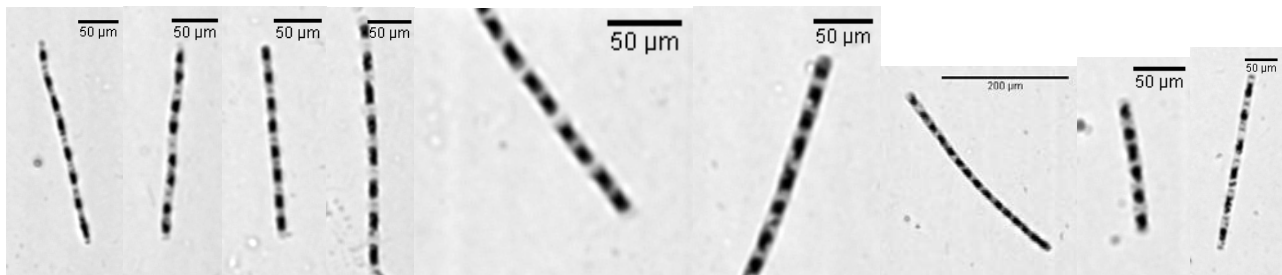
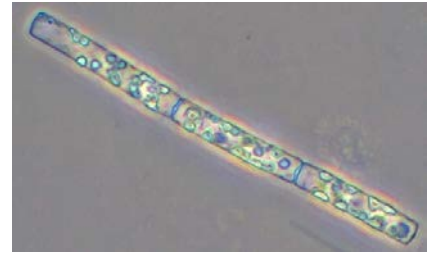
NT



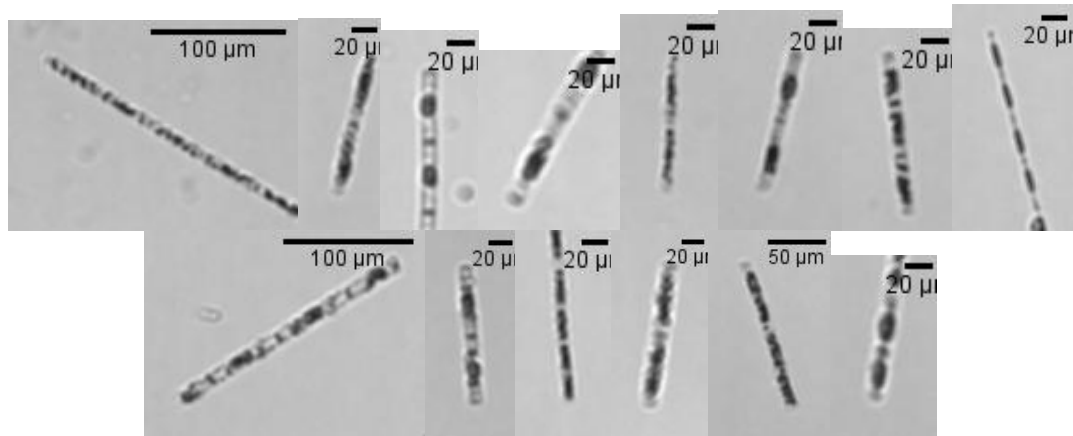
AR

5.28. *Leptocylindrus*

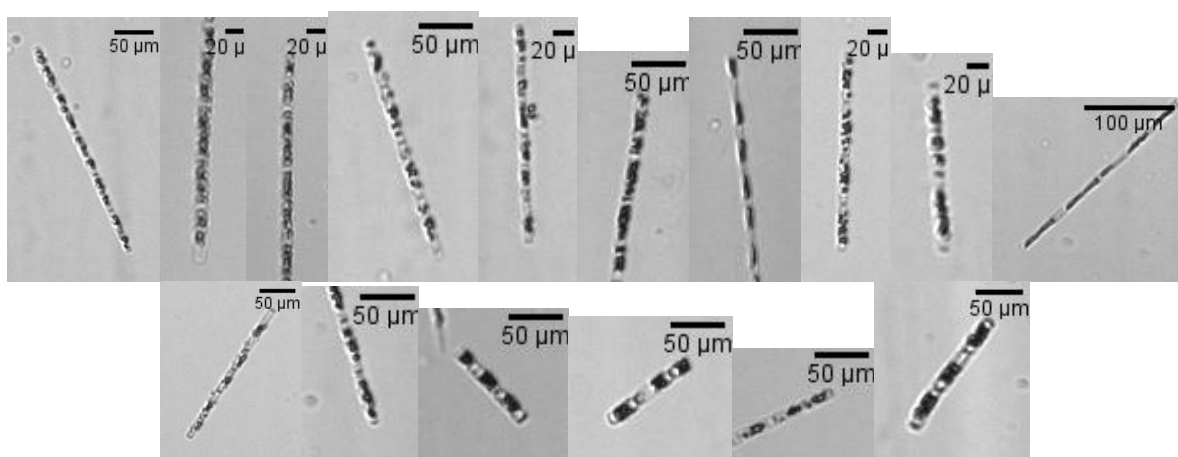
Ochrophyta\Bacillariophyceae\Leptocylindraceae\Leptocylindrus



BL



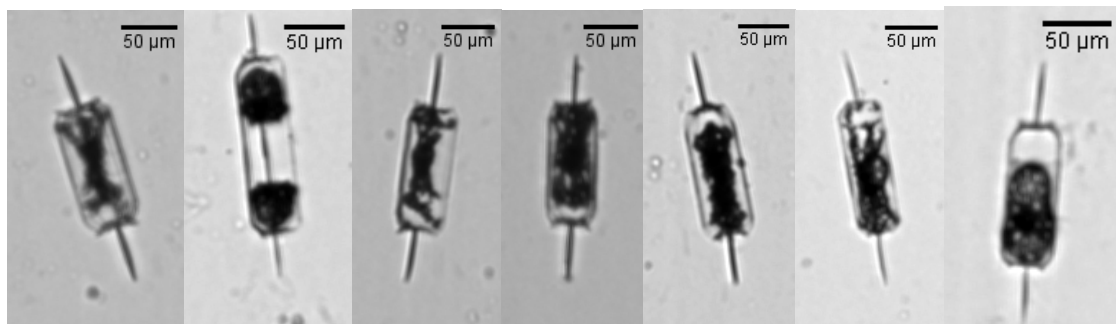
NT



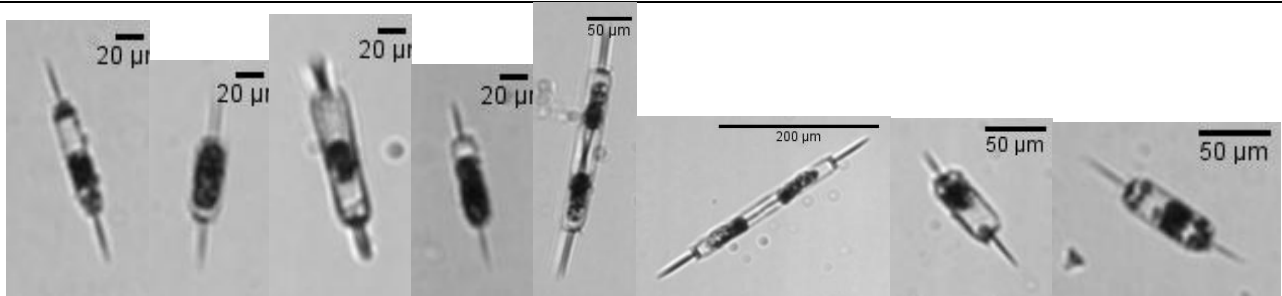
AR

5.29. *Ditylum brightwellii*

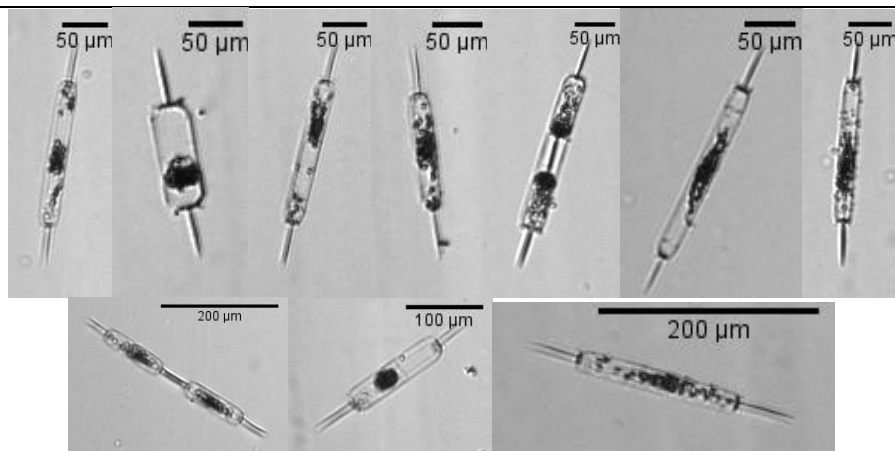
Ochrophyta\Bacillariophyceae\Lithodesmiaceae\Ditylum brightwellii



BL



NT

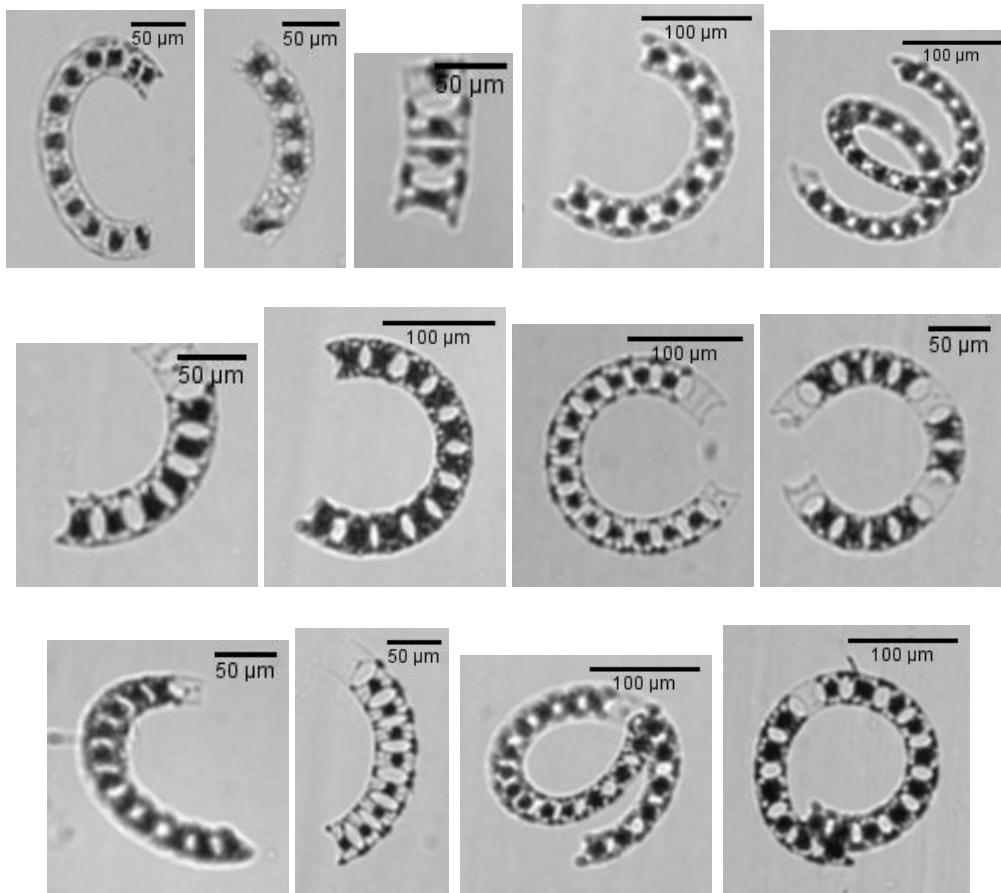


AR



5.30. *Eucampia*

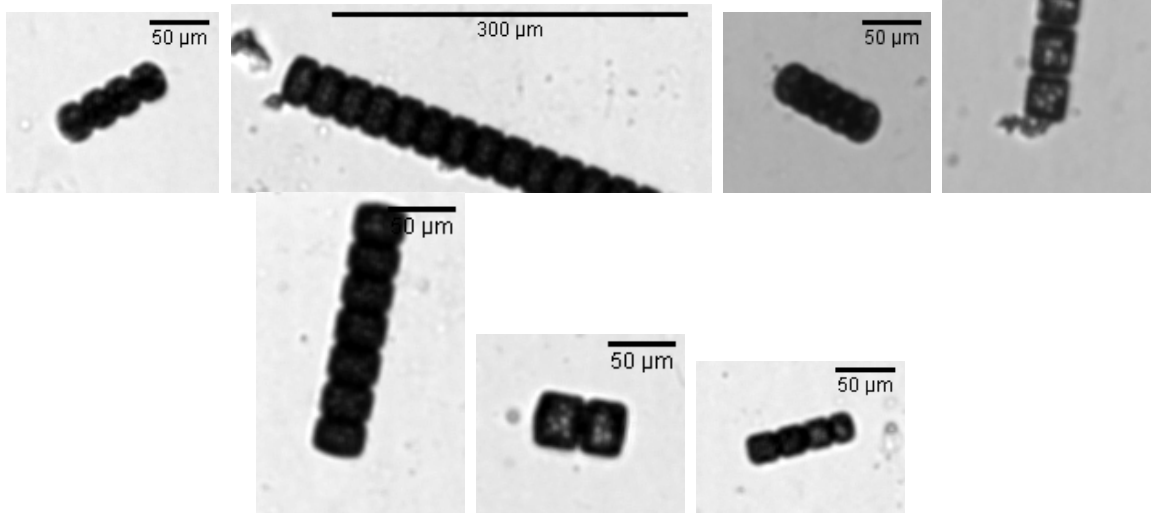
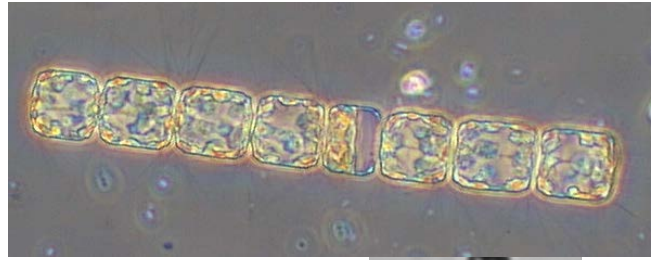
Ochrophyta \ Bacillariophyceae \ Hemiaulaceae \ *Eucampia*



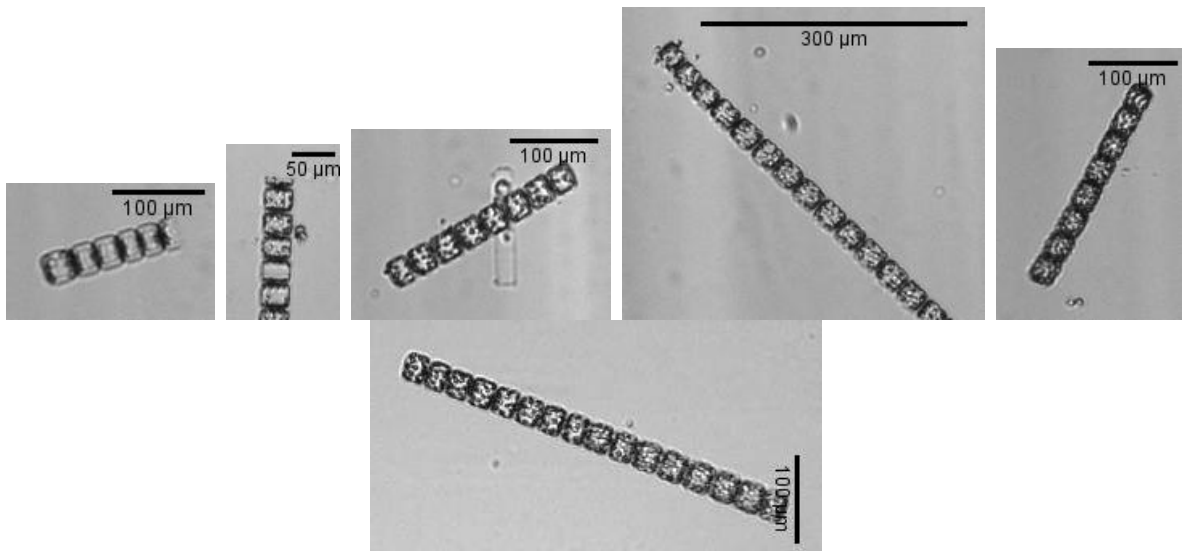
NT

5.31. *Lauderia*

Ochrophyta \ Bacillariophyceae \ Lauderiaceae \ *Lauderia*



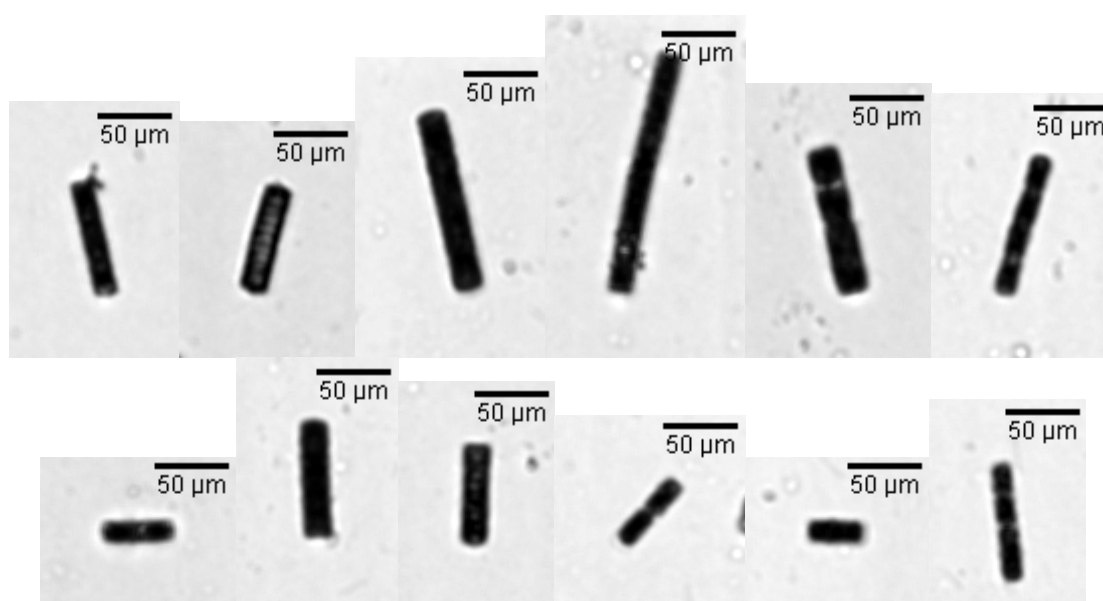
BL



NT

5.32. *Paralia*

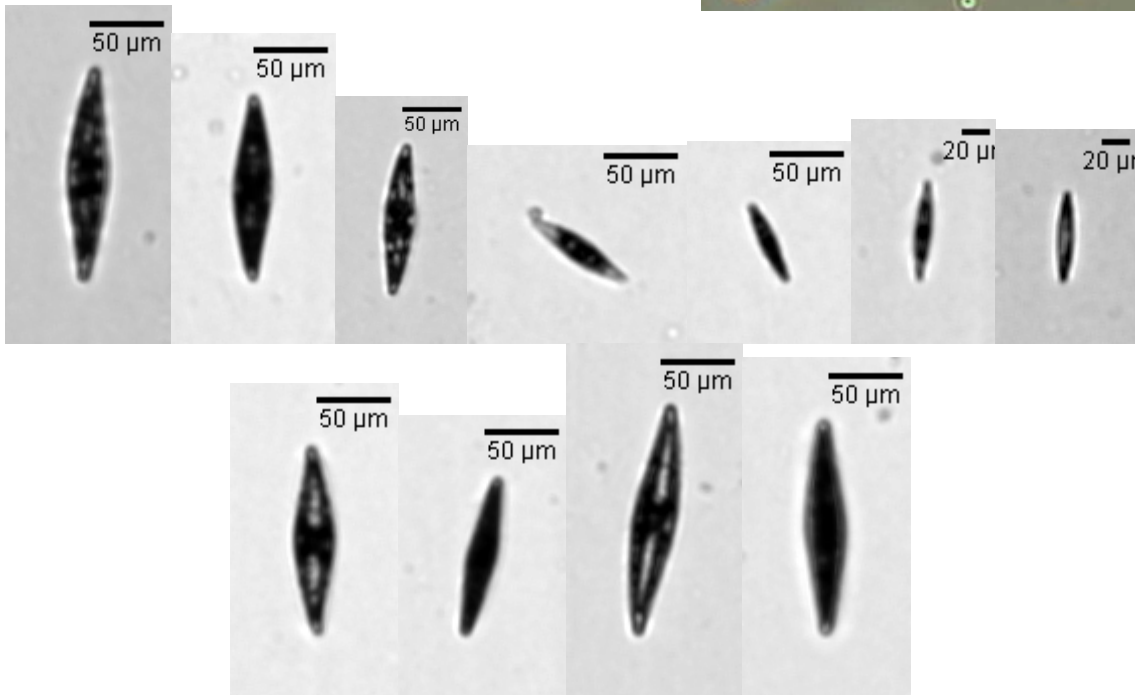
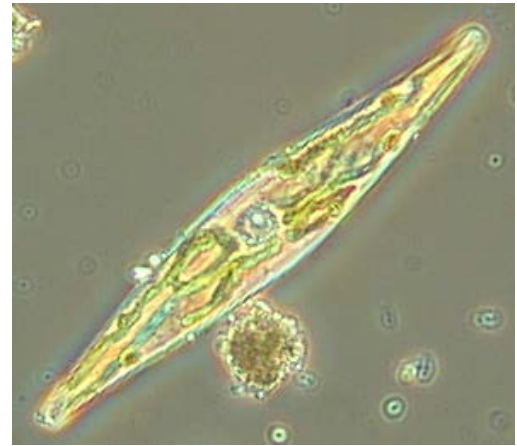
Ochrophyta\Bacillariophyceae\Paraliaceae\Paralia



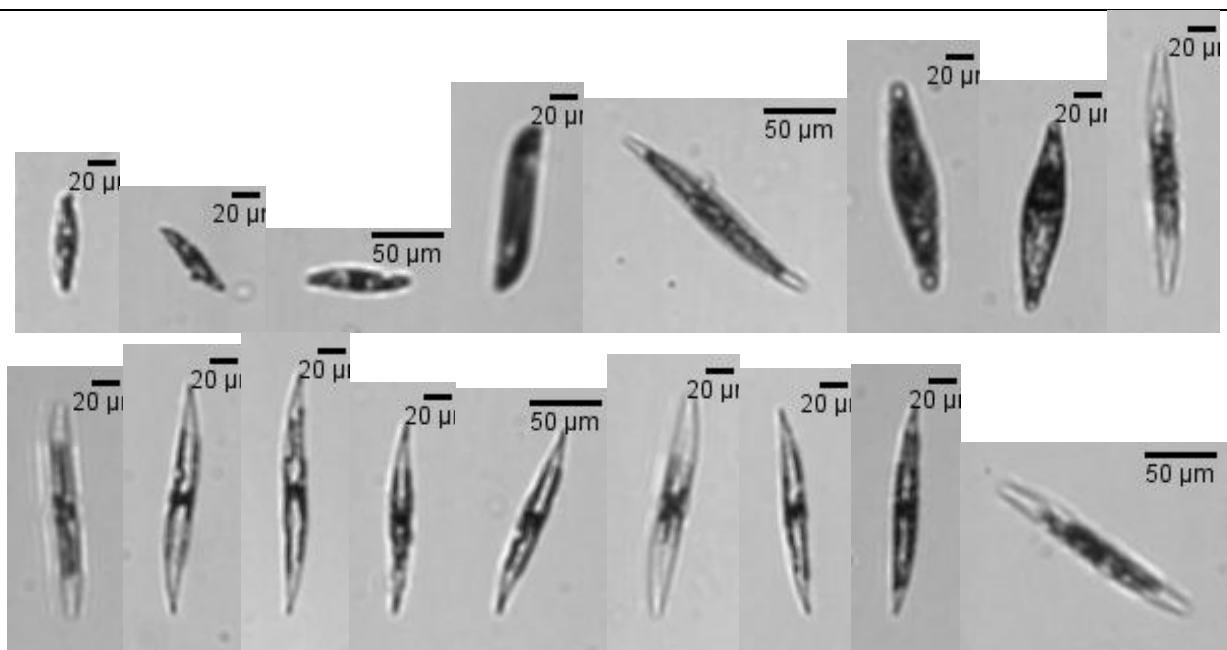
BL

5.33. *Gyrosigma Pleurosigma*

Ochrophyta\Bacillariophyceae\Pleurosigmataceae
\Gyrosigma Pleurosigma



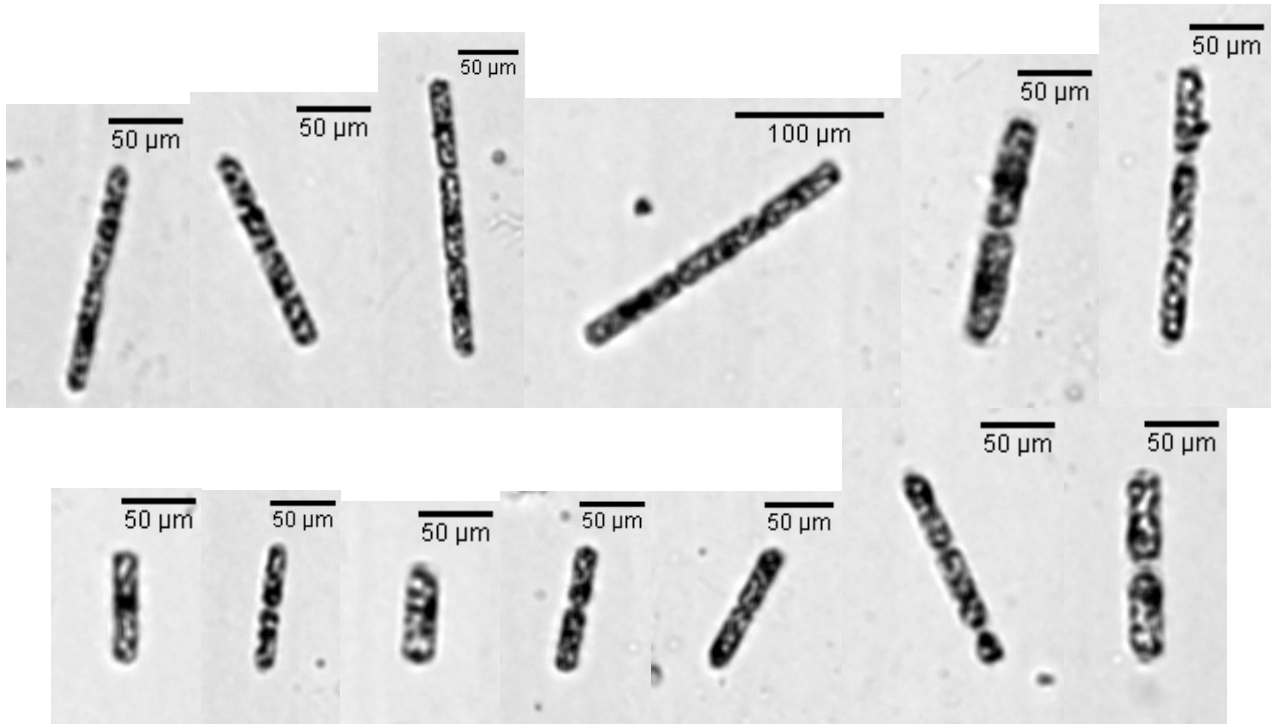
BL



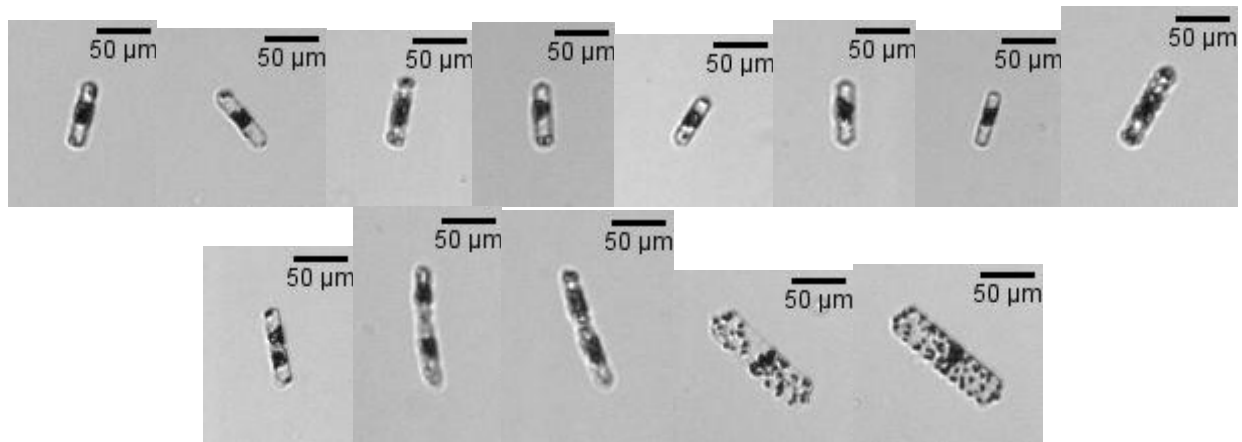
NT

5.34. *Dactyliosolen fragilissimus*

Ochrophyta\Bacillariophyceae\rhizosoleniaceae\Dactyliosolen fragilissimus



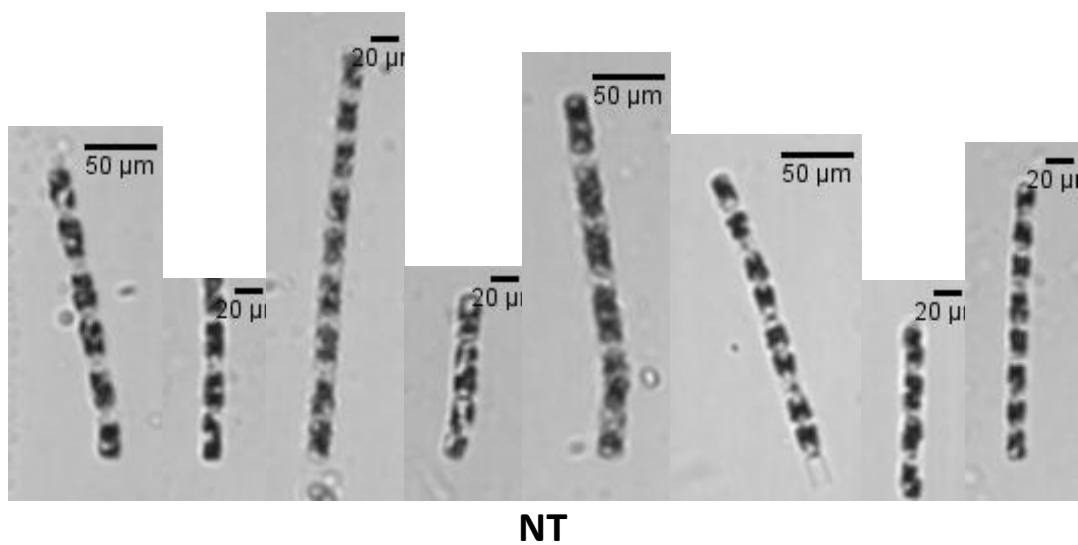
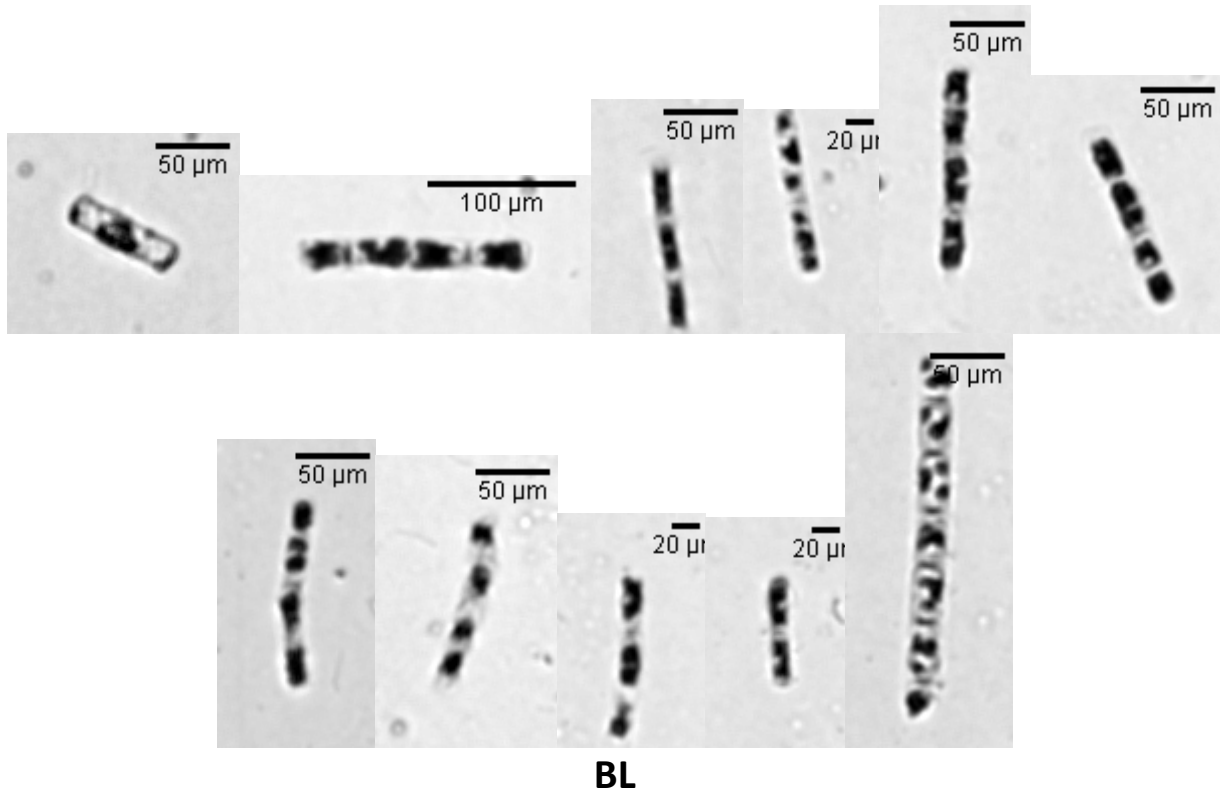
BL



AR

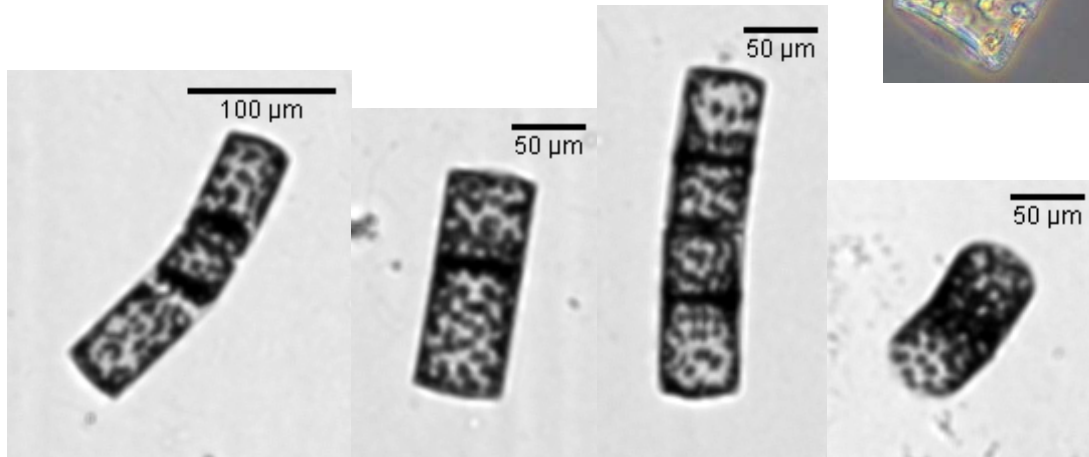
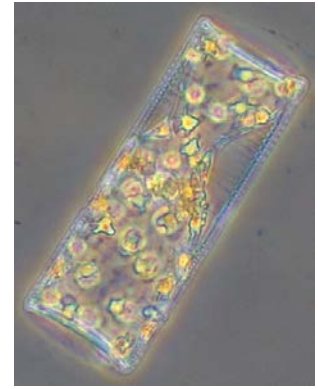
5.35. *Guinardia delicatula*

Ochrophyta\Bacillariophyceae\rhizosoleniaceae\Guinardia delicatula

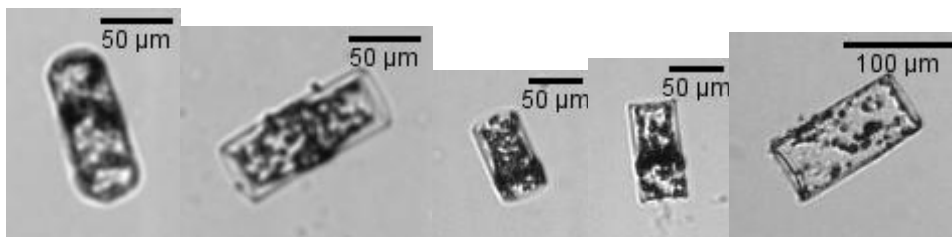


5.36. *Guinardia flaccida*

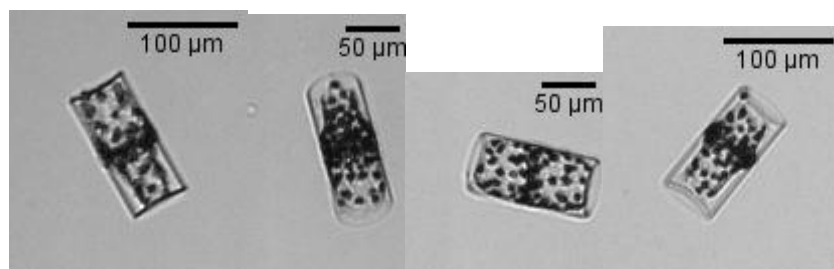
Ochrophyta\Bacillariophyceae\rhizosoleniaceae\Guinardia flaccida



BL



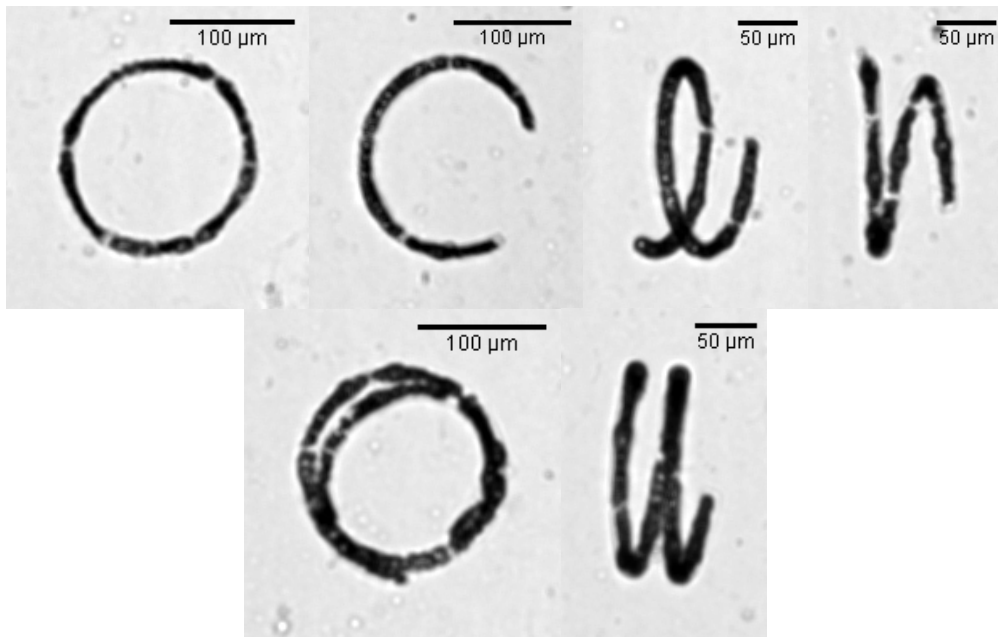
NT



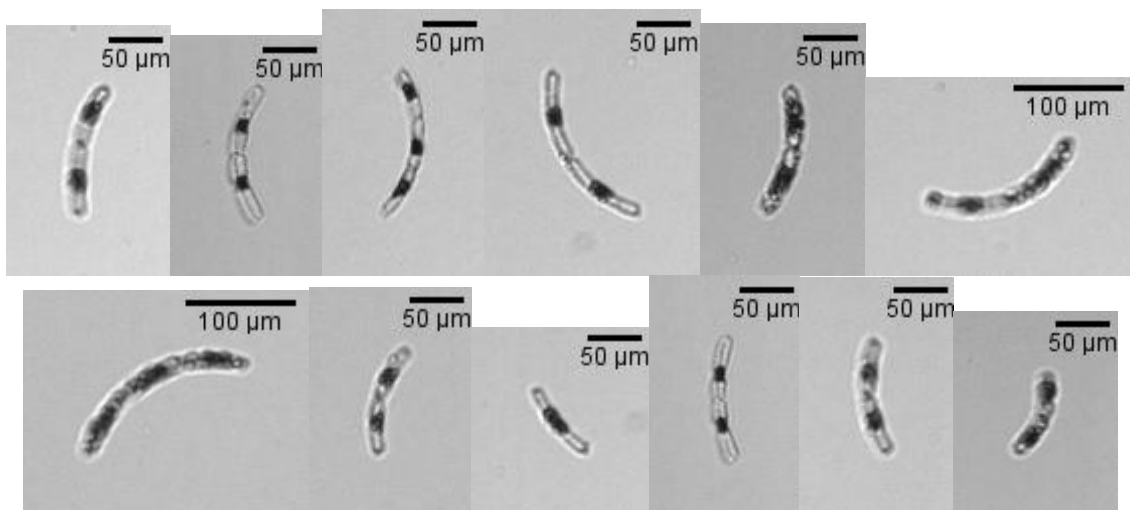
AR

5.37. *Guinardia striata*

Ochrophyta\Bacillariophyceae\Rhizosoleniaceae\Guinardia striata



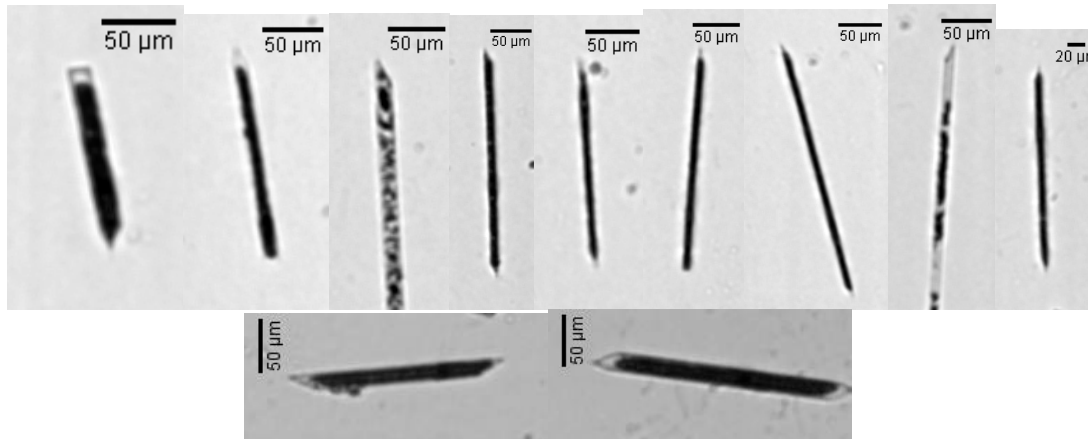
BL



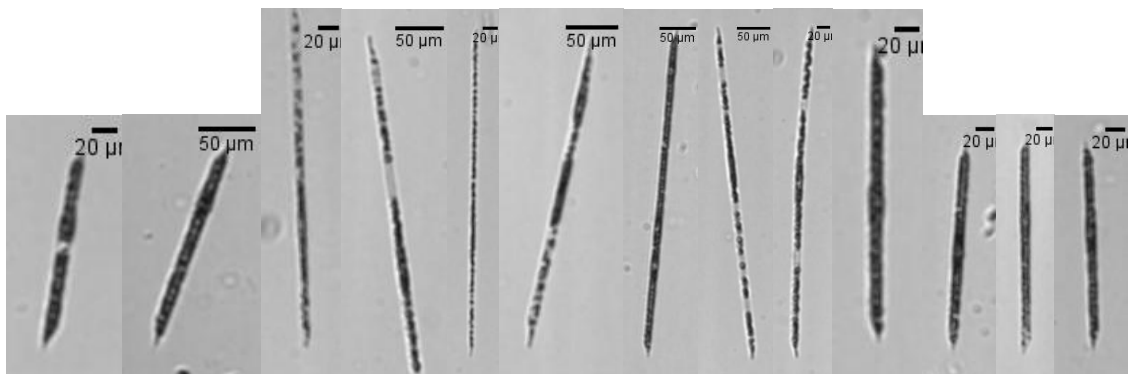
AR

5.38. *Proboscia Rhizolenia*

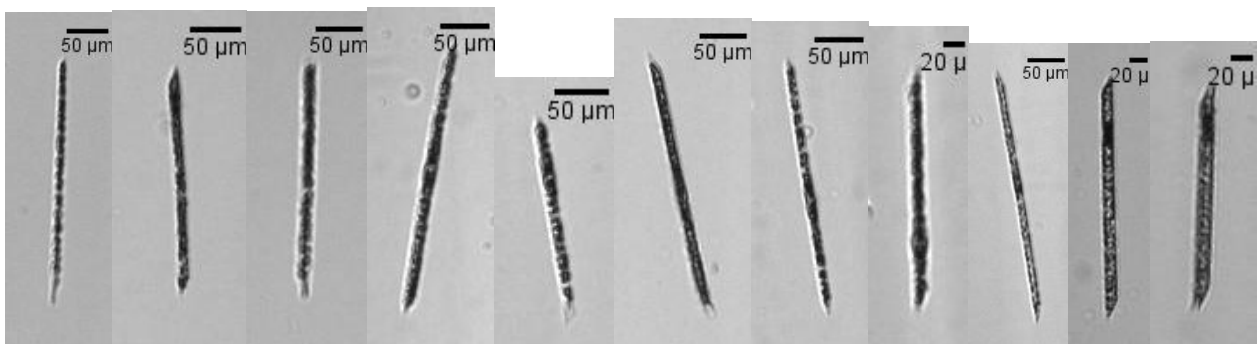
Ochrophyta\Bacillariophyceae\Rhizoleniaceae\Proboscia Rhizolenia



BL



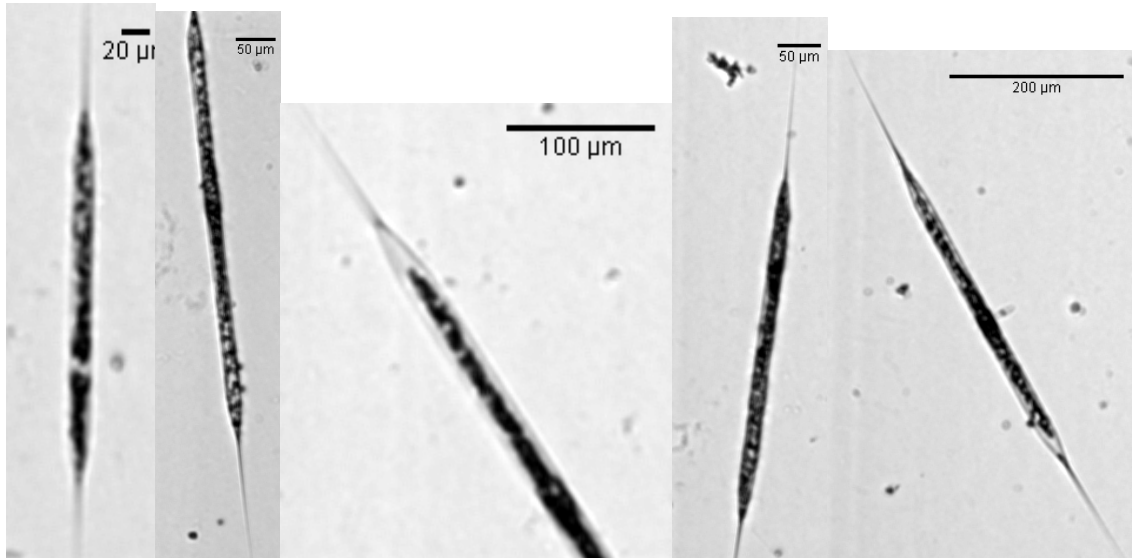
NT



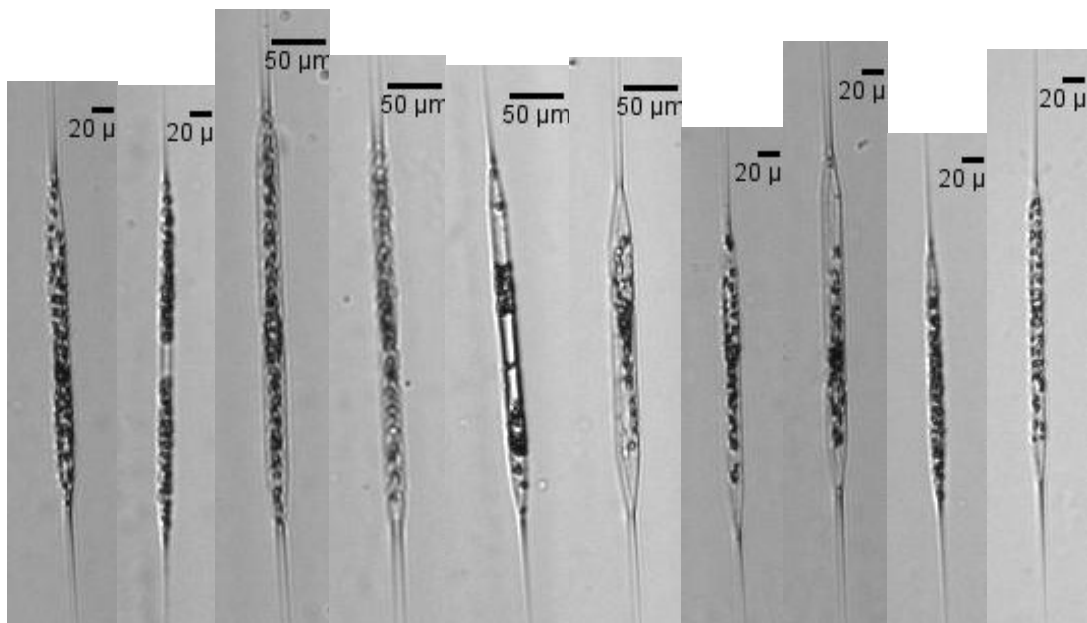
AR

5.39. *Rhizosolenia setigera*

Ochrophyta\Bacillariophyceae\rhizosoleniaceae\rhizosolenia setigera



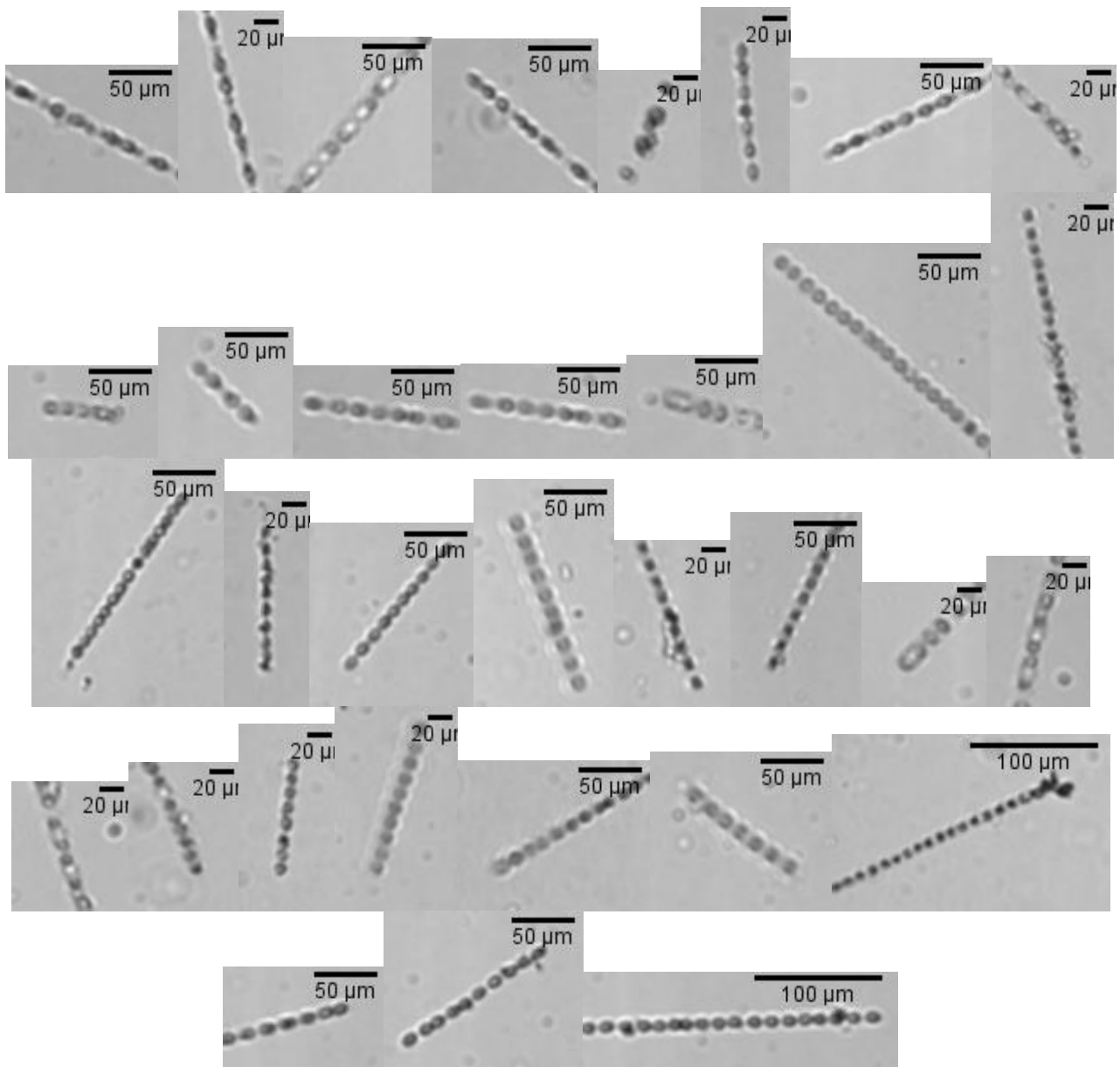
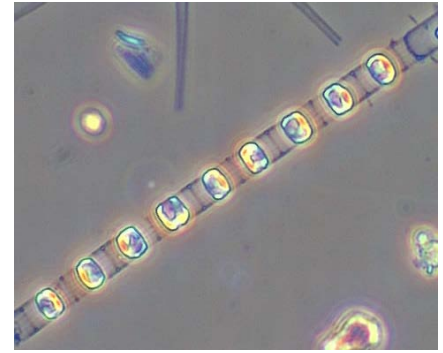
BL



AR

5.40. *Skeletonema*

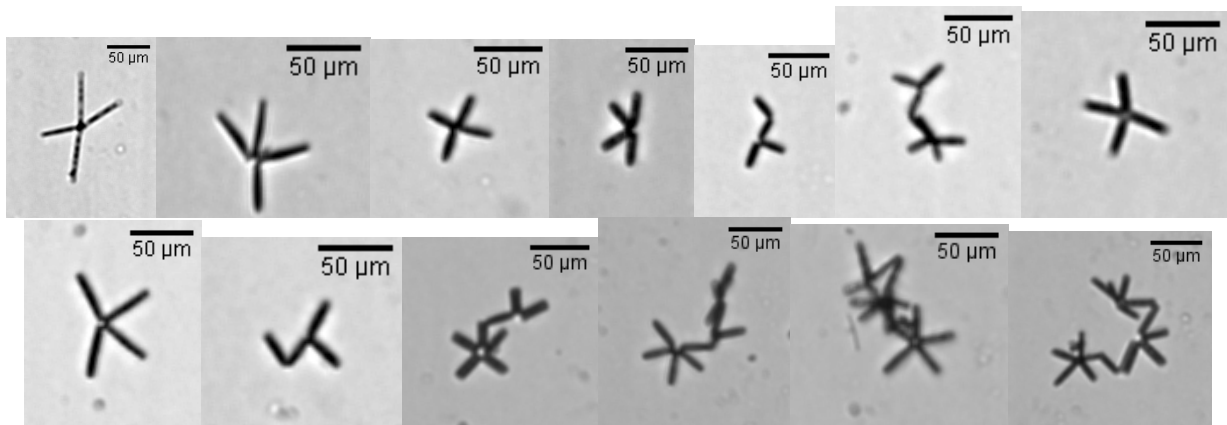
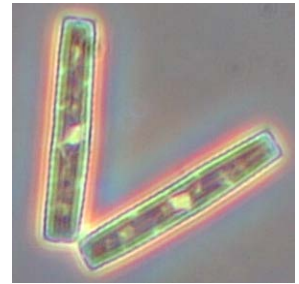
Ochrophyta\Bacillariophyceae\Skeletonemaceae\Skeletonema



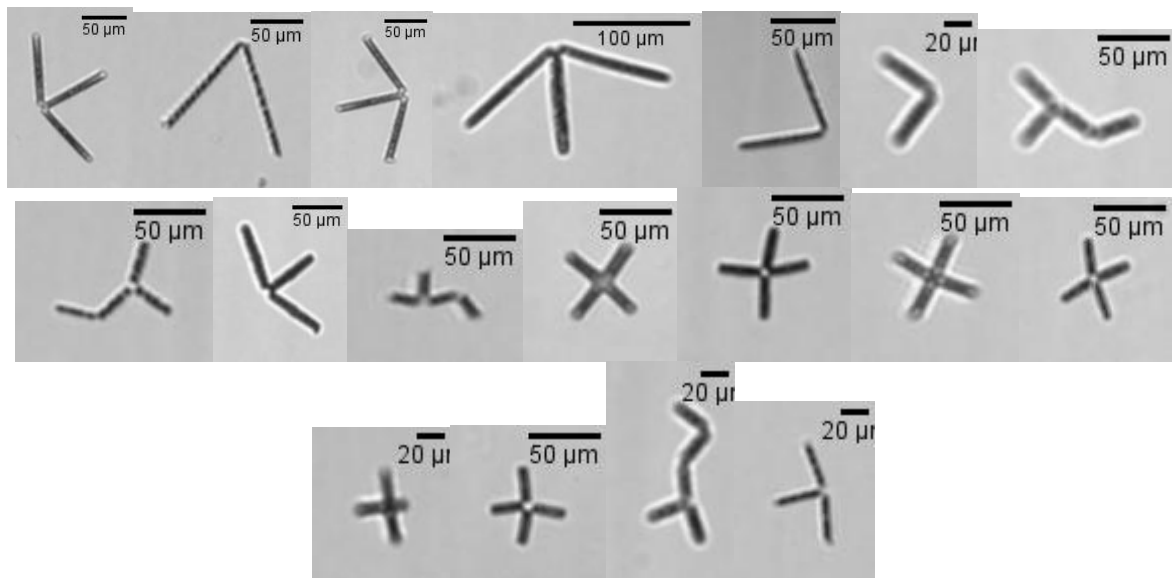
NT

5.41. *Thalassionema*

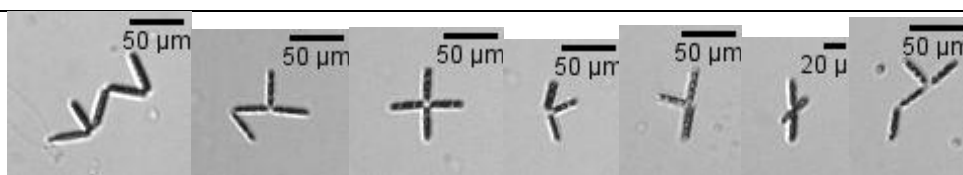
Ochrophyta\Bacillariophyceae\Thalassionemataceae\Thalassionema



BL



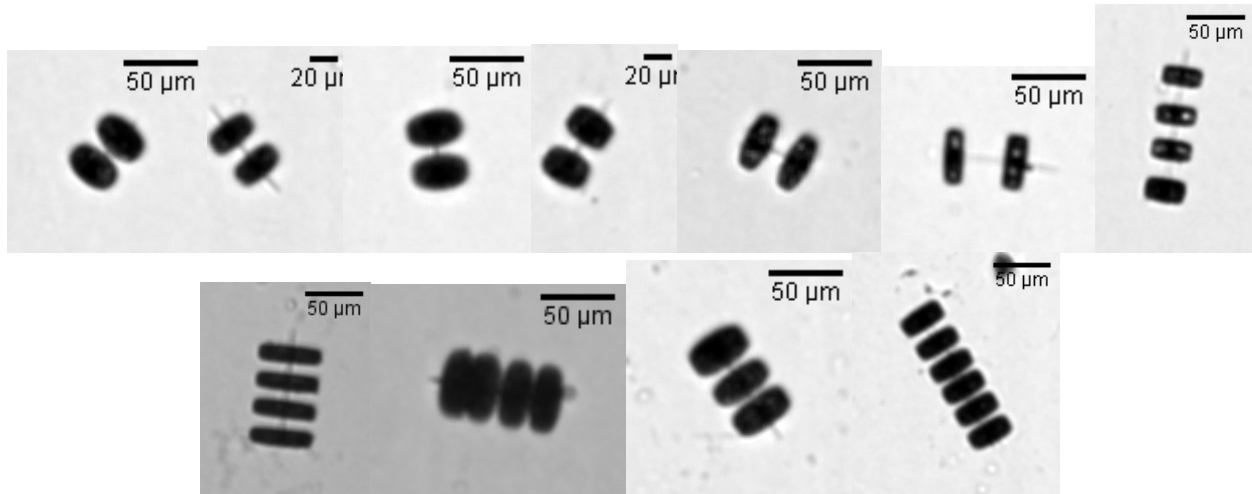
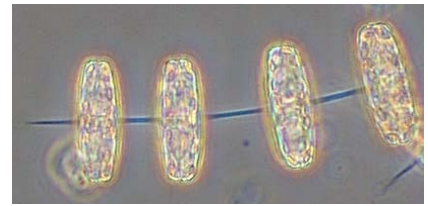
NT



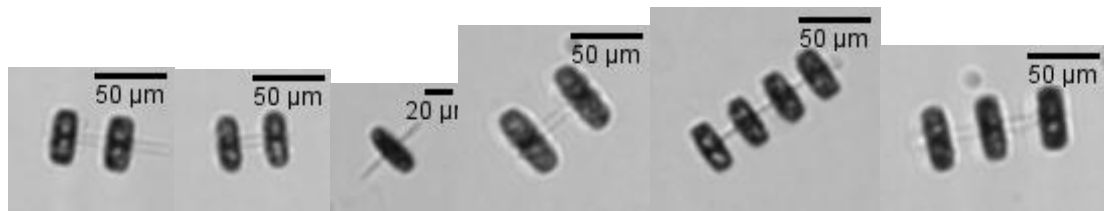
AR

5.42. *Thalassiosira big chaines*

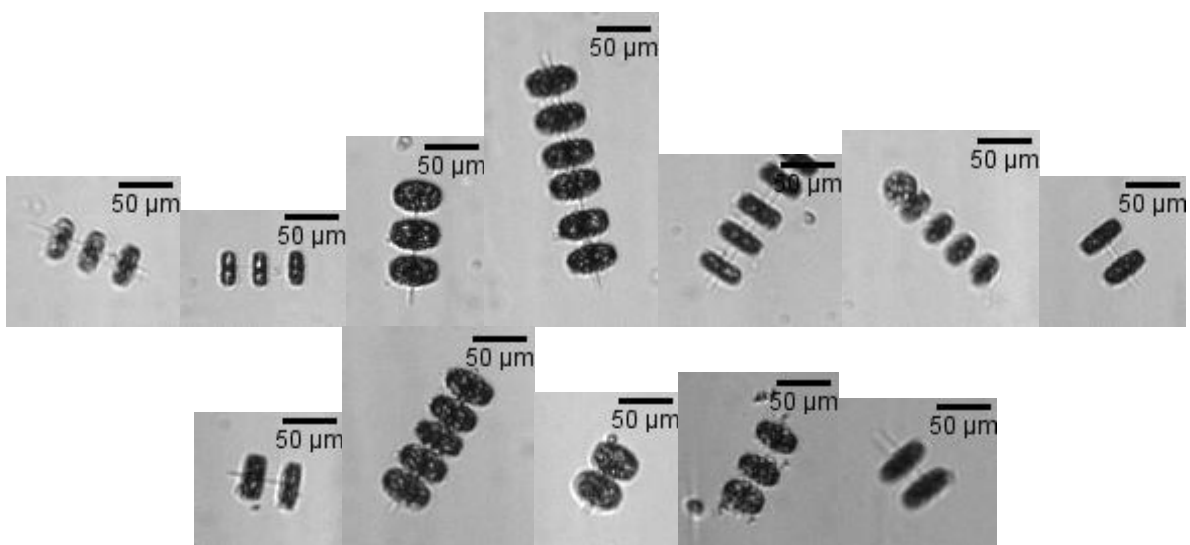
Ochrophyta\Bacillariophyceae\Thalassiosiraceae\Thalassiosira big chaines



BL



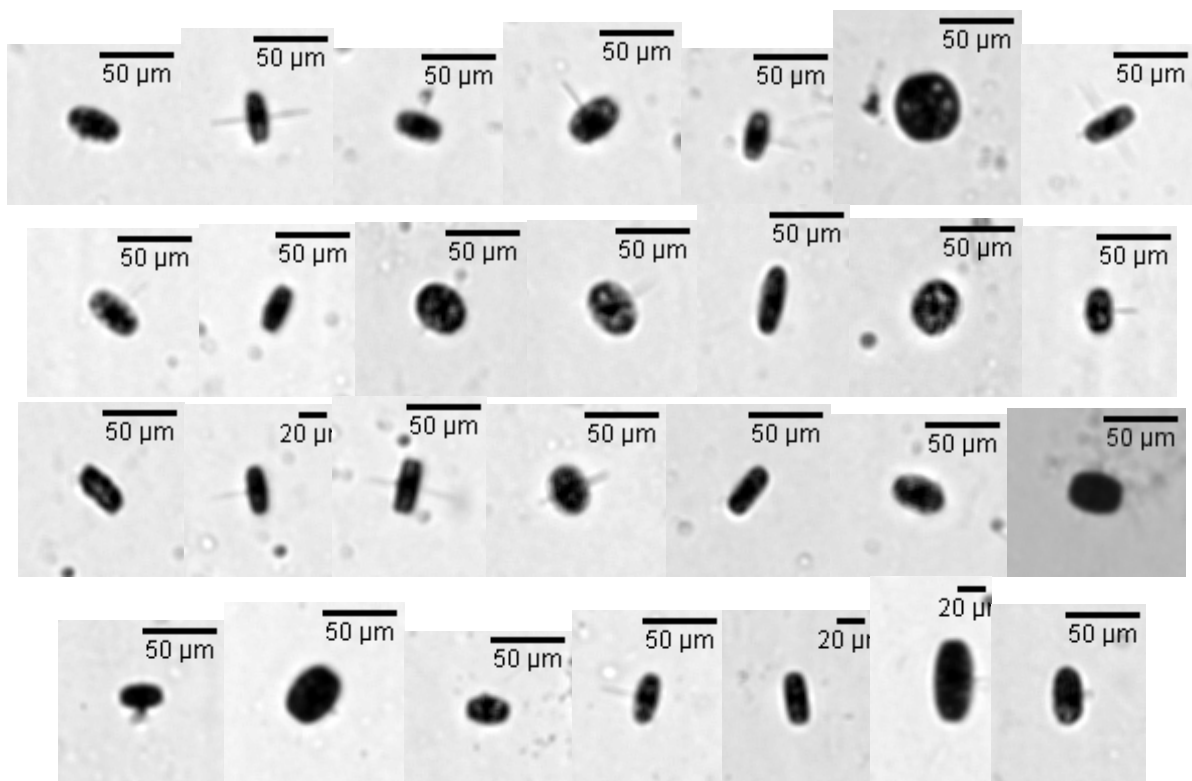
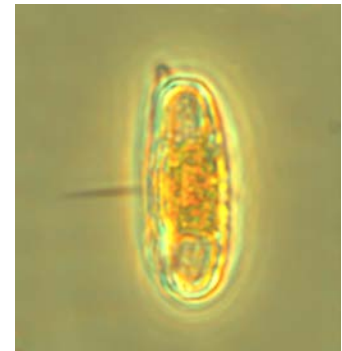
NT



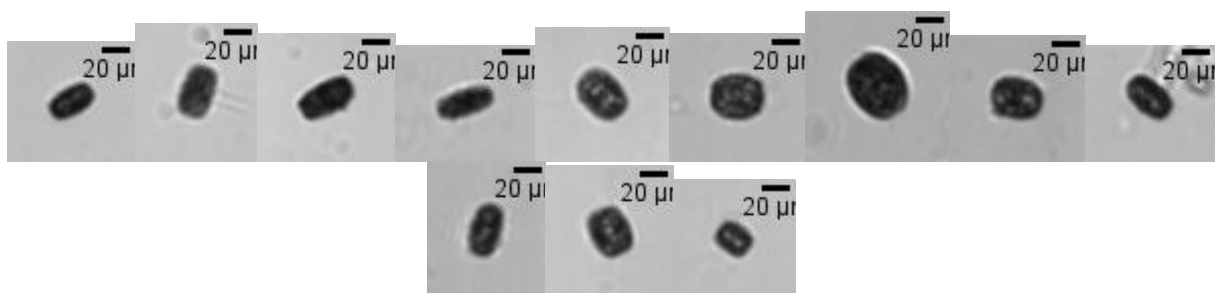
AR

5.43. *Thalassiosira* cells

Ochrophyta\Bacillariophyceae\Thalassiosiraceae\Thalassiosira cells



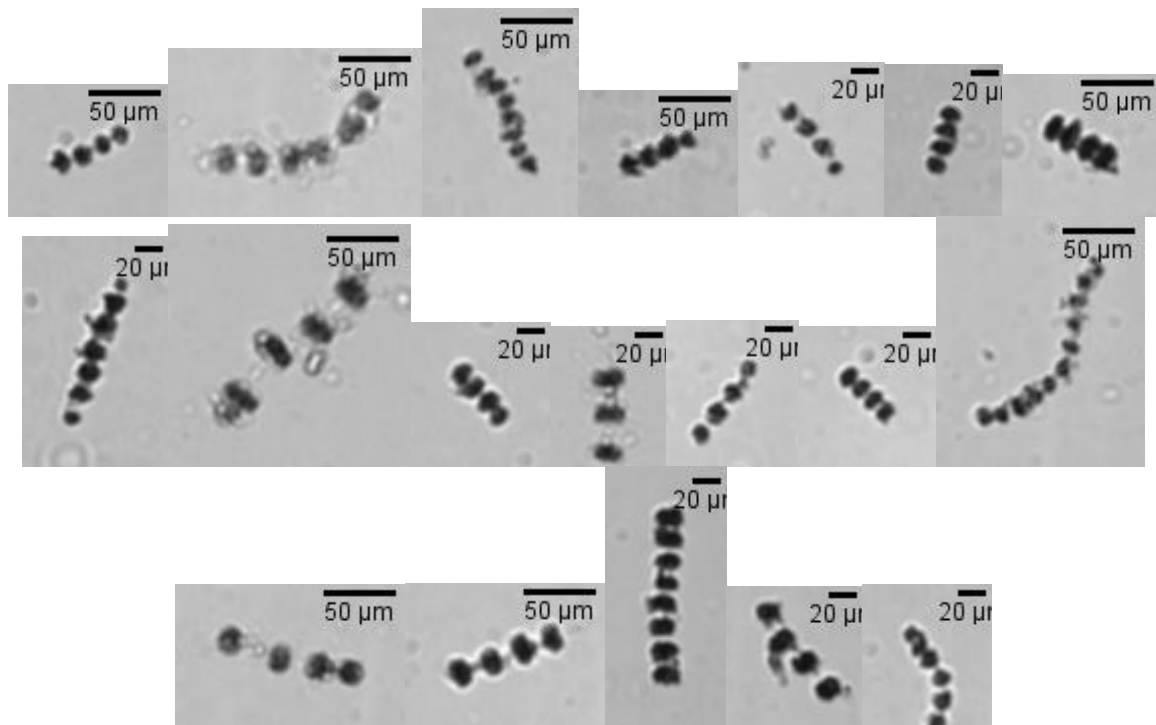
BL



NT

5.44. *Thalassiosira small chaines*

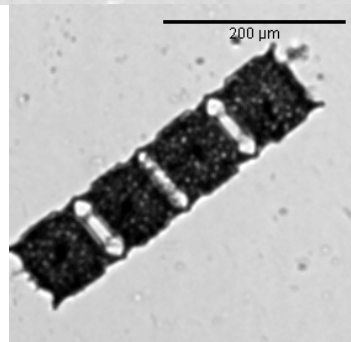
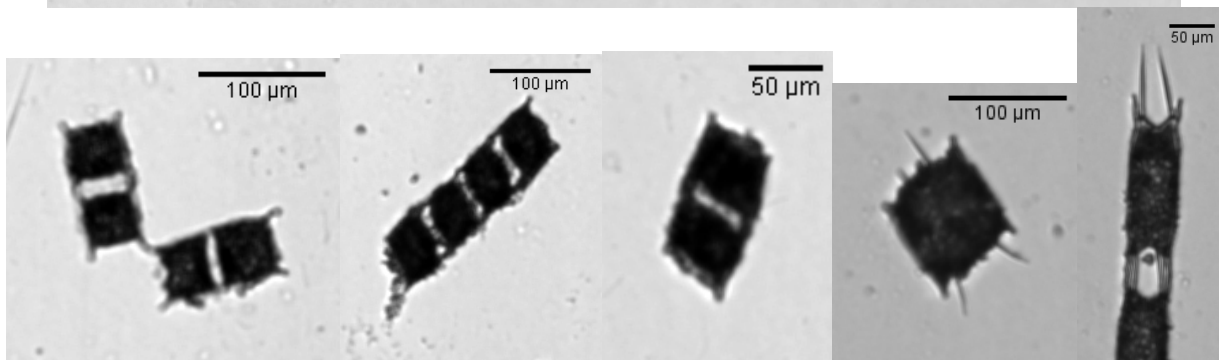
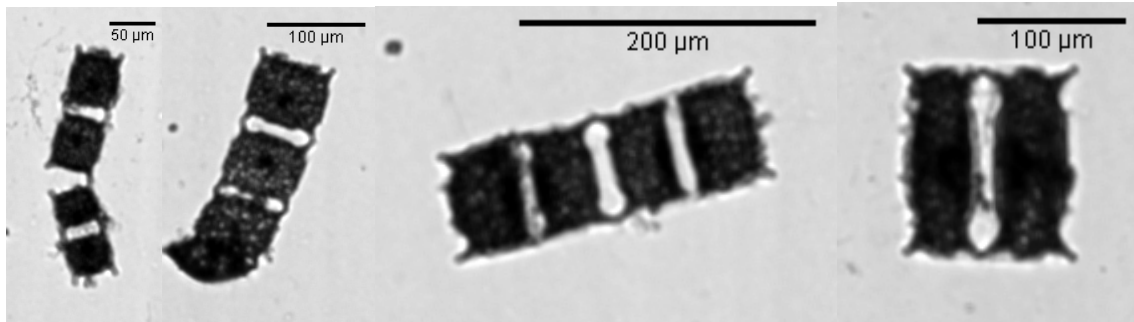
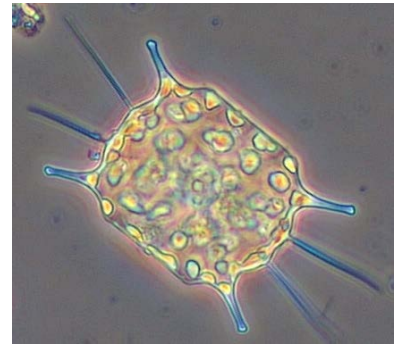
Ochrophyta\Bacillariophyceae\Thalassiosiraceae\Thalassiosira small chaines



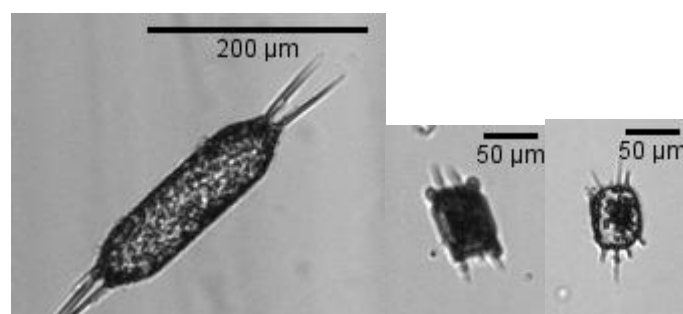
NT

5.45. *Odontella*

Ochrophyta\Bacillariophyceae\Triceratiaceae\Odontella



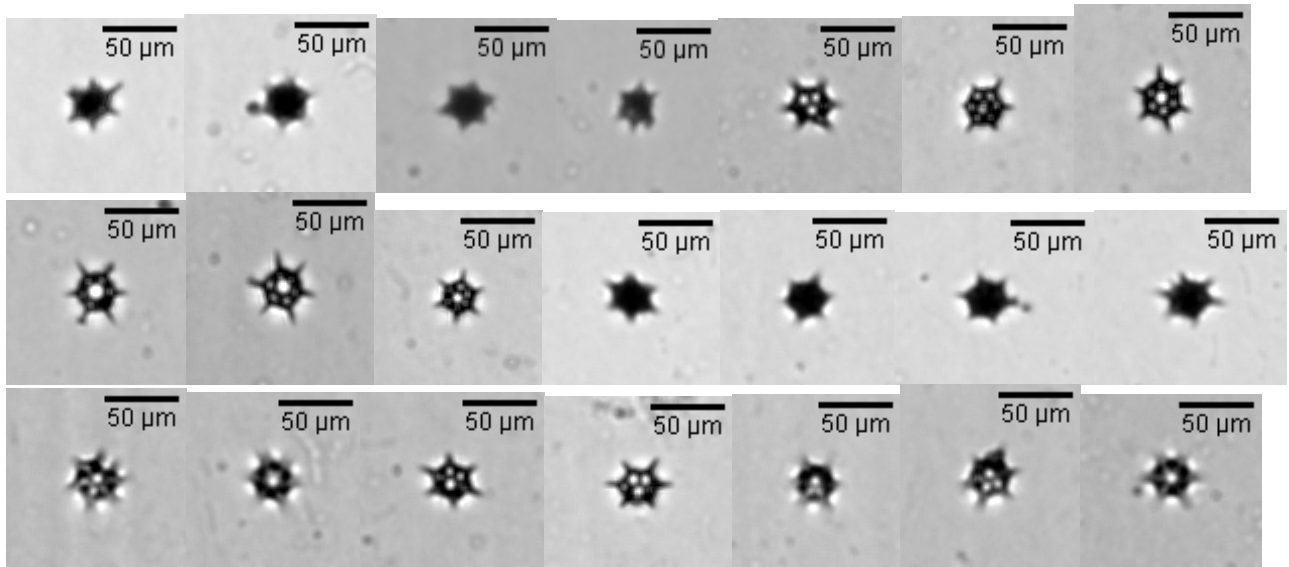
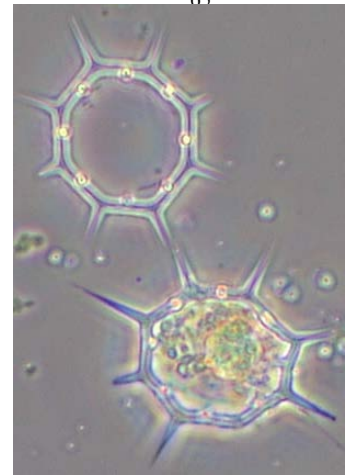
BL



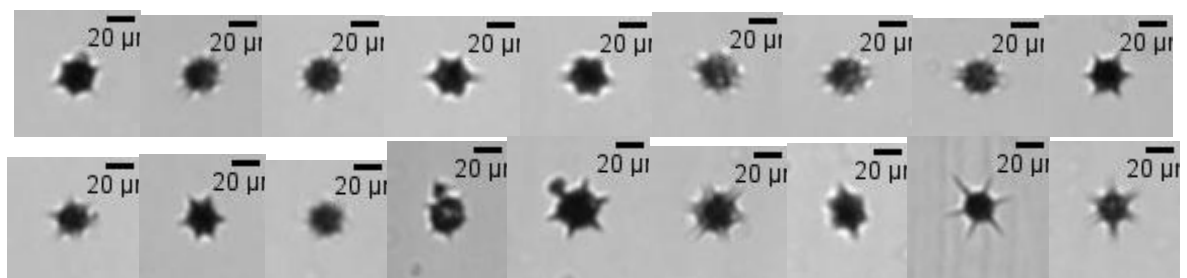
AR

5.46. *Dictyocha*

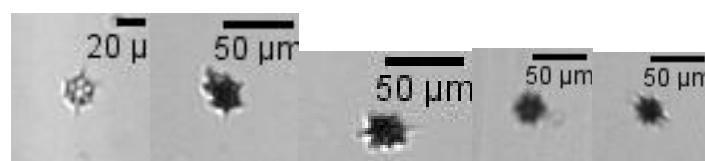
Ochrophyta\Dictyochophyceae\Dictyochaceae\Dictyocha



BL



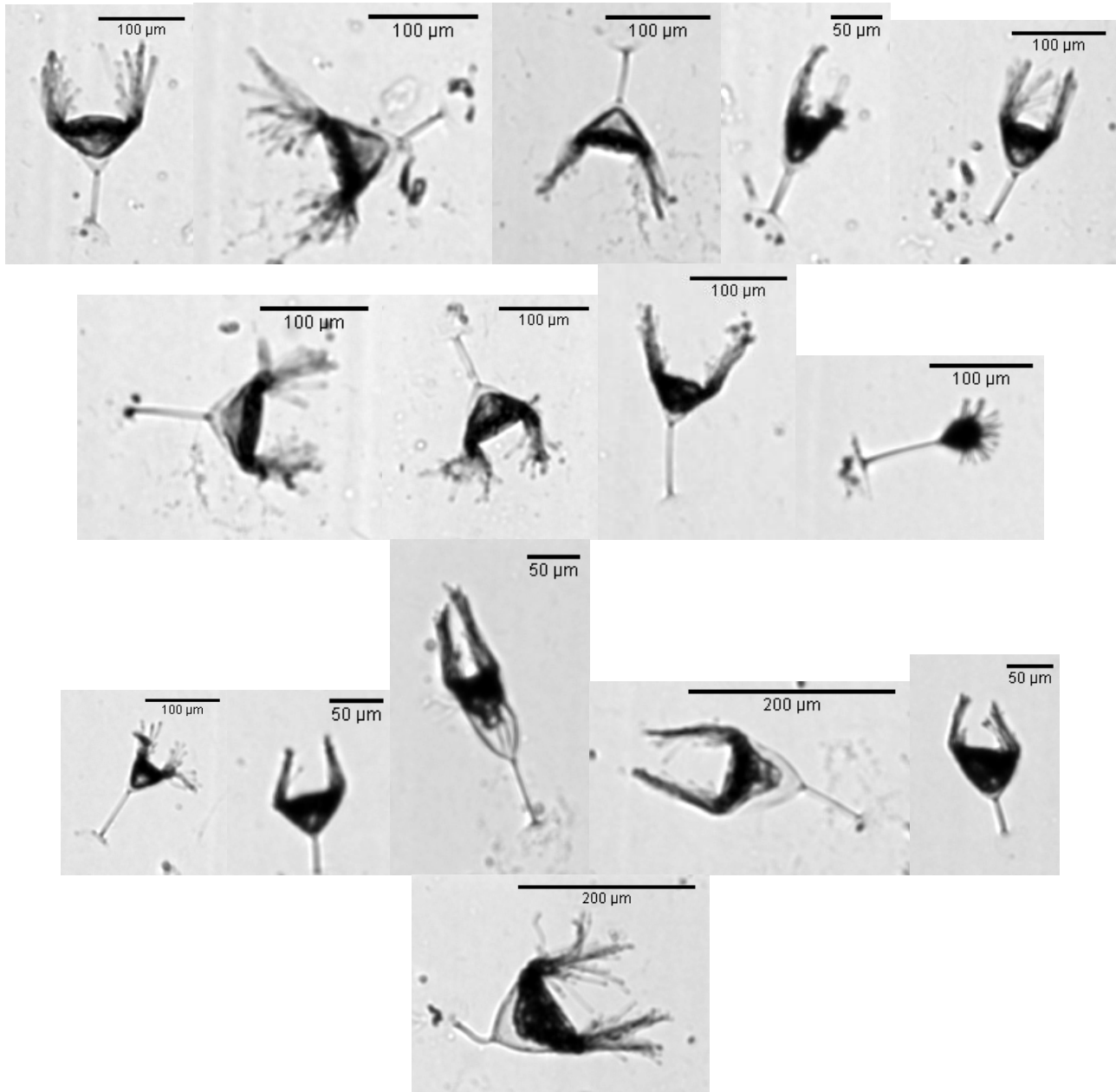
NT



AR

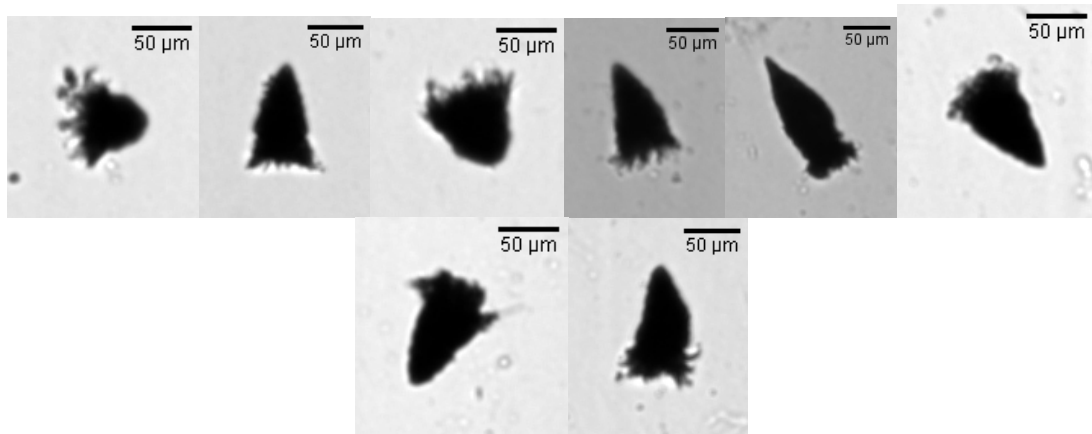
6. Classes du zooplancton intégrées au set

6.1. *Ciliophora* big

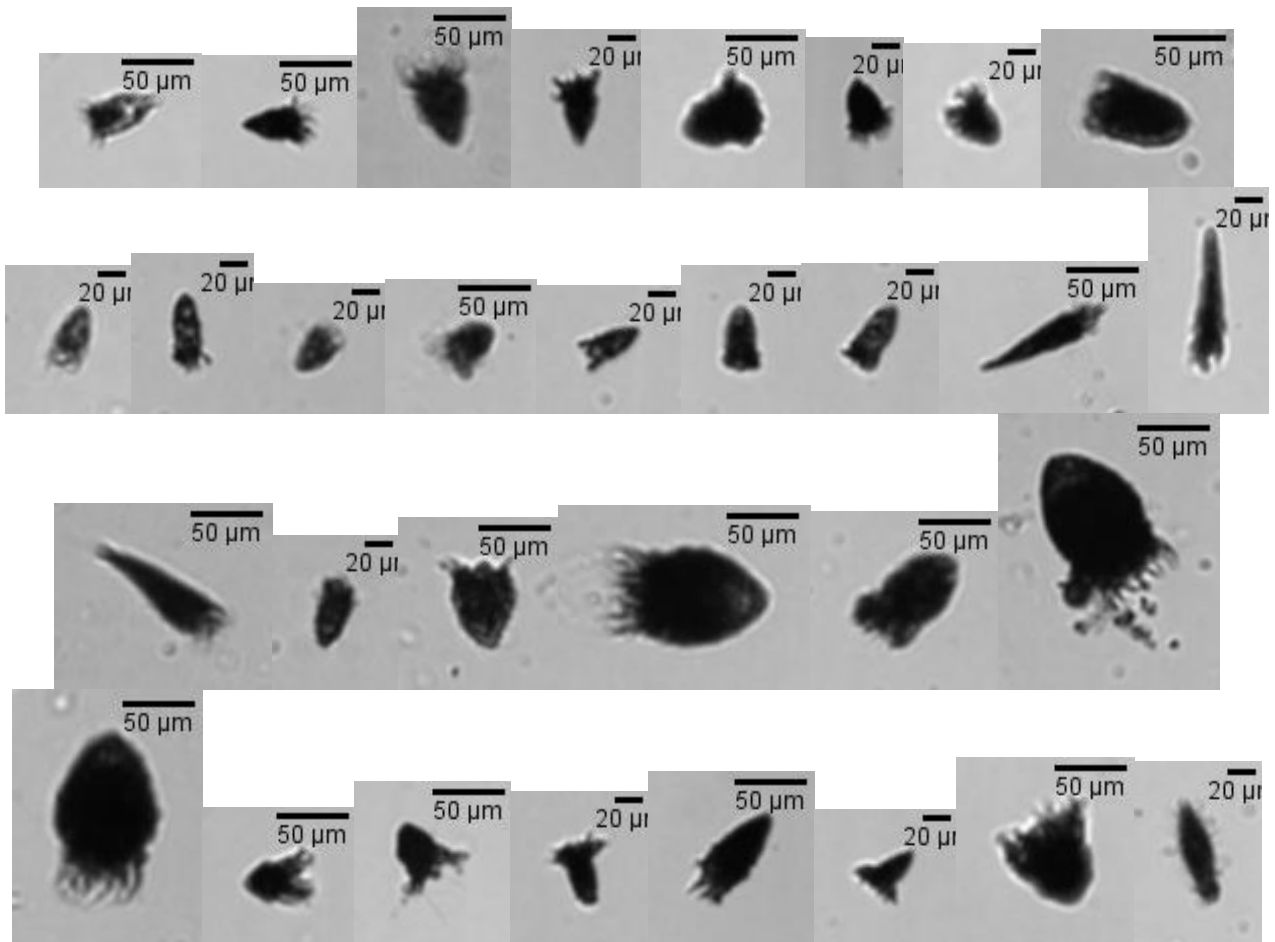


BL

6.2. Ciliophora cone

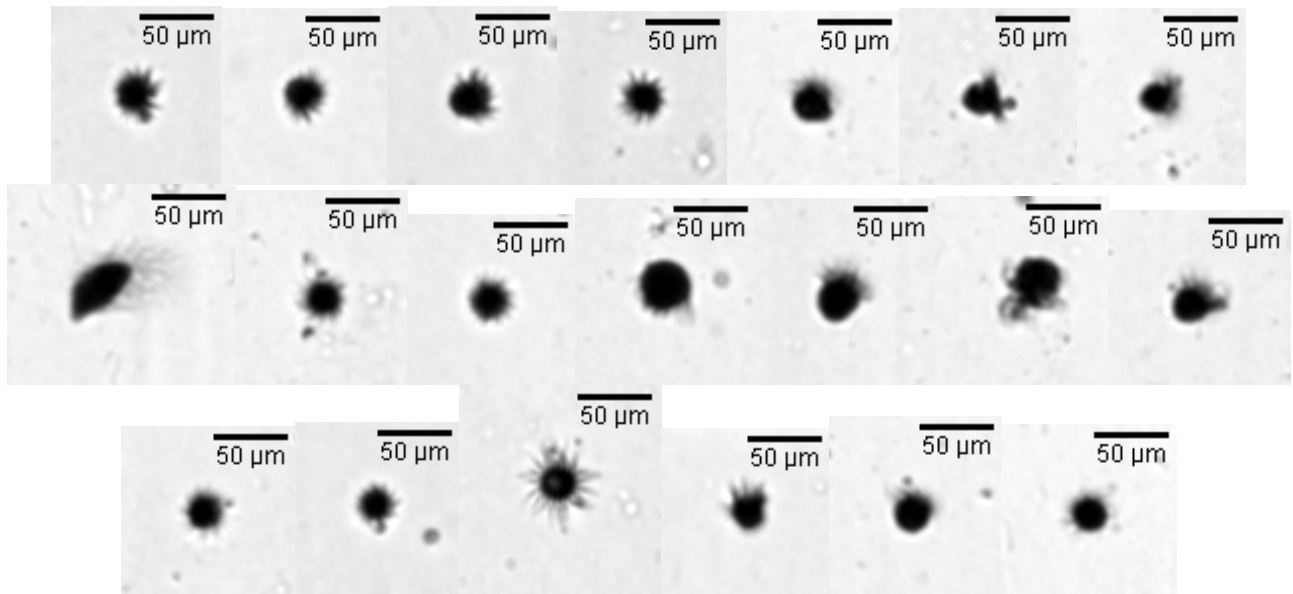


BL

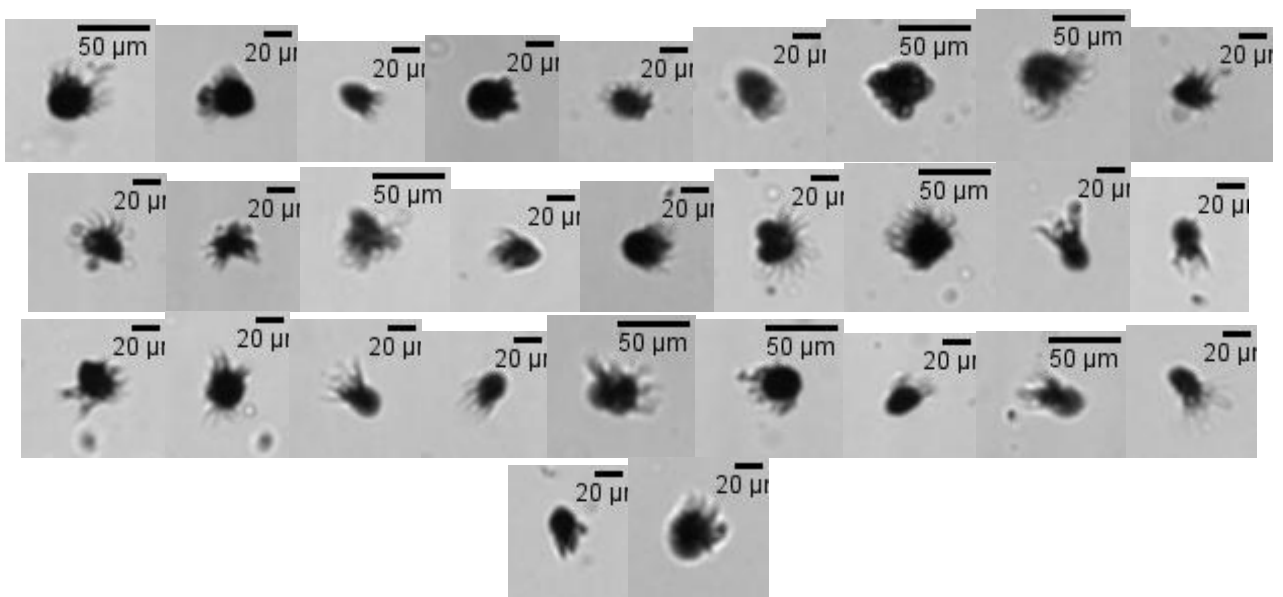


NT

6.3. *Ciliophora* small

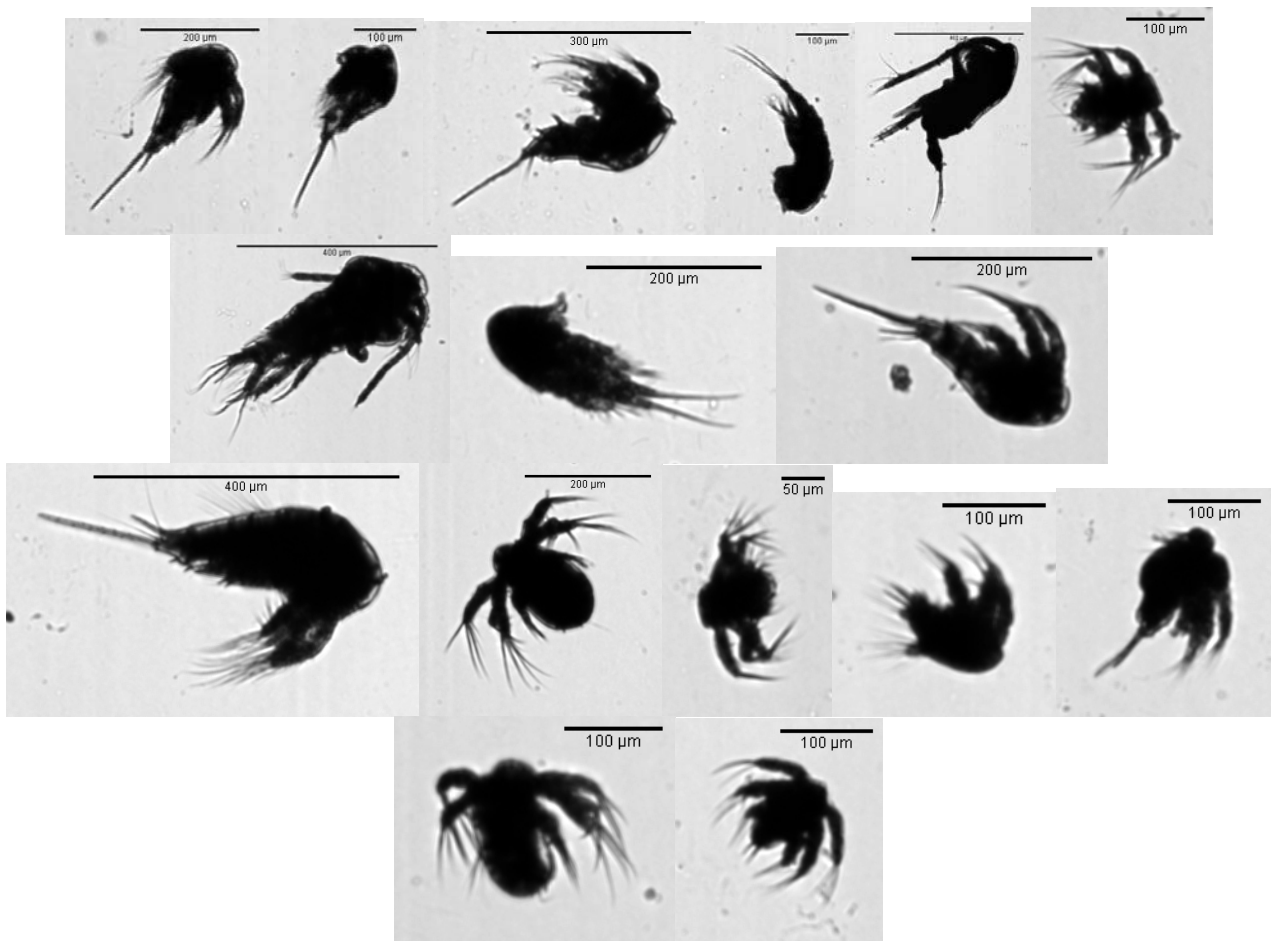


BL

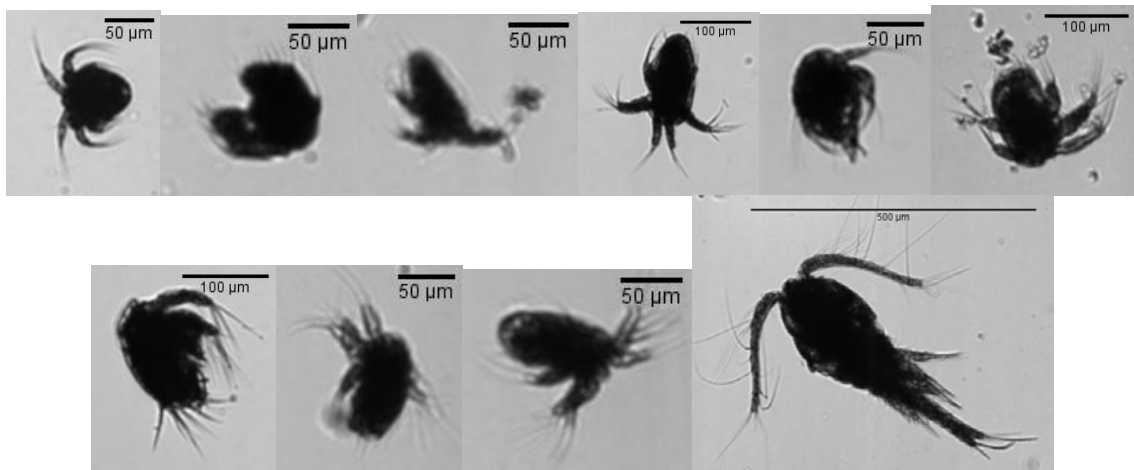


NT

6.4. Zooplankton spp



BL



NT



Bancarisation des données Zoo/PhytoImage

Auteurs et contributeurs

Ifremer

Emilie GAUTHIER

Catherine BELIN

Antoine HUGUET

Nadine NEAUD-MASSON

Stéphane BOCANDE

Dominique SOUDANT

Alice LAMOUREUX

Gaetane DURAND

Catherine BORREMANS

Umanis

Antoine SZIDON

Arnaud LEFEBVRE

Université de Mons

Philippe GROSJEAN

Guillaume WACQUET



Synthèse de l'étude sur la banclarisation des données ZooPhytoImage

Emilie Gauthier

Février 2016

Bancarisation des données ZooPhytoImage Synthèse des travaux

Liste des contributeurs à l'étude

Emilie GAUTHIER (Ifremer/ODE/VIGIES)

Nadine NEAUD-MASSON (Ifremer/ODE/VIGIES)

Catherine BELIN (Ifremer/ODE/VIGIES)

Stéphane BOCANDE (Ifremer/IDM/ISI)

Dominique SOUDANT (Ifremer/ODE/VIGIES)

Antoine HUGUET (Ifremer/ODE/VIGIES)

Alice LAMOUREUX (Ifremer/ODE/VIGIES)

Catherine BORREMANS (Ifremer/IDM/SISMER)

Antoine SZIDON (Umanis)

Arnaud LEFEBVRE (Umanis)

Philippe GROSJEAN (Université de MONS, Belgique)

Guillaume WACQUET (Université de MONS, Belgique)

Rappel du contexte

L'analyse du phytoplancton via la numérisation par FlowCAM, suivie du traitement des résultats via le logiciel ZooPhytoImage, produit des données d'identification, d'abondance, de biovolume, et d'autres métriques quantitatives sur les particules contenues dans les échantillons. Nombre de ces données sont nouvelles par rapport aux résultats de la surveillance du phytoplancton effectuée dans le cadre du réseau REPHY et obtenus par analyse des échantillons au microscope optique.

Cette nouveauté a deux composantes :

	Microscopie optique	FlowCam / ZooPhytoImage
Format des données	Fiches paillasse papier transcrites numériquement par saisie manuelle dans la base Quadrige ²	Fichiers texte (.csv) contenant l'ensemble des données
Métriques mesurées	Nombre de cellules / L par taxon identifié	Par classe de particules (dont les taxons) : nombre de cellules / L et biovolume Par particule : mesures quantitatives de taille, caractéristiques physiques, etc. (plus de 60 colonnes dans le fichier de données)

Ces différences soulèvent des questions :

- 1) Format : comment bancariser les données issues de ZooPhytoImage dans Quadrigé² avec les autres données « classiques » du REPHY, sans ressaisir manuellement le contenu des fichiers générés ?
- 2) Métriques : quelles sont les informations pertinentes à bancariser ? Quelles métriques retenir ? A quelle échelle les bancariser (particule ? classe de particule ? échantillon total ?) ? Si elles sont agrégées, quels éléments statistiques bancariser (médiane et/ou moyenne, écart-type, min/max, etc.) ?

D'autre part, ce nouveau protocole d'analyse permet de conserver à la fois une image des particules contenues dans les échantillons, et l'ensemble des mesures effectuées sur toutes ces particules. Ainsi, au gré de l'évolution des connaissances et de l'amélioration des techniques d'identification (e.g. set d'apprentissage du logiciel ZooPhytoImage), il est possible : (i) de rejouer les analyses via ZooPhytoImage pour améliorer la qualité des données (i.e. classification taxinomique des particules et donc révision des abondances et biovolumes par taxon), (ii) d'accéder aux métriques les plus fines pour répondre à des questions relatives à une expertise ou à une zone géographique particulière (e.g. turbidité du milieu).

Ainsi, des réflexions ont été menées en interne pour proposer des solutions de bancarisation des données (plusieurs réunions en interne Ifremer depuis septembre 2014). Une prestation d'expertise a également été menée pour proposer une architecture technique de l'ensemble de la chaîne d'acquisition et définir des spécifications d'intégration des données dans la base Quadrigé. Tous ces éléments sont détaillés ci-après.

Architecture du système

Architecture actuelle

Actuellement, les échantillons destinés à être traités par l'outil FlowCAM / ZooPhytoImage sont prélevés par l'ensemble des LERs métropolitains dans leurs périmètres géographiques respectifs. Les échantillons sont ensuite envoyés vers un des laboratoires disposant d'un FlowCAM (actuellement trois appareils en service : Boulogne, Nantes et Arcachon) (Figure 1).

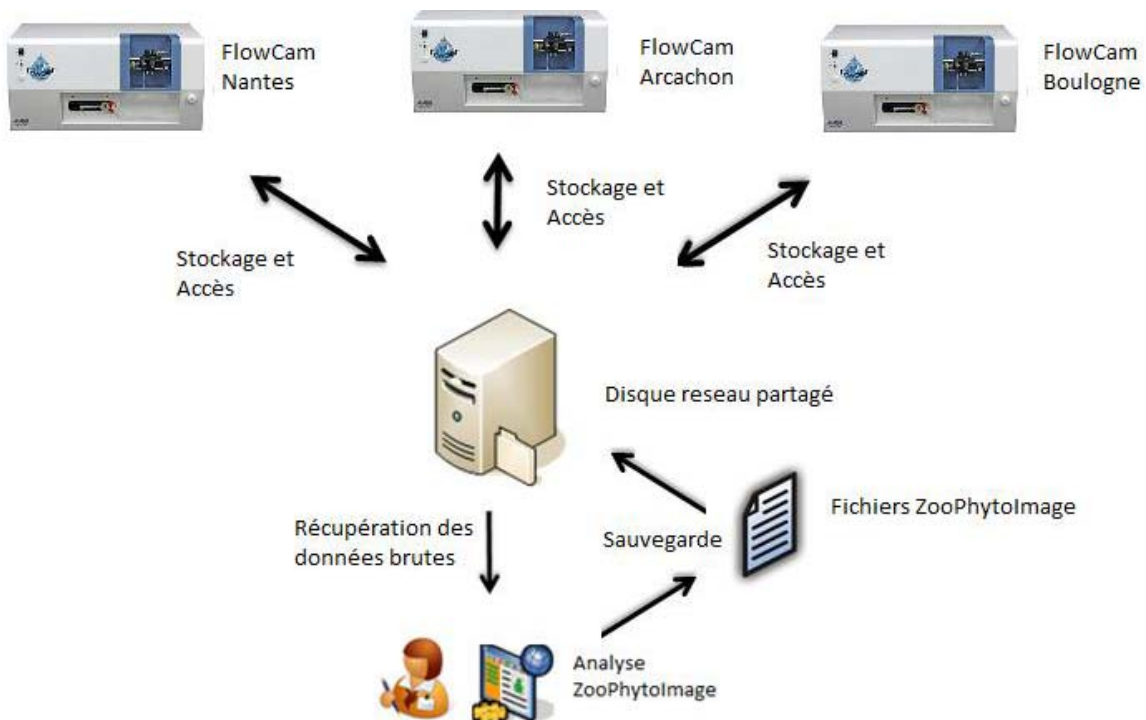


Figure 1 : Architecture actuelle pour la sauvegarde des données brutes (source : rapport d'étude « Bancarisation des données ZooPhytoImage », Umanis, janvier 2016).

Un disque réseau partagé a été alloué à ces données. Ainsi, au début de la numérisation d'un échantillon par le FlowCAM, l'utilisateur indique l'emplacement des répertoires de sauvegarde des résultats sur cet espace disque. Lorsque le réseau de laboratoires du REPHY sera formé à l'utilisation du logiciel, le fournisseur de l'échantillon sera prévenu que l'analyse a été réalisée. Ce dernier pourra alors récupérer les fichiers issus du FlowCAM pour les traiter avec le logiciel ZooPhytoImage. Il indiquera à son tour l'emplacement des fichiers de résultats de ce traitement dans le même répertoire que les données brutes initiales.

Cette solution permet de limiter l'accès aux données via une connexion sécurisée par mot de passe, tout en permettant le partage des données entre les utilisateurs autorisés. Le disque partagé est géré par les infrastructures informatiques de l'Ifremer, garantissant une sauvegarde quotidienne des données et la maintenance des supports techniques. L'étude d'Umanis a montré la pertinence de cette architecture pour les besoins actuels (*i.e.* en phase de test du protocole d'analyse). Toutefois, l'espace disque nécessaire va rapidement augmenter : en routine, à raison de 4 000 échantillons par an, produisant chacun 1 Go de données environ, l'espace de stockage actuel sera rapidement insuffisant. D'autre part, une ouverture du système vers des utilisateurs extérieurs à l'Ifremer pourrait être envisagée dans le cadre de projets de labellisation des réseaux de surveillance et de coopération avec d'autres instituts de recherche (*e.g.* SOMLIT, RESOMAR). Or le disque réseau actuel n'est pas accessible hors Ifremer.

Future organisation des données

Un projet d'équipement mutualisé de calcul (supercalculateur) et de gestion de données de gros volume nommé DATARMOR est en cours de spécification au sein de l'Ifremer. Cet outil, destiné à l'Ifremer et à ses partenaires institutionnels (IRD, SHOM, CNRS...), permettra le stockage de données et le déploiement de services d'accès et de traitement de ces données. Cette architecture est la cible visée pour la sauvegarde des données FlowCAM et ZooPhytoImage à l'horizon de quelques années (Figure 2). Les analyseurs FlowCAM possèdent déjà les infrastructures technologiques nécessaires (*i.e.* unité centrale connectée au réseau sous système d'exploitation Windows).

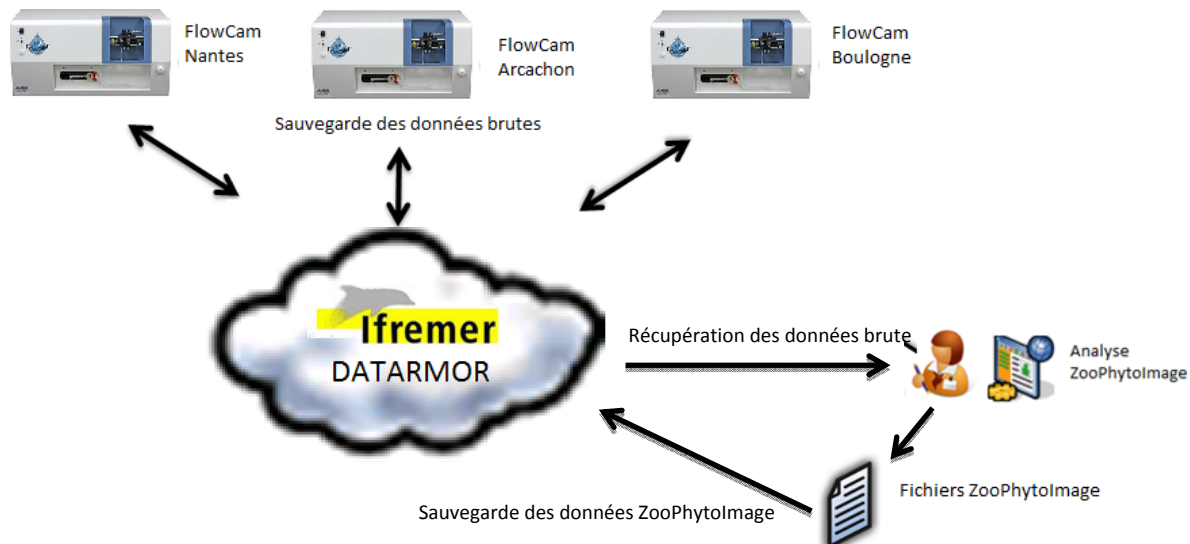


Figure 2 : Future architecture de sauvegarde des données ZooPhytoImage (d'après Etude de la bancarisation des données ZooPhytoImage. Umanis, janvier 2016).

Les détails des configurations techniques et services nécessaires pour cette architecture ont été fournis par Umanis dans le cadre de l'expertise réalisée. L'équipe Ifremer en charge des spécifications DATARMOR peut ainsi les intégrer dès maintenant à son cahier des charges.

Cette solution permet également d'envisager l'installation du logiciel directement sur DATARMOR. Cela permettrait de répercuter les évolutions logicielles et les mises à jour du set d'apprentissage en un seul endroit, au lieu de répéter l'opération sur chaque poste utilisateur. Des gains de performance

de l'outil sont probables : des tests de comparaison entre l'utilisation réseau actuelle et une utilisation en local sur une machine puissante doivent être réalisés pour évaluer ces améliorations.

Bancarisation des données dans Quadrige

Procédure d'intégration des données

Les fichiers de résultats produits par ZooPhytoImage sont des fichiers .RData, lisibles uniquement via la console du logiciel R dans lequel est développé ZooPhytoImage. Un script R de transformation de ces fichiers .RData en fichiers .csv a été réalisé par l'équipe Ifremer VIGIES.

La procédure retenue pour l'intégration de ces fichiers dans Quadrige est de les convertir au format Quadrilabo (format standard opérationnel depuis 2011, basé sur le référentiel national SANDRE, cf. le document « fichier Quadrilabo » dans le présent livrable). Ainsi, à l'issue de l'analyse des données via ZooPhytoImage :

- 1) l'utilisateur choisit les fichiers à intégrer dans Quadrige. Ces fichiers ont des règles de nommage prédéfinies, contenant le mnémonique de l'échantillon Quadrige qui recevra ces résultats.
- 2) il les dépose sur un répertoire spécifique via des interfaces web accessibles en Internet sur le site de l'assistance Quadrige (http://wwz.ifremer.fr/quadrige2_support)
- 3) Un job Talend (script informatique développé via l'ETL¹ Talend) lit les fichiers, les transforme en fichier Quadrilabo, puis lance le script d'intégration Quadrilabo déjà existant.
- 4) Les résultats d'analyse sont ainsi intégrés sur un échantillon préalablement créé dans Quadrige (la saisie en est faite manuellement par le LER ayant réalisé le prélèvement).
- 5) L'utilisateur reçoit un mail de notification l'informant du chargement correct des données.

Cette solution permet de garantir l'évolutivité des formats ZooPhytoImage, sans impacter l'application Quadrige ni le script Quadrilabo existant. Elle permet aussi de limiter le chargement des résultats à un échantillon à la fois, évitant la saisie à retardement de gros lots de données.

Toutefois, l'outil à développer devra permettre une intégration en masse de fichiers, réservée à un profil administrateur, pour faciliter les reprises de données historiques et le rattrapage d'éventuels retards de saisie (en cas de problèmes techniques temporaires empêchant l'intégration des données, par exemple).

Umanis a rédigé des spécifications générales de conversion des fichiers ZooPhytoImage au format Quadrilabo. Ces spécifications pourront être utilisées pour développer le script Talend d'intégration des données, dès que des jeux de données seront prêts.

D'autre part, le référentiel Quadrige doit être complété pour intégrer les paramètres correspondant aux nouvelles métriques à intégrer, les méthodes d'analyse (des propositions ont été faites et doivent être testées), les engins d'analyse et autres informations nécessaires. Les programmes et stratégies REPHY dans Quadrige devront ensuite être mis à jour avec ces nouveaux référentiels.

Réflexions sur les données à intégrer

Le nombre de résultats générés par les analyses ZooPhytoImage est important : pour un échantillon contenant environ 6000 particules par exemple, si seulement 10 métriques sont bancarisées, cela fait 60 000 résultats à visualiser dans l'interface graphique Quadrige. D'une part l'application Quadrige actuelle risque d'avoir des performances insuffisantes pour assurer la visualisation et l'éventuelle mise à jour de ces données, et d'autre part cette granularité des données n'est pas jugée nécessaire pour le traitement et la valorisation des résultats.

¹ ETL : Extract Transform Load

Il est donc décidé de ne bancariser que des métriques calculées à l'échelle du taxon, et non pas à l'échelle de la particule. Par ailleurs, les informations relatives aux particules n'appartenant pas aux taxons phytoplanctoniques, ne seront pas bancarisées dans Quadrige. Les abondances et biovolumes bruts sont fournis directement par ZooPhytoImage (1^{er} type de fichier généré). Les autres caractéristiques, mesurées pour chaque particule appartenant à un taxon phytoplanctonique, (longueur, largeur, diamètre, surface, etc.), seront agrégées par taxon. La liste précise des métriques utiles reste à définir, et les paramètres statistiques à bancariser doivent être validés (*a priori* médiane et valeurs minimale et maximale).

Perspectives et actions à poursuivre

Les besoins pour la sauvegarde des données sur DATARMOR ont été remontés, et intégreront le cahier des charges en cours d'élaboration. En attendant la mise en production de ce supercalculateur, la sauvegarde des données ZooPhytoImage continue sur le disque réseau partagé. La procédure de numérisation rédigée par Nadine Neaud-Masson (cf. livrable n°2 pour l'année 2014) devra être mise à jour avec les consignes de nommage des fichiers (*i.e.* codification des noms de fichiers, à saisir dans Quadrige² dans le mnémonique de l'échantillon qui accueillera les résultats).

Des échanges doivent être poursuivis avec Philippe Grosjean de l'université de Mons (Belgique), développeur de ZooPhytoImage, pour identifier les métriques pertinentes à intégrer dans Quadrige² (*i.e.* quelles colonnes sur les 60 présentes dans le fichier de détail des particules). Une fois la liste des métriques arrêtée, le référentiel Quadrige² devra être complété, et les programmes et stratégies mis à jour pour préparer l'intégration des données. Des consignes de saisie Q² des métadonnées devront être fournies aux producteurs de données (*e.g.* création d'un échantillon spécifique dans Q², règles de nommage de l'échantillon).

Les spécifications de conversion des fichiers ZooPhytoImage au format Quadrilabo, préparées par Umanis dans le cadre de l'étude « Bancarisation », devront être détaillées. Le script correspondant devra être développé et mis à disposition des utilisateurs.

Enfin, les résultats de l'ensemble de ce travail alimenteront le cahier des charges Quadrige3 en prenant en compte l'intégration des données ZooPhytoImage dans le développement d'interfaces thématiques sur le phytoplancton.



IFREMER

Etude bancarisation des données ZooPhytoImage

ETUDE



REFERENCE : F005236C02

VERSION : 3.1

DATE DE LA VERSION : 02/02/2016

ETAT DU DOCUMENT : FINAL

Diffusion			
Périmètre de diffusion : <input type="checkbox"/> Interne <input checked="" type="checkbox"/> Contrôlé <input type="checkbox"/> Libre			
Nom	Société	Nombre	Support
Equipe Umanis	UMANIS	1	Electronique
Equipe Ifremer	IFREMER	1	Electronique

Historique des évolutions	
Version	Objet de la version et liste des modifications
1.0	Initialisation du document
2.0	Mise à jour avec les retours de la réunion du 15-12-2015
3.0	Mise à jour avec les retours de la réunion du 07-01-2016
3.1	Version finale mise à jour avec les retours de la réunion du 02-02-2016

Validations						
Version	Emetteur/Rédacteur		Vérificateur		Approbateur	
	Nom	Date/Visa	Nom	Date/Visa	Nom	Date/Visa
2.0	Antoine Szidon	04-01-2016/ASD	Arnaud Lefebvre	04-01-16/ALB	Frédéric Vernet	04-01-16/FVT
3.0	Antoine Szidon	27-01-2016/ASD	Arnaud Lefebvre	27-01-2016/ALB	Frédéric Vernet	27-01-2016/FVT
3.1	Antoine Szidon	02-02-2016/ASD	Arnaud Lefebvre	02-02-2016/ALB	Frédéric Vernet	02-02-2016/FVT

SOMMAIRE

1	Contexte et objectifs	5
2	Définitions	6
2.1	FlowCam	6
2.2	ZooPhytoImage.....	6
2.3	DATARMOR.....	6
2.4	La sauvegarde	6
2.5	La bancarisation.....	6
2.6	QuadriLabo	6
2.7	Talend	6
2.8	ZooCam.....	7
3	Solutions	8
3.1	Solutions de sauvegarde.....	8
3.1.1	Disque réseau partagé.....	8
3.1.2	OwnCloud	9
3.1.3	DATARMOR.....	16
3.1.4	ZooPhytoImage installé sur DATARMOR.....	16
3.2	Solutions de bancarisation	18
3.2.1	Intégration des données.....	18
3.2.2	Format QuadriLabo	19
3.2.3	Intégration automatique unitaire.....	20
3.2.4	Intégration automatique en masse	20
3.2.5	Détail de l'intégration dans Q ²	21
3.2.6	Gestion du classifieur et du set d'apprentissage.....	23
3.2.7	ZooCam.....	24
4	Solutions exclues	25
4.1	Solutions de sauvegarde.....	25
4.1.1	Sauvegarde sur disque réseaux individuels.....	25
4.1.2	Sauvegarde sur le CLOUD	25
4.2	Solutions de bancarisation	25
4.2.1	Bancarisation depuis ZooPhytoImage	25
4.2.2	Récupération des ZID/ZIDB pour traitement et bancarisation	25

4.2.3	Bancarisation directe depuis Q ²	25
4.2.4	Format EdiLabo.....	25
5	Conclusion	26

1 CONTEXTE ET OBJECTIFS

Dans le cadre du REPHY, des données de composition et d'abondance du phytoplancton sont acquises via l'outil « FlowCAM / ZooPhytoImage ». Cet outil permet de numériser chaque particule (cellule vivante ou débris) présente dans un échantillon. Le logiciel ZooPhytoImage, développé par Philippe Grosjean de l'université de Mons en Belgique est un outil de reconnaissance des cellules phytoplanctoniques qui fournit une identification taxinomique ainsi que des mesures quantitatives (comptage en nombre de cellules, biovolumes, taille de cellule, etc.). Ces données doivent être sauvegardées (données brutes) puis bancarisées dans la base de données Quadrigé 2 (Q²) (résultats de l'analyse de l'échantillon). Une architecture de ce processus et des solutions techniques doivent être proposées à l'Ifremer pour réaliser cette bancarisation.

Le but de ce rapport est de proposer différents processus de sauvegarde et de bancarisation. Le choix du support de sauvegarde n'a pas d'incidence sur la méthode de bancarisation, ces deux parties sont indépendantes.

2 DEFINITIONS

2.1 FlowCam

Le FlowCAM (Flow Cytometer And Microscope) est un outil d'analyse qui combine deux technologies, la cytométrie en flux d'une part et la microscopie photonique reliée à une caméra numérique d'autre part.

2.2 ZooPhytoImage

ZooPhytoImage a pour objectif de fournir une solution puissante et riche en fonctionnalités logicielles pour utiliser les images de zooplancton ou phytoplancton provenant d'origines diverses et les transformer en une table de mesures utilisables (c'est-à-dire, les abondances, les spectres de taille totaux et partiels, les biomasses totales et partielles, ..)

2.3 DATARMOR

DATARMOR est un projet d'équipement mutualisé de calcul (supercalculateur) et de gestion de données de gros volumes. Il sera géré par l'IFREMER à destination de la communauté scientifique (Ifremer, Shom, IRD, IUEM, ect....). La mise en service de DATARMOR est planifiée pour 2016.

2.4 La sauvegarde

La sauvegarde consiste à archiver les données brutes (fichiers + arborescence des dossiers) générées par FlowCam ainsi que les fichiers générés par ZooPhytoImage. Ces fichiers sont conservés pour pouvoir refaire les analyses dans le cas où une évolution logicielle, matérielle ou scientifique permettrait de refaire des analyses plus précises. La sauvegarde n'est pas liée à la bancarisation, ces deux processus sont techniquement indépendants. La méthode de sauvegarde doit seulement permettre à un utilisateur de récupérer les fichiers sauvegardés via l'explorateur de fichier Windows. Fonctionnellement il faut pouvoir relier les données entre elles, donc par exemple depuis les données brutes on doit pouvoir retrouver les données ZooPhytoImage puis les données Q² et inversement.

2.5 La bancarisation

La bancarisation consiste à intégrer après retraitement les fichiers générés par ZooPhytoImage dans Q² pour ensuite pouvoir exploiter ces données de la même manière que les autres données issues de Q². La bancarisation est indépendante techniquement de la sauvegarde, le seul impératif étant de pouvoir récupérer les fichiers sauvegardés via l'explorateur Windows pour pouvoir les retraiter et les bancariser. Fonctionnellement, les contraintes sont les mêmes que pour la sauvegarde.

2.6 QuadriLabo

QuadriLabo est un format Excel normalisé contenant des données à intégrer dans Q². Actuellement un fichier au format QuadriLabo peut être uploadé via le site suivant :

http://wwz.ifremer.fr/quadrige2_support/Outils/Quadrilabo-integration-donnees

2.7 Talend

Talend est un ETL (Extract Transform Load) qui peut être utilisé pour la synchronisation de données entre applicatifs d'un système d'information.

2.8 ZooCam

Le ZooCam est un appareil « similaire » au FlowCam utilisé pour l'identification automatisée du mésozoo plancton. La présente Etude intéresse également les utilisateurs de cet appareil.



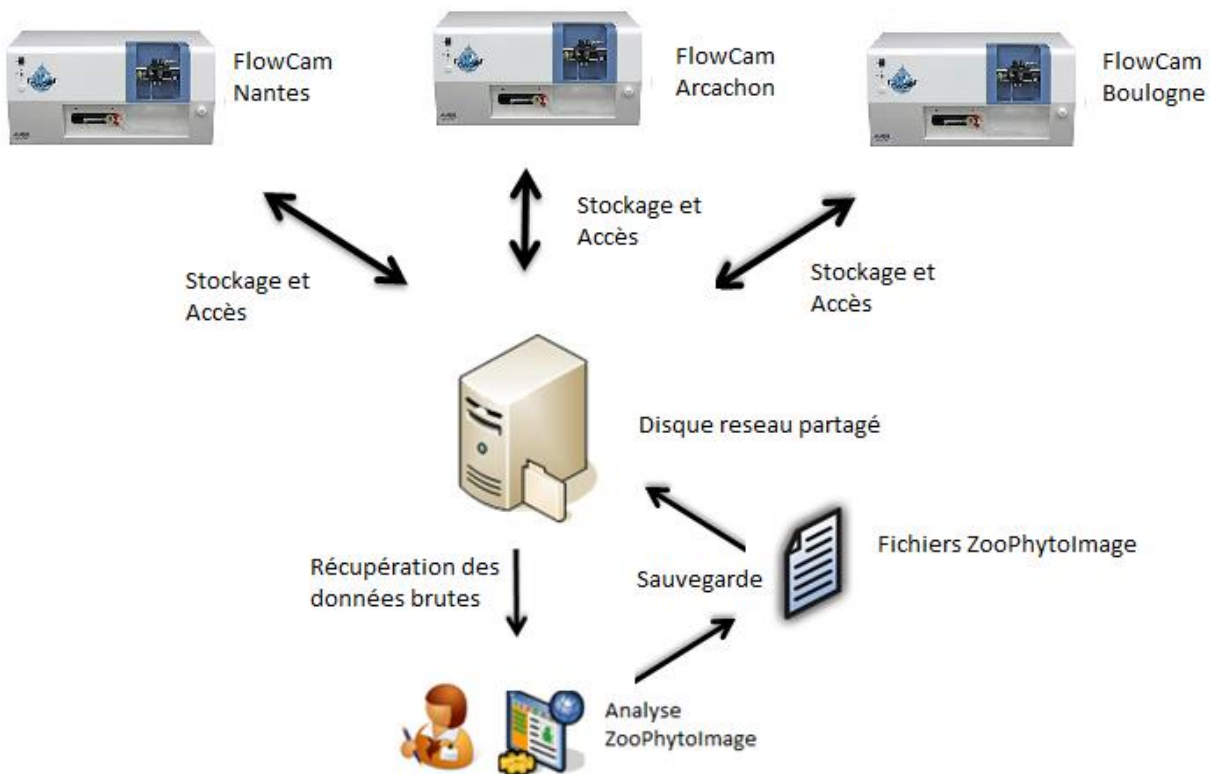
3 SOLUTIONS

3.1 Solutions de sauvegarde

Les solutions de sauvegarde ne doivent pas impliquer de faire évoluer FlowCam ou ZooPhytoImage car l'IFREMER n'a pas la main sur le développement de ces applications. La seule évolution tolérée pour ZooPhytoImage est l'ajout d'une ligne de commande permettant de générer des fichiers csv pour les résultats d'analyse.

3.1.1 Disque réseau partagé

C'est le fonctionnement actuel. Les résultats de la numérisation sont stockés sur un même serveur accessible à tous les LERs (Laboratoires Environnement Ressources). Chaque LER ne peut accéder qu'aux résultats de ses propres numérisations. Le disque partagé n'est pas ouvert vers l'extérieur (université, ...)



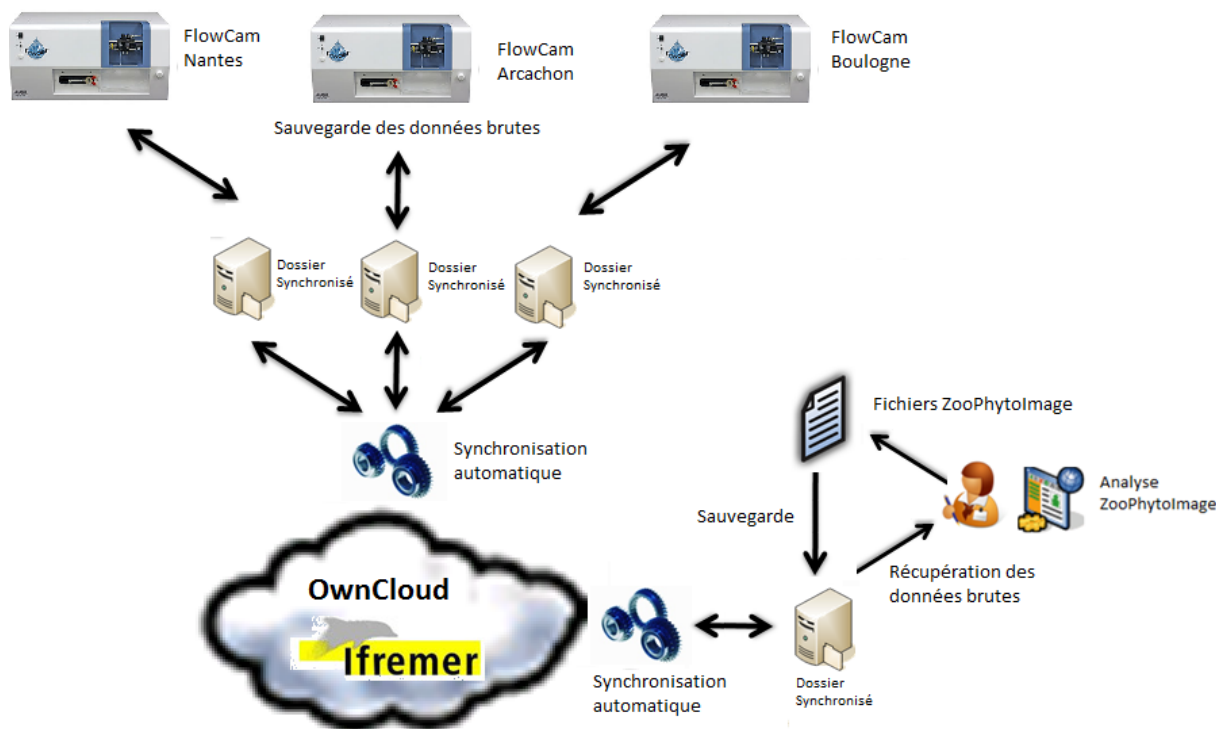
Cette solution bien que fonctionnelle, ne répond pas entièrement aux besoins de l'IFREMER.

Les fonctionnalités manquantes sont les suivantes :

- Ce mode de stockage ne permet pas de diffuser les données sauvegardées vers l'extérieur (exemple : une université partenaire, une personne réalisant un doctorat...)
- Gestion de droit difficile à mettre en œuvre, aucun outil spécifique disponible
- Gestion des actions utilisateur impossible, on ne peut pas savoir qui ajoute, supprime ou modifie des fichiers.

3.1.2 OwnCloud

Cette solution permet de commencer à archiver les données FlowCam/ZooPhytoImage sur un système équivalent à DATARMOR (Cf §2.3) pour la partie stockage. Il faut prévoir une gestion des droits d'accès équivalente à celle du disque réseau partagé. Il est également possible d'étendre cette gestion de droits pour une ouverture vers l'extérieur. OwnCloud utilise un système de synchronisation de dossiers ce qui permet de ne pas changer les méthodes de récupération de fichier.

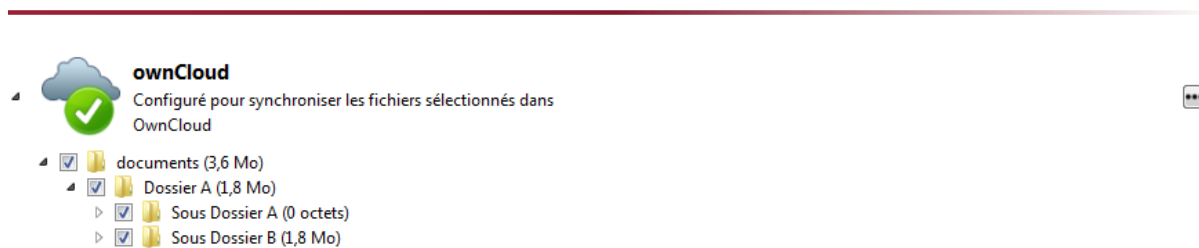


OwnCloud apporte donc des fonctionnalités supplémentaires qui comblent une partie des lacunes de la solution de stockage locale.

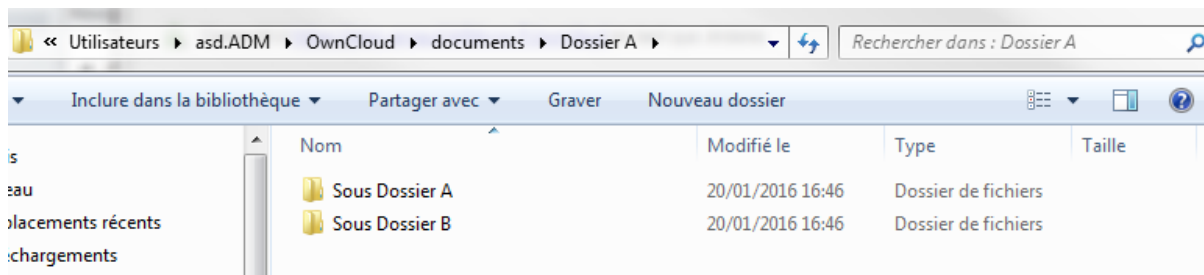
- Diffusion des données vers un intervenant extérieur simplifiée
- Interface d'administration pour gérer les droits des utilisateurs/groupes d'utilisateurs
- Log des actions utilisateur
- Protection renforcée, possibilité de verrouiller les données par mot de passe au cas par cas
- Accès aux données depuis n'importe quel poste ayant une connexion internet

3.1.2.1 Problématique de synchronisation des répertoires

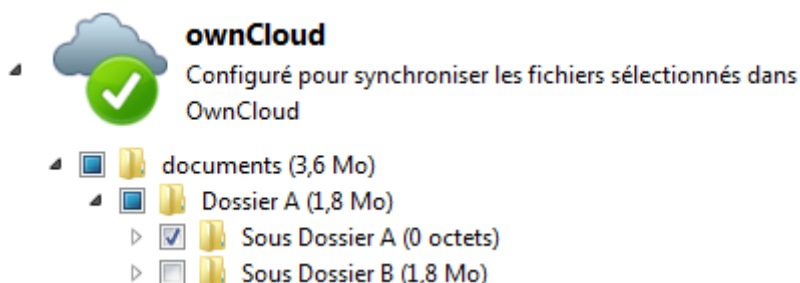
OwnCloud permet de synchroniser une partie ou tous les répertoires avec l'ordinateur sur lequel l'utilisateur souhaite travailler. OwnCloud propose 3 composants principaux, l'application installée sur l'ordinateur, les répertoires synchronisés avec l'ordinateur et une interface web.



Ici, On souhaite avoir sur l'ordinateur le dossier « documents » ainsi que tous les sous dossiers.



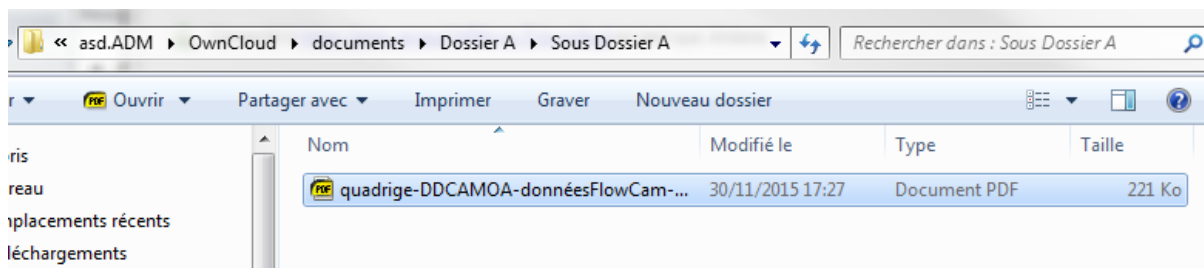
On retrouve tous les dossiers avec leur contenu sur l'ordinateur



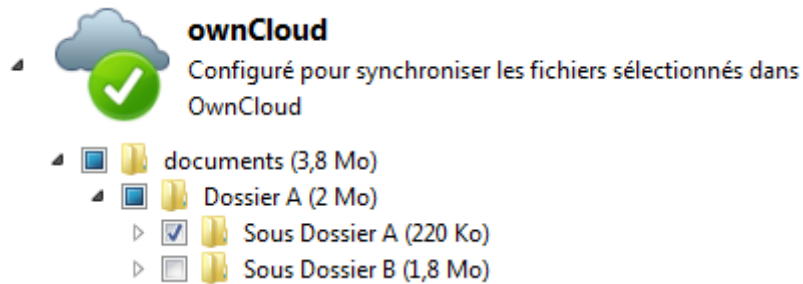
Ici, on ne souhaite garder sur l'ordinateur que le répertoire de travail « Sous Dossier A »



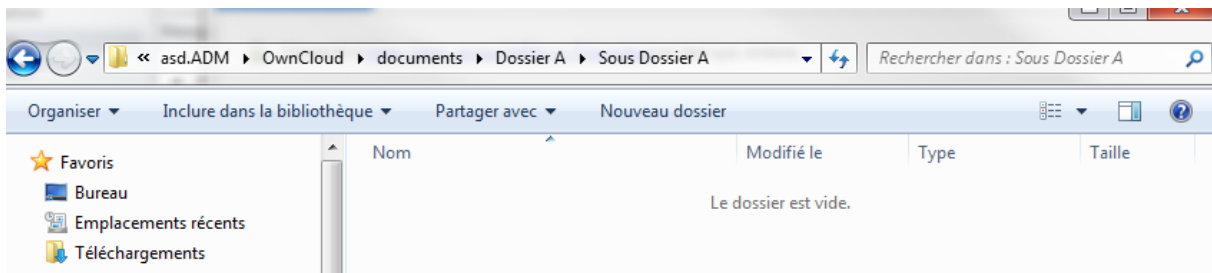
L'arborescence est automatiquement mise à jour sur l'ordinateur de l'utilisateur. Il n'a plus accès qu'au répertoire «Sous Dossier A ». Le contenu des autres dossiers ainsi que les autres dossiers sont toujours présent sur OwnCloud.



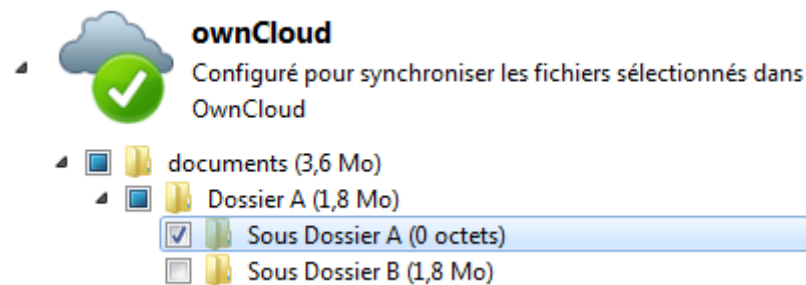
Tout fichier qui est mis dans le dossier « Sous Dossier A » est automatiquement synchronisé avec le server OwnCloud. On ajoute un PDF pour l'exemple.



On constate que « Sous Dossier A » contient maintenant des données (220ko au lieu de 0 précédemment).



On vide le dossier « Sous Dossier A »



Le répertoire « Sous Dossier A » ne contient plus rien.

Il est important de noter qu'il faut décocher le répertoire sur OwnCloud pour supprimer le contenu local de l'ordinateur tout en le conservant sur OwnCloud. Par default, si la personne qui a initiée le partage supprime un fichier partagé de son ordinateur cela le supprime sur OwnCloud.

Activité du serveur
 Protocole de synchronisation
 Non Synchronisé

Protocole de synchronisation locale

Heure	Fichier	Dossier	Action	Taille
20/01/2016 10:26:15	documents/Dossier A/tutu/test	OwnCloud	Téléversé	
20/01/2016 10:21:06	documents/Dossier A/Nouveau dossier	OwnCloud	Effacé	
20/01/2016 10:21:06	documents/Dossier A/tata	OwnCloud	Téléversé	
20/01/2016 10:21:02	documents/Dossier A/tutu	OwnCloud	Téléversé	
20/01/2016 10:20:57	documents/Dossier A/Nouveau dossier	OwnCloud	Téléversé	
20/01/2016 10:20:07	documents/Dossier A/Nouveau dossier	OwnCloud	Déplacé vers documents/Dossier A/Sous Dossier B	
20/01/2016 10:20:01	documents/Dossier A/Nouveau dossier	OwnCloud	Téléversé	
20/01/2016 10:19:54	documents/Dossier A/Nouveau dossier	OwnCloud	Déplacé vers documents/Dossier A/Sous Dossier A	
20/01/2016 10:19:49	documents/Dossier A/Nouveau dossier	OwnCloud	Téléversé	
20/01/2016 10:19:45	documents/Nouveau dossier	OwnCloud	Déplacé vers documents/Dossier A	
20/01/2016 10:19:40	documents/Nouveau dossier	OwnCloud	Téléversé	
20/01/2016 10:09:53	ownCloudUserManual.pdf	OwnCloud	Déplacé vers documents/ownCloudUserManual.pdf	
20/01/2016 10:08:38	ownCloudUserManual.pdf	OwnCloud	Téléchargé	1,8 Mo
20/01/2016 10:08:23	documents/example.odt	OwnCloud	Téléchargé	23 Ko
20/01/2016 10:08:23	documents	OwnCloud	Téléchargé	

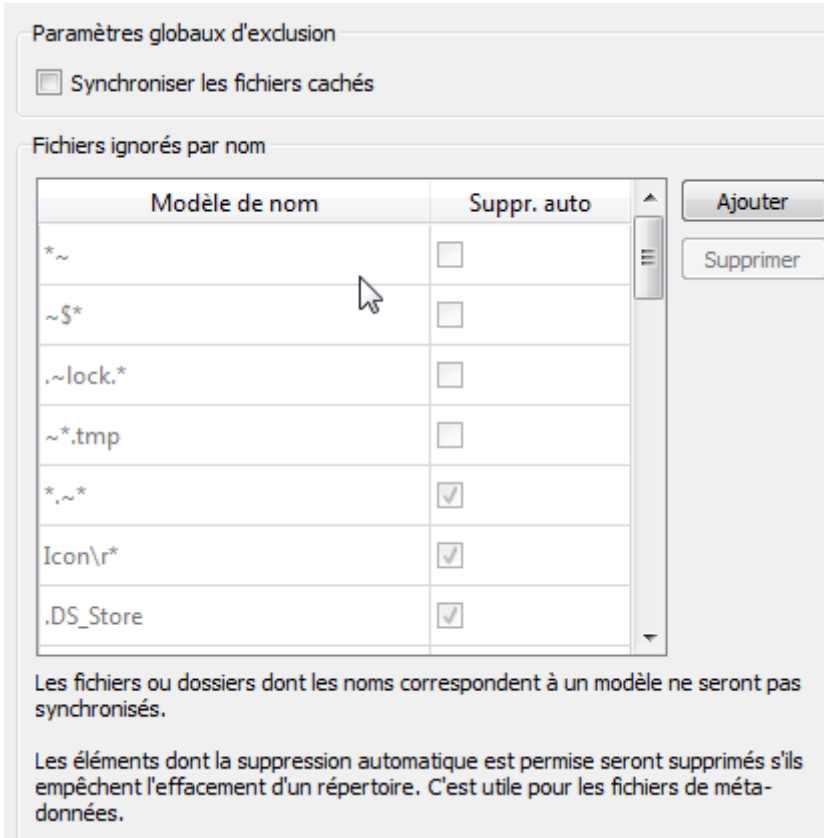
Copier

Un journal des actions effectuées est disponible.

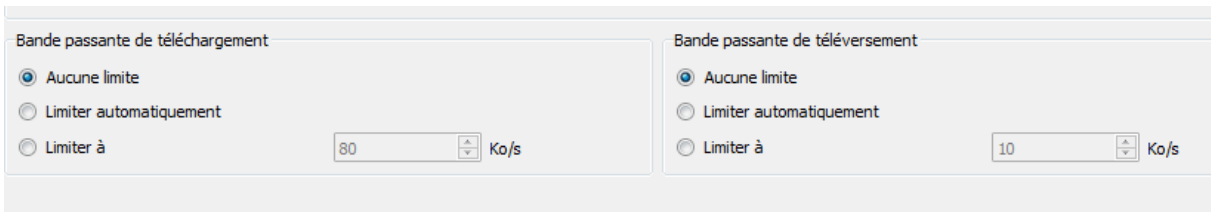
Demander confirmation avant de télécharger les dossiers de taille supérieure à Mo

Il est possible de demander à OwnCloud de proposer de synchroniser ou non un dossier en fonction de sa taille.

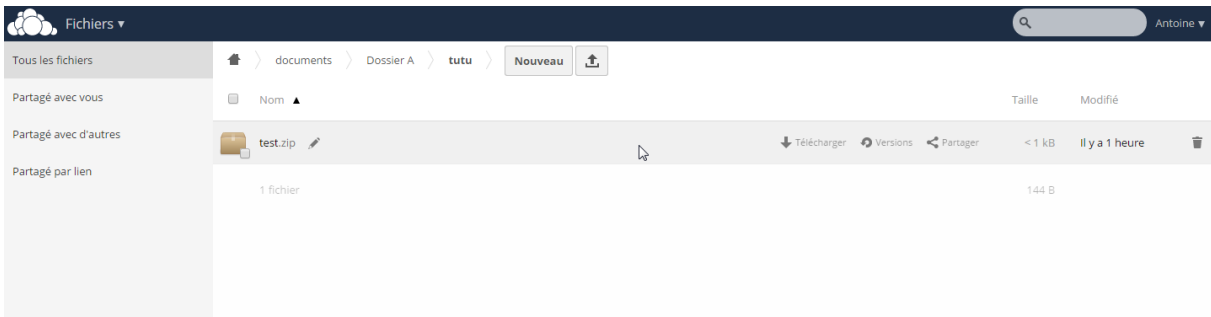




Il est possible d'exclure systématiquement des fichiers ou dossiers en fonction de leur nom, extension, ...

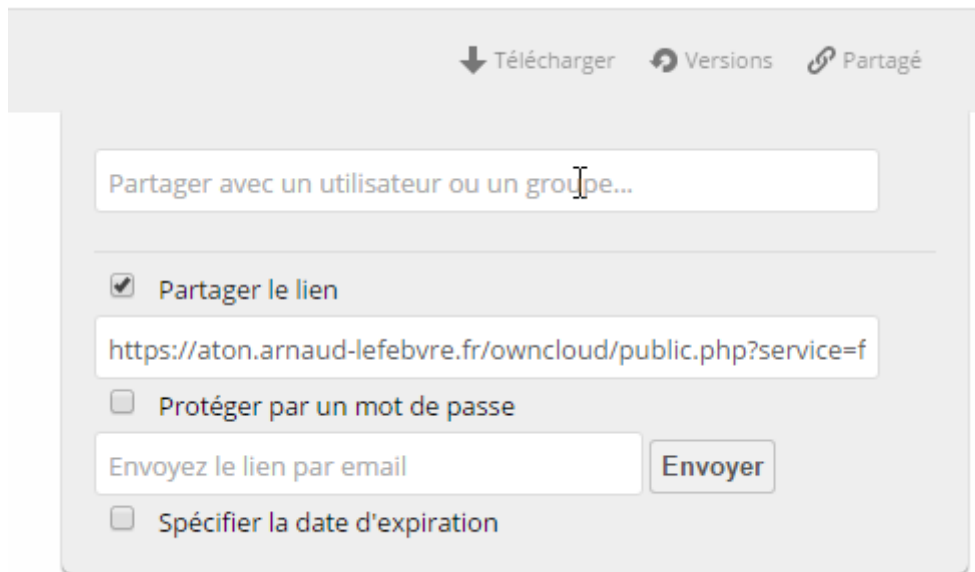


Il est possible de fixer des limites dans la vitesse de téléchargement ou de dépose des fichiers sur le serveur afin de partager la bande passante si besoin. Cela permet d'éviter que d'autres utilisateurs du réseau voient leurs performances réduite par un transfert de donnée vers OwnCloud.



Il existe également une interface web permettant elle aussi de gérer les fichiers et dossiers.





↓ Télécharger Versions Partagé

Partager avec un utilisateur ou un groupe...

Partager le lien

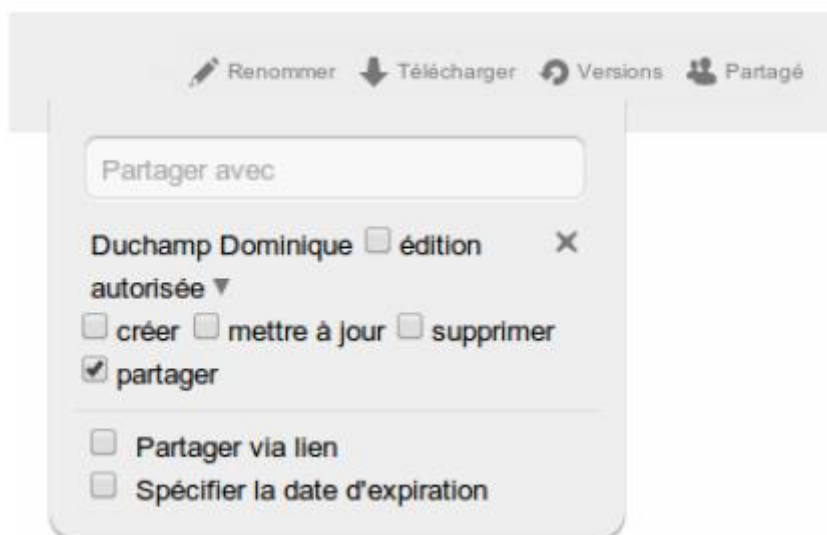
<https://aton.arnaud-lefebvre.fr/owncloud/public.php?service=f>

Protéger par un mot de passe

Envoyez le lien par email Envoyer

Spécifier la date d'expiration

Cette interface web permet de définir les options de partages des fichiers/dossiers. Soit pour un utilisateur en particulier, soit pour un groupe d'utilisateur.



Renommer ↓ Télécharger Versions Partagé

Partager avec

Duchamp Dominique édition ✕

autorisée ▼

créer mettre à jour supprimer

partager

Partager via lien

Spécifier la date d'expiration

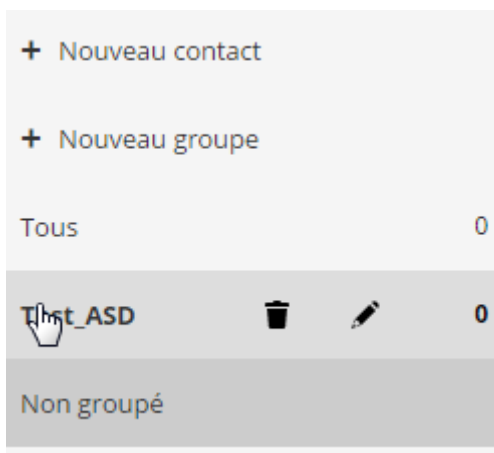
Par défaut un fichier est partagé en lecture seule. Il est par contre possible de spécifier le mode de partage afin de donner plus de droits aux utilisateurs.

Protéger par un mot de passe

Il est même possible de protéger un fichier sensible par mot de passe.



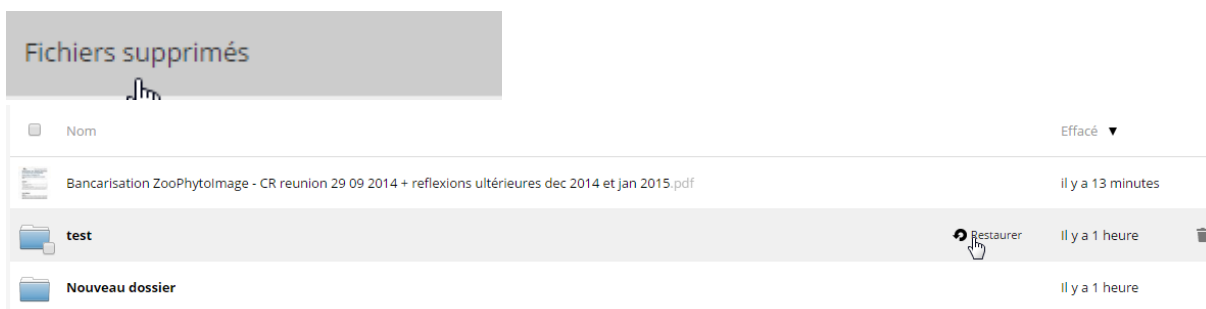
Il est possible de retrouver rapidement tout ce qui est partagé.



Afin de gérer le partage il est possible de créer des contacts et des groupes de contacts.

**Vous n'avez aucun contact dans votre carnet d'adresses ou votre carnet d'adresses est désactivé.
Add a new contact or import existing contacts from a file (VCF, CSV or LDIF).**

Il est possible d'importer des contacts selon des formats prédéfinis pour gagner du temps



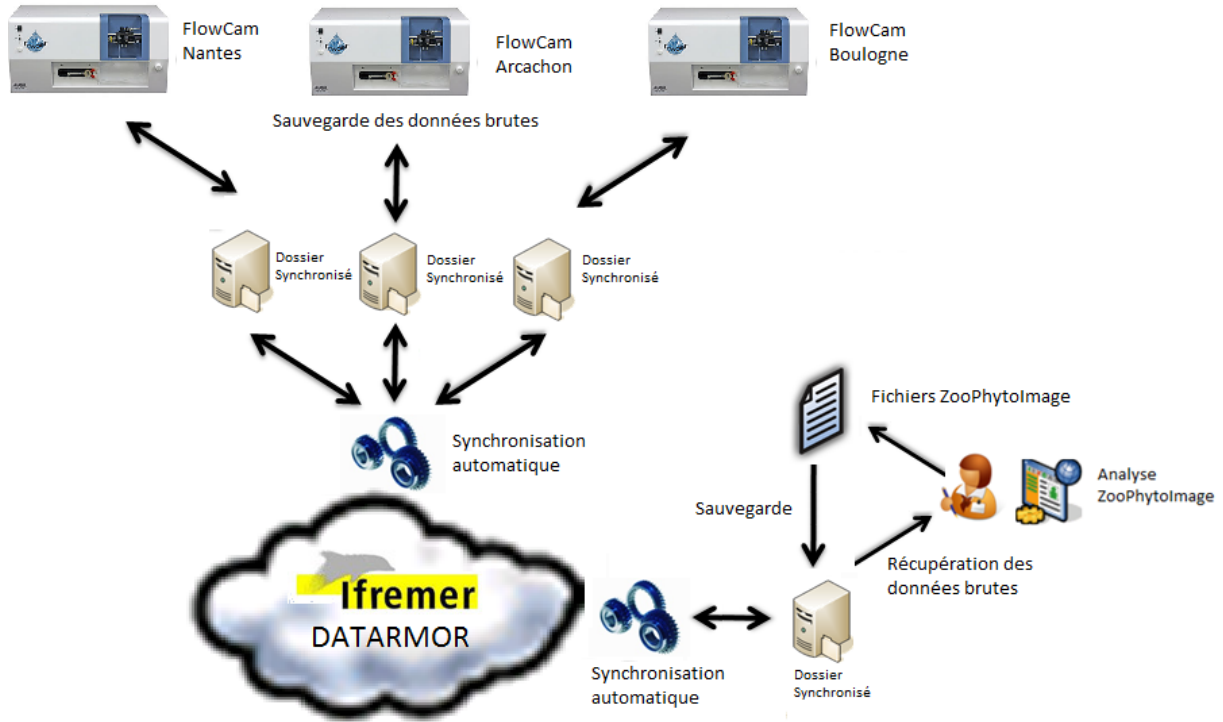
Le site web permet également de consulter les fichiers supprimés et de les restaurer si besoin.

OwnCloud peut donc se substituer à la solution de stockage local de manière presque transparente pour l'utilisateur. Des manipulations pour choisir quel répertoire synchroniser et quand seront tout de même nécessaires chaque fois qu'un utilisateur souhaite travailler. Les manipulations restent simples et peu consommatrices en temps.



3.1.3 DATARMOR

Cette solution est à mettre en œuvre après le passage sur OwnCloud. Il sera intéressant de noter les éventuels défauts d’OwnCloud lors de son utilisation afin de voir s’il est possible de les corriger avec DATARMOR.

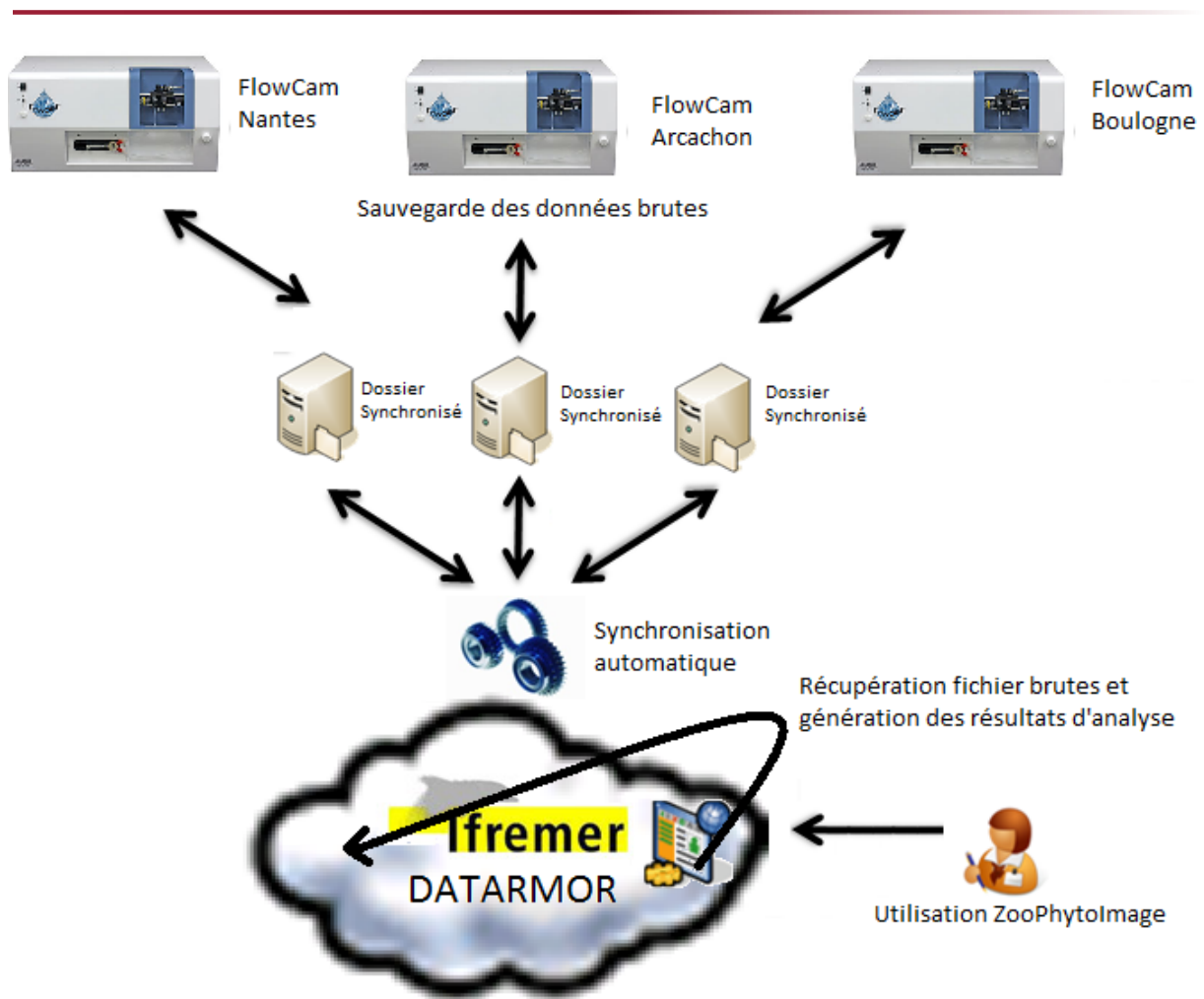


3.1.3.1 Problématique de synchronisation des répertoires

DATARMOR n’étant pas encore disponible on ne peut que supposer que le fonctionnement sera au moins équivalent à celui d’OwnCloud. Comme indiqué précédemment il faut noter ce qui ne convient pas ou peut être amélioré dans l’utilisation d’OwnCloud pour que ces fonctionnalités soit éventuellement développées sur DATARMOR.

3.1.4 ZooPhytoImage installé sur DATARMOR

Cette solution n’est permise que s’il est possible d’installer ZooPhytoImage directement sur DATARMOR. Cette solution est très semblable à celle où DATARMOR et ZooPhytoImage sont utilisé de manière séparée mais permet des performances d’accès supérieurs au prix de développement supplémentaires.



3.1.4.1 Nécessaire pour l'installation de ZooPhytoImage sur DATARMOR.

- DATARMOR devra allouer plusieurs threads processeur à ZooPhytoImage pour garantir une bonne vitesse de traitement
- DATARMOR devra allouer à minima 4GB de ram à ZooPhytoImage
- DATARMOR devra proposer une plateforme 64 bits pour installer ZooPhytoImage
- La version 5.4-0 de ZooPhytoImage demande la version 3.0.x ou supérieur de R
- DATARMOR doit permettre d'ouvrir Safari ou Google Chrome car l'interface web est nécessaire pour l'utilisation de ZooPhytoImage
- Il devra être possible de lancer RGui (l'interface graphique de R)
- Il devra être possible d'installer les package R « shiny », « tiff » et « zooimage » depuis RGui soit en ligne de commande soit via le menu « package »

Bien que la solution ne soit pas encore disponible, si l'on ne souhaite pas faire de développement spécifique pour ZooPhytoImage, cette liste de prérequis impliquent la présence d'une Machine Virtuelle (VM) a minima sous Windows 7 64 bits avec au moins 4 cœur et de 8 à 16GB de ram afin de permettre l'installation et l'utilisation de ZooPhytoImage dans de bonnes conditions. Cette VM sera reliée directement aux répertoires de stockage DATARMOR (on peut imaginer qu'il sera possible de les monter comme des répertoires réseau). Cette solution implique également une connexion de rapide et stable entre l'ordinateur de l'utilisateur et la VM pour ne

pas dégrader l'expérience d'utilisation. Il faudra également prévoir plusieurs VM pour permettre à plusieurs utilisateurs de travailler en même temps.

3.2 Solutions de bancarisation

Les solutions de bancarisation ne doivent pas impliquer de faire évoluer FlowCam ou ZooPhytoImage car l'IFREMER n'a pas la main sur le développement de ces applications. La seule évolution tolérée pour ZooPhytoImage serait l'ajout d'une ligne de commande permettant de générer des fichiers csv pour les résultats d'analyse.

3.2.1 Intégration des données

Actuellement, ZooPhytoImage génère des fichiers avec l'extension .RDATA. Ces fichiers ne sont pas exploitables en l'état, il faut donc les convertir en CSV. Cette fonctionnalité peut être intégrée directement au processus de génération des fichiers avec une commande R :

```
write.csv(`NT.058.2014-04-07.300A4X.03_valid`, file="NT.058.2014-04-07.300A4X.03_valid.csv")
```

Actuellement un script à exécuter après la génération des fichiers RDATA est disponible pour effectuer la conversion en CSV. Il est également possible de gérer cette conversion directement avec R.

- Les fichiers qui seront bancarisés dans Q² sont les suivants :
 - ..._res.RDATA à convertir en fichier .csv
 - ..._valid.RDATA à convertir en fichier .csv
- Les fichiers qui ne seront pas bancarisés dans Q² sont les suivants :
 - notes.txt (FlowCam)
 - run_summary.txt(FlowCam)
 - ..._valid.txt (ZooPhytoImage)
 - Les graphiques statistiques sur les corrections effectuées dans ZooPhytoImage

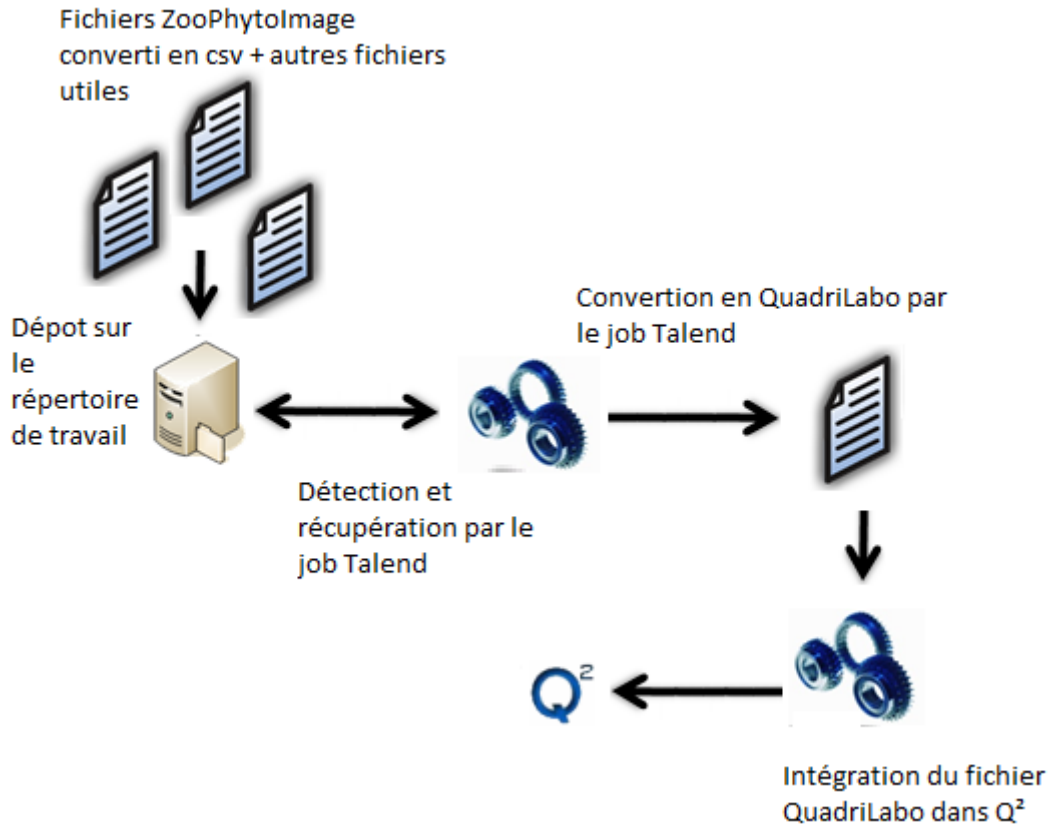
A noter que le contenu de ces fichiers peut être utilisé pour la bancarisation des données. Par exemple le fichier ..._valid.txt contient les informations sur le set d'apprentissage, le classifieur, l'algorithme, le nom de l'analyste, la date de l'analyse...

Une fois converti en .csv il sera possible d'intégrer les données contenues dans ces fichiers en les déposant sur un répertoire de travail via une des méthodes abordées dans les points suivants de ce document. Ils seront ensuite détectés par le job Talend qui les convertira en format QuadriLabo et lancera en suite l'intégration dans Q². Attention aux fichiers non bancarisés comme ..._valid.txt qui nécessiteront quand même d'être déposés sur le répertoire de travail pour permettre au job Talend d'en extraire les informations utiles pour l'intégration. Le mécanisme d'intégration de fichier QuadriLabo existe déjà et fonctionne. Aucun développement à faire de ce côté.

La nomenclature actuelle des fichiers générés par ZooPhytoImage du type « NT.058.2014-04-07.300A4X.03 » devra être modifiée afin de conserver des informations importantes dans le nom de fichier et faire le lien entre le mnémonique Q² d'un échantillon et tous les fichiers à charger. Le

nouveau format sera CODE_FLOWCAM-MNEMO_LIEU-DATE_PASSAGE(aaaammjj)_n°implémentation ce qui donnera par exemple « BL.001-P-058.2015-11-25.4X.01 »

Actuellement, seuls les taxons seront bancarisés dans Q². Les déchets sont ignorés (classe « other » et « total » des fichiers csv.)



3.2.2 Format QuadriLabo

Les informations dont nous disposons actuellement nous permettent de traiter uniquement le fichier ..._res.csv. L'équivalence QuadriLabo de ..._valid.csv devra être complétée par l'IFREMER. Le fichier ..._res.csv peut tout de même être bancarisé sans ..._valid.csv.

Le fichier d'équivalence QuadriLabo - ..._res.csv :



Fichier_QuadriLabo_r
emplissage_ZooPhytc

Comme toutes les données nécessaires à la création d'un fichier QuadriLabo ne sont pas disponibles dans les résultats d'analyse de FlowCam/ZooPhytoImage, le job Talend de conversion devra aller récupérer ces informations directement dans Q² au niveau des métadonnées.

3.2.3 Intégration automatique unitaire

Cette solution est à mettre en place pour les utilisateurs de ZooPhytoImage. Pour chaque échantillon, l'utilisateur déclenche le processus de bancarisation des résultats. Pour se faire, la solution retenue sera d'utiliser une interface web du même type que QuadriLabo :

QuadriLabo intégration données



L'outil permet d'intégrer vos données dans QuadriLabo.

L'outil vérifie si votre fichier respecte le format QuadriLabo et si les données sont intégrables à QuadriLabo.

Si vos données respectent le format préconisé, vos données sont intégrées.

Vous recevrez un mail de notification.

Reprise des résultats au format QuadriLabo

Adresse e-mail :

Fichier : Aucun fichier choisi

Le fichier doit être au format excel (.xls, .xlsx) ou format .csv avec séparateur ";"

Cette solution est la plus intéressante car elle dissocie l'outil de chargement de donnée des évolutions de l'application Q². Une fois uploadé via l'interface web, ces fichiers seront détectés par un job Talend qui assurera la conversion au format QuadriLabo et lancera l'intégration dans Q². La condition au bon fonctionnement de ce processus étant que les métadonnées nécessaires soit présente dans Q² avant l'intégration.

L'interface web devra proposer les mêmes fonctionnalités que l'interface web QuadriLabo mais en plus étoffée :

- Prévoir un champ « CODE RESEAU » qui sera dans un premier temps non modifiable et forcé à REPHY
- Un contrôle sur le champ « adresse e-mail » afin que seuls des adresses ayant une correspondance avec un « CODE SANDRE SAISISSEUR » ne soit utilisables.
- Un contrôle sur le nom des fichiers afin de vérifier que leurs noms sont correctement formatés.
- Un contrôle sur les types et la tailles des fichiers uploadé peut aussi être mis en place.

Ces contrôles seront également présent dans le job Talend mais il peut être intéressant de les dédoublés sur l'interface web pour le confort de l'utilisateur.

3.2.4 Intégration automatique en masse

Cette fonctionnalité doit rester à la main des administrateurs. Cette fonctionnalité sera mise en place sous la forme d'un job Talend qui prend en paramètre des fichiers déposés au préalable dans un répertoire spécifique pour l'intégration des données. Il sera donc possible de déposer le résultat de plusieurs analyses en même temps, la seule condition étant que les métadonnées nécessaires soient présentes dans Q².

Une interface graphique pour les administrateurs n'est pas nécessaire pour ce processus car la reprise de données peut être déléguée à la cellule Q² qui s'occupe déjà des reprises de données. Comme il n'y a pas d'interface pour l'import en masse, c'est l'analyste qui sera reporté comme saisisseur de la donnée dans Q².

3.2.5 Détail de l'intégration dans Q²

3.2.5.1 Métadonnées nécessaires

Afin de réaliser l'import QuadriLabo, des métadonnées doivent être créées dans Q² en amont. Il s'agit des PSFM associés ainsi que les passages/prélèvement/échantillons.

Quelques contraintes spécifiques au processus s'appliquent :

- Le libellé de la méthode du PSFM doit contenir une concaténation du set d'apprentissage de ZooPhytoImage, l'algorithme ainsi que le classifieur utilisé pour l'analyse et la version de ZooPhytoImage.
- Le mnémonique de l'échantillon sera à remplir selon la même nomenclature que celle des fichiers ZooPhytoImage soit `CODE_FLOWCAM-MNEMO_LIEU-DATE_PASSAGE(aaaammjj)_n°implémentation` ce qui donnera par exemple « BL.001-P-058.2015-11-25.4X.01 »

Il existe des contraintes spécifiques aux métadonnées pour éviter que QuadriLabo les recrée. Ces champs doivent être créés dans Q² avant l'intégration des résultats d'analyse :

- Passage
 - Identifiant du lieu de surveillance auquel est rattaché le passage
 - Date du passage
 - Heure du passage
 - Mnémonique du passage
- Prélèvement
 - Identifiant du passage auquel est rattaché le prélèvement
 - Identifiant de l'engin de prélèvement
 - Identifiant du niveau du prélèvement
 - Heure du prélèvement
 - Mnémonique prélèvement
- Echantillon
 - Identifiant du prélèvement
 - Identifiant du support
 - Mnémonique de l'échantillon
 - Identifiant interne du taxon référent.
 - Identifiant interne du groupe de taxon.

Afin de gérer le stockage des résultats de plusieurs analyses du même échantillon où réplicats, deux méthodes peuvent être utilisées :



- On peut créer plusieurs échantillons différents avec un réplicat par échantillon.
- On peut créer plusieurs réplicats sur un même échantillon mais il faudra faire attention à pouvoir identifier le résultat provenant de chaque analyse.

3.2.5.2 Gestion des droits

La gestion de droits Q² implique que la personne qui insère les données dans Q² (via l'interface web pour un utilisateur ou via dépôt de fichier pour un administrateur) soit reconnue au niveau Q² comme saisisseur ou responsable du programme des données insérées.

Pour ce faire, il faut que le login saisi dans ZooPhytoImage et reporté en suite dans ..._valid.txt ou que le login saisi dans l'interface web utilisé lors de l'intégration automatique unitaire ait une correspondance dans Q².

L'interface web pourra facilement contrôler la validité de l'email utilisé lors de l'upload des fichiers (comme pour le « CODE SANDRE SAISISSEUR »).

Pour ce qui est du login dans ..._valid.txt, étant donné qu'on ne doit pas faire évoluer ZooPhytoImage, il n'y aura pas de contrôle. La charge est donc à l'utilisateur de faire très attention lors de la saisie de son nom d'utilisateur.

3.2.5.3 Contrôles nécessaires

Des contrôles seront à mettre en place coté Q² pour le formatage du libellé de la méthode ainsi que du mnémonique de l'échantillon afin d'éviter d'avoir des problèmes de rapprochement.

Cependant, ces contrôles ne doivent s'exercer que dans le cas d'une méthode ou d'un échantillon destiné à être lié à une analyse ZooPhytoImage. Il faudrait donc prévoir un moyen de l'indiquer pour déclencher les contrôles. Il est possible par exemple de proposer une simple « check box » pour indiquer l'usage qui sera fait de l'échantillon ou de la méthode.

3.2.5.4 Gestion du multiprogramme

Pour gérer le multiprogramme on peut :

- Mettre le code programme en dur sur REPHY
- Laisser le choix du code lors de l'upload des fichiers à bancariser

3.2.5.5 Gestion des doublons

Afin d'éviter que le job Talend n'intègre des fichiers en double il faut prévoir en plus du répertoire de travail, un répertoire pour les fichiers intégrés ainsi qu'un répertoire pour les fichiers rejetés.

Il faudra donc prévoir une purge régulière de ces fichiers afin qu'ils ne prennent pas trop d'espace disque. Les fichiers originaux au format RDATA étant sauvegardés, il n'y a pas de risque de perte de donnée et il sera toujours possible de convertir de nouveau un fichier RDATA en csv pour retenter une intégration si besoin.

3.2.5.6 Impact sur les performances de la base de donnée

Le fichier ..._valid.csv peut poser problème lors de son intégration si on intègre toute les lignes qu'il contient. En effet, un fichier fait en moyenne 6000 lignes, si on intègre 10 résultats par ligne il y aura 60 000 lignes en base de données Q².

Pour illustrer ce problème de performance potentiel une simple requête SQL a été exécutée sur un environnement de teste sur une table de 54669 lignes et 18 colonnes avec affichage des résultats à l'écran. Les performances de l'environnement de test ne sont pas les mêmes qu'en production mais cela permet de se faire une idée des impacts :

```
select * from TAXON_NAME
```

```
54669 ligne(s) selectionnee(s).
```

```
Ecoule : 00 :11 :28.50
```

Il a donc fallu 11 minutes et 28 secondes pour exécuter la requête et afficher les résultats. Le temps de traitement coté Q² sera « largement » plus long car les données seront de nouveau traitées dans le code Java. La consommation en mémoire de l'application Q² risque elle aussi d'augmenter de manière conséquente ce qui peut avoir des impacts sur la vitesse d'utilisation de l'application voir la rendre inopérante sur certains postes utilisateur peu pourvu en mémoire vive. On pourra rajouter à ces problèmes le fait que la quantité massive de donnée à remonter de la base de données peut mettre du temps à être récupéré coté Q² si les performances réseau sont insuffisantes.

3.2.5.7 Transcodification

Les noms de colonne des fichiers CSV générés sont attribués par ZooPhytoImage, et ne seront pas modifiés. On peut donc les utiliser pour le transcodage.

Le mécanisme à mettre en place pour le transcodage sera le même que celui qui a été mis en place dans BDRecif et s'effectuera au niveau de la base de donnée avec des tables spécifiques.

Une table de transcodage sera également nécessaire pour les taxons.

3.2.5.8 Problématique de validation Q² des données

Lors d'une intégration via QuadriLabo les résultats sont systématiquement validés.

3.2.5.9 Gestion des vignettes

Si les vignettes sont bancarisées dans Q² cela permettra de les retrouver via une extraction. Cette bancarisation serait possible dans les photos des échantillons :

- Dans le libellé de la photo avec le numéro échantillon et le numéro vignette
- Dans la légende de la photo avec la classe égale au taxon : dans ce cas, avec les critères d'extraction Q² actuels, on ne pourrait pas mettre de critère sur le taxon car le champ « légende » ne fait pas partie des critères d'extraction Q².

3.2.6 Gestion du classifieur et du set d'apprentissage

Le set d'apprentissage et le classifieur seront sauvegardés de la même manière que le reste des données à sauvegarder c'est-à-dire selon la même arborescence que celle présente sur le disque réseau actuel. Dans sa première version le set d'apprentissage fait 412mo. Cette taille peut augmenter ou diminuer en fonction de ce qu'on inclut dans le set.

Arborescence des dossiers utilisée actuellement :



Arborescence_analyse-echantillon.docx

3.2.7 ZooCam

Les solutions DATARMOR/OwnCloud sont aussi intéressantes pour les utilisateurs de ZooCam car ils ont également des besoins d'espace de stockage collaboratif pour permettre à plusieurs experts de participer à l'analyse d'un même échantillon. Par contre ZooCam n'utilise pas R, il faut donc prévoir une méthode d'intégration particulière pour les résultats issus de ZooCam. Il est a priori possible de convertir les résultats issus de ZooCam en fichier CSV.

4 SOLUTIONS EXCLUES

4.1 Solutions de sauvegarde

4.1.1 Sauvegarde sur disque réseaux individuels

Cette solution permettait notamment de dimensionner différemment les disques des LERs selon le volume des données, mais :

- Pas besoin de dimensionner les disques des LERs (charge d'analyse répartie équitablement sur l'ensemble du littoral)
- Le partage des données entre les LERs deviendrait trop complexe
- Les autres contraintes liées au stockage sur disque réseau sont toujours présente.

4.1.2 Sauvegarde sur le CLOUD

Cette solution est l'équivalent d'OwnCloud/DATARMOR en moins bien car externalisé et donc moins maîtrisé. Elle ne présente donc pas d'intérêt.

4.2 Solutions de bancarisation

4.2.1 Bancarisation depuis ZooPhytoImage

Cette solution a été exclue car l'IFREMER n'a pas la main sur ZooPhytoImage, il n'est donc pas possible de le faire évoluer pour permettre ce système de bancarisation

4.2.2 Récupération des ZID/ZIDB pour traitement et bancarisation

Toutes les solutions impliquant une utilisation de ces archives ont été exclues car la gestion des ZID/ZIDB est complexe. De plus, tous les éléments contenu dans ces archives ne sont pas utiles pour la bancarisation et il est plus simple de convertir les résultats « utiles » (fichier « .RDATA ») de ZooPhytoImage en fichier csv facilement exploitable.

4.2.3 Bancarisation directe depuis Q²

Les solutions impliquant une récupération des données à bancariser directement depuis Q² ont été exclues car elles impliquent de nombreux développements dans Q² alors que Q³ va le remplacer. De plus, cela rend Q² dépendant de la modélisation des données produites par ZooPhytoImage et tout changement impliquerait de faire évoluer Q².

4.2.4 Format EdiLabo

Ce format a été abandonné au profit de QuadriLabo qui est plus abouti et qui est déjà utilisé pour l'intégration de donnée dans Q².

5 CONCLUSION

L'étude montre l'intérêt de sauvegarder les données FlowCam et ZooPhytoImage avec la solution OwnCloud Ifremer afin de se familiariser avec le fonctionnement d'une sauvegarde sur Cloud, de partager plus facilement les données sauvegardées avec des intervenants extérieurs et d'attendre que DATARMOR soit disponible.

En fonction de l'évolution de DATARMOR il sera ensuite possible de migrer les données sauvegardée sur ce support puis au final d'installer ZooPhytoImage directement sur DATARMOR afin de profiter de la puissance de calcul de cette solution tout en s'affranchissant des contraintes de sauvegarde sur support distant. La méthode de stockage DATARMOR est donc la cible du fait de ses multiples avantages. L'intérêt de cette solution réside dans son intégration dans l'écosystème IFREMER, les facilités de diffusion des données qu'elle propose ainsi que ces éventuelles connexions avec d'autres données pouvant servir dans le cadre de ces analyses.

Le fait d'utiliser le format QuadriLabo et des jobs Talend permet :

- La mise en place de solutions indépendantes
- La non adhérence des processus de sauvegarde, de bancarisation
- L'application Q2 ne dépend pas des solutions retenues
- La seule contrainte imposée par cette solution est de créer des métadonnées avant de faire l'intégration.

Cela permet de réduire les coûts de développement tout en dé-corrélant la bancarisation des données des évolutions futures de Quadri 2. Ce mode de fonctionnement regroupe les fonctionnalités d'intégration de masse et d'intégration unitaire selon le niveau d'habilitation des utilisateurs et permet donc de faire des reprises massives si besoin.



Arborescence de travail pour
l'analyse d'un échantillon par
ZooPhytoImage

Nadine Neaud-Masson

Décembre 2015

Arborescence de travail pour analyse d'échantillon par ZooPhytoImage

Contient les fichiers zidb générés par ZooPhytoImage lors du traitement des fichiers bruts sortis FlowCAM. Fichiers compressés contenant les vignettes individuelles des particules (jpg) + les mesures + les métadonnées:

BN.072.2015-04-14.300A4X.01_882.jpg	Image JPEG	1 544
BN.072.2015-04-14.300A4X.01_883.jpg	Image JPEG	1 421
BN.072.2015-04-14.300A4X.01_884.jpg	Image JPEG	1 300
BN.072.2015-04-14.300A4X.01_dat1.zim	Zoolmage Metadata	509 043
BN.072.2015-04-14.300A4X.01_dat1.RData	R Workspace	183 516

Contient le fichier méthode d'analyse :

Rephy 4X lugol v.1.0.R

Contient les fichiers résultats d'analyse :

N.309.2015-08-10.300A4X.01_res.RData	15/12/2015 17:18	R Workspace	2 Ko
N.309.2015-08-10.300A4X.01_valid.RData	15/12/2015 17:18	R Workspace	951 Ko
N.309.2015-08-10.300A4X.01_valid.txt	15/12/2015 17:18	Document texte	1 Ko

Contient les fichiers R de l'outil de reconnaissance généré à partir du set d'apprentissage stocké dans le sous répertoire : trainRephy_4Xlugol.01 :

classrfRephy_4Xlugol.01.RData	14/12/2015 12:54	R Workspace	5 759 Ko
trainRephy_4Xlugol.01.RData	14/12/2015 12:52	R Workspace	4 981 Ko

Echantillon analysé, contient les fichiers bruts sortis du Flowcam :

cal_image_000001.tif	02/09/2015 14:05	Image TIFF	1 208 Ko
N.309.2015-08-10.300A4X.01.csv	02/09/2015 14:19	Fichier CSV Micro...	2 295 Ko
N.309.2015-08-10.300A4X.01.ctx	02/09/2015 14:18	Fichier CTX	3 Ko
N.309.2015-08-10.300A4X.01.edg	02/09/2015 14:18	Fichier EDG	1 304 Ko
N.309.2015-08-10.300A4X.01.lst	02/09/2015 14:18	Fichier LST	2 779 Ko
N.309.2015-08-10.300A4X.01_000001.tif	02/09/2015 14:05	Image TIFF	1 208 Ko
N.309.2015-08-10.300A4X.01_000002.tif	02/09/2015 14:05	Image TIFF	1 208 Ko
N.309.2015-08-10.300A4X.01_000003.tif	02/09/2015 14:05	Image TIFF	1 208 Ko
N.309.2015-08-10.300A4X.01_000004.tif	02/09/2015 14:05	Image TIFF	1 208 Ko
N.309.2015-08-10.300A4X.01_notes.txt	02/09/2015 14:18	Document texte	1 Ko
N.309.2015-08-10.300A4X.01_run_summa...	02/09/2015 14:18	Document texte	1 Ko
N.309.2015-08-10.300A4X.01_summary.csv	02/09/2015 14:19	Fichier CSV Micro...	1 Ko



Ifremer

Éléments pour remplir le fichier Quadrilabo,
servant d'intermédiaire entre les fichiers csv en
provenance de ZooPhytoImage, et Quadrige

Emilie Gauthier

Février 2016

Nom de la colonne	Définition	Pour ZooPhytoImage - fichier ...res.csv
NUMERO_LIGNE	Numéro de ligne incrémenté : pas de doublon!	Numéro incrémenté par le script d'intégration
CODE_LIEU_DE_SURVEILLANCE	Code Sandre du lieu de surveillance	Récupérer dans Q ² à partir de l'échantillon
CODE_RESEAU	Code Sandre du dispositif de collecte (= programme Quadrige ²)	REPHY : sélectionné par défaut (et non modifiable dans un premier temps) dans l'interface de chargement des fichiers
CODE_SANDRE_SAISSISSEUR	Code Sandre de l'intervenant saisisseur de la donnée	Code Sandre du service auquel appartient la personne qui charge le fichier (identification à partir du mail dans l'interface)
ZONE_DESTINATION_DRAGAGE	Zone de destination dragage Quadrige ²	Vide
CAMPAGNE	Libellé de la campagne Quadrige ²	Vide
SORTIE	Libellé de la sortie Quadrige ²	Vide
DATE_PRLVMT	Date à laquelle a été effectué le prélèvement terrain ou l'observation	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
HEURE_PASSAGE	Heure à laquelle a été visité le lieu de surveillance	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
SONDE	Hauteur d'eau sur le lieu au moment du passage	Vide
UNITE_SONDE	Code Sandre de l'unité dans laquelle est exprimée la hauteur d'eau (mètre)	Vide
MNEMONIQUE_PASSAGE	Mnémonique du passage	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
COMMENTAIRES_PASSAGE	Commentaires sur le passage	Vide
LATITUDE_PASSAGE	Latitude en WGS84 degrés décimaux du passage : à ne remplir que si différent du lieu de surveillance	Vide
LONGITUDE_PASSAGE	Longitude en WGS84 degrés décimaux du passage : à ne remplir que si différent du lieu de surveillance	Vide
POSITIONNEMENT_PASSAGE	Façon dont le passage a été positionné : code Sandre du positionnement dans la nomenclature [598] "Type d'acquisition des coordonnées du lieu de surveillance"	Vide
NOMBRE_INDIVIDU_PASSAGE	Nombre d'individus observés / analysés sur le passage	Vide
ENGIN_PRLVMT	Code Sandre de l'engin de prélèvement dans la nomenclature [603] "Engin de prélèvement"	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
NIVEAU_PRLVMT	Code Sandre du niveau auquel est effectué le prélèvement : nomenclature [202] "Niveau d'immersion"	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
CODE_SANDRE_PRELEVEUR	Code Sandre de l'organisme préleveur (référentiel Intervenant du Sandre)	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
MNEMONIQUE_PRLVMT	Mnémonique du prélèvement	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
IMMERSION_PRLVMT	Valeur de la profondeur d'immersion à laquelle est effectuée le prélèvement	Vide
IMMERSION_MAX_PRLVMT	Valeur maximale de la profondeur d'immersion	Vide
IMMERSION_MIN_PRLVMT	Valeur minimale de la profondeur d'immersion	Vide
UNITE_IMMERSION	Code Sandre de l'unité dans laquelle est exprimée l'immersion	Vide
TAILLE_PRLVMT	Quantité de "matrice" prélevée ou observée	Vide
UNITE_TAILLE_PRLVMT	Code Sandre de l'unité dans laquelle est exprimée la taille du prélèvement	Vide
HEURE_PRLVMT	Heure à laquelle a été effectué le prélèvement	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
COMMENTAIRES_PRELEVEMENT	Commentaires sur le prélèvement	Vide
LATITUDE_PRELEVEMENT	Latitude en WGS84 degrés décimaux du prélèvement : à ne remplir que si différent du passage	Vide
LONGITUDE_PRELEVEMENT	Longitude en WGS84 degrés décimaux du prélèvement : à ne remplir que si différent du passage	Vide
POSITIONNEMENT_PRELEVEMENT	Façon dont le prélèvement a été positionné : code Sandre du positionnement dans la nomenclature [598] "Type d'acquisition des coordonnées du lieu de surveillance"	Vide
NOMBRE_INDIVIDU_PRLVMT	Nombre d'individus observés / analysés sur le prélèvement	Vide
LOT_AQUACOLE	Lot aquacole	Vide

Nom de la colonne	Définition	Pour ZooPhytoImage - fichier ...res.csv
CODE SANDRE SUPPORT_ECHANTILLON	Code Sandre du support de l'échantillon analysé	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
TAXON_SUPPORT_ECHANTILLON	Code Sandre du taxon correspondant à l'échantillon	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
GROUPE_TAXON_SUPPORT_ECHANTILLON	Code Sandre du groupe de taxon correspondant à l'échantillon	Récupérer dans Q ² à partir de l'échantillon (si utilisé pour vérification unicité)
NUM_ECHANTI	Numéro de l'échantillon (mnémonique Quadrige ²)	Nom du fichier qui est chargé (mnémo de l'échantillon)
TAILLE_ECHANTILLON	Quantité de "matrice" composant l'échantillon	Vide
UNITE_TAILLE_ECHANTILLON	Code Sandre de l'unité dans laquelle est exprimée la taille de l'échantillon	Vide
COMMENTAIRE_ECHANTILLON	Commentaires sur l'échantillon	Vide
NOMBRE_INDIVIDU_ECHANTILLON	Nombre d'individus composant l'échantillon	Vide
NIVEAU_SAISIE_RESULTAT	Niveau de données Quadrige ² sur lequel est créé le résultat	"ECHANT" (données sur échantillon)
CODE SANDRE PARAMETRE	Code Sandre du paramètre analysé	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
LIBELLE SANDRE	Libellé Sandre du paramètre analysé	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
CODE SANDRE SUPPORT	Code Sandre du support analysé	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
CODE SANDRE FRACTION	Code Sandre de la fraction analysée	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
CODE SANDRE METHOD	Code Sandre de la méthode d'analyse	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
NUMERO_INDIVIDU	Numéro de l'individu analysé	Vide (sera rempli pour valid.csv avec le numéro de la particule)
TAXON_RESULTAT	Code Sandre du taxon correspondant au taxon identifié	Code Sandre du taxon correspondant à la classe ZooPhytoImage
GROUPE_TAXON_RESULTAT	Code Sandre du groupe taxon correspondant au taxon identifié	Code Sandre du groupe de taxon correspondant à la classe ZooPhytoImage
RESULTAT	Valeur numérique du résultat	Valeur contenue dans la deuxième ligne du fichier (valeurs d'abondance et de biovolume)
RESULTAT_QUALITATIF	Code Sandre de la valeur qualitative du paramètre	Vide
CODE SANDRE UNITE	Code Sandre de l'unité de mesure du résultat	Selon tables de transcodage : PSFMU correspondant aux 3 premières lettres de l'entête de colonne (Abd = abondance, Bio = biovolume)
CODE SANDRE ANALYSTE	Code Sandre de l'organisme ayant réalisé l'analyse	Identifiant de l'utilisateur identifié dans le fichier valid.txt
ENGIN_ANALYSE	Code Sandre de l'engin d'analyse utilisé (nomenclature [719] "Engin d'analyse")	FlowCam : identifié dans le nom du fichier, à convertir en code Sandre engin d'analyse via table de transcodage située dans le modèle Quadrige
CODE REMARQUE	Code Sandre de la précision de la valeur numérique des résultats (< LD : 2 < LQ : 10)	Vide
PRECISION	Incertitude sur la valeur numérique du résultat	Vide
TYPE_PRECISION	Code Sandre du type de précision du résultat (nomenclature [602] : 1 pour "%", 2 pour "unité de la valeur mesurée")	Vide
COMMENTAIRE	Commentaire sur le résultat	Mentionner : "Taxon classé initialement dans " + nom de la classe, suivi d'une mention " - issu de l'intégration ZooPhytoImage de l'échantillon"+nom du fichier chargé + "du "date de l'intégration de données

Bancaisation des données provenant du traitement par ZooPhytoImage

Réunion à Nantes, 29 septembre 2014

Réflexions ultérieures, décembre 2014 et janvier 2015

Sommaire

Préambule	2
Réunion à Nantes, 29 septembre 2014.....	2
Objectifs de la réunion	2
Premières spécifications de la bancaisation	2
Éléments supplémentaires apportés lors du COPIL FlowCAM / ZooPhytoImage le 2 décembre 2014..	4
Réunion du 15 janvier, Nantes : tentative de formalisation de la bancaisation et premiers choix.....	5

Liste de diffusion

Participants

Membres du COPIL : Philippe Grosjean et Guillaume Wacquet (Univ. Mons), Felipe Artigas (LOG CNRS ULCO Wimereux), Denis Hamad (ULCO LISIC Calais), Alain Lefebvre et Elvire Antajan (Ifremer Boulogne), Florent Colas, Luis Lampert, Michel Lunven, Michel Répécaud, Jean-François Rolin, Raffaele Siano (Ifremer Brest), Catherine Belin, Nadine Neaud-Masson, Dominique Soudant (Ifremer Nantes), Danièle Maurer (Ifremer Arcachon)

Autres : Jean François Cadiou (ODE/DIR), René Robert (ODE/UL/DIR), Chantal Compère (REM/RDT), Antoine Huguet et Laurence Miossec (DYNECO VIGIES), Pascale Hébert et Camille Blondel (LER BL), Claire Méteigner et Myriam Rumèbe-Perrière (LER AR), Martin Plus, Marie Madeleine Danielou, Marie Pierre Crassous (DYNECO PELAGOS), tous responsables LERs

Préambule

La bancaisation des données provenant du traitement par ZooPhytoImage des images numérisées par FlowCAM est une action prévue dans la fiche n° 9 des conventions Ifremer / ONEMA 2014 puis 2015, fiche intitulée « Optimisation de l'identification et du dénombrement du micro-phytoplancton avec le système couplé de numérisation et d'analyse d'images FlowCAM / ZooPhytoImage (système innovant) ».

Réunion à Nantes, 29 septembre 2014

Participants

Philippe Grosjean (Univ. Mons)

Antoine Huguet, Dominique Soudant, Catherine Belin (Ifremer/ DYNECO / VIGIES)

Partie corrigée et validée par Ph. Grosjean

Objectifs de la réunion

Définir les formats et les volumes nécessaires à la bancaisation des données acquises par le logiciel ZooPhytoImage, incluant les images.

Premières spécifications de la bancaisation

La bancaisation peut et doit être un point commun aux deux outils que sont FlowCAM/ZooPhytoImage d'une part, ZooScan/Plankton-Identifieur d'autre part. Actuellement à Ifremer : (i) le premier outil est utilisé pour le traitement des données phytoplancton numérisées par un FlowCAM dans le cadre du REPHY, (ii) le deuxième outil est utilisé pour le traitement des données zooplancton acquises par le LER Boulogne à l'aide d'un ZooScan.

Le FlowCAM produit des fichiers ZIDB (qui remplacent les ZID). Le ZooScan produit des fichiers similaires à l'ancien format ZID. Et le FastCAM (prototype en cours à REM/RDT, pour remplacer potentiellement à terme le FlowCAM) peut produire des fichiers ZID (vérifier avec Florent Colas, s'il peut produire des ZIDB)

Le format ZIDB contient : des fichiers images (jpeg ou png, une par particule), un fichier contenant les métadonnées, un fichier csv contenant les attributs mesurés sur chaque particule, etc

Les images brutes du FlowCAM sont-elles à garder ou non ? Actuellement, on ne peut pas rejouer l'analyse, sauf si on décide de passer sur des vignettes au format png ; il faudrait alors modifier les

outils existants pour pouvoir ré-analyser ces vignettes, et aujourd'hui ces outils n'existent pas. Mais l'important est de se donner la possibilité de ré-analyser les images si c'est nécessaire à l'avenir.

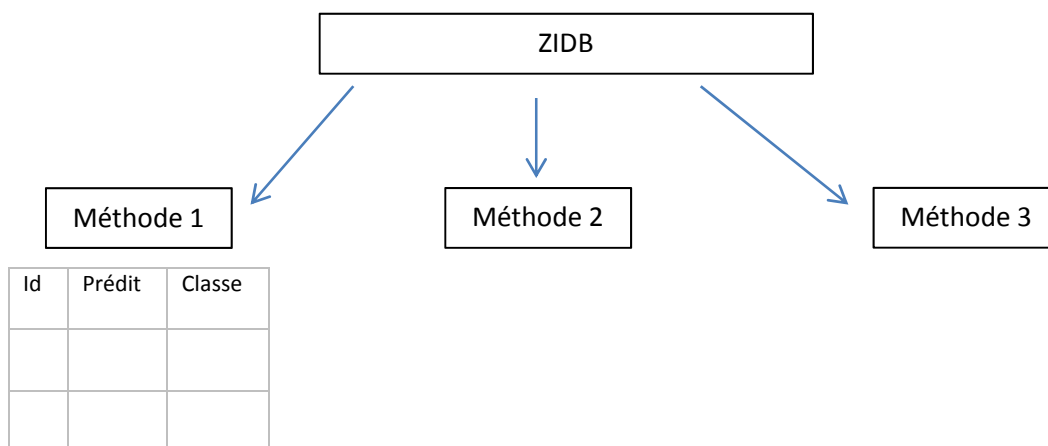
Les sets d'apprentissage (Zitrain), nécessaires pour utiliser ZooPhytoImage, sont des fichiers R-data (format de stockage des données dans R). Ils sont liés aux ZIDB par la colonne CLASSE.

Des données sont ou pourront être présentes dans un fichier texte généré après analyse complète de l'échantillon, comme les abondances, les biovolumes, les spectres de taille, le nombre de colonies, le nombre de cellules par colonie, etc

Dans l'interface conviviale de ZooPhytoImage (V5), l'utilisateur ne verra rien des traitements, il validera seulement des images. La notion de set d'apprentissage local disparaît au profit d'un apprentissage actif pour tous. [procédure restant à valider à plus large échelle]

Un script R contiendrait tous les critères et déterminerait une méthode. Il faudrait historiser la méthode utilisée : le set d'apprentissage actif serait figé pour un temps et chaque échantillon serait associé à une méthode comprenant le set d'apprentissage à un moment donné.

Il y aurait un lien entre le ZIDB et Quadriges² : au niveau du prélèvement ? (à voir)



Les données de ce tableau seraient dans Quadriges² : au niveau de l'échantillon ? (à voir)

A partir de là, on aurait les abondances, les biovolumes, etc

Pour une traçabilité et reproductibilité optimales des analyses, on propose dans un premier temps d'entreposer les statistiques par échantillon (abondances, biovolumes, etc.) reliées à une analyse particulière dont il faut aussi conserver le script R, et le lien vers le ZIDB et l'objet Ziclass ou ZITrain, c'est à déterminer encore.

Éléments supplémentaires apportés lors du COPIL FlowCAM / ZooPhytoImage le 2 décembre 2014

Une discussion sur ce sujet a eu lieu ultérieurement à cette réunion du 29 septembre, lors du COPIL FlowCAM / ZooPhytoImage le 2 décembre. Cette partie du CR est ici reprise in extenso :

Antoine Huguet (AH) a transmis à Philippe Grosjean (PG) un modèle de format QuadriLabo, sur lequel travaille Guillaume Wacquet (GW). Il faudrait fournir à GW une liste à jour du référentiel taxinomique QuadriLabo pour le phytoplancton : action Nadine Neaud-Masson (NNM) : [c'est fait].

Le contenu de ce qui sera disponible directement dans Q², et de ce qui sera disponible dans Q² au travers d'un lien vers des fichiers hébergés ailleurs, n'est pas encore complètement défini. En particulier, actuellement les données phytoplancton acquises par microscope sont constituées, pour chaque taxon identifié dans un échantillon, d'un seul résultat qui est le nombre de cellules par litre. Avec ZooPhytoImage, en plus du nom du taxon (qui s'appuiera sur le référentiel Q²) et du nombre de cellules par litre, de nouvelles mesures seront disponibles, mais à un niveau différent du taxon, c'est à dire soit au niveau de chacune des particules numérisées, soit au niveau de l'échantillon, par exemple :

- *la taille de chaque particule est calculée actuellement dans ZooPhytoImage comme le Diamètre Equivalent Sphérique (ECD)*
- *il en est déduit au niveau de l'échantillon un spectre de tailles, dont les classes sont à définir (par défaut, les classes sont actuellement définies de 10 en 10 μm)*
- *le biovolume est disponible pour chaque particule*

Les réponses à apporter rapidement sont :

- *doit-on stocker dans Q² les infos sur les particules, qui incluent de nombreuses particules n'appartenant pas au phytoplancton, ou bien les stocker dans un fichier externe ? (avec un lien par échantillon)*
- *dans le deuxième cas, quels sont les paramètres qui seront régulièrement utilisés et qui méritent donc d'être gardés dans Q², au niveau de chaque taxon ou au niveau de l'échantillon ?*
- *pour les paramètres supplémentaires qui seraient retenus, comment agréger au niveau d'un taxon les informations disponibles au niveau des particules qui le composent, par exemple pour le biovolume, cela pourrait être le min et le max des biovolumes de ces particules, mais est-ce satisfaisant et suffisant ?*
- *pour la taille, doit-on garder une information au niveau de chaque taxon ou au niveau de l'échantillon, ou bien aux deux niveaux ?*
 - *dans le premier cas (taxon), quelle est l'info à garder : l'ECD ou la plus grande longueur ? voir aussi la question de l'agrégation de ce résultat*
 - *dans le deuxième cas (échantillon), un graphique de distribution des spectres de taille est actuellement disponible, qui pourrait être attaché à l'échantillon. Il serait intéressant d'avoir en plus le résultat du nombre de particules par classe de taille, ce qui fait autant de paramètres que de classes de taille, et dans ce cas veut-on voir la répartition des spectres de taille de l'ensemble des particules ou bien seulement de celles appartenant au phyto ?*

Catherine Belin (CB) fera circuler un questionnaire pour recueillir les avis des membres du COPIL.

Par ailleurs, pour un échantillon, les informations doivent être liées à une méthode : soit parce que la méthode a évolué dans le temps, soit parce que l'on doit pouvoir utiliser au choix une ou une autre méthode (par ex une méthode classerait les vignettes en seulement dix groupes, une autre classerait selon des regroupements beaucoup plus précis, une autre ne classerait que les particules marquées FISH, etc). Dans l'interface ZooPhytoImage, la possibilité de faire ceci existe, et dans Q² cela ne pose pas de problème car un échantillon est toujours associé à un PSFM¹ qui contient la méthode. Il faudrait cependant ne pas multiplier les méthodes possibles et surtout ne pas faire évoluer les méthodes trop souvent, ce qui poserait un problème de mise à jour du référentiel Q².

Avec le set d'apprentissage actif, il faudra stocker le set en lien avec l'échantillon, car celui-ci évolue à chaque échantillon : cela doublera le volume de données à stocker, et encore seulement si on ne stocke pas les vignettes de l'échantillon. Ce problème est donc à considérer.

N.B. *le set d'apprentissage actif est une des améliorations en cours de ZooPhytoImage, qui consistera à améliorer le set commun à tous les utilisateurs, en fonction des corrections que chacun d'entre eux aura apporté aux échantillons traités.*

Réunion du 15 janvier, Nantes : tentative de formalisation de la bancaisation et premiers choix

Participants

Antoine Huguet & Catherine Belin

Pour un échantillon analysé, on considèrera successivement :

- les particules (ou vignettes) numérisées (comprenant du phytoplancton et tout le reste) : jusqu'à plusieurs milliers de vignettes par échantillon.
- les taxons, dont la liste est inféodée au référentiel Q², et correspondant à au moins une des particules de l'échantillon
- l'échantillon dans son ensemble

Particules / Vignettes

De nombreuses variables sont mesurées pour chaque particule, par ex l'ECD, le biovolume, etc.

Stockage souhaité de l'ensemble des particules (phyto et non phyto), mais peut-être pas dans Q² (choix à faire par Antoine : avantages et inconvénients). Une option pourrait-elle être un stockage dans Q² des seules vignettes phyto et est ce souhaitable ?

Les particules phyto sont regroupées par taxon Q² : le lien vers Q² se fait ici.

¹ PSFM : Paramètre-Support-Fraction-Méthode, entité du référentiel Q² nécessaire pour saisir tout résultat

PSFMs et Taxons

Il est envisageable de créer plusieurs PSFMs selon la précision voulue :

- un PSFM pour un classement fin, se rapprochant des observations du type flores totales faites dans le REPHY
- un PSFM pour un classement grossier, en quelques catégories seulement
- des PSFMs intermédiaires ?

Pour chaque taxon, les variables mesurées sur chacune des particules constituant le taxon doivent être agrégées. Actuellement, dans un échantillon lié au PSFM dit « Flore Totale », le seul résultat bancaisé est le nombre de cellules par litre pour ce taxon. Des résultats supplémentaires pourraient être associés à chacun des taxons, par exemple :

- Nb cellules par litre
- [ECD min, ECD max]
- ECD moyen ou médian
- [longueur min, longueur max]
- longueur moyenne ou médiane
- [largeur min, largeur max]
- largeur moyenne ou médiane
- [biovolume min, biovolume max]
- biovolume moyen ou médian
- Nb colonies
- Nb particules par colonie
- Etc. **A compléter / corriger par les utilisateurs (de l'outil et des données)**

Echantillon

Le set d'apprentissage sera stocké en lien avec chaque échantillon. Attention : le set actuel = plusieurs centaines de Mo.

Actuellement il est prévu, au niveau de l'échantillon, un graphique de distribution des spectres de taille, de 10 cm en 10 cm, qui concerne toutes les particules, y compris non phyto.

Est-il possible d'envisager en plus :

- le même graphique seulement pour les particules phyto
- le nombre de particules phyto par classe de taille
- **à compléter par les utilisateurs (de l'outil et des données)**

Sujet : Etude bancarisation des données ZooPhytoImage

Objet : Compte-rendu de la réunion de lancement de la prestation par Umanis

Participants

Catherine BELIN	Arnaud LEFEVRE (Umanis)
Catherine BORREMANS	Antoine SZIDON (Umanis)
Stéphane BOCANDE	
Emilie GAUTHIER*	
Nadine NEAUD-MASSON	

* Rédacteur

Conclusions :

Des données de surveillance du phytoplancton sont acquises par imagerie (appareil FlowCam), puis analysées via un logiciel ZooPhytoImage permettant d'identifier les particules contenues dans les échantillons, et de les quantifier (nombre de cellules par taxon, biovolumes, taille des particules, etc.). Des analyses similaires sont effectuées sur le zooplancton (ZooCam).

Ces données doivent être sauvegardées, et bancarisées pour permettre leur traitement et leur valorisation comme les autres données du réseau REPHY, déjà présentes dans Quadrigé².

L'équipe Ifremer sollicite la société Umanis pour fournir un rapport d'expertise technique proposant différents processus de bancarisation, avec une analyse critique des technologies utilisables à chaque étape. Ces propositions devront tenir compte le plus possible des outils existant à l'Ifremer.

Le processus d'acquisition des données a été présenté à Umanis. Des documents de référence restent à fournir, ainsi que des jeux de données en exemple.

La dernière version du logiciel ZooPhytoImage présente des bugs bloquants. Dès qu'une version déboguée sera disponible, N. Masson fera une démonstration à l'équipe Ifremer.

L'équipe Umanis enverra des premières propositions de solutions techniques au plus tôt, afin d'écarter rapidement les solutions inappropriées, et de permettre à Umanis de détailler au maximum les solutions les plus pertinentes. Un bilan sera fait lors de la prochaine réunion prévue le mardi 15 décembre 2015.

Liste de diffusion

Participants	
Antoine HUGUET	
Dominique SOUDANT	

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 2/15			

1 Contexte

Dans le cadre du REPHY, des données de composition et d'abondance du phytoplancton sont acquises via l'outil « FlowCAM / ZooPhytoImage ». Cet outil permet de numériser chaque particule (cellule vivante ou débris) présente dans un échantillon. Le logiciel ZooPhytoImage, développé par Philippe Grosjean de l'université de Mons en Belgique, est un outil de reconnaissance des cellules phytoplanctoniques qui fournit une identification taxinomique ainsi que des mesures quantitatives (comptage en nombre de cellules, biovolumes, taille de cellule, etc.).

L'intérêt de ces outils pour la surveillance REPHY est double :

- Archivage des données : PhytoImage permet de conserver les photos (vignettes) des cellules identifiées de chaque échantillon, et donc de rejouer les analyses a posteriori pour bénéficier de l'amélioration des connaissances scientifiques sur les jeux de données historiques. Ce qui n'est pas possible avec les échantillons analysés par microscopie.
- PhytoImage apporte des informations complémentaires difficilement accessibles par microscopie (biovolumes et tailles notamment), et très importantes pour la compréhension de la dynamique du phytoplancton.

Ainsi, l'Ifremer utilise le logiciel ZooPhytoImage et participe à son développement (tests utilisateurs, spécifications). La version 5.4-4 vient d'être livrée par Philippe Grosjean.

Les données issues de ces outils doivent être bancarisées, au même titre que l'ensemble des données REPHY. Le système d'information choisi est Quadrigé, pour assurer la cohérence avec les autres données REPHY.

Ce travail se compose de trois actions :

- 1) Sauvegarde des données brutes : enregistrement sur des serveurs spécifiques des données issues du FlowCam et/ou de ZooPhytoImage, selon une arborescence et une codification à valider (un serveur avec une procédure de stockage sont déjà définis)
- 2) Sauvegarde des paramètres ZooPhytoImage : conservation versionnée des sets d'apprentissage + classifier + algorithme ZooPhytoImage, et référencement de la version utilisée sur les résultats d'analyse
- 3) Transfert des résultats des analyses ZooPhytoImage sous forme de données dans la base Quadrigé².

Ces trois types de données doivent être reliables entre elles (ex : à partir d'un jeu de données brutes, retrouver les données Q² qui en sont issues, ou à l'inverse depuis des données dans Q², retrouver le set d'apprentissage utilisé et les données brutes).

L'Ifremer a sollicité l'expertise de la société Umanis via un appel d'offre lancé en octobre 2015 pour formuler des recommandations techniques sur l'organisation et les outils utilisables pour ces tâches.

Cette prestation consiste en des échanges entre les équipes thématiques et techniques de l'Ifremer (REPHY, Quadrigé) et les prestataires Umanis, aboutissant à un rapport de recommandations. Ces échanges se feront par mail, téléphone et par 3 réunions en séances, dont une de restitution.

Afin de préparer l'intervention d'Umanis, l'équipe Ifremer s'est réunie pour cadrer le contenu de la prestation et lister les actions à mener pour regrouper toutes les informations nécessaires au prestataire.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 3/15			

Des réflexions du même type étant en cours pour les données zooplancton acquises notamment dans le cadre de la surveillance DCSMM, Catherine Borremans, administratrice des données au SISMER, participe aux échanges pour évaluer les possibilités de mutualisation des solutions envisagées.

2 Déroulement de la prestation

2.1 Planning prévisionnel

Nom de tâche	Début	Fin	Visio	Nb j	Remarques
Réunion de lancement / atelier fonctionnel	Mar 01/12/15	Mar 01/12/15	Nantes1 -Brest1	1	9h30-17h30 mais visioconférence que jusqu'à 15h30
Préparation atelier	mer 02/12/15	Lun 14/12/15		5	
Atelier fonctionnel	Mar 15/12/15	Mar 15/12/15	Non	1	Fin à 16h30
Préparation atelier	Mer 16/12/15	Mer 06/01/16		5	
Atelier fonctionnel	jeu 07/01/16	jeu 07/01/16	Nantes1 -Brest1	1	9h30-17h30
Rédaction document de synthèse / préconisations Umanis	Ven 08/01/16	Lun 18/01/16		6	
Présentation solutions	Lun 25/01/2016	Lun 25/01/2016	Nantes1 -Brest1	1	En séance à Nantes? Visio réservée au cas où...
				20	

2.2 Déroulement de la réunion de lancement

L'équipe Ifremer a présenté le projet ZooPhytoImage (C. Belin, N. Masson). Un document listant les éléments de réflexion sur la bancarisation des données issues de ZooPhytoImage a été lu en séance, et complété avec les informations fournies par les différents participants.

La suite de ce compte-rendu détaille ces éléments.

2.3 Retour attendu

Le résultat de cette prestation est un rapport d'expertise montrant, si possible avec le support de schémas :

- Les différentes étapes du cycle de vie des données, de la numérisation des échantillons avec le FlowCam à la diffusion des données, en passant par le logiciel ZooPhytoImage et la bancarisation dans Quadrige²
- Les acteurs impliqués à chaque étape en précisant les conditions d'accès aux données (connexions authentifiées ou pas, périmètres accessibles, notion d'intranet/extranet)
- Des propositions d'architecture technique pour chaque étape (outils utilisables, technologies possibles, etc.) avec analyse comparative des coûts et des caractéristiques techniques et fonctionnelles (que pourrait-on faire ou ne pourrait-on pas faire avec chaque solution possible).

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 4/15			

Selon l'avancement des réflexions, les travaux pourraient aller jusqu'à débiter des spécifications de migration des fichiers .csv issus de ZooPhytoImage sous forme de résultats dans Quadriga.

2.4 Travaux lors de la prochaine réunion

L'équipe Ifremer doit fournir des documents source à l'équipe Umanis, et des exemples de jeux de données.

La prochaine réunion – atelier est prévue le mardi 15 décembre à l'Ifremer de Nantes. Umanis présentera une première synthèse des informations recueillies avec un premier lot de solutions techniques envisageables pour une partie des étapes du processus.

Si des échanges par mail avant cette réunion permettent d'éliminer certaines solutions, elles pourront toutefois être mentionnées dans le rapport final avec l'argumentation ayant abouti à leur exclusion.

Une expertise technique sur le format ZIDB pourra être présentée (évaluation des caractéristiques techniques de ce format, possibilités de lecture par d'autres applications que PhytoImage).

3 Volume de données :

3.1 Quel « poids » font les différents types de fichiers à traiter pour 1 échantillon (images / csv / métadonnées) ?

- Fichiers bruts sortie FlowCam (TIF + csv + txt) : le nombre de fichiers varie en fonction de la charge en particules de l'échantillon. Cela peut aller de 10 à 200 fichiers, avec en général 25 à 35 fichiers (à multiplier par 10 pour convertir en Mo, soit 30 fichiers = 30 Mo) Fichiers pré-traités sortie ZooPhytoImage (jpg + csv) : plus de 7000 fichiers (images de chaque particule phytoplanctonique) pour environ 15 Mo.

Un premier traitement des fichiers bruts sortie FlowCAM par ZooPhytoImage génère des fichiers compressés .ZID (Zoo Image data) ou .ZIDB. Ces fichiers contiennent les images de chaque individu (vignette comprenant une échelle, format .jpg), leurs mesures et les métadonnées (format.csv). **Ce sont ces fichiers ZID qui sont utilisés pour créer ou enrichir un set d'apprentissage.**

L'analyse d'un échantillon inconnu à l'aide de l'outil de reconnaissance (classifier) créé par ZooImage en fonction d'un set d'apprentissage donné, génère par échantillon deux fichiers R et un fichier txt (environ 300 Ko pour un échantillon de 30 Mo sorti FlowCAM).

Ces données comprennent toutes les particules, y compris les particules non phytoplanctoniques (débris, zooplancton...).

Il faut aussi bancariser le set d'apprentissage (environ 400 Mo) : le logiciel PhytoImage analyse les échantillons avec un classifieur (fichier R) issu d'un set d'apprentissage, et selon une méthode précise (algorithme).

Actuellement, le set d'apprentissage contient 61 classes contenant au total 7061 vignettes (ex : 1 taxon = 1 classe, classes de débris : bubble, clear, fiber, etc.). Le logiciel utilise ce set pour élaborer un outil de reconnaissance (classifier) Il permet au logiciel de reconnaître les particules et de les classer par catégories (Figure 1).

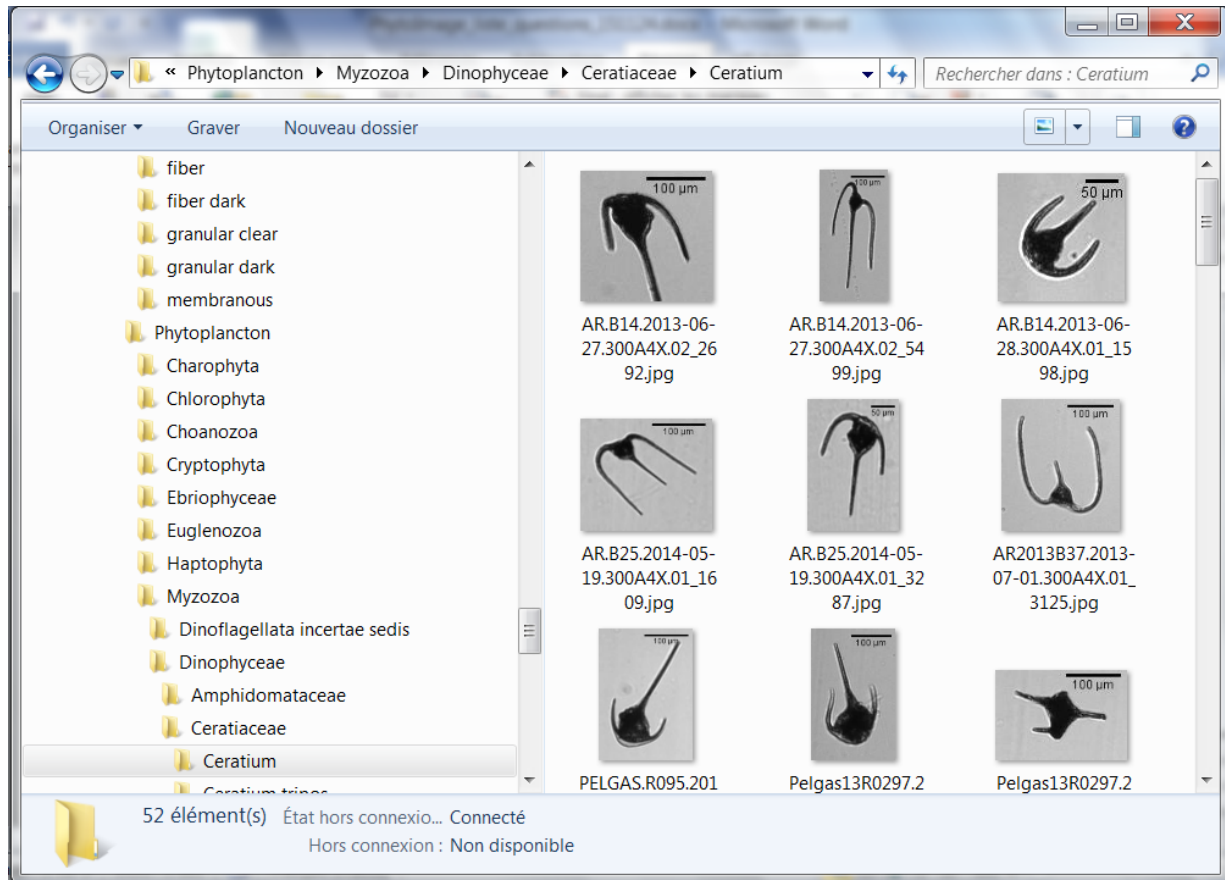


Figure 1 : Exemple de classes de particules issues du logiciel ZooPhytolmage v5.4.0, avec pré-visualisation dans l'explorateur Windows des imagerie d'une des classes (la classe Ceratum).

Proposition : chaque version de l'ensemble set d'apprentissage + classifier + algorithme pourrait correspondre à une méthode au sens Quadrige. Chaque classifier a ses propres performances analysables par les outils de ZooPhytolmage qui génèrent des graphes (matrice de confusion, taux de précision, F-score) : ce sont ses capacités à discriminer les particules. L'analyse des performances d'un classifier n'est pas à bancariser.

L'analyse du zooplancton semble suivre le même processus (numérisation des échantillons via le ZooCam, puis traitement des fichiers numérisés via un logiciel classant les particules en différentes classes sur la base d'un set d'apprentissage).

3.2 Combien d'échantillons sont acquis par an, par lieu et par laboratoire ?

Environ 4000 échantillons phytoplanktoniques par an au niveau national (dont 1000 flores totales).

Chaque échantillon est **numérisé 3 fois (réplicas d'analyse)**. Cela implique d'avoir 3 jeux de données brutes par échantillon. En effet, lors des numérisations, il arrive que des particules se superposent, ou qu'une même particule reste coincée dans la cellule de passage et soit prise x fois en photo. Les réplicas d'analyse permettent d'optimiser l'analyse d'un échantillon. **Les résultats des différents réplicas doivent être agrégés** pour obtenir une seule série de résultats par échantillon. Les fichiers

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 6/15			

issus de l'analyse de chaque réplica par ZooPhytoImage (txt, RData) pourront en revanche être conservés pour chaque réplica sous forme de fichiers de mesure.

Il y a 3 appareils FlowCam en service (Nantes, Boulogne, Arcachon). Tous les fichiers issus de l'utilisation de ces 3 appareils sont actuellement stockés sur le serveur <\\iota1\zoophytoimage> (369 Go actuellement sur 4 To, dont d'autres fichiers que les fichiers bruts du FlowCam).

En 2015, 50 Go sont occupés par les fichiers des numérisations des échantillons au FlowCAM de Boulogne + Nantes + Arcachon.

BOULOGNE : 118 numérisations occupent 32.3 G. De nombreuses numérisations contiennent plus de 100 fichiers.

NANTES : 211 numérisations occupent 12.6 Go

ARCACHON : 110 numérisations occupent 5.2 Go

Donc pour un total de 439 numérisations (correspondrait à 146 échantillons de 3 réplicas chacun), la place occupée serait de 50 Go. Soit 0.34 Go par échantillon.

Si 4000 échantillons sont numérisés par an, la place nécessaire pour stocker les fichiers bruts sortis du flowCAM serait de 1360 Go. Mais ces fichiers pourraient être zippés. Mais ce ne sont pas ces fichiers qui devront être bancarisés dans quadrigé.

Pour les données à bancariser dans la base Quadrigé², c'est-à-dire les résultats des analyses d'un échantillon, pour l'instant ces données sont sauvegardées par ZooPhytoImage dans 3 fichiers (<echantillon>_res.RData, <echantillon>_Valid.Rdata et <echantillon>_Valid.txt). Reste à voir comment manipuler les fichiers R vers un format adapté à l'intégration dans quadrigé (format Quadrilabo possible).

Enfin, concernant les fichiers ZID ou ZIDB, ce sont des fichiers compressés qui contiennent toutes les vignettes des particules de l'échantillon + les mesures + les métadonnées.

Les échantillons actuellement analysés pour test sont représentatifs des échantillons qui seront acquis en routine à terme (même volume analysé, une année entière, etc.).

Le volume de données généré augmentera certainement progressivement, car les trois appareils en service devront monter en charge avant de couvrir toutes les analyses, et le personnel devra être formé à l'utilisation du logiciel PhytoImage.

3.3 Quelle est la fréquence de la bancarisation (une fois par an ? au fur et à mesure des acquisitions : fréquence à préciser) ?

La bancarisation se fera en continu au gré des prélèvements (les échantillons sont fixés au Lugol en attendant l'analyse, et peuvent être conservés au maximum 1 mois pour une analyse optimale).

Le set d'apprentissage (accompagné du classifieur et de l'algorithme) se ferait **annuellement**. Les sets d'apprentissage peuvent évoluer au cours du temps selon les décisions des experts (ex : un nouveau taxon est détecté dans une zone géographique, et intervient fréquemment ou abondamment dans la composition des échantillons : il faut l'ajouter au set d'apprentissage pour que le classifieur les reconnaisse et les classe dans ce nouveau taxon, au lieu de les laisser dans la classe « other » qui doit être examinée manuellement par les experts).

Cependant, ces évolutions se feront en concertation avec la coordination nationale de l'expertise (N. Masson), et permettra de diffuser nationalement les évolutions apportées localement.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 7/15			

4 Format ZID/ZIDB

4.1 A quoi correspond ce format (standard informatique ou spécifique de certains logiciels ?)

Ce format correspond aux Zoolmage Data. C'est un format compressé spécifique du logiciel ZooPhytoImage.

4.2 Quelles technologies peuvent les utiliser (logiciels qui peuvent générer, stocker, et/ou lire ce format) ?

Florent Colas (Ifremer) travaille sur ces outils d'imagerie, et peut générer ce format à partir d'autres logiciels.

- ➔ Des précisions doivent être demandées à Florent Colas et Philippe Grosjean.
- ➔ Fournir un exemple de fichier à Umanis.

5 Liste des acteurs impliqués tout au long de la chaîne d'acquisition / traitement des données

Pour chaque acteur, définir le nombre d'utilisateur, leurs organismes d'appartenance (Ifremer ou hors Ifremer), les besoins d'authentification (accès sécurisé ou non).

5.1 Qui analyse les échantillons d'eau avec le FlowCam ?

Il y a 3 appareils en service à l'Ifremer (Boulogne, Nantes, Arcachon). Les échantillons seront prélevés sur la côte par chaque LER, puis envoyés dans les implantations disposant des FlowCam. Les échantillons seront analysés sur ces 3 sites, et les résultats de la numérisation seront stockés sur un serveur accessible à tous les LERs (chaque laboratoire pourra accéder à son espace disque pour y retrouver les fichiers bruts à traiter avec ZooPhytoImage).

5.2 Qui utilise le logiciel PhytoImage ? Avec quels fichiers en entrée (liste des fichiers avec leur format et leur taille) ?

Chaque LER traite ses propres fichiers issus du FlowCam. (TIF + csv données). Chaque LER génère ainsi des fichiers issus du traitement par ZooPhytoImage de ses échantillons au format jpg + csv (pour les résultats). Chaque LER bancariserait ensuite ses propres données.

5.3 Qui les consulte(raît) ?

Données brutes originelles (issues du FlowCam) : chaque LER doit pouvoir accéder à ses propres données (pour rejouer les analyses si besoin, comptabilisation pour les rapports d'activité, etc.).

Données issues de ZooPhytoImage : idem

Données bancarisées dans Quadrigé : tous les utilisateurs Quadrigé + diffusion via les outils classiques Quadrigé (Survall).

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
				Page 8/15

6 Décrire le déroulement du traitement des données

Pour tous ces points, différencier ce qui doit être fait depuis Quadrige de ce qui est fait hors Quadrige.

6.1 Pour chacun des acteurs listés au 3., décrire (ou schématiser) les outils utilisés (une application utilisateur ? une copie de données sur un serveur ? une interface web ? etc.)

La numérisation des échantillons avec le FlowCam fait l'objet d'une procédure (à fournir à Umanis). Elle décrit l'arborescence des fichiers bruts sur les serveurs actuels de stockage, la nomenclature utilisée, etc. Le FlowCam est connecté en réseau et crée les fichiers de numérisation sur le répertoire [\\iota1\zoophytoimage](#). C'est au moment de la numérisation que le FlowCam ouvre l'explorateur et demande l'endroit où déposer les fichiers. Les fichiers ainsi créés sur le répertoire commun sont accessibles à tous les LERs qui sont déclarés dessus.

Le manuel ZooPhytoImage décrit notamment la procédure à suivre pour le traitement des données numérisées par le FlowCam (à fournir à Umanis).

6.2 Décrire les fonctionnalités nécessaires :

6.2.1 Clic sur un bouton pour accéder à la fonctionnalité X ?

La bancarisation dans Quadrige se ferait via un import des fichiers issus de ZooPhytoImage.

Un 1^{er} scénario serait que l'utilisateur ouvre l'application Quadrige². Il demande alors l'import des données ZooPhytoImage. La question est :

- Import échantillon par échantillon, ou
- Import de plusieurs échantillons.







Un 2^{ème} scénario serait de sélectionner un ensemble de répertoires contenant les données ZooPhytoImage de plusieurs échantillons, et de les « envoyer vers » Q². Un programme informatique lirait alors ces répertoires, identifierait les données in situ Q² correspondantes (échantillons), puis transformerait le contenu de ces fichiers en résultats d'analyse dans Q².

Fichiers sortis de ZooPhytoImage :

- <echantillon>_res.RData : contient les abondances, les biovolumes et les spectres de taille pour chaque groupe taxinomique dans l'échantillon.
- <echantillon>_Valid.Rdata : contient les mesures sur chacune des particules, plus une colonne "Class" correspondant à la classification finale des particules de l'échantillon

Ces fichiers peuvent être chargés et manipulés dans R.

- <echantillon>_Valid.txt : ce fichier contient les informations essentielles à l'interprétation des résultats, comme le nom de l'ensemble d'apprentissage, le nom de l'outil de reconnaissance, l'algorithme utilisé, etc., ainsi que les métadonnées telles que le nom de l'analyste, la date d'analyse, etc.

Nom	Modifié le
 NT.058.2014-04-07.300A4X.03_res.RData	16/02/2015 11:28
 NT.058.2014-04-07.300A4X.03_res.RData.bak	16/02/2015 09:18
 NT.058.2014-04-07.300A4X.03_valid.RData	16/02/2015 11:28
 NT.058.2014-04-07.300A4X.03_valid.RData.bak	16/02/2015 09:18
 NT.058.2014-04-07.300A4X.03_valid.txt	16/02/2015 11:28
 NT.058.2014-04-07.300A4X.03_valid.txt.bak	16/02/2015 09:18

Actuellement, Nadine qui ne sait pas manipuler les fichiers Rdata, copie les résultats qui s'affichent dans la fenêtre de R lors d'une session d'analyse d'un échantillon par ZooPhytoImage et le colle dans un fichier Excel qu'elle retravaille. C'est ce fichier Excel final qui contient les données qu'elle utilise.

6.2.2 Nécessité de visualiser les vignettes ? Dans quel type d'application ? Pour faire quoi dessus (consultation / modification / suppression / ajout) ?

Les imageries n'ont pas besoin d'être visualisées de manière systématique après bancarisation. Il peut être intéressant de trouver des échantillons d'intérêt, des images de certains taxons, des illustrations, etc. Pour cela, les laboratoires analystes (LERs) auront accès aux données sauvegardées et consulteront les imageries (formats TIF et jpg) via les logiciels de lecture d'image classiques.

Besoin de consultation par l'extérieur : données Quadriges diffusées à l'extérieur (Surval, utilisateurs hors Ifremer inscrits dans Q²), mais les fichiers situés en amont de la bancarisation (TIFF, jpg, fichiers txt et csv globaux) ne seront pas accessibles hors Ifremer car fichiers bruts non conviviaux.

6.2.3 Validation des données

Les données brutes doivent-elles être « validées » ? Comment gère-t-on les droits sur les données ? Faut-il bloquer l'écriture des données ?

NM : oui, il faut empêcher la modification et la suppression des données brutes enregistrées. Autant les résultats des analyses dans Q² doivent suivre le cycle de vie Q², autant les données brutes doivent être protégées.

7 Gestion des données d'imagerie à l'Ifremer : état de l'art

7.1 Quelles sont les sources de données de type imagerie à l'Ifremer (photos, vidéos, imagerie type sonar/sondeur, données satellite) ?

Vidéo et photo :

- Les engins sous-marins : Victor6000, Nautilie, Scampi,...
- Les équipements particuliers : caméras tractées, mini ROV, Pagure,... et plongeurs humains
- Les observatoires fond de mer : EMSO Azores,...
- Les stations en point fixe : STAVIRO-MICADO

Imagerie acoustique : à voir avec DYNECO/Benthos et GM

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 10/15			

7.2 Comment sont-elles bancarisées ?

La vidéothèque scientifique contient actuellement les données vidéo récentes (postérieures à 2010) acquises au cours des plongées des submersibles de l'Ifremer (ROV Victor 6000).

Dans cette vidéothèque, les vidéos sont organisées par campagne océanographique, par engin sous-marin, par plongée et par caméra. Ces données sont accompagnées par des métadonnées relatives aux plongées des engins sous-marins et aux conditions d'acquisition : positionnement et attitude de l'engin, données de température et salinité, position et paramétrage de la caméra. Les commentaires enregistrés en parallèle du tournage vidéo sont également accessibles lorsqu'ils existent (cf. <http://video.ifremer.fr/>).

En plus des vidéos, sont proposées des images acquises au cours des mêmes événements :

- "Aperçus" : également appelés « mini-films », il s'agit d'un échantillonnage automatique de la vidéo à une fréquence donnée.
- "Images" : il s'agit des photos en Haute Définition prises à l'aide de l'appareil photo numérique de l'engin et des captures de la vidéo.

En pratique les images se trouvent sur des espaces disque « web » et leur URL est renseignée en base de données.

La bancarisation des autres types de vidéo et photos n'est pas encore réalisée mais devrait s'appuyer sur la vidéothèque existante, sauf peut-être pour les vidéos en point fixe.

7.3 Y a-t-il un lien vers les métriques qui en sont tirées ? Si oui, comment ?

L'association des commentaires (métadonnées) aux vidéos est possible dans la mesure où chaque donnée de la plongée est horodatée à l'aide d'une base de temps commune.

Des annotations réalisées par les scientifiques après dépouillement des vidéos des engins sous-marins ont été bancarisées dans le système Biocean pour une certaine série de jeux de données. A nouveau, c'est le **time code** qui permet de faire le lien entre l'observation et l'image.

8 Gestion des campagnes SISMER :

8.1 Où / comment / par qui sont-elles bancarisées ?

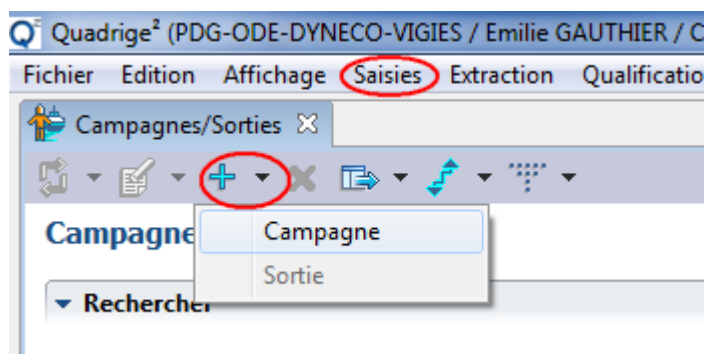
Les campagnes sont gérées par le service SISMER dans la base des campagnes (par raccourci souvent appelée « base SISMER »). Les campagnes sont créées avec leur identifiant propre au moment où elles sont prévues au calendrier de la Flotte. Elles sont ensuite mises à jour après la campagne, après réception par le SISMER de la fiche CSR (Cruise Summary Report) renseignée par le chef de mission.

8.2 Lien SISMER – Quadrige

Un lien a initialement été prévu dans l'application Quadrige. Le principe est le suivant :

- 1) Dans l'application Quadrige, un utilisateur demande à ajouter une campagne.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			Page 11/15



2) Dans la fenêtre de saisie de la campagne, des informations sont à renseigner :

Les champs en orange sont obligatoires.

3) L'utilisateur peut accéder au site Web des campagnes SISMER (<http://www.ifremer.fr/sismerData/jsp/campagnesALaMer.jsp>) pour consulter le catalogue des campagnes et trouver la référence de la campagne qu'il souhaite saisir dans Quadrigé.

Accès aux Données | Banques Nationales | Présentation | Assurance quali

PRECED. REHONT. SUIV.

Accueil Email Nouv. Contrib Login

ifremer © 2005

Campagnes à la mer

Zone Géographique

RÉSULTAT Campagnes à la mer : 28 résultat(s)

REBENT 2014 LEG2 F1352014012200 - (Thalia)

-- Lorient-Les Glénans-Trévignon-Baie de Concarneau du 28-03-2014 au 30-03-2014

4) L'utilisateur saisit la référence SISMER de la campagne et demande sa vérification :

*Nouvelle campagne

Campagne

Service Saisisseur : PDG-ODE-DYNECO-VIGIES

Référence SISMER : [Page WEB des campagnes SISMER](#)

Vérifier la référence F1352014012200

Libellé : Nouvelle campagne

Date de début : [calendar icon] Date de fin : [calendar icon]

Responsable : [dropdown menu]

Navire : [dropdown menu]

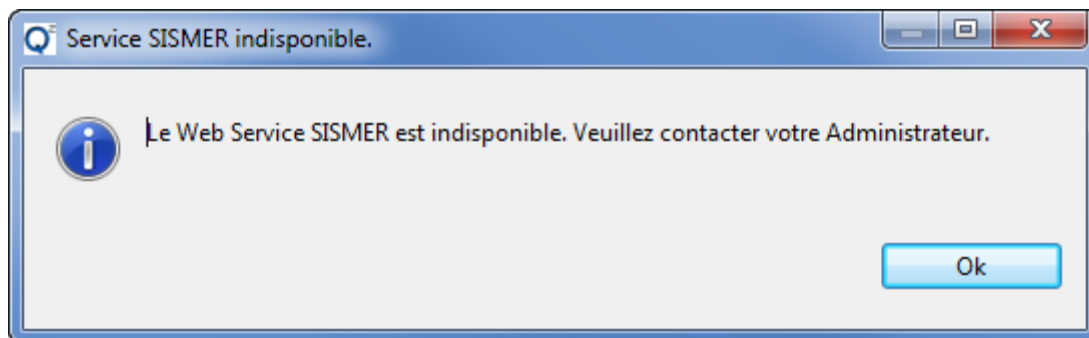
5) Le système récupère les informations liées à la campagne :

- Libellé
- Dates de début et de fin
- Nom du responsable
- Navire utilisé.

6) Ces informations se pré-remplissent dans l'interface Quadrige². L'utilisateur peut alors enregistrer sa campagne.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 13/15			

Or, actuellement, lorsqu'on clique sur le bouton « Vérifier la référence », on obtient le message d'erreur suivant :



Une fiche d'anomalie est recensée dans notre forge, mais toujours en attente :

ID	Catégorie	Sévérité	Reproductibilité	Date de soumission	Dernière mise à jour
0004183	[quadrige2] Evolution	mineur	n'a pas essayé	30-03-2010 13:58	05-10-2012 16:49
Rapporteur	Christian Bennet	Afficher l'état	public		
Assigné à					
Priorité	basse	Résolution	ouvert		
État	en attente			Version du produit	2.1.24
Résumé	0004183: ajouter_campagne > Test 001_2 - Pb appel WS campagne SISMER				
Description	1. assert-user [pass] : 2. assert-user [pass] : 3. assert-user [pass] : 4. assert-user [inconclusive] ----- On obtient une erreur lors de l'appel au service SISMER (apparemment problème appel web service Ifremer)				

9 Stockage DATARMOR

DATARMOR est un projet d'équipement mutualisé de calcul (supercalculateur) et de gestion de données de gros volumes. Il sera géré par l'IFREMER à destination de la communauté scientifique (Ifremer, Shom, IRD, IUEM, ect....). La mise en service de DATARMOR est planifiée pour 2016.

Le stockage des données scientifique volumineuses ouvertes au public y sera gratuit.

Ci-dessous un lien vers le Pôle de Calcul Intensif pour la Mer :

http://www.ifremer.fr/institut_eng/The-Institute/Schedule-of-events/Evenements-passes/Pole-de-Calcul-Intensif-pour-la-mer-PCIM-Journee-de-rencontre-des-utilisateurs

Discussion des possibilités de gestion d'images côté IDM et il en est ressorti la proposition suivante:

- stockage des images sur DATARMOR (qui faciliterait, à terme, des retraitements sur les images par exemple)
- la donnée pourra être uploadée via un mécanisme particulier à définir (peut-être via OwnCloud/Datacloud)
- la donnée pourrait être accessible via Quadrige par un lien de type URL fixe
- il faut réfléchir à la gestion des droits

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 14/15			

Donc, côté Zoocam, si Quadrige s'avérait ne pas convenir, on pourrait s'appuyer sur les mêmes services (DATARMOR, Owncloud...) tout en faisant un lien vers une autre base de données (par exemple la base des campagnes Simer).

Un des intérêts d'utiliser un DataCloud, serait de permettre à des agents à l'extérieur IFREMER de télécharger les données brutes.

Le DataCloud permet

- de gérer la sécurité
- de gérer les Entrées/Sorties fichiers
- des fonctionnalités avancées (Montage de disque distant équivalent à DROPBOX).


	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
				Page 15/15

Tableau récapitulatif des actions

N°	Libellé de l'action	Porteur	Dû pour le
1	Préciser le nombre d'échantillons analysés et dont la numérisation est stockée sur \\iota1	N. Masson	9 décembre 2015
2	Démonstration de la chaîne de traitement des données	N. Masson, E. Gauthier	9 décembre 2015
3	Fournir un exemple de fichier ZIDB à l'équipe Umanis	E. Gauthier	4 décembre 2015
4	Fournir la procédure de numérisation des échantillons ainsi que le manuel ZooPhytoImage à Umanis	E. Gauthier	4 décembre 2015
5	Répondre aux questions DATARMOR	S. Bocandé	9 décembre 2015
6	Rédiger une première synthèse des informations échangées, si possible sous forme de schémas	Umanis	15 décembre 2015
7	Analyse technique du format ZIDB	Umanis	15 décembre 2015

Q3

COMPTE RENDU DE REUNION

Lieu : Ifremer Nantes

Date : 07/01/2016

Objet : Etude bancarisation des données ZooPhytoImage – réunion 3

Page 1/6

Sujet : Etude bancarisation des données ZooPhytoImage

Objet : Compte-rendu de la réunion de lancement de la prestation par Umanis

Participants

Catherine BELIN	Arnaud LEFEVRE (Umanis)
Catherine BORREMANS	Antoine SZIDON (Umanis)
Stéphane BOCANDE	
Emilie GAUTHIER*	
Nadine NEAUD-MASSON	

* Rédacteur

Conclusions :

Des données de surveillance du phytoplancton sont acquises par imagerie (appareil FlowCam), puis analysées via un logiciel ZooPhytoImage permettant d'identifier les particules contenues dans les échantillons, et de les quantifier (nombre de cellules par taxon, biovolumes, taille des particules, etc.). Des analyses similaires sont effectuées sur le zooplancton (ZooCam).

Ces données doivent être sauvegardées, et bancarisées pour permettre leur traitement et leur valorisation comme les autres données du réseau REPHY, déjà présentes dans Quadrigé².

L'équipe Ifremer sollicite la société Umanis pour fournir un rapport d'expertise technique proposant différents processus de bancarisation, avec une analyse critique des technologies utilisables à chaque étape. Ces propositions devront tenir compte le plus possible des outils existant à l'Ifremer.

Le processus d'acquisition des données a été présenté à Umanis. Des documents de référence restent à fournir, ainsi que des jeux de données en exemple.

La dernière version du logiciel ZooPhytoImage présente des bugs bloquants. Dès qu'une version déboguée sera disponible, N. Masson fera une démonstration à l'équipe Ifremer.

L'équipe Umanis enverra des premières propositions de solutions techniques au plus tôt, afin d'écarter rapidement les solutions inappropriées, et de permettre à Umanis de détailler au maximum les solutions les plus pertinentes. Un bilan sera fait lors de la prochaine réunion prévue le mardi 15 décembre 2015.

Liste de diffusion

Participants	
Antoine HUGUET	
Dominique SOUDANT	

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	07/01/2016
	Objet : Etude bancarisation des données ZooPhytoImage – réunion 3			
				Page 2/6

1 Contexte

Dans le cadre du REPHY, des données de composition et d'abondance du phytoplancton sont acquises via l'outil « FlowCAM / ZooPhytoImage ». Cet outil permet de numériser chaque particule (cellule vivante ou débris) présente dans un échantillon. Le logiciel ZooPhytoImage, développé par Philippe Grosjean de l'université de Mons en Belgique, est un outil de reconnaissance des cellules phytoplanctoniques qui fournit une identification taxinomique ainsi que des mesures quantitatives (comptage en nombre de cellules, biovolumes, taille de cellule, etc.).

L'intérêt de ces outils pour la surveillance REPHY est double :

- Archivage des données : PhytolImage permet de conserver les photos (vignettes) des cellules identifiées de chaque échantillon, et donc de rejouer les analyses a posteriori pour bénéficier de l'amélioration des connaissances scientifiques sur les jeux de données historiques. Ce qui n'est pas possible avec les échantillons analysés par microscopie.
- PhytolImage apporte des informations complémentaires difficilement accessibles par microscopie (biovolumes et tailles notamment), et très importantes pour la compréhension de la dynamique du phytoplancton.

Ainsi, l'Ifremer utilise le logiciel ZooPhytoImage et participe à son développement (tests utilisateurs, spécifications). La version 5.4-4 vient d'être livrée par Philippe Grosjean.

Les données issues de ces outils doivent être bancarisées, au même titre que l'ensemble des données REPHY. Le système d'information choisi est Quadrigé, pour assurer la cohérence avec les autres données REPHY.

Ce travail se compose de trois actions :

- 1) Sauvegarde des données brutes : enregistrement sur des serveurs spécifiques des données issues du FlowCam et/ou de ZooPhytoImage, selon une arborescence et une codification à valider (un serveur avec une procédure de stockage sont déjà définis)
- 2) Sauvegarde des paramètres ZooPhytoImage : conservation versionnée des sets d'apprentissage + classifier + algorithme ZooPhytoImage, et référencement de la version utilisée sur les résultats d'analyse
- 3) Transfert des résultats des analyses ZooPhytoImage sous forme de données dans la base Quadrigé².

Ces trois types de données doivent être reliables entre elles (ex : à partir d'un jeu de données brutes, retrouver les données Q² qui en sont issues, ou à l'inverse depuis des données dans Q², retrouver le set d'apprentissage utilisé et les données brutes).

L'Ifremer a sollicité l'expertise de la société Umanis via un appel d'offre lancé en octobre 2015 pour formuler des recommandations techniques sur l'organisation et les outils utilisables pour ces tâches.

Cette prestation consiste en des échanges entre les équipes thématiques et techniques de l'Ifremer (REPHY, Quadrigé) et les prestataires Umanis, aboutissant à un rapport de recommandations. Ces échanges se feront par mail, téléphone et par 3 réunions en séances, dont une de restitution.

Afin de préparer l'intervention d'Umanis, l'équipe Ifremer s'est réunie pour cadrer le contenu de la prestation et lister les actions à mener pour regrouper toutes les informations nécessaires au prestataire.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	07/01/2016
	Objet : Etude bancarisation des données ZooPhytoImage – réunion 3			
				Page 3/6

Des réflexions du même type étant en cours pour les données zooplancton acquises notamment dans le cadre de la surveillance DCSMM, Catherine Borremans, administratrice des données au SISMER, participe aux échanges pour évaluer les possibilités de mutualisation des solutions envisagées.

2 Déroulement de la réunion

Début : 10h

Présentation de l'ordre du jour :

- Démo Nadine
- Relecture rapport expertise
- Travail intégration données dans Q²

Tâches prévues lors du dernier CR : OK

3 Démo Nadine

Login utilisateur saisi en début d'analyse (stocké dans le fichier ..._valid.txt)

Ouverture navigateur Safari : interface utilisateur (utilise le navigateur par défaut, ne fonctionne pas avec FireFox).

Travaux de correction des vignettes assez long selon les exigences de taux d'erreur (ex : 20 sets de correction de 10 min chacun).

Fichier résultats donne le nombre de particules par litre et par classe de particule.

Le fichier valid.csv contient 1 ligne par vignette avec en colonne les paramètres mesurés sur chaque vignette, avec en dernière colonne la classe attribuée.

Les noms de colonne du fichier valid.csv sont attribués par ZPI, et ne seront pas modifiés : on peut les utiliser pour le transcodage.

Pour une étude Dinophysis (comparaison des résultats du FlowCam avec analyses microscope), Nadine a eu besoin de rechercher toutes les vignettes attribuées à Dinophysis dans les échantillons. Pour ce faire, elle a sollicité Alice qui a intégré une commande DOS dans le fichier valid.csv permettant d'identifier le chemin du répertoire de stockage des vignettes Dinophysis, et de les recopier dans un répertoire spécifique.

- ➔ La bancarisation des vignettes dans Q² permettrait de faire cette sélection via une extraction. Cela relance la question de la bancarisation des vignettes (nécessaire ou pas) ?

Cette bancarisation serait possible dans les photos des échantillons :

- Libellé photo : numéro échantillon + numéro vignette

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	07/01/2016
	Objet : Etude bancarisation des données ZooPhytoImage – réunion 3			
				Page 4/6

- Légende photo : classe (= taxon) : dans ce cas, avec les critères d'extraction Q² actuels, on ne pourrait pas mettre de critère sur le taxon (le champ « légende » ne fait pas partie des critères d'extraction Q²).

4 Présentation Umanis

L'intégration en masse nécessite un outil spécifique mais pas forcément d'interface graphique pour les administrateurs (la reprise de données fait partie des missions de la cellule Q² qui pourra s'en charger à la demande des administrateurs).

SB : quelle dynamique de saisie ? Est-ce que l'utilisateur peut être amené à créer plusieurs échantillons dans Q², puis, dans un second temps charger les fichiers ZPI ? Dans ce cas, un bouton dans Q² n'est pas indispensable. Une autre interface serait envisageable.

NN : Oui, une solution de dépose par fichier pourrait être adaptée car les échantillons sont déjà créés pour saisie des autres résultats.

AS : est-ce le même user qui crée l'échantillon et qui intègre les données phyto ?

→ NN : oui.

5 Bilan des solutions

5.1 Sauvegarde

Disque réseau partagé : solution actuelle.

Data cloud : permettrait de commencer à archiver les données sur un système similaire à DATARMOR. Faciliterait la migration vers DATARMOR. Problème de gestion des droits d'accès : il faudra reproduire la gestion des droits du disque réseau partagé. Il faut que le Data Cloud ait un service (type Webservice, API) qui permette, depuis ZPI, de désigner un répertoire de stockage des résultats.

→ Pb : une API nécessite de faire évoluer ZPI pour l'utiliser, ce qui est proscrit car nous n'avons pas la main sur ces développements

→ Pb : un système type DropBox implique une réplication en local, ce qui est proscrit car trop lourd.

Arnaud fait une démo d'un cloud qui permet de cocher dans une arborescence les éléments à synchroniser ou non.

DATARMOR : objectif à plus ou moins long terme. L'idéal serait dans un 1^{er} temps de sauvegarder les fichiers sur DATARMOR, puis dans un 2nd temps d'y déployer ZPI.

Quid du set d'apprentissage et du classifier ? Il sera sauvegardé de la même façon selon la même arborescence que sur le disque réseau partagé actuel.

ZooCam irait dans le même sens du cloud, car gros besoin d'espace collaboratif où plusieurs experts pourraient participer à l'analyse d'un même échantillon.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	07/01/2016
	Objet : Etude bancarisation des données ZooPhytoImage – réunion 3			
				Page 5/6

5.2 Bancarisation Q²

Après discussion, les fichiers suivants ne sont pas bancarisés : les fichiers issus du FlowCam (notes.txt, run_summary.txt, etc.) et le fichier _valid.txt issu de ZPI. Les graphiques statistiques sur les corrections effectuées dans ZPI ne sont pas bancarisés.

Les fichiers à bancariser sont :

- ..._res.RData converti en .csv
- ..._valid.RData converti en .csv

La solution de bancarisation directe depuis ZPI est abandonnée. La solution retenue est de convertir des fichiers de données .csv en fichier Quadrilabo.

Intégration en masse :

Réservée à des administrateurs. Possible avec un job détectant la présence de fichiers dans un répertoire spécifique à l'intégration de données.

Intégration unitaire : pour les utilisateurs ZPI. Pour chaque échantillon, l'utilisateur déclenche le chargement des résultats. Reste à voir comment :

- un bouton dans Q² sur l'échantillon ? NON pour limiter les évolutions Q² car Q3 arrive.
- par dépôt des fichiers de résultat dans un répertoire ?
- par chargement via une interface web type quadrilabo ? OUI

La solution d'un chargement depuis une interface web semble adaptée, d'autant qu'elle dissocie cet outil de chargement de données des évolutions de l'application Q².

Le nom d'échantillon FlowCam peut être stocké dans le mnémonique d'échantillon Q² et faire le lien avec tous les fichiers à charger. Le mnémo d'échantillon pourrait être reformaté : il pourrait être composé de la façon suivante :

CODE FLOWCAM_MNEMO LIEU_DATE PASSAGE (aaaammjj)_n°implémentation.

Actuellement, la nomenclature type « NT.058.2014-04-07.300A4X.03 » peut être modifiée selon cette codification (ex : « BL.001-P-058.2015-11-25.4X.01 »).

Cela permettrait d'identifier le programme correspondant. Une autre solution serait de récupérer les programmes de l'échantillon et de les appliquer systématiquement à tous les résultats : Nadine indique que ça ne fonctionne pas car l'échantillon peut être multiprogramme

Les classes autres que phytoplancton ne sont pas à reprendre (« other », « total »).


	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
				Page 6/6

Tableau récapitulatif des actions

N°	Libellé de l'action	Porteur	Dû pour le
1	Préciser le nombre d'échantillons analysés et dont la numérisation est stockée sur \\iota1	N. Masson	9 décembre 2015
2	Démonstration de la chaîne de traitement des données	N. Masson, E. Gauthier	9 décembre 2015
3	Fournir un exemple de fichier ZIDB à l'équipe Umanis	E. Gauthier	4 décembre 2015
4	Fournir la procédure de numérisation des échantillons ainsi que le manuel ZooPhytoImage à Umanis	E. Gauthier	4 décembre 2015
5	Répondre aux questions DATARMOR	S. Bocandé	9 décembre 2015
6	Rédiger une première synthèse des informations échangées, si possible sous forme de schémas	Umanis	15 décembre 2015
7	Analyse technique du format ZIDB	Umanis	15 décembre 2015

Sujet : Etude bancarisation des données ZooPhytoImage

Objet : Compte-rendu de la réunion de restitution par Umanis

Participants

Catherine BELIN (CBe)	Arnaud LEFEVRE (Umanis)
Catherine BORREMANS (CBo)	Antoine SZIDON (Umanis)
Stéphane BOCANDE (SB)	Dominique SOUDANT (DS)
Emilie GAUTHIER* (EG)	Gaétane DURAND (GD)
Nadine NEAUD-MASSON (NN)	

* Rédacteur

Conclusions :

L'architecture de sauvegarde actuelle est efficace et sera conservée, en attendant le transfert vers DATARMOR lorsque cette architecture sera disponible.

Les fichiers de résultats d'analyse à bancariser dans Q² ont été identifiés. Ils seront chargés unitairement par les utilisateurs via une interface Web (site cellule Q²). Le chargement des fichiers sur cette interface lancera un job Talend de conversion des fichiers au format Quadrilabo, puis le job Talend de chargement Quadrilabo existant. Le module de conversion ZooPhytoImage – Quadrilabo sera ainsi le seul à évoluer en cas de modification des outils ZooPhytoImage. Cela évite des développements dans Q² alors que le projet Q3 démarre.

Des spécifications ont été rédigées pour réaliser cette conversion ZooPhytoImage – Quadrilabo.

L'ensemble des livrables est préparé pour l'Onema (fin du financement du projet ZooPhytoImage).

Des réflexions restent à mener sur les résultats acquis pour chaque particule (quelles métriques retenir ? Comment les agréger ? Quelles informations statistiques à conserver ?). Le processus dans son ensemble devra être documenté et transmis aux utilisateurs sous forme de consignes ou manuel (ex : codification des échantillons et saisie dans le mnémonique échantillon Q²).

Les résultats de cette étude seront présentés à l'équipe en charge des suivis de zooplancton pour évaluer les possibilités de réutilisation du processus.

Un gros travail sera à réaliser sur le référentiel Q² (création des méthodes et PSFMUs, tables de transcodage, création des programmes / stratégies). Cette étape sera réalisée lorsque les choix de métriques à intégrer seront faits.

Les développements d'outils de chargement des données seront ensuite effectués.

Liste de diffusion

Participants	
Antoine HUGUET	
Bruno ANDRAL	
Jean-Paul LECOMTE	

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	02/02/2016
	Objet : Etude bancarisation des données ZooPhytoImage - Restitution			
	Page 2/8			

1 Contexte

Dans le cadre du REPHY, des développements méthodologiques sont en cours pour l'acquisition de données de composition et d'abondance du phytoplancton via l'outil « FlowCAM / ZooPhytoImage ». Les données issues de ces outils doivent être bancarisées, au même titre que l'ensemble des données REPHY. Le système d'information choisi est Quadrige, pour assurer la cohérence avec les autres données REPHY.

L'Ifremer a sollicité l'expertise de la société Umanis via un appel d'offre lancé en octobre 2015 pour formuler des recommandations techniques sur l'organisation et les outils utilisables pour la sauvegarde et la bancarisation dans Quadrige de ces données, sous forme d'un rapport de recommandations.

La réunion de restitution de l'étude finale a lieu le 2 février 2016 à l'Ifremer de Nantes. Le présent document en fait le compte-rendu.

2 Déroulement de la prestation

2.1 Planning

Nom de tâche	Début	Fin	Visio	Nb j
Réunion de lancement / atelier fonctionnel	Mar 01/12/15	Mar 01/12/15	Nantes1-Brest1	1
Préparation atelier	mer 02/12/15	Lun 14/12/15		5
Atelier fonctionnel	Mar 15/12/15	Mar 15/12/15	Non	1
Préparation atelier	Mer 16/12/15	Mer 06/01/16		5
Atelier fonctionnel	jeu 07/01/16	jeu 07/01/16	Nantes1-Brest1	1
Rédaction document de synthèse / préconisations Umanis	Ven 08/01/16	Lun 18/01/16		6
Présentation solutions	Mar 02/02/2016	Mar 02/02/2016	Nantes1-Brest1	0,5
Révision du rapport final	Mer 03/02/2016	Ven 12/02/2016		0,5

20

2.2 Liste des livrables

- Bancarisation ZooPhytoImage - CR reunion 29 09 2014 + reflexions ultérieures dec 2014 et jan 2015.pdf
- 20151201_PhytoImage_CR_reunion_V3.pdf : réponses Ifremer aux questions posées par l'étude
- Arborescence_analyse-echantillon.pdf : présentation des fichiers générés tout au long de l'acquisition des données
- Rapport d'étude final (Umanis)
- fichier_quadrilabo_remplissage_PhytoImagexlsx.pdf : spécifications de conversion des fichiers de résultats ZPI en format Quadrilabo

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	02/02/2016
	Objet : Etude bancarisation des données ZooPhytoImage - Restitution			
	Page 3/8			

3 Compte-rendu final

ZooPhytoImage (ZPI) est un projet lancé il y a plusieurs années, sur financement de l'Onema pour la période 2013-2015, incluant un volet sauvegarde des données jusqu'à leur bancarisation dans Quadrige. Les questions de bancarisation incluent des réflexions sur la granularité des informations ZPI à bancariser (comment agréger les mesures de chaque vignette par taxon, voire par échantillon ?).

3.1 Présentation du rapport par Umanis

Deux volets : sauvegarde d'une part, et bancarisation d'autre part.

3.1.1 Sauvegarde

1 – Sauvegarde actuelle sur disque réseau partagé

- + solution déjà opérationnelle
- Données non partagées

La cible est un équivalent avec un protocole http. L'utilisation des données par des personnes extérieures à l'Ifremer n'est pas encore planifiée. Il faut l'envisager d'ici 3 ou 4 ans. La solution actuelle est donc pour l'instant adaptée.

2 – Own Cloud Ifremer

Cette solution nécessite d'installer un client supplémentaire sur le poste des utilisateurs. C'est un système type DropBox avec une gestion des droits assez fine. Mais cela implique une réplication de données sur les postes utilisateurs, ce qui pose un problème de stockage en local sur les postes utilisateurs (gros volume de données à répliquer).

Une autre solution est un système équivalent à un disque réseau partagé, mais sur un espace réseau dans Ifremer/Cloud. Il faudrait mettre en place un protocole type Bagdad qui permettrait de passer une extension du protocole http. Des choses existent à l'Ifremer mais pas de façon opérationnelle. C'est une cible vers laquelle on se dirige.

Question : est-ce que les flowcam peuvent fonctionner ainsi ? Le FlowCam contient une unité centrale connectée au réseau, sous système d'exploitation Windows, ce qui permet d'enregistrer les fichiers sur tout répertoire réseau spécifié. La solution d'un répertoire réseau partagé dans le cloud est donc possible.

3 - DATARMOR

C'est la même solution que l'Own Cloud mais avec l'ajout de services et de fonctionnalités supplémentaires. Il serait notamment possible de lancer ZPI depuis DATARMOR pour optimiser les performances.

Le Cahier des Charges DATARMOR est en cours de rédaction. L'ouverture vers l'extérieur est incluse, notamment pour permettre l'utilisation de logiciels depuis l'extérieur et avec des performances bien supérieures à l'utilisation en local. Le stockage sur DATARMOR est gratuit.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	02/02/2016
	Objet : Etude bancarisation des données ZooPhytoImage - Restitution			
	Page 4/8			

Pour utiliser ZPI depuis DATARMOR, il faut pouvoir connecter plusieurs utilisateurs simultanément. Plusieurs architectures techniques sont possibles :

- Plusieurs Virtual Machine (1 / utilisateur) : attention au coût (2000€ / VM, sauf si facturation uniquement si les VM sont réellement utilisées)
- 1 machine / service
- Etc.

Tout dépendra du coût alloué à l'exploitation de ce système.

L'utilisation de DATARMOR est un besoin pour l'amélioration des performances. Est-ce que ZPI pose des problèmes de performance actuellement ? Si oui, pourquoi ?

NN : manque de recul pour savoir si ça pourrait être mieux. Il y a des phases de travail du logiciel qui sont longues, mais NN a du mal à estimer les gains possibles. Ex : le traitement des vignettes est assez long (2-3 min), d'autant plus que le nombre de particules dans l'échantillon est grand. Quand on analyse un échantillon, il y a ce prétraitement des vignettes, puis les interfaces s'ouvrent quasi instantanément. Si on souhaite utiliser un échantillon type pour analyser, le logiciel charge des informations pendant un temps assez long. NN s'est organisée pour faire autre chose pendant les phases de traitement.

➔ Il faudrait faire un test d'utilisation en local par rapport au réseau, pour voir si les temps de traitement sont optimisables.

Le gros intérêt du fonctionnement de ZPI sur DATARMOR est notamment de partager un logiciel unique dans la même version avec un set d'apprentissage commun géré à un seul endroit. Les classifieurs en revanche auront peut-être une déclinaison locale. De même, si des évolutions du logiciel sont proposées, elles sont déployées à un seul endroit pour tous les utilisateurs, au lieu de mettre à jour le logiciel sur chaque poste de travail.

3.1.2 Bancarisation

Les fichiers de résultats bruts sont au format RData, et doivent être convertis en .csv. Pour l'instant, cette conversion est faite manuellement par NN via un script R développé par Alice Lamoureux.

L'idéal serait de demander à P. Grosjean d'ajouter ce script au logiciel ZPI. Mais ça doit rester un choix utilisateur : il faut que l'utilisateur puisse demander le résultat en .csv uniquement si l'analyse semble pertinente.

Question : est-ce que Talend peut lancer des scripts R ?

Attention ! Il y a un besoin utilisateur de générer les .csv sans intégrer systématiquement les résultats dans Q². Le processus est en phase d'étude, donc il faut pouvoir gérer chaque étape séparément, notamment l'obtention des résultats en .csv pour les analyser et évaluer leur exactitude.

La bancarisation des données se limite aux informations associées aux particules phytoplanctoniques. Les données relatives aux débris et autres particules ne sont pas bancarisées.

Pour le fichier _valid.RData, une réflexion est à mener en interne Ifremer pour identifier les informations pertinentes à intégrer dans Q². Il faudra limiter le périmètre dans un 1^{er} temps, pour avoir les informations minimales. Dans tous les cas, il est décidé de ne pas stocker les données de chaque particule, et ceci pour aucun des paramètres (ex : 6000 particules x 10 paramètres = 60 000

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	02/02/2016
	Objet : Etude bancarisation des données ZooPhytoImage - Restitution			
				Page 5/8

lignes de résultats ! Trop important par rapport aux performances d'affichage dans Q²). Les informations liées aux particules devront donc être agrégées par taxon : cette agrégation doit faire l'objet d'une réflexion ultérieure, mais D. Soudant propose *a minima* que soient stockées les informations classiques pour définir la position (*e.g.* la médiane) et la variabilité (*e.g.* le min et le max), c'est-à-dire la médiane. En fonction du nombre de paramètres qui seraient retenus (C. Belin rappelle que les paramètres permettant d'estimer les biovolumes par taxon sont cruciaux), il faudra voir si d'autres agrégations pourraient être intéressantes au niveau d'un taxon. Il faudra voir aussi si des informations agrégées au niveau de l'échantillon pourraient également être intéressantes (par exemple répartition des classes de taille des particules phytoplancton sur l'ensemble de l'échantillon).

Rattachement multi-programme : la situation est hétérogène au niveau national (ex : toutes les données sont double programme en Normandie, mais pas à Arcachon où les échantillons peuvent être communs REPHY-ARCHYD, mais les flores uniquement REPHY). Donc il faut bien que le programme « REPHY » soit spécifié.

Attention ! Si des informations sont saisies dans l'interface Web pour le chargement unitaire, il faut prévoir comment les renseigner pour le chargement en masse sans passer par l'interface (reprise de données historiques). Cela concerne le programme et le saisisseur.

Le groupe propose, si le lancement est effectué en masse sans passer par l'interface, de :

- Ecrire le programme REPHY en dur dans le script Talend
- Mettre en saisisseur le même code que l'analyste.

La codification du mnémonique des échantillons doit être suffisamment simple pour éviter les erreurs de saisie dans Q². NN devra tester les éléments figés dans ce code par le logiciel ZPI, et les éléments libres. Malgré la complexité de ce code échantillon, le gain en saisie par chargement de fichier (au lieu de la saisie des taxons et dénombrements et biovolumes manuellement dans Q²) compense la vigilance nécessaire sur la saisie de ces codes dans les mnémoniques des échantillons Q².

Job Talend : le mnémonique échantillon doit contenir le code du fichier, mais pas forcément être strictement égal au code.

Une proposition d'évolution de Q² est faite : ajouter une case à cocher sur les échantillons pour identifier ceux destinés à recevoir les données ZPI. Le cochage de cette case permettrait le lancement de contrôles (ex : mnémonique rempli, structure du mnémo conforme à la codification, etc.). Rejeté par le groupe, car il est considéré que le rejet des échantillons dont le mnémo ne serait pas conforme ou inexistant, fait au moment du chargement du fichier Quadrilabo, est suffisant.

Pour le stockage des vignettes : pas de double sauvegarde DATARMOR / Q² ! Il vaut mieux une fonctionnalité d'extraction qui permette de récupérer les données Q² et de cocher une case demandant la récupération des vignettes associées situées sur DATARMOR.

Set d'apprentissage et classifier : sauvegardés sur les répertoires réseaux, pas dans Q² (pas de double sauvegarde). La mise à jour de ces informations étant très ponctuelle, il n'y a pas besoin d'outil de gestion spécifique.

<h1>Q3</h1>	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	02/02/2016
	Objet : Etude bancarisation des données ZooPhytoImage - Restitution			
	Page 6/8			

3.2 Conclusion sur le rapport d'étude

Cette étude a bien répondu au besoin de l'Ifremer car les solutions techniques possibles ont été évaluées et décrites, et les avantages et inconvénients pris en compte. L'architecture de sauvegarde et de bancarisation des données issues de ZPI est aujourd'hui clarifiée.

En revanche, le rapport d'étude final doit être complété :

- Rappeler, pour chaque solution, quel est le problème avec le fonctionnement actuel, et quels sont les apports des solutions proposées (ex : déployer ZPI sur DATARMOR. OK, mais pourquoi ? Parce que les performances peuvent être améliorées, et que l'outil serait commun à tous les utilisateurs avec set d'apprentissage commun, mises à jour uniques au lieu de chaque poste utilisateur, etc.).
- il manque les solutions écartées avec l'argumentaire ayant abouti au rejet des solutions. L'Ifremer souhaite tracer ces réflexions pour ne pas se reposer les questions ultérieurement.
- Rajouter des captures d'écran des fichiers de résultats.
- Préciser que les métadonnées sont saisies dans Q² préalablement.

Une nouvelle et dernière version du rapport d'étude sera fournie à l'Ifremer pour le 12 février au plus tard.

4 Poursuite des actions

Cf. tableau suivant.

Les solutions de type « cloud » sont pour l'instant écartées tant que les réflexions ne sont pas plus avancées sur l'opérationnalité des outils.

Pour l'instant, la méthodologie ZPI est en cours de test pour une mise en œuvre opérationnelle. Donc il n'y a pas de jeu de données à intégrer (si ce n'est des tests). Une analyse fine est en cours par l'Ifremer pour évaluer le fonctionnement de l'outil ZPI.

L'intégration des données via un job Talend nécessite le remplissage des tables de transcodage. Ce remplissage nécessite d'avoir les informations ZooPhytoImage à transcoder. Des réflexions sont encore à mener sur les paramètres utiles à migrer dans Q² (classes de taille, biovolumes, etc. d'après le fichier ...valid.RData). Le référentiel Quadrige² sera complété à l'issue de ces réflexions, et le job Talend d'intégration pourra alors être développé.

Durant le premier semestre 2016, les expertises seront poursuivies pour évaluer les possibilités de mise en routine de ZPI, d'intégration dans Q², d'architecture technique à mettre en place (DATARMOR), équivalences de processus avec le suivi du Zooplancton.

A l'automne 2016 (septembre), une réunion du présent groupe permettra de faire le bilan des actions menées en 2016, et des actions à poursuivre.



	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
	Page 7/8			

Tableau récapitulatif des actions

N°	Libellé de l'action	Porteur	Dû pour le
1	Rédiger une synthèse de l'étude pour le livrable Onema	EG	18 février
2	Faire un test d'utilisation de ZPI en local pour voir le gain de temps par rapport au fonctionnement en réseau	Nadine Neaud-Masson	1 ^{er} trimestre 2016
3	Etudier la possibilité pour Talend de lancer des programmes R (script Talend qui convertirait les données RData en .csv)	Cellule Q ²	1 ^{er} trimestre 2016
4	Demander à P. Grosjean l'ajout du script de conversion des RData en .csv	Equipe ZPI	Si Talend ne peut pas convertir les RData en .csv
5	Récupérer les informations détaillées sur le contenu du fichier ...valid.RData	NN	1 ^{er} trimestre 2016
6	Réflexion sur les informations détaillées des particules à migrer dans Q ²	Equipe ZPI + équipe stat Q ² + PELAGOS + LER/BO	1 ^{er} semestre 2016
7	Finaliser la codification des noms d'échantillon (<i>i.e.</i> nom des fichiers), après tests dans ZPI (certains éléments du code sont fixés par ZPI, d'autres sont libres)	NN	1 ^{er} semestre 2016
8	Présentation de l'étude ZPI à l'équipe ZooCam + Données Zooplancton Quadrige	C. Borremans + Cellule Q ²	1 ^{er} semestre 2016

	COMPTE RENDU DE REUNION			
	Lieu :	Ifremer Nantes	Date :	01/12/2015
	Objet : Etude bancarisation des données ZooPhytoImage			
				Page 8/8

N°	Libellé de l'action	Porteur	Dû pour le
9	Démonstration ZooCam + Zooldenier (?)	Equipe ZooCam + C. Borremans	1 ^{er} semestre 2016
10	Remonter les contraintes ZPI à DATARMOR	SB	2016
11	Intégrer ce processus de gestion des données ZPI dans le projet Q3	EG	2016
12	Création du référentiel PSFM + engins d'analyse dans Q ²	Cellule Q ²	Quand les données seront prêtes
13	Créer les programme / stratégies correspondant	Coordination REPHY	Après création du référentiel par la cellule
14	Remplir les tables de transcodage	Cellule Q ²	Après création des référentiels. Cette étape nécessite aussi que les évolutions de modèle sur les PSFMUs soient réalisées.
15	Saisir les échantillons spécifiques à l'intégration des données, en spécifiant leur mnémo	NN + Cellule Q ²	Quand les données seront prêtes
16	Développer le job Talend d'intégration des fichiers ZPI : ...res.RData (Lot1) puis ...valid.RData (Lot2)	Cellule Q ²	Quand les choix d'intégration de données seront effectués
17	Rédaction des consignes utilisateurs ZooPhytoImage intégrant la bancarisation dans Q ²	Coordination REPHY	Quand le processus sera prêt à devenir opérationnel



MASTER 2 ECOLOGIE

SPECIALITE

**FOnctionnement et Gestion des Ecosystèmes Marins
(FOGEM)**

**Co-habilité entre l'Université des Sciences et Technologies de Lille
(Lille 1) et l'Université du Littoral Côte d'Opale (ULCO)**

Année 2014-2015

Contribution à l'optimisation d'un réseau d'observation et de
surveillance du phytoplancton par analyses d'images

Rémi CUVELLIEZ

Alain LEFEBVRE

Ifremer



Remerciements

Je tiens tout d'abord à remercier mon maitre de stage, Alain LEFEBVRE, pour son suivi, ses conseils et son aide pendant ces 6mois de stage. Je souhaite également remercier toute l'équipe du LER qui a su m'accueillir chaleureusement.

Un grand merci à Nadine et Guillaume pour leur aide et leur collaboration au travail de ce stage. Merci également à Franck pour le temps qu'il a passé sur mes problèmes informatiques.

Je voudrai aussi remercier tous les partenaires participant au projet FlowCAM/ZooPhytoImage (Ifremer Nantes, Arcachon et Boulogne, Université de Mons et ULCO).

Je remercie mes collègues, Maud, Morgane et Benoit qui ont su garder une ambiance conviviale dans le bureau.

Enfin, je souhaite remercier tout le personnel de l'Ifremer pour m'avoir intégré au sein de cette structure comme un membre à part entière.

Sommaire

Remerciements	2
Introduction	4
Matériel et Méthodes :	5
REPHY et points de prélèvements	5
Le FlowCAM	6
Zoo/PhytoImage	8
Interprétation des résultats	10
Modules complémentaires	15
Résultats	16
Performances du set Manche-Atlantique	16
<i>Manche-Atlantique V0</i>	16
<i>Manche-Atlantique V1</i>	20
<i>Manche-Atlantique V2</i>	23
Performances du set Manche	25
Module de dénombrement des colonies	28
Discussion	34
Conclusion	40
Bibliographie	41

Introduction

Depuis plusieurs années, la Manche orientale fait l'objet d'un suivi régulier des communautés phytoplanctoniques qui y sont présentes. En effet, l'étude du phytoplancton est d'un intérêt majeur puisque celui-ci est responsable de la fixation de la moitié du carbone planétaire. De plus, des blooms successifs de *Phaeocystis globosa* (avec plus ou moins d'importance) sont apparus ces dernières années. Cette espèce est responsable du phénomène appelé « mousse de mai » qui se produit à la fin du bloom, lors de la destruction des colonies. Cette mousse est due à un mélange de la matrice permettant la cohésion des colonies, avec le battement des vagues. Ce phénomène n'est pas nuisible pour la santé mais à des enjeux économiques et touristiques. En effet, cette mousse dégage une forte odeur et recouvre les plages, voir même les rues pendant les années à forte concentration en *P. globosa*. Or ce phénomène, comme son nom l'indique, apparaît au printemps à la reprise des activités touristiques. Ces dernières sont donc dépendantes de l'intensité de ce bloom. Une deuxième espèce responsable de blooms est présente dans l'écosystème de la Manche. Cette espèce est *Pseudo-nitzschia*. Elle suscite un intérêt du fait qu'elle peut potentiellement être la cause de blooms toxiques affectant certaines activités commerciales. En effet cette espèce produit des toxines amnésiantes pouvant s'accumuler dans certains fruits de mer comme les moules. Même si cette toxicité est peu fréquente, il est essentiel d'avoir un suivi régulier afin d'anticiper toute contamination chez l'homme par ingestion de fruits de mer contaminés (Lefebvre *et al.*, 2011 ; Guiselin, 2010).

C'est dans ce contexte que s'insère le REPHY (réseau de surveillance du phytoplancton et des phycotoxines). Ce réseau national, créé à la suite de nombreuses intoxications alimentaires, a deux objectifs. D'une part, l'observation de la dynamique des communautés phytoplanctoniques dans un but de recherche et de suivi pour la DCSMM (Directive Cadre Stratégie pour le Milieu Marin) et d'autre part, la mise en œuvre d'un réseau de suivi des phycotoxines afin de prévenir d'éventuels risque de contaminations.

C'est donc pourquoi le système FlowCAM/ZooPhytoImage est développé. En effet, celui-ci pourrait permettre d'augmenter la fréquence d'échantillonnages du fait de sa rapidité d'analyse. Ce système pourrait donc permettre d'avoir une connaissance plus précise de la dynamique du phytoplancton sur la façade Manche-Atlantique et donc de prévenir plus efficacement l'apparition de blooms et/ou l'accumulation de toxines dans le milieu et les organismes présents.

L'objectif de ce stage de 6 mois sera donc de choisir et développer un outil de reconnaissance automatique ayant une efficacité optimale dans le secteur de la Manche orientale (le laboratoire de Boulogne-sur-mer étant responsable de ce secteur pour le REPHY). Le stage s'intégrant à la partie recherche et développement du projet, un deuxième objectif consistera à contribuer à l'optimisation du fonctionnement du logiciel ZooPhytoImage, de son efficacité et des méthodes utilisées afin de pouvoir entrer en phase opérationnelle avant la fin d'année 2015.

Matériel et Méthodes :

REPHY et points de prélèvements

L'utilisation du FlowCAM pour l'IFREMER s'intègre dans le cadre du REPHY qui est le réseau de surveillance du phytoplancton et des phycotoxines. Il a été créé en 1984 suite à de nombreuses intoxications alimentaires. C'est un réseau national, dont la couverture est assurée par plusieurs laboratoires côtiers qui se partagent le littoral français. Les objectifs du REPHY sont l'observation de l'ensemble des espèces phytoplanctoniques, le recensement des efflorescences exceptionnelles, ainsi que la détection de prolifération d'espèces toxiques ou nuisibles pour les consommateurs de coquillages. En effet, la dynamique des communautés phytoplanctoniques suscite un fort intérêt depuis plusieurs années. De plus, la Manche orientale a un intérêt particulier puisque des changements de dominances d'espèces ainsi que des blooms successifs ont pu être répertoriés depuis le début des années 1990 (Lefebvre *et al.*, 2011 ; Guiselin, 2010 ; Hernandez-Farinas, 2014).

Les prélèvements se faisant sur les points REPHY/SRN, ceux-ci vont donc servir à différentes analyses comme les MES, la turbidité, la chlorophylle et le pH. Une sonde CTD est aussi utilisée pour caractériser la structure verticale des masses d'eaux au niveau de chaque point de prélèvements (température, salinité, profondeur...).

Le laboratoire de Boulogne-sur-mer est responsable de 3 sites de prélèvements (REPHY et SRN confondus). Ces sites comportant eux-mêmes plusieurs points de prélèvements (Figure 2). Le premier site est celui de Dunkerque qui possède 3 stations de prélèvements (DK1, DK3, DK4). Le site suivant est Boulogne-sur-mer, avec lui aussi 3 points de prélèvements (BL1, BL2, BL3). Enfin, le dernier site se situe en Baie de Somme. Il comporte 5 stations de prélèvements (ATSO, BIF, MIM, ME1, ME2).

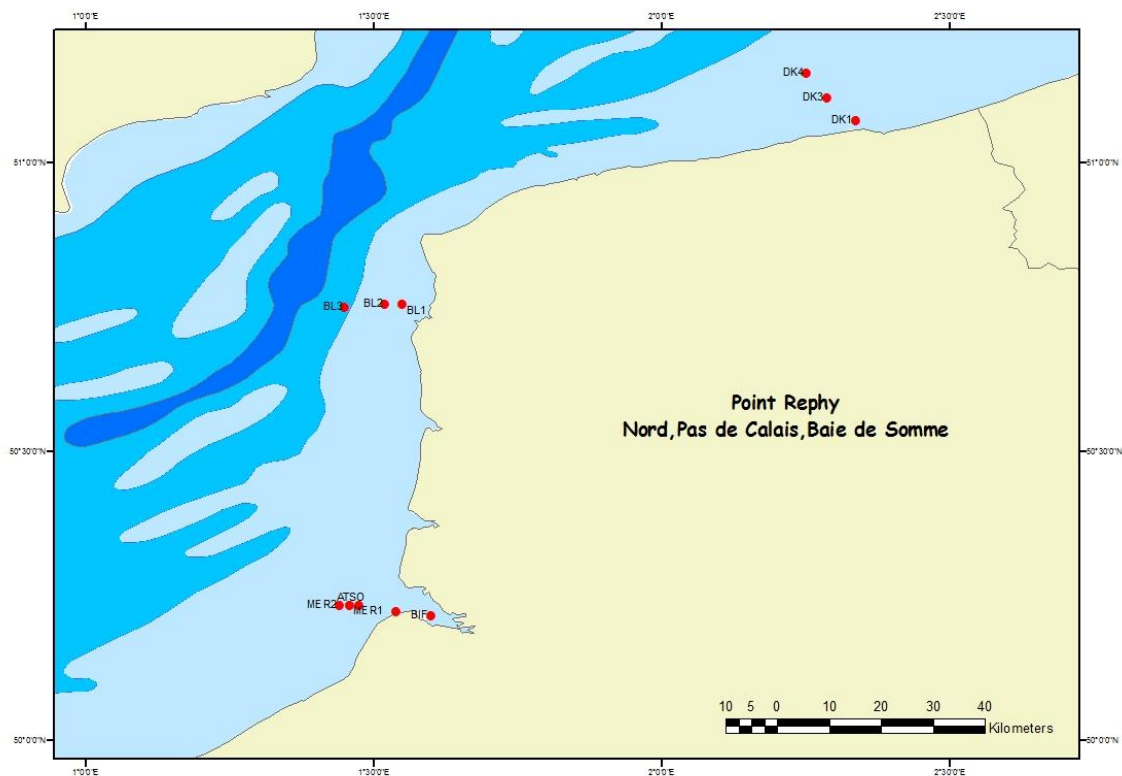


Figure 2. Localisation des points de prélèvements REPHY/SRN attribués au centre de Boulogne-sur-Mer (© Rémi CUVELLIEZ/LER-BL)

Les prélèvements se font en sub-surface à l'aide d'une bouteille Niskin. Ils sont ensuite stockés dans des flacons et sont passés au FlowCAM dès le retour au laboratoire (si le passage n'est pas possible le jour ou le lendemain du prélèvement alors celui est fixé au lugol pour le préserver avant utilisation). Attention, il faut prendre en compte que la fixation au lugol est une méthode destructrice. En effet, cette méthode change la couleur de l'échantillon, détruit certaines cellules ou en déforme d'autres. Elle est donc choisie en dernier recours et l'utilisation d'échantillons vivants est privilégiée. Parallèlement à l'utilisation du FlowCAM, une lecture de la flore planctonique est effectuée par microscopie inversée. Cela permet, entre autre d'avoir un point de comparaison pour les différents types d'identifications.

Le FlowCAM

Le FlowCAM (Flow Cytometer And Microscope) est un outil d'imagerie en flux de particules, développé par la société Fluid Imaging Technologies. Il est généralement composé d'un laser pour la détection de la fluorescence (absent sur le FlowCAM du LER Boulogne) et d'une caméra couplée permettant l'acquisition d'images, qui seront par la suite utilisées par l'outil de classification automatisée. Cet appareil est conçu pour fonctionner avec des échantillons liquides et visqueux, permettant l'analyse du plancton dans la colonne d'eau (Tunin-Ley et Maurer, 2011).

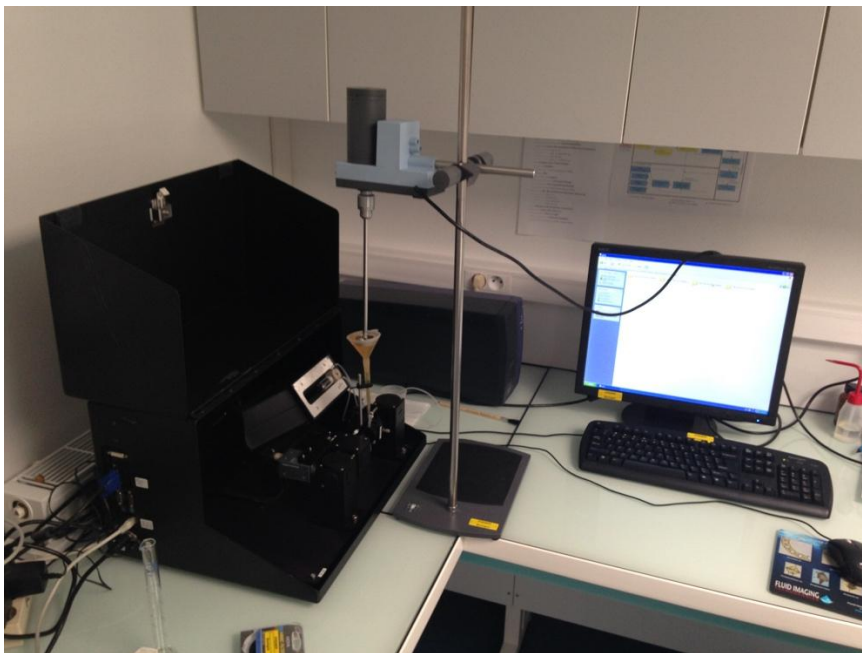


Figure 3. FlowCAM utilisé à l'IFREMER Boulogne (© Rémi CUVELLIEZ/LER-BL)

Le système fonctionne de la façon suivante :

Tout d'abord, le liquide est entraîné dans la chambre de flux (chambre en verre de dimensions parfaitement connues et positionnée devant l'objectif du microscope) par le biais d'une pompe péristaltique à 27 vitesses (cela permet de régler la pompe afin d'éviter une destruction des particules lors de leur passage dans la chambre de flux). Pour éviter la sédimentation des particules et l'obstruction de la pompe, l'échantillon est agité par une hélice pendant toute la phase d'acquisition. Une fois le liquide arrivé dans la chambre, celui est filmé par une caméra haute résolution (la chambre étant éclairée par une LED). Ensuite, en fonction du mode choisi, l'outil

effectuera des captures d'images et des mesures à partir de celles-ci (la classification se faisant grâce au logiciel Zoo/PhytoImage). Au LER Boulogne, le mode « AutoImage » est utilisé pour l'acquisition des images. Celui-ci définit l'acquisition d'un nombre fixe d'images par intervalles de temps réguliers (ici, 22 images.s⁻¹). Ce mode est privilégié car il offre une estimation très précise de la concentration cellulaire (Tunin-Ley et Maurer, 2011). Les photos sont par la suite classées dans des fichiers appelés « collages » afin d'optimiser les ressources mémoires de l'ordinateur. Finalement les images obtenues seront identifiées et classées par un logiciel de classification supervisée qui est développé par l'université de Mons en Belgique (fonctionnement développé dans la section suivante) (Sieracki et al. 1998 ; Lepage et al. 2004 ; See et al. 2005).

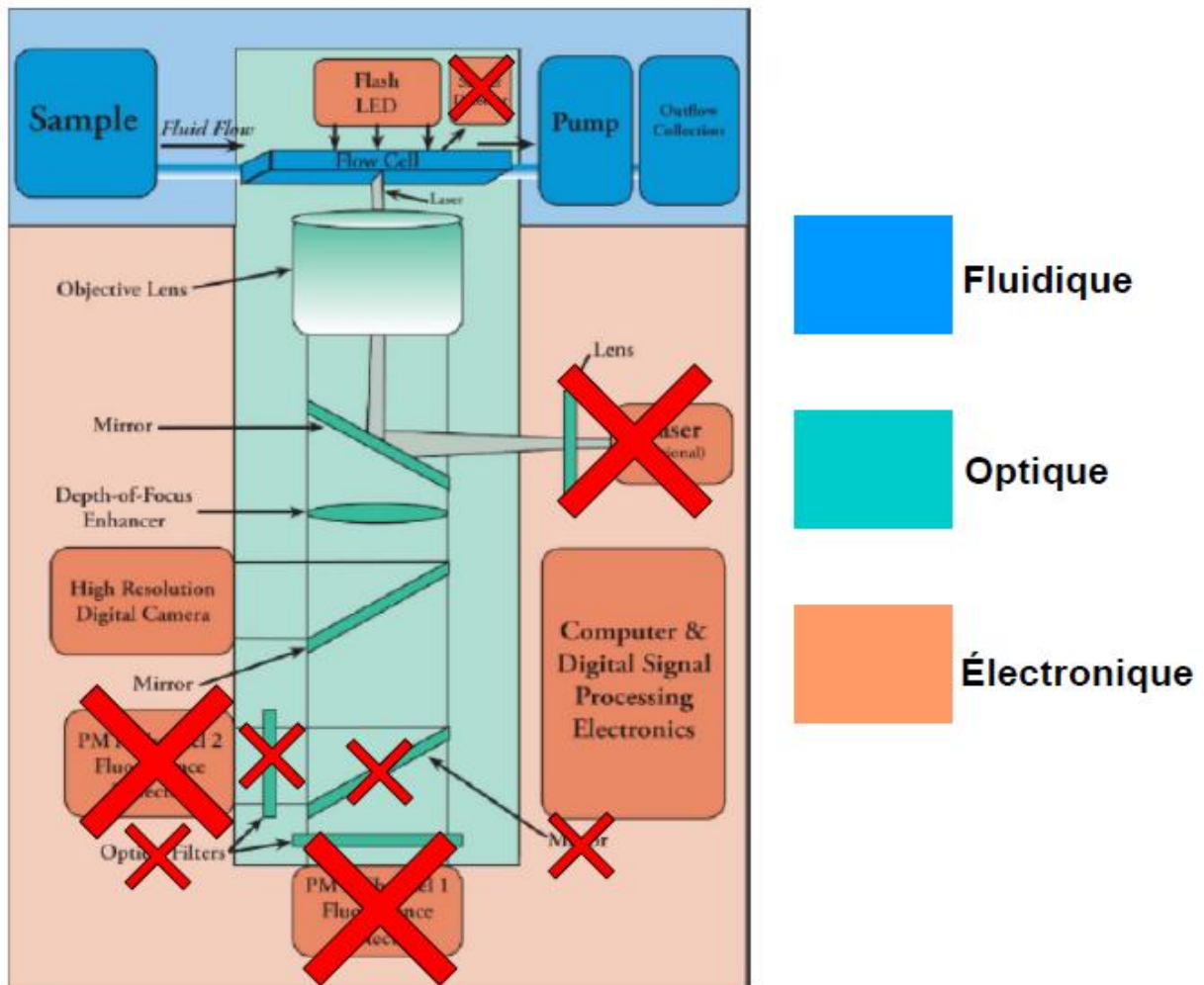


Figure 4. Schéma représentatif du FlowCAM du centre IFREMER Boulogne (© Guillaume Wacquet/LER-BL)

De plus, le fait d'avoir différentes combinaisons de chambres et d'objectifs, permet l'analyse de particules dont la taille varie entre 3 µm et 1mm. A l'Ifremer Boulogne, la chambre utilisée a une profondeur de 300 µm et l'objectif a un grossissement X4. Cela permet d'étudier une gamme de particules comprise entre 30 et 300 µm (Buskey et al. 2006).

Zoo/PhytoImage

Zoo/PhytoImage est un outil développé par l'université de Mons (Belgique) permettant aux biologistes d'étudier des séries spatio-temporelles d'échantillons de planctons (que ce soit du zooplancton ou du phytoplancton) à partir d'images. C'est un logiciel gratuit et « open source ». Il fonctionne comme une boîte à outils, pilotées par une interface graphique développée sous R. Il a été conçu pour traiter des images provenant de différents appareils (ex : FloxCAM, ZooScan, scanner, loupe binoculaire, etc.). L'utilisation de Zoo/PhytoImage permet de pallier aux limites de Visual SpreadSheet (logiciel fourni avec le FlowCAM) concernant la reconnaissance des particules.

Son utilisation peut être découpée en 3 parties :

- Traitement d'images
- Set d'apprentissage
- Outils de reconnaissance et analyse spatio-temporelle de séries

→ Traitement d'images :

Avant de récolter les informations relatives aux différents paramètres, il est nécessaire d'effectuer un traitement de l'image. Cela permet de discriminer la particule du fond de la cuve et donc de ne faire les mesures que sur celles-ci.

Pour effectuer le traitement, le logiciel va réaliser un processus de binarisation des images (on parle aussi de segmentation des images). C'est à dire qu'il va utiliser un masque binaire pour découper la région d'intérêt dans l'image d'origine (*i.e.* : image brute). Les silhouettes obtenues sont ensuite remplies par les différents niveaux de gris présent sur l'image d'origine. Les images obtenues après ce processus sont appelées « blobs » (*i.e.* : Binary Large OBjectS). C'est sur ces blobs que les mesures seront faites par le logiciel. Les valeurs obtenues seront utilisées lors de la création de l'outil de reconnaissance et pendant l'identification des particules. Enfin, une image est créée pour chacune des particules. Celle-ci reprend l'image de la particule analysée, à laquelle est ajoutée une échelle en haut à droite. L'image obtenue est appelée « vignette ».

C'est cette vignette qui sera utilisée pour la classification manuelle lors de la création du set d'apprentissage. (Wacquet *et al.*)

→ Set d'apprentissage :

La confection du set d'apprentissage est une étape cruciale de la reconnaissance automatisée du phytoplancton. En effet, il correspond à une base de données qui sera utilisée pour l'élaboration de l'outil de reconnaissance automatisée. Si plusieurs couplages objectif/chambre de flux sont utilisés (ex : 4X/300µm et 10X/100µm) alors il sera nécessaire de créer autant de set que de couplages.

Lors de la création du set d'apprentissage, l'utilisateur va devoir classer manuellement un ensemble de vignettes dans différents groupes (ces derniers correspondant à un/plusieurs genres ou à une espèce). Les dossiers obtenus après cette répartition seront eux même classés sous la forme d'une arborescence de dossiers représentant la classification taxonomique du phytoplancton.

Avant de commencer la création du set d'apprentissage, il faudra éliminer les groupes minoritaires en milieu naturel car il y aura un manque de données et ceux-ci seront mal reconnus lors des analyses. Il est aussi nécessaire que les images du set d'apprentissage représentent au mieux la

variabilité naturelle d'un taxon donné, que celle-ci soit intrinsèque (était physiologique, longueur des colonies, taille des cellules, etc.) ou extrinsèque (position de la cellule sur l'image, netteté de l'image, etc.). Le set aura également besoin d'une arborescence de dossiers représentant les débris susceptibles d'être trouvés dans les échantillons et qui sont généralement source de confusion. Enfin, il est indispensable de créer des groupes zooplanctoniques qui contiendraient des espèces aptes à passer dans la chambre de flux du FlowCAM.

Idéalement, il faudrait que chaque groupe ai le même nombre de vignettes. En effet, il semblerait que le nombre d'images donne proportionnellement plus de poids à un groupe au niveau de la classification des taxons par l'algorithme. Néanmoins, l'ajout de vignettes dans certains groupes permet de diminuer le pourcentage d'erreur (tout comme, dans certains cas, la fusion de groupes auparavant confondus ou la fission d'autres groupes en plusieurs dossiers). En effet, le seuil de vignettes par groupes reste tout de même différent, du fait que la variabilité (intrinsèque ou extrinsèque) des vignettes est différente pour chaque taxon. (Tunin-Ley et Maurer, 2011)

Outil de reconnaissance :

Il existe 2 méthodes de classification utilisables :

La classification non supervisée → ici les classes ne sont pas reconnues *a priori*.

La classification supervisée, aussi appelé « machine learning » → les différentes classes sont connues *a priori* et ont une nomenclature qui leur est associée. Le concept de « machine learning » repose sur l'utilisation d'un algorithme qui est capable de classer les particules à l'aide de mesures associées. L'outil effectue l'apprentissage nécessaire sur un jeu de « données exemples » où l'identification des particules est préalablement connue (*i.e.* : le set d'apprentissage). L'algorithme va donc développer une série de règles qui vont se baser sur les mesures effectuées auparavant. L'outil de reconnaissance sera finalement capable de prédire la classe de particules indépendantes au set, uniquement à partir de mesures effectuées sur celles-ci.

C'est cette dernière méthode qui sera utilisée pour créer l'outil de reconnaissance relatif aux travaux de l'IFREMER.

Avant de créer l'outil de reconnaissance, il est indispensable de choisir correctement l'algorithme qui va gérer la classification des particules.

Pour cela, l'équipe de l'université de Mons a procédé à un test de validation croisée sur les différents algorithmes proposés. C'est l'algorithme « Random forest » qui a été sélectionné. En effet, celui-ci a fourni le meilleur taux global de reconnaissance (*i.e.* 77,59 %), suivi de la LDA (71,76%). De plus, les autres algorithmes ne fournissent pas un taux suffisamment élevé pour envisager leur utilisation (Grosjean et Denis, 2010). « Random forest » est aussi appelé « forêt d'arbres décisionnels » et a été proposé par Léo Breiman (Breiman *et al.*, 2001). Il s'agit d'un algorithme qui va générer une forêt comportant un nombre important d'arbres décisionnels différents (ces derniers étant indépendants). Chacun de ces arbres va donc fournir une classification indépendante des autres arbres. L'algorithme va ensuite classer les particules du set dans chaque arbre en fonction des informations contenues dans ceux-ci. Puis, une fois toutes les particules classées, l'algorithme choisira l'arbre permettant la meilleure classification des données.

Une fois l'algorithme choisi, l'utilisateur pourra enfin générer l'outil de reconnaissance à partir de celui-ci et du set d'apprentissage.

Interprétation des résultats

Performance de l'outil de reconnaissance

Avant de commencer toute analyse d'échantillons, il est indispensable de tester les performances de l'outil de reconnaissance. Cela se fait par le biais d'une méthode statistique qui utilise uniquement le set d'apprentissage : la validation croisée.

Pour ce faire, la méthode mélange aléatoirement le set d'apprentissage en 10 sous-ensembles de même effectif. Ensuite, un apprentissage sur 9 des 10 sous-ensembles créés est réalisé. Puis la reconnaissance va s'opérer sur le sous-ensemble restant. Cette action est répétée 10 fois (à chaque répétition, un sous-ensemble différent est utilisé pour le test de reconnaissance). On aboutit donc à l'obtention de 10 outils de reconnaissances ignorant chacun une fraction différente du set d'origine. Enfin, les identifications obtenues sont regroupées et comparées aux comptages traditionnels par microscopie inversée.

Les résultats de ce test sont visibles par le biais de différentes figures décrites ci-dessous.

Matrice de confusion :

Suite au test de validation croisée, 2 vecteurs sont créés. Le premier contient le nombre prédit de chaque classe et le second correspond au vecteur original de la classification manuelle. La matrice de confusion est donc un tableau de contingence qui regroupe ces deux vecteurs au sein de l'ensemble des classes représentées. Elle confronte l'identification traditionnelle des organismes classés manuellement, dans le set d'apprentissage, à l'identification automatique de ces mêmes images par l'outil de reconnaissance. Les organismes correctement identifiés (par l'outil de reconnaissance) vont se situer sur la diagonale de la matrice. Tandis que les individus localisés hors de la diagonale sont ceux qui n'ont pas été reconnus correctement. La matrice nous permet également de savoir combien de confusions ont été faites et entre quels groupes.

La lecture en ligne de la matrice nous informe sur la façon dont ont été classés les vignettes du groupe donné. Alors que la lecture en colonne nous informe sur le contenu du groupe donné après classification par l'outil (Lefebvre et Wacquet, 2014).

Groups clustering :

D'autres représentations graphiques peuvent aider l'utilisateur à interpréter les résultats de classification obtenus comme le dendrogramme appelé « Groups clustering ». Il illustre les séparations des groupes du set selon les difficultés de discriminations qu'a rencontré l'outil de reconnaissance. Il nous donne deux grandes informations : les liens entre les différents groupes et aussi leur niveau de proximité (par le biais de la hauteur des branches). Il est donc possible de retrouver les confusions qu'il est possible d'observer entre les différentes particules.

Taux global de reconnaissance et taux global d'erreur :

Le taux global de reconnaissance (aussi appelé « accuracy ») est une statistique globale qui renseigne sur la proportion de tous les organismes correctement classés par l'outil de reconnaissance. Il correspond tout simplement au pourcentage d'individus situés dans la diagonale de la matrice de confusion.

On a donc : $TGR = 100 \times \frac{\text{diagonale}}{N}$ (N étant le nombre total d'individus, qu'il soit ou non sur la diagonale).

Le taux global d'erreur (TGE), lui, renseigne sur la proportion de tous les organismes qui ont été incorrectement identifiés par l'outil.

Ces taux sont complémentaires. On obtient donc $TGR + TGE = 1$

Représentations du « Recall » et de la « Precision » :

Ces représentations sont utiles à la compréhension des phénomènes conduisant aux résultats obtenus. Par exemple, lors de travaux sur un set antérieur, il a pu être remarqué que l'abondance des groupes *Pseudo-nitschia_spp* et *Pleuro-Gyro_empty* était correctement estimée malgré leur taux d'erreur respectifs égaux à 7,692 et 18,333 % (Wacquet et Lefebvre, 2014). Grâce au Recall et à la Precision, il sera donc possible d'expliquer le fait que l'estimation de l'abondance reste correcte.

Pour créer ces représentations graphiques, 4 taux sont utilisés (Wacquet et Lefebvre, 2014) :

- Le taux de faux positifs (FP) :
Il représente la proportion d'organismes d'un groupe classés à tort dans un groupe considéré. Ces erreurs augmentent l'estimation de l'abondance du groupe considéré. Ce taux est calculé à partir des colonnes de la matrice de confusion.
- Le taux de faux négatifs (FN) :
Il correspond à la proportion d'organismes d'un groupe considéré qui sont classés à tort dans un autre groupe par l'outil de reconnaissance. L'augmentation du taux de faux négatifs va entraîner une diminution de l'abondance mesurée du groupe. Ce taux est calculé selon les lignes de la matrice de confusion.
- Le « Recall » :
Il est aussi appelé « Sensitivity » ou taux de vrai positif. Il correspond à la proportion d'organismes correctement identifiés dans un groupe donné par rapport au nombre total d'organismes appartenant effectivement à ce groupe.
Si on a $\text{Recall} = 1$, alors cela signifie que chaque objet d'un groupe donné a bien été classé dans ce même groupe (*i.e.* Tous les organismes devant être placés dans ce groupe l'ont été). Toutefois, le Recall n'indique pas le nombre d'individus incorrectement classés dans ce groupe.
En d'autres termes, le Recall mesure la capacité de l'outil de reconnaissance à classer des individus appartenant à un groupe, dans ce même groupe, quel que soit le nombre d'autres individus classés dans ce groupe.

$$\text{Recall} = \frac{\text{Nb de particules correctement attribuées au groupe}}{\text{Nb de particules appartenant réellement au groupe}} = \frac{TP}{TP + FN}$$

- La « Precision » :
Ce taux est une mesure de l'exactitude ou de la fidélité d'un outil de reconnaissance. La Precision est décrite comme la proportion d'objets d'un groupe donné, correctement reconnus. Pour une Precision égale à 1, toutes les particules attribuées à ce groupe appartiennent effectivement à celui-ci. En revanche, elle n'indique pas le nombre de particules devant être placées dans ce groupe mais qui ne l'ont pas été.

Autrement dit, c'est une mesure de la capacité d'un outil de reconnaissance à fournir majoritairement de bonnes identifications, quelle que soit par ailleurs la fraction des individus de ce groupe effectivement classés comme tels.

$$Precision = \frac{Nb \text{ de particules correctement attribuées au groupe}}{Nb \text{ de particules attribuées au groupe}} = \frac{TP}{TP + FP}$$

→ Le F-score

$$Fscore = 2 \times TP$$

Le F-score est aussi appelé « Precision versus Recall ». Il nous donne des informations sur le poids des faux positifs et des faux négatifs et donc indirectement du Recall et de la Precision. Plus la valeur du F-score sera élevée, plus la reconnaissance sera correcte. De plus, le F-score se déplace vers la droite ou la gauche par rapport au centre du graphique, en fonction des valeurs du Recall et de la Precision. En effet, un déplacement du F-score vers la droite nous indique qu'il y a plus de faux positifs que de faux négatifs. Le Recall est donc plus élevé et il faut s'attendre à une surestimation de l'abondance du groupe donné. Inversement, si le F-score tend vers la gauche cela nous indique une Precision plus élevée et le groupe aura donc tendance à être sous-estimé.

→ Groups comparison

Cette représentation graphique nous permet de visualiser directement et simplement le Recall et la Precision. Les deux taux sont représentés sous forme de demi-cercles (demi-cercle vert pour le Recall et demi-cercle bleu pour la Precision). Le diamètre des demi-cercles pour chacune des statistiques est relatif à leur valeur (*i.e.* plus le diamètre est élevé, plus le taux est élevé). Dans cette représentation, le but est d'avoir des diamètres les plus grands et les plus proches possibles (*i.e.* identification et estimation de l'abondance optimales).

Analyse des échantillons

La phase d'analyse des échantillons se fait via une interface graphique. Celle-ci fonctionne dans le navigateur internet par défaut (Figure 5).

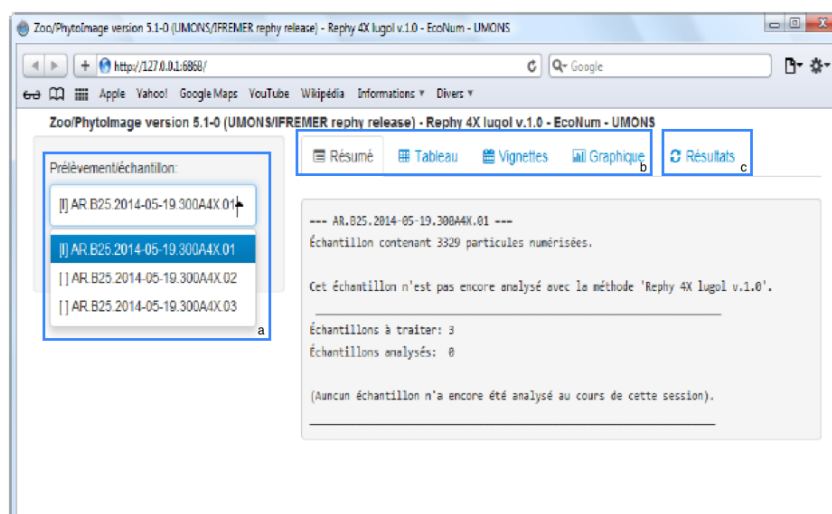


Figure 5. Interface graphique du logiciel ZooPhytoImage (a : module d'importation des données brutes, b : module de visualisation des données importées, c : onglet résultats)

Cette interface est composée de plusieurs modules :

Module d'importation des données brutes dans ZooPhytoImage (Figure 5a) :

Comme son nom l'indique, ce module permet l'importation des échantillons présents dans le répertoire de travail. Chaque étape de l'importation est caractérisée par un code entre crochets. Pour un échantillon qui n'est pas encore importé le code sera « [] ». Si l'échantillon est importé ce sera « [I] » et si celui-ci est analysé le code sera « [A] ». Dans cette version du logiciel, l'importation se fait automatiquement. Il suffit de sélectionner l'échantillon souhaité et le logiciel créera les fichiers nécessaires.

Module de visualisation des données importées (Figure 5b) :

Ce module comprend 4 onglets différents.

→ **Résumé** : cet onglet fournit les informations générales sur l'échantillon sélectionné et importé (*i.e.* nombre de particules numérisées, état de l'analyse avec la méthode sélectionnée, nombre d'échantillons à traiter, nombre d'échantillons analysés).

→ **Tableau** : cet onglet offre une visualisation rapide et simple des différentes mesures pour chaque individu de l'échantillon. Il est possible de trier les valeurs par ordre croissant ou décroissant.

→ **Vignettes** : cet onglet affiche aléatoirement 30 vignettes de l'échantillon sélectionné.

→ **Graphique** : cet onglet permet de visualiser la distribution de taille des particules sous la forme d'un graphique (en abscisse : taille des particules ; en ordonnée : fréquence).

Module d'analyse des échantillons et de visualisation des résultats :

Ce module peut être divisé en 3 parties.

→ **(Re)analyser** : une fois qu'un échantillon est importé, il est possible de l'analyser en le sélectionnant et ensuite en utilisant la commande « (Re)analyser ».

→ **Correction d'erreur** : une fois l'analyse lancée, le logiciel va ouvrir le module de correction d'erreur dans une nouvelle fenêtre du navigateur internet. Dans cette fenêtre est affiché un ensemble (par défaut) représentant 1/20 des vignettes de l'échantillon.

Il faudra opérer une correction sur cette partie du set qui est sélectionnée aléatoirement. Pour cela, on pourra reclasser les vignettes incorrectement identifiées dans leur colonne respective (Figure 6).

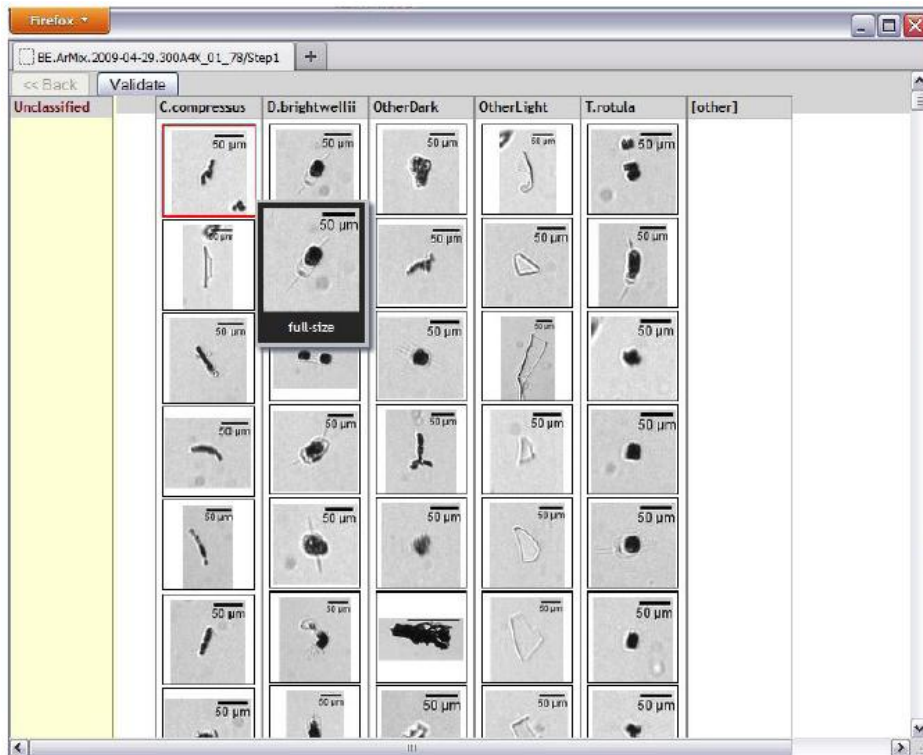


Figure 6. Module de correction d'erreur du logiciel ZooPhytoImage

Une fois que la correction de cette partie de l'échantillon est validée, un rapport sur le processus de validation effectué est affiché (Figure 7).

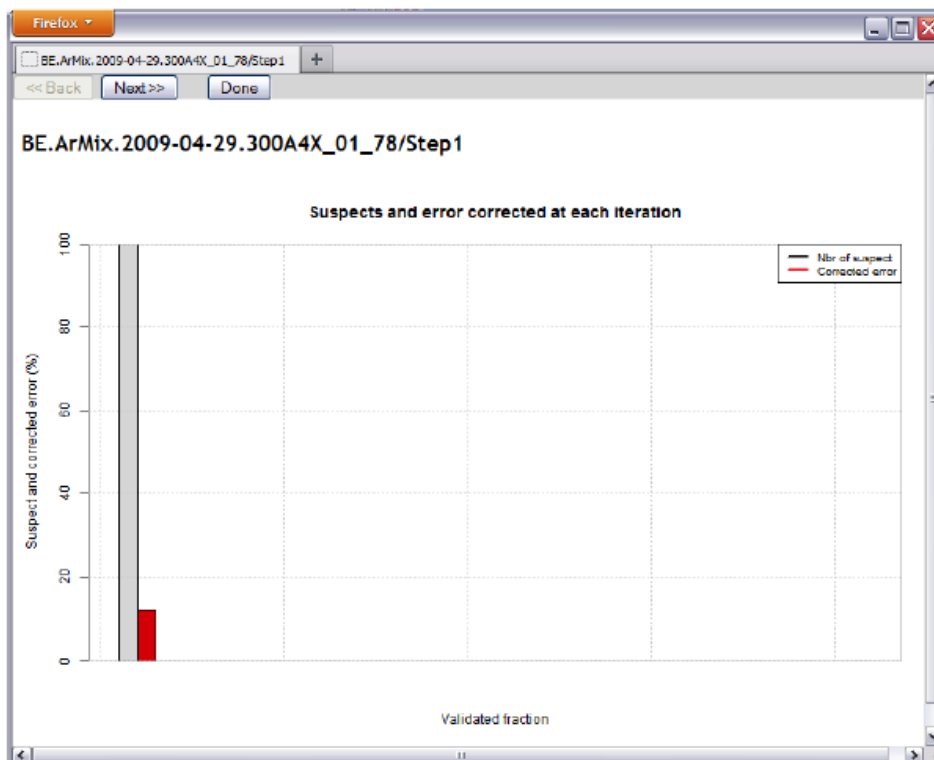


Figure 7. Rapport du processus de validation

Ce rapport est visualisable sous la forme d'un diagramme bâtons (en gris, proportion d'objets suspects dans la fraction validée et en rouge, fraction de vignettes incorrectement identifiées et corrigées).

Ensuite, deux choix sont possibles :

- Cliquer sur « Next » :
Le logiciel va présenter un deuxième sous-ensemble de particules de l'échantillon étudié. En revanche, cette fois il n'est pas choisi aléatoirement. Il est confectionné à partir de « vignettes suspectes » (*i.e.* Au vu des corrections faites lors de la première correction d'erreur, l'outil de reconnaissance va en déduire des vignettes suspectes dont l'emplacement après identification n'est pas certain). Le taux d'erreur est donc plus élevé. Cela permet de se concentrer essentiellement sur les vignettes qui posent problème et donc d'être plus efficace lors de la correction. Suite à ce deuxième traitement, un nouveau rapport du processus de validation est présenté.
Il est possible de réitérer cette action jusqu'à avoir corrigé l'entièreté de l'échantillon mais ici, le but est d'avoir des valeurs de la proportion d'objets suspects et de vignettes incorrectement identifiées les plus proches possible et les plus petites possibles.
- Cliquer sur « Done » :

Une fois que les valeurs du rapport de processus de validation correspondent aux attentes, on utilise la commande « Done » pour finaliser l'analyse sur cet échantillon (que la correction soit faite entièrement ou non). A ce stade, il n'est plus possible de reprendre une correction des erreurs.

→ Résultats :

Une fois l'analyse terminée, on peut remarquer que sur l'interface graphique, le code devant l'échantillon a changé (*i.e.* le code est passé de [I] à [A]). Cela signifie que l'échantillon a bien été analysé et que les résultats sont disponibles.

Dans l'onglet résultats (Figure 5c), on obtient les abondances ainsi que les biovolumes et les spectres de tailles de chaque groupe.

De plus, les résultats obtenus sont automatiquement sauvegardés sur le disque dans un nouveau dossier qui se trouve dans « _analyses ». Ce nouveau dossier comporte 3 fichiers :

- <echantillon>_res.RData :
C'est un fichier manipulable sous R. Il contient les résultats pour les abondances, biovolumes et spectres de taille.
- <echantillon>_valid.RData :
Egalement manipulable sous R, ce fichier comprend toutes les mesures effectuées sur chaque particule.
- <echantillon>_valid.txt :
Ce fichier contient les informations essentielles à l'interprétation des résultats (nom de l'ensemble d'apprentissage et de l'outil de reconnaissance, algorithme utilisé, métadonnées, etc.).

Modules complémentaires

Afin d'améliorer les performances de l'outil de reconnaissance, 2 modules complémentaires sont en cours de développement.

Le premier est un module de dénombrement des colonies. Il s'agit d'un petit programme permettant le comptage manuel des cellules dans le but d'en déduire par la suite un algorithme d'estimation du nombre de cellules au sein des colonies. Celui-ci est actuellement en développement et sera intégré à la version 5 du logiciel dès que possible.

Le second module est un module d'apprentissage actif. Celui-ci devrait pouvoir permettre à l'outil de reconnaissance d'apprendre de ses erreurs à l'aide de la correction manuelle, ainsi que d'ajouter des vignettes essentielles à l'amélioration du set.

Résultats

Performances du set Manche-Atlantique

Manche-Atlantique V0

Le training set Manche-Atlantique a été développé dans le but d'éviter toutes variabilités d'analyse dues aux différences de conception des sets entre les centres Ifremer. Il a été créé sur la base des numérisations effectuées dans les laboratoires de Boulogne-sur-mer, Nantes et Arcachon. Il comprend 9788 vignettes classées dans 67 groupes différents, eux même étant répartis de la façon suivante :

- 29 groupes phytoplanctoniques intégrés à la classe des *Bacillariophyceae*.
- 16 groupes phytoplanctoniques de la classe des *Dinophyceae*.
- 5 groupes phytoplanctoniques appartenant à d'autres classes.
- 5 groupes zooplanctoniques représentant les individus visibles sur les numérisations FlowCAM.
- 12 groupes détritiques représentatifs des débris/particules inertes présents dans les échantillons.

Les groupes sont présentés en annexe 1.

La première étape dans l'obtention des résultats a été de calculer les taux d'erreur grâce à la validation croisée, afin de connaître les groupes pour lesquels l'outil de reconnaissance est le moins performant. En couplant ces résultats au nombre de vignettes par groupe, il sera possible de discriminer ceux dont le nombre de vignettes à disposition (<20), ne permet pas d'améliorer les performances de la reconnaissance. Ceux-ci seront par la suite supprimés du set d'apprentissage. Ils sont visibles dans le Tableau 1 ci-dessous.

Tableau 1. Synthèse des groupes écartés du set d'apprentissage après première analyse des performances.

Nom du groupe	Nombre de vignettes	Taux d'erreur
<i>polen</i>	16	44
<i>Empty frustule</i>	15	100
<i>Lauderia</i>	12	92
<i>Helicotheca</i>	7	100
<i>Pisces larva</i>	7	100
<i>Diplopsalis Diplopelta Oblea</i>	6	100
<i>Asteroplanus karianus</i>	5	100
<i>Eucampia</i>	3	100
<i>Pyromonas</i>	3	100
<i>Hemiaulus</i>	2	100
<i>Heterocapsa</i>	2	100

Au total, 11 groupes ont été écartés du set d'apprentissage et placés dans le dossier nommé « _unused ». Cela permet de les conserver en vue d'un enrichissement de ceux-ci au fur et à mesure des numérisations et de les réintégrer au set ultérieurement. Le groupe *Noctiluca* ne possède que 16 vignettes mais est conservé tel quel car le taux d'erreur associé est tout à fait acceptable (13%). Un groupe a aussi été ajouté : « Hantzschia Nitzschia ».

Un nouveau set est alors créé. Il est composé de 57 groupes, et correspond à la version 0.1 du training set. Les TGE des versions 0 et 0.1 peuvent donc être comparés. On remarque une diminution du taux d'erreur de 8,34 % entre les deux versions ce qui est non négligeable (TGE = 46,43 % pour la version 0 et TGE = 38,09 % pour la version 0.1). Attention, cela n'indique pas une amélioration des performances sur tous les groupes, c'est une statistique qui nous donne une information globale des performances. En effet, on note que 28 groupes ont vu leur taux d'erreur augmenter, 17 ont diminué et 11 ont gardé la même valeur. 50 % des groupes ont donc vu leurs performances diminuer par rapport à l'ancienne version. Néanmoins ces variations restent faibles et la suppression de certains groupes engendrera moins d'interférences avec d'autres une fois l'optimisation terminée. Les résultats de l'évaluation de l'outil de reconnaissance après écartement des groupes à faible effectif sont visibles dans l'annexe 1.

Une nouvelle validation croisée a également été effectuée afin de pouvoir analyser les performances de la version 0.1 à l'aide des figures disponibles dans la version 5 du logiciel ZooPhytoImage.

Que ce soit pour le F-score (Figure 8) ou le taux d'erreur, il est difficile de choisir un seuil à partir duquel le taux devient mauvais. En effet, théoriquement il n'y a pas de seuil puisque le but est d'optimiser au maximum le set et l'outil de reconnaissance (i.e. pour le F-score on cherche à atteindre 100 % et pour les taux d'erreur, l'objectif est de s'approcher le plus possible des 0 %). Pour avoir une idée des performances, nous avons donc décidé de classer les groupes en 3 catégories pour l'estimation de l'abondance et de la reconnaissance (Tableau 2). Attention, ces seuils sont donc totalement arbitraires et sont utiles pour comparer les différentes versions des sets entre elles.



Figure 8. F-score pour chaque groupe composant le set Manche-Atlantique V0.1.

Il est a noté que la mauvaise qualité des figures est due à l’affichage fait par le logiciel.

Nous avons donc choisi de classer les valeurs de F-score en fonction de 3 catégories :

- Bon : valeurs supérieures ou égales à 70 %
- Moyen : valeurs comprises dans l’intervalle [50 ; 70[
- Mauvais : valeurs inférieures à 50 %

On a alors 25 groupes dont le F-score est bon, 16 avec un F-score moyen et 16 mauvais. Si l’on considère que seuls les F-score bons sont acceptables, alors une majorité des groupes est à retravailler (*i.e.* 32 groupes).

Les estimations d’abondances sont classées selon ces 3 catégories :

- Estimation correcte : F-score centré
- Sous-estimation : F-score décalé vers la gauche
- Surestimation : F-score décalé vers la droite

22 groupes sont alors en sous-estimation et 15 en surestimation, tandis que seulement 19 groupes ont une estimation correcte. Il est a noté que le groupe « Pisces larva » n’a pas pu être classé puisque sa reconnaissance est nulle.

Globalement, il y a donc une tendance à la sous-estimation et celles-ci semblent être plus fortes en termes d’intensité (les déplacements vers la gauche sont plus prononcés que ceux vers la droite). Il est aussi possible de remarquer que les groupes de la catégorie « débris » sont potentiellement source de problèmes. En effet, sur 10 groupes, 6 sont en sous-estimation, 3 en surestimation et seulement 1 groupe dont l’estimation de l’abondance est correcte. De plus, 9 de ces groupes sont en dessous du seuil acceptable pour le F-score (*i.e.* 70 %).

La catégorie « zooplancton » peut également être source d'erreurs. En effet, sur les 6 groupes la composant, seulement 1 a une valeur de F-score acceptable. De plus, tous les groupes sont en sous-estimation, sauf « Pisces larva » qui ne peut pas être classé.

Pour une visualisation simple et rapide des performances de reconnaissance et d'estimation de l'abondance, il est possible d'utiliser le graphique « groups comparison » (Figure 9).

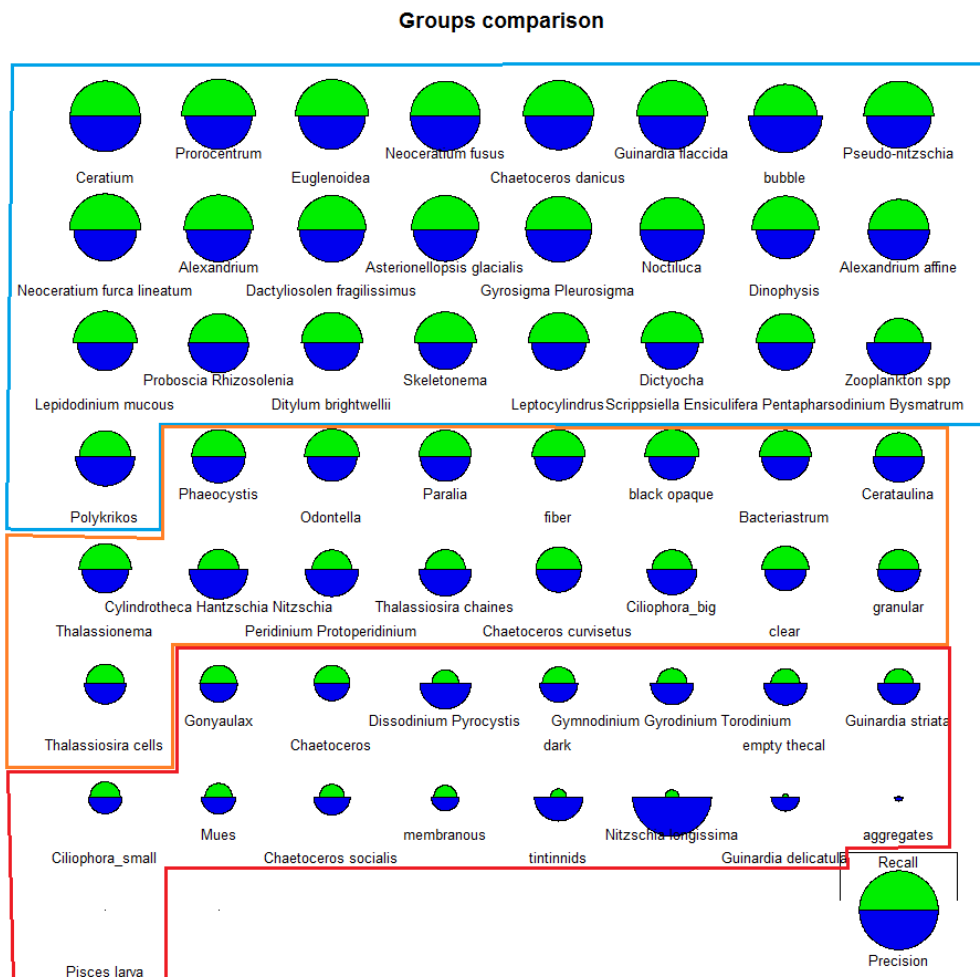


Figure 9. Visualisation des performances de reconnaissance et d'estimation de l'abondance pour le set Manche-Atlantique V0.1 sous forme de demi-cercles (en bleu : les groupes dont la valeur du F-score est considérée comme bon ; en orange : les groupes dont le F-score est moyen ; et en rouge : les groupes avec de mauvais résultats).

A la vue de ces performances, la décision d'une nouvelle optimisation vers une version 1 du set Manche-Atlantique a été prise.

Manche-Atlantique V1

Optimisation vers une version 1

Lorsque le nombre de vignettes pour un groupe dépasse les 200, il a pu être remarqué une stagnation de l'augmentation des performances de reconnaissance (Wacquet *et al.*, 2014). Cela est dû au fait que la quasi-totalité de la variabilité du groupe est représentée avec environ 200 vignettes. Néanmoins certains groupes ont plus de variabilité que d'autres. C'est pourquoi, il a été décidé de garder au maximum 260 vignettes afin de ne pas surcharger le set inutilement. Les vignettes à supprimer seront sélectionnées dans le but de garder un maximum de variabilité, un minimum de redondances et de garder des vignettes provenant des 3 laboratoires.

Dans le but d'optimiser le set Manche-Atlantique, il est nécessaire de pouvoir visualiser les confusions existantes entre les différents groupes. Pour une visualisation rapide des confusions il est possible d'utiliser le graphique appelé « Groups clustering » (Figure 10) et pour une analyse plus détaillée des confusions il est nécessaire d'utiliser la matrice de confusion (Figure 11).

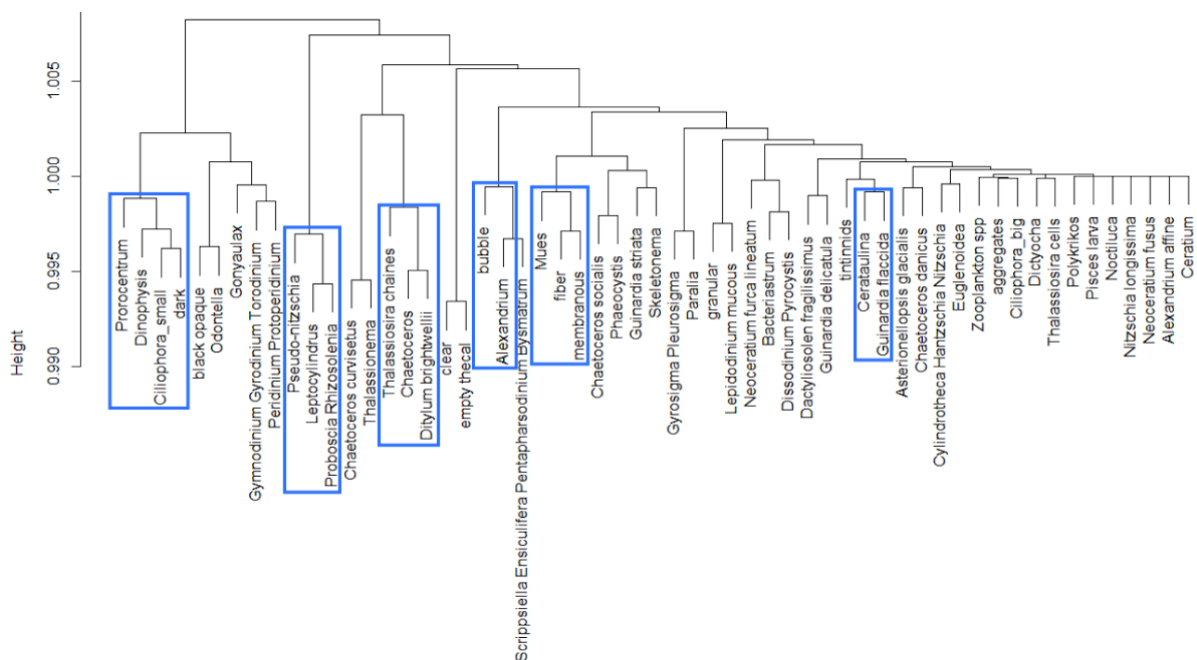


Figure 10. Dendrogramme des confusions entre les groupes du set Manche-Atlantique V0.1 (en bleu : les confusions en lien avec les groupes dont le nombre de vignettes dépasse 260).

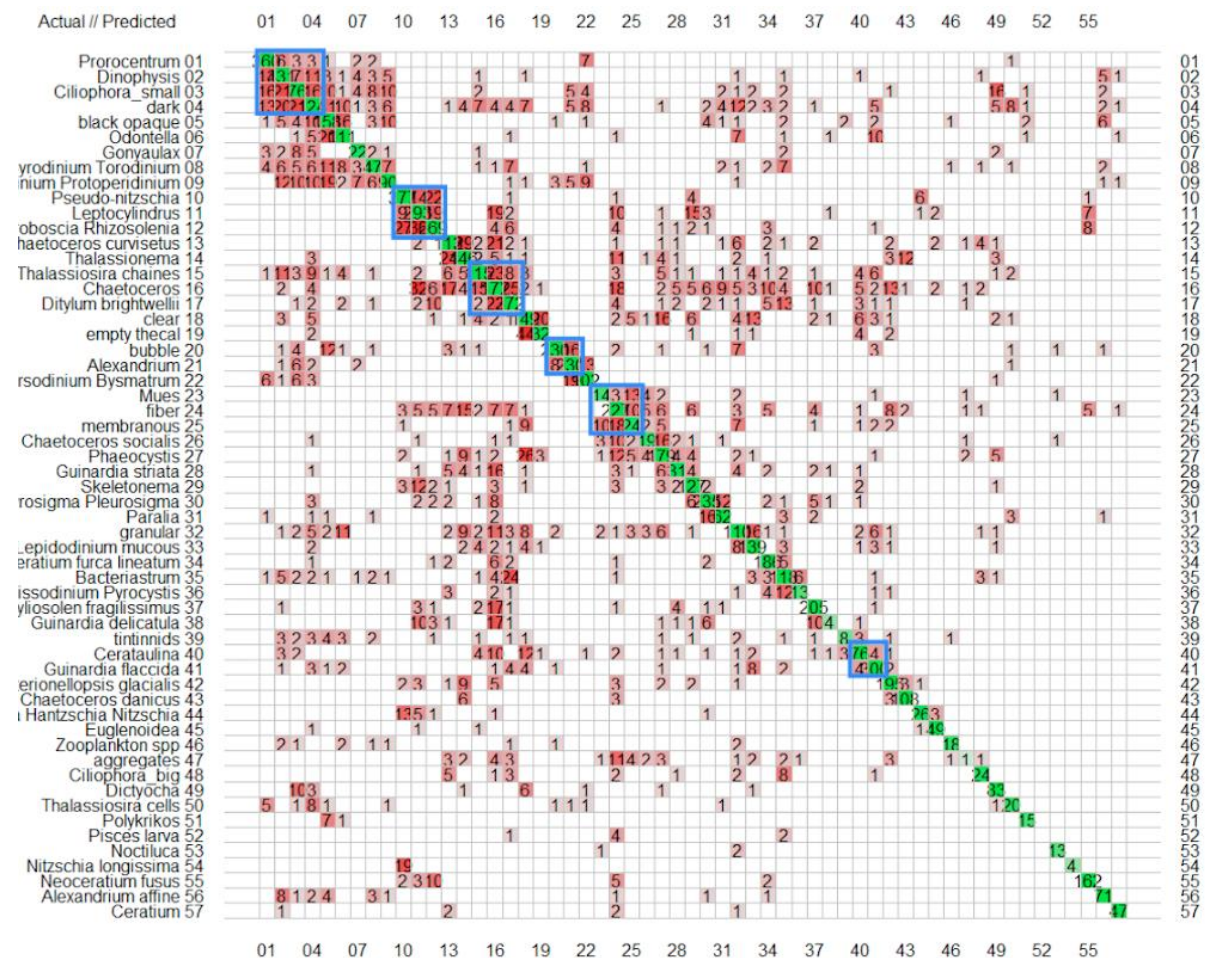


Figure 11. Matrice représentant les confusions entre les groupes du set Manche-Atlantique V0.1 (en bleu : les confusions en lien avec les groupes dont le nombre de vignettes dépasse 260).

Travail effectué :

- Pseudo-nitzschia :

Ce groupe possède de fortes confusions avec les groupes « Leptocylindrus » et « Proboscia Rhizosolenia ». Afin de diminuer ces confusions, les vignettes ressemblant à celles de ces 2 groupes vont être supprimées. Pour atteindre les 260 vignettes, les vignettes dites « doublons » sont supprimées.

Au total, 166 vignettes sont retirées du groupe « Pseudo-nitzschia ».

- Dinophysis :

Ici, il y a de fortes confusions avec les groupes « Proocentrum », « Ciliophora_small » et « dark ». Pour éliminer ces confusions, la même procédure que précédemment est effectuée.

Au total, 134 vignettes sont écartées du groupe, 36 reclassées dans un nouveau groupe appelé « Dinophysis tripos » et 5 dans « Dinophysis acuta ». Ce dernier étant placé directement dans le dossier « <LowAbundances » qui permet à ce groupe de ne pas être pris en compte par l’outil de reconnaissance mais qui est conservé en vue d’une implémentation de vignettes.

- Chaetoceros :

Ce groupe possède des confusions assez fortes avec les groupes « *Ditylum brightwellii* » et « *Thalassiosira chaines* ». 6 vignettes sont reclassées vers les groupes « *Chaetoceros curvisetus* » et « *Chaetoceros danicus* ». Les vignettes facilement confondues sont celles dont les soies ne sont pas visibles. Le choix a donc été de supprimer celles-ci.

Au total 123 vignettes ont été supprimées.

- Prorocentrum :

Ce groupe a des confusions avec les groupes « *Dinophysis* », « *Ciliophora_small* » et « dark ». La plupart des confusions ont été supprimées lors de l'optimisation du groupe « *Dinophysis* ». Le travail pour ce groupe a donc été essentiellement de retirer les vignettes en surplus.

Au total, 125 vignettes ont été supprimées.

- Leptocylindrus :

Les confusions avec les groupes « *Pseudo-nitzschia* » et « *Proboscia Rhizosolenia* » ont en partie été résolues pendant l'optimisation de « *Pseudo-nitzschia* ». Néanmoins une vérification des 3 groupes a été faite.

Au total, 122 vignettes ont été supprimées du groupe.

- *Proboscia Rhizosolenia* :

Dans un premier temps 41 vignettes sont reclassées dans le groupe « *Rhizosolenia setigera* » et 4 dans « *Proboscia indica* ». Ce dernier est mis directement dans « <LowAbundance ».

Ensuite 58 vignettes sont supprimées.

- *Ditylum brightwellii* :

Pour ce groupe, une simple suppression aléatoire de vignettes a été effectuée afin d'alléger celui-ci.

Au total, 90 sont effacées.

- Fiber :

Ce groupe possède des confusions avec « membranous » et « mues ». La sélection de vignettes se fait selon les mêmes procédures qu'avec les groupes précédents.

Au total 77 vignettes sont retirées.

- *Guinardia flaccida* :

Il a des confusions avec « *Cerataulina* ». Le groupe a subi une suppression des vignettes contaminées par d'autres éléments, des vignettes avec des cellules coupées et des vignettes ambiguës.

Au final, 20 vignettes sont supprimées puis 55 effacées aléatoirement.

- dark :

Ce groupe a subi une suppression des vignettes qui engendre une confusion avec « *Ciliophora_small* » puis une suppression des vignettes pouvant ressembler à du phytoplancton. Enfin, des vignettes sont supprimées aléatoirement afin d'atteindre 200 vignettes.

- bubble :

Ce groupe a essentiellement des confusions avec « *Alexandrium* ». On supprime donc les vignettes qui y ressemblent.

Les nouveaux effectifs sont disponibles en annexe 2.

Performances

Les évolutions des performances entre la V0.1 et la V1 sont visibles dans le tableau 2.

L'efficacité de l'outil de reconnaissance est ensuite testée en analysant 5 échantillons (inconnus du set). Il a pu être noté que l'outil de reconnaissance semblait avoir des difficultés avec 4 groupes (Thalassionema, Thalassiosira chaines, Dictyocha et Ciliophora_small). De fortes confusions entre les groupes « Alexandrium » et « Scrippsiella » ont pu également être remarqué. Cependant ce problème existe aussi en microscopie.

Tout d'abord, le groupe « Thalassiosira chaines » a subi un nettoyage (*i.e.* suppression des vignettes dites « sales » ayant trop de débris fixés sur la chaîne et/ou un fond sale). Ensuite ce groupe a été divisé en deux (« Thalassiosira big chaines » et « Thalassiosira small chaines »). Le groupe « Ciliophora_small » a également été nettoyé et divisé en deux (« Ciliophora cone » et « Ciliophora small »). Enfin, le groupe Thalassionema a simplement subi un nettoyage.

Ce travail a donc abouti à la création d'une version 2 du set Manche-Atlantique. Celle-ci est composée de 8508 vignettes réparties dans 59 groupes (les effectifs et taux d'erreurs sont disponibles en annexe 3).

Manche-Atlantique V2

Afin de connaître les performances du set pour la région de la Manche orientale, nous allons nous focaliser sur les genres/espèces majeures de cette région.

Celles-ci sont : (Lefebvre et Flesch, 2015 ; Lefebvre et Mégret, 2014 ; Guiselin, 2010 ; Breton *et al.*, 2000)

- *Chaetoceros*
- *Guinardia*
- *Pseudo-nitzschia delicatissima* et *Pseudo-nitzschia seriata*
- *Rhizosolenia imbricata*
- *Thalassionema nitzschioides*
- *Paralia*
- *Skeletonema costatum*
- *Asterionellopsis glacialis*
- *Leptocylindrus*
- *Phaeocystis globosa*

Les évolutions des performances entre la V1 et la V2 sont visibles dans le tableau 2.

Tableau 2. Evolution des performances de reconnaissance et d'estimation d'abondance entre la V1 et la V2 du set Manche-Atlantique

Nom du groupe	F-score V1 (%)	F-score V2 (%)	Remarques sur l'évolution de la reconnaissance	Remarques sur l'estimation de l'abondance
<i>Chaetoceros socialis</i>	58	63	Diminution de la reconnaissance pour ce groupe	Sous-estimation de l'abondance
<i>Phaeocystis</i>	39	32	Augmentation de la reconnaissance pour ce groupe	Estimation de l'abondance plutôt correcte
<i>Proboscia Rhizosolenia</i>	8	26	Diminution de la reconnaissance pour ce groupe	Estimation de l'abondance plutôt correcte
<i>Pseudo-nitzschia</i>	24	12	Augmentation de la reconnaissance	Estimation correcte de l'abondance
<i>Guinardia delicatula</i>	88	89	Très mauvaise reconnaissance	Forte sous-estimation
<i>Guinardia striata</i>	12	58	Forte diminution de la reconnaissance	Sous-estimation
<i>Chaetoceros</i>	65	63	Légère augmentation de la reconnaissance pour ce groupe	Légère sous-estimation
<i>Chaetoceros curvisetus</i>	43	42	Pas d'amélioration	Bonne estimation de l'abondance
<i>Thalassionema</i>	30	36	Légère diminution de la reconnaissance	Bonne estimation
<i>Paralia</i>	49	35	Amélioration de la reconnaissance	Bonne estimation
<i>Leptocylindrus</i>	23	21	Pas d'amélioration	Surestimation de l'abondance
<i>Rhizosolenia setigera</i>	24	34	Légère diminution de la reconnaissance	Sous-estimation
<i>Skeletonema</i>	17	22	Légère diminution de la reconnaissance	Bonne estimation
<i>Asterionellopsis glacialis</i>	14	15	Légère diminution de la reconnaissance	Estimation de l'abondance correcte
<i>Guinardia flaccida</i>	58	13	Forte augmentation de la reconnaissance	Estimation correcte
<i>Chaetoceros danicus</i>	11	8	Légère amélioration	Estimation correcte

Pour ce qui est des confusions, les graphiques sont disponibles en annexes 4 et 5.

Il est à noter que tous les groupes ont des confusions sauf « *Chaetoceros danicus* ». Un travail de réduction des confusions devra être fait.

Néanmoins, on peut remarquer une amélioration du TGE par rapport aux versions précédentes (TGE = 30,6 %). Cependant pour des échantillons composés de nombreuses vignettes, ce taux reste élevé.

Performances du set Manche

Ce set correspond à une version « spéciale » Manche dont seules des espèces recensées en Manche Orientale y sont représentées. Avant analyse des performances, il a fallu corriger les erreurs de nomenclature dû aux changements de noms des espèces phytoplanctoniques. Ce travail a été réalisé sur la base de la classification WORMS. Attention, dans cette première version du set, le groupe « *Skeletonema* » n'est pas présent. Il sera donc à rajouter dans une éventuelle version 2 du set.

Dans un premier temps, notons que le TGE = 20,63 %, ce qui n'est pas mauvais en soit. Néanmoins, cela représente 1 vignette incorrectement classée sur 5. Si l'on prend par exemple certains échantillons classiques du laboratoire de Boulogne, cette valeur de TGE reste tout de même élevée (*i.e.* pour un échantillon de 50 000 vignettes, 10 000 sont mal classées) et engendre une augmentation du temps d'analyse pour le technicien. D'autre part, le TGE ne donne pas d'indication sur la variabilité des taux d'erreur entre les différents groupes composant le set (les taux d'erreurs par groupe étant visibles en annexe 5). En effet, les taux d'erreurs par groupes sont très variables dans ce set. Ils vont de 4,975 % pour le groupe « *Asterionellopsis glacialis* » à 100 % pour le groupe « Larva ». Un travail sur les groupes les moins optimisés devra donc être fait.

En ce qui concerne le nombre de vignettes pour chaque groupe, il est à noter qu'aucun n'atteint le seuil de 260 vignettes déterminé précédemment pour le set Manche-Atlantique. Il sera donc possible d'ajouter des vignettes si besoin est, lors de la phase d'optimisation du set.

Concernant les confusions sur les groupes sélectionnés dans le set Manche-Atlantique, celles-ci vont tout d'abord être identifiées à l'aide du dendrogramme « groups clustering » (Figure 12) et leur intensité sera mesurée à l'aide de la matrice de confusion (Figure 13).

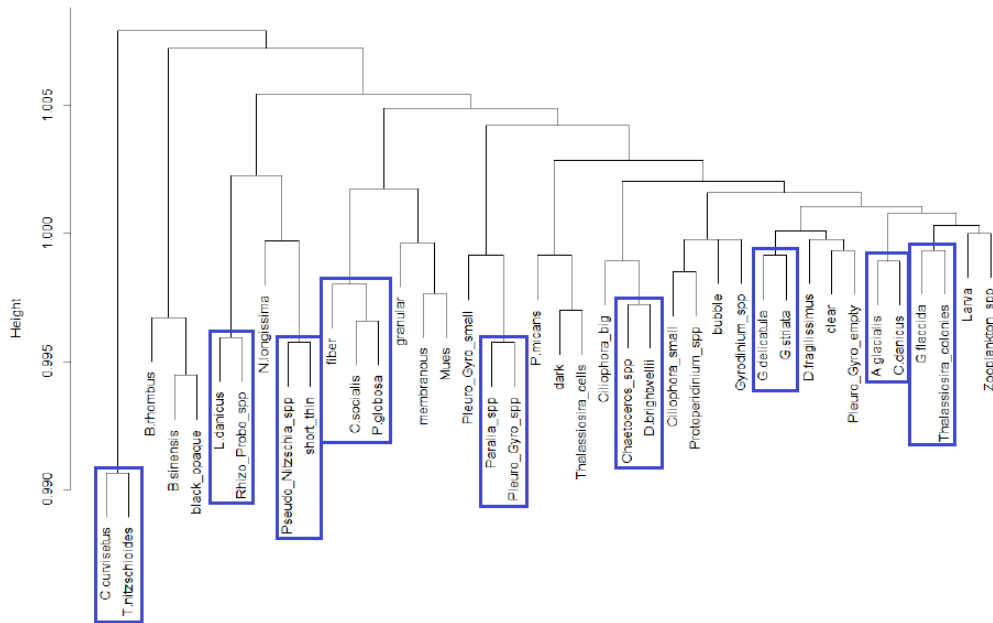


Figure 12. Dendrogramme représentant les confusions d'intérêt dans le set Manche V1 (en bleu : les confusions en lien avec les groupes d'intérêt).

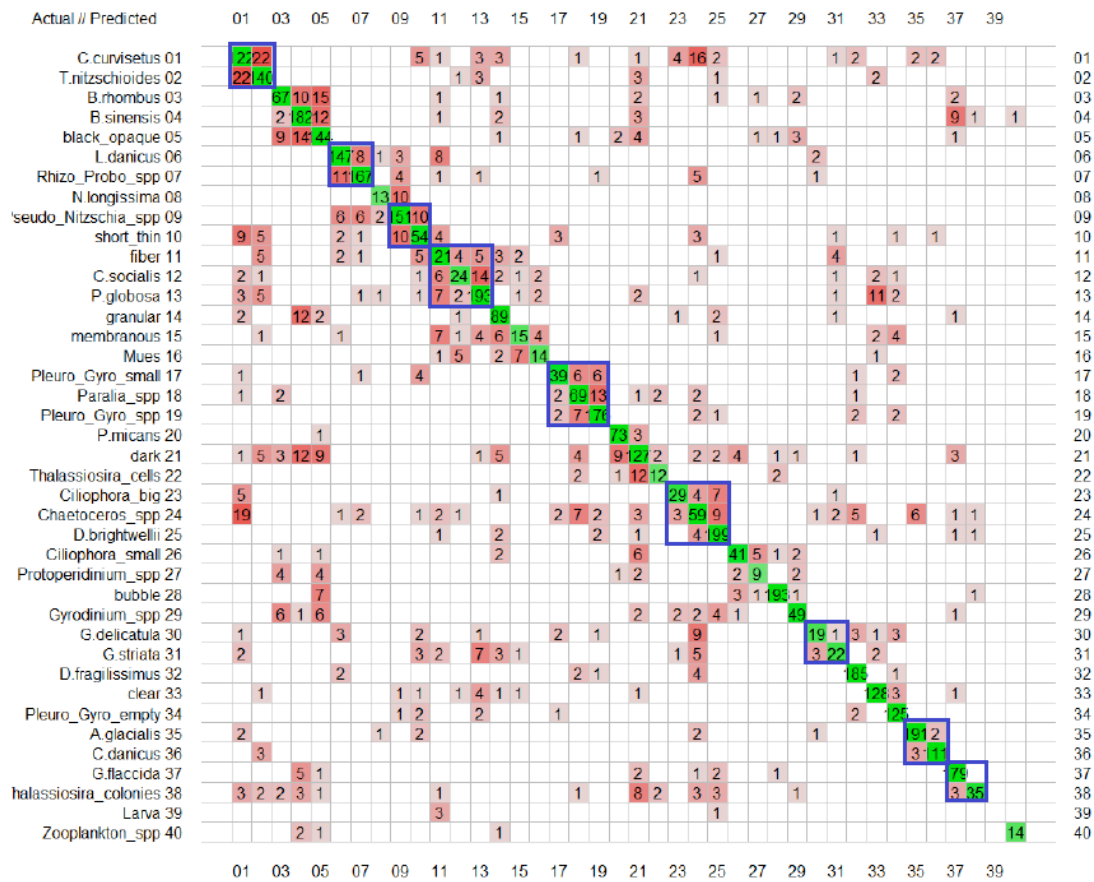


Figure 13. Matrice permettant de visualiser l'intensité des confusions choisies (en bleu : les confusions en rapport avec les groupes d'intérêt).

Les confusions relevées sont les suivantes :

- « C. curvisetus » et « T. nitzschioides »
Les confusions entre ces deux groupes sont assez fortes.
- « L. danicus » et « Rhizo_Probo_spp »
Ici, les confusions sont moyennement importantes.
- « Pseudo_Nitzschia_spp » et « short_thin »
Pour ces groupes les confusions restent également de moyenne intensité.
- « fiber », « C. socialis » et « P. globosa »
Les confusions entre les deux premiers groupes restent assez faibles. Néanmoins celles entre « C. socialis » et « P. globosa » sont plus fortes. En effet, la deuxième partie des confusions peut être due à la forme des colonies et des cellules qui en font parties (celles-ci peuvent comporter des ressemblances).
- « Pleuro_Gyro_small », « Paralia_spp » et « Pleuro_Gyro_spp »
Les conclusions des intensités de confusions pour ces 3 groupes sont les mêmes que précédemment.
- « Ciliophora_big », « Chaetoceros_spp » et « D. brightwellii »
Ici, c'est encore une fois le même cas.
- « G. delicatula » et « G. striata »
Les confusions entre ces deux groupes sont assez faibles.
- « A. glacialis » et « C. danicus »
Pour ces groupes, c'est le même cas que précédemment.
- « G. flaccida » et « Thalassiosira_colonies »
Ici, les confusions sont très faibles par rapport aux autres groupes.

Afin d'avoir une vue d'ensemble sur les performances du set Manche V1, il est nécessaire d'étudier les taux de F-score ainsi que les sur/sous-estimations de l'abondance des groupes (Figure 14). Les mêmes seuils que pour le training set Manche-Atlantique seront utilisés.

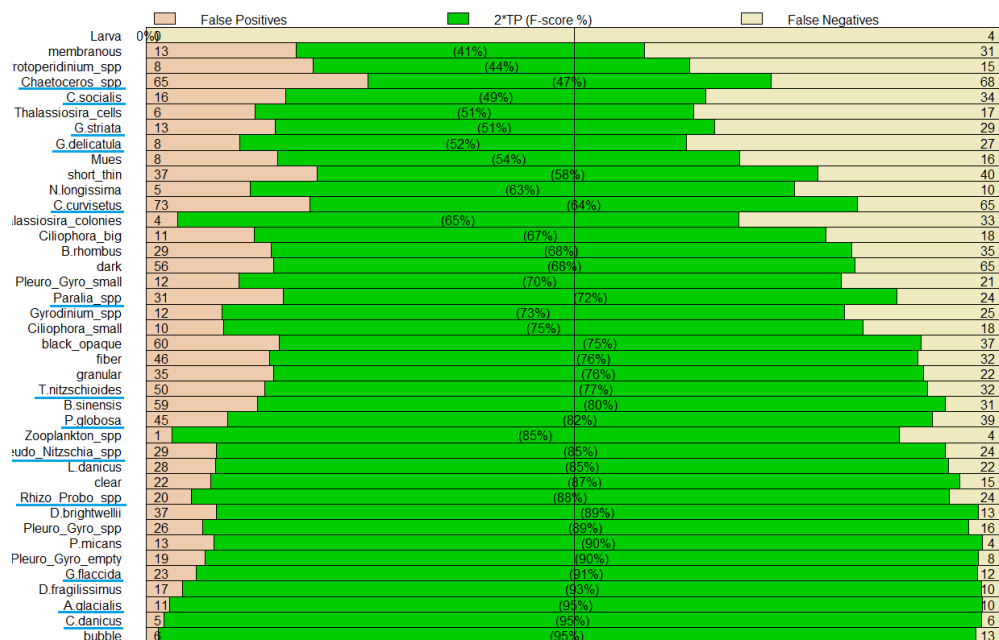


Figure 14. Représentation graphique des taux de F-score pour chaque groupe du set Manche version 1 (les groupes soulignés en bleu sont ceux considérés comme les plus représentatifs de la Manche orientale)

Sur les 13 groupes sélectionnés, 8 ont un F-score > 70 %. Ils ont donc un taux de reconnaissance plutôt bon. En revanche, 3 groupes ont un F-score moyen (entre 50 et 70 %) et 2 en ont un mauvais (< 50 %). Quant à l'estimation de l'abondance, celle-ci est correcte pour 8 groupes sur les 13. De plus, 3 groupes sont en sous-estimation et 2 en surestimation. Les résultats détaillés sont disponibles dans le tableau 3.

Tableau 3. Informations sur les taux de F-score et les estimations d'abondances pour les groupes sélectionnés, caractéristiques de la Manche orientale.

Nom du groupe	Valeur F-score (%)	Remarques sur le F-score (bon, moyen, mauvais)	Estimation de l'abondance
<i>C. danicus</i>	95	bon	estimation correcte
<i>A. glacialis</i>	95	bon	estimation correcte
<i>G. flaccida</i>	91	bon	estimation correcte
<i>Rhizo_Probo_spp</i>	88	bon	estimation correcte
<i>Pseudo_Nitzschia_spp</i>	85	bon	estimation correcte
<i>P. globosa</i>	82	bon	estimation correcte
<i>T. nitzschioides</i>	77	bon	légère surestimation
<i>Paralia_spp</i>	72	bon	légère surestimation
<i>C. curvisetus</i>	64	moyen	estimation correcte
<i>G. delicatula</i>	52	moyen	sous-estimation prononcée
<i>G. striata</i>	51	moyen	sous-estimation prononcée
<i>C. socialis</i>	49	mauvais	sous-estimation prononcée
<i>Chaetoceros_spp</i>	47	mauvais	estimation correcte

De même que pour le set Manche-Atlantique, on remarque que les sous-estimations semblent être plus prononcées que les surestimations. En revanche, contrairement au set Manche-Atlantique, les catégories « débris » et « zooplancton » ont des valeurs beaucoup plus variables. En effet, dans le précédent set, les groupes de ces deux catégories étaient classés dans les moins performants. Or dans le set Manche certains groupes sont très performants (« bubble » ou « clear » par exemple) alors que d'autres ne le sont pas du tout (comme « larva » ou « mues »).

Module de dénombrement des colonies

L'outil de reconnaissance ne faisant pas la différence entre 1 cellule et 1 colonie, un outil d'estimation de l'abondance de cellules par colonies, est en développement. C'est le set Manche-Atlantique V2 qui est utilisé. Au total, 22 groupes formant potentiellement des colonies sont choisis. Ce qui représente 3787 vignettes. Les groupes et les nombres de vignettes retenus sont visibles dans le Tableau 4.

La procédure de comptage est simple. Pour chaque vignette, il suffit de cliquer sur les cellules visibles. Dès lors, un point rouge apparaît permettant d'éviter de compter deux fois la même cellule par erreur (Figure 15).

Pour distinguer une cellule d'une autre, les méthodes étaient relativement les mêmes entre les groupes :

- Utilisation de la paroi, ainsi que du léger renforcement entre deux cellules qui est dû à leur forme.
- Il est aussi possible d'utiliser la « fenêtre » visible entre deux cellules (*i.e.* léger vide au milieu de la jonction entre les cellules).
- Utilisation du contenu cellulaire qui en général est visible sous la forme d'une tache noire.
- On peut utiliser également les soies qui permettent à certains genres comme *Chaetoceros* de se maintenir en chaîne (celles-ci délimitent la séparation entre les cellules).

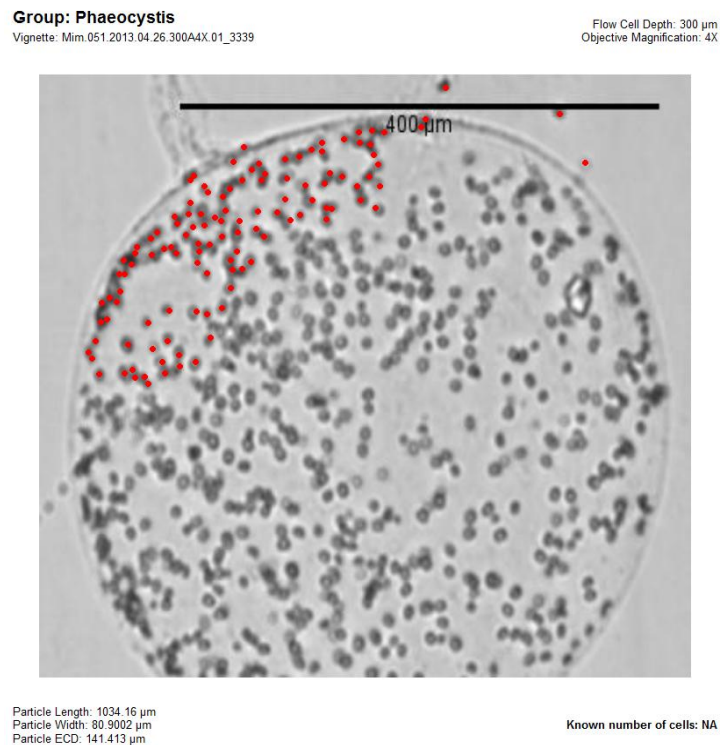


Figure 15. Colonie de *Phaeocystis* en court de comptage (en rouge : les cellules déjà comptées).

Néanmoins, des difficultés sont apparues au court du comptage, obligeant les analystes à se concerter pour certains types de vignettes. En effet, certaines cellules étaient en court de division lors de la numérisation de l'échantillon (Figure 16). Il a donc été décidé qu'une cellule en division représentait 2 individus.



Figure 16. Exemple de cellule en court de division (*Ditylum brightwellii*).

De plus, il arrive qu'il soit impossible de visualiser le nombre de cellules d'une colonie, la faisant accessoirement passer pour un seul individu (Figure 17).



Figure 17. Colonie de *Paralia spp* dont il est impossible de distinguer le nombre de cellules.

Un autre problème apparaît pour les grosses colonies comme *Phaeocystis* par exemple. Il arrive que celles-ci se retrouvent sur deux vignettes, engendrant par conséquent des erreurs de comptage du fait que certaines parties de la colonie sont visibles sur les deux vignettes. D'autre part, il est possible que les soies ne soient pas visibles. Dans ce cas, il est indispensable d'utiliser une autre technique pour discerner les cellules. L'orientation de la colonie a aussi un rôle important. Par exemple si une colonie en chaîne est prise en photo par le dessus, l'analyste est dans l'incapacité de connaître le nombre de cellule (Figure 18).

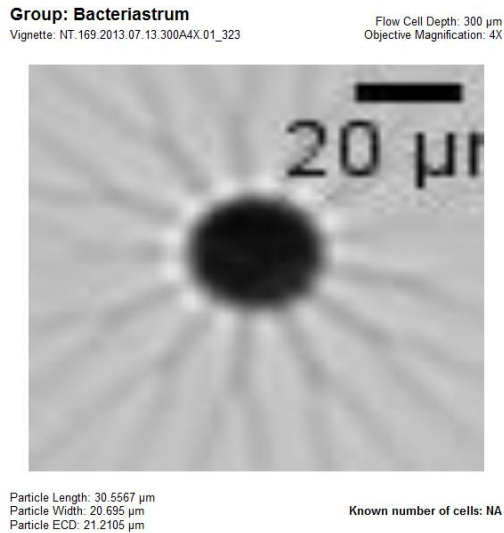


Figure 18. Colonie de *Bacteriastrum* dont la capture de l'image s'est faite du dessus. Il est impossible de connaître le nombre de cellules.

Enfin, un problème récurrent est survenu avec les colonies ne formant pas de chaînes plates. Celles-ci étant en 3 dimensions sur les numérisations, il est impossible de distinguer les cellules qui se chevauchent. De plus, la mise au point de l'objectif de la caméra sur les cellules du fond de la cuve est mauvaise, les rendant floues et peu discernables (Figure 19).

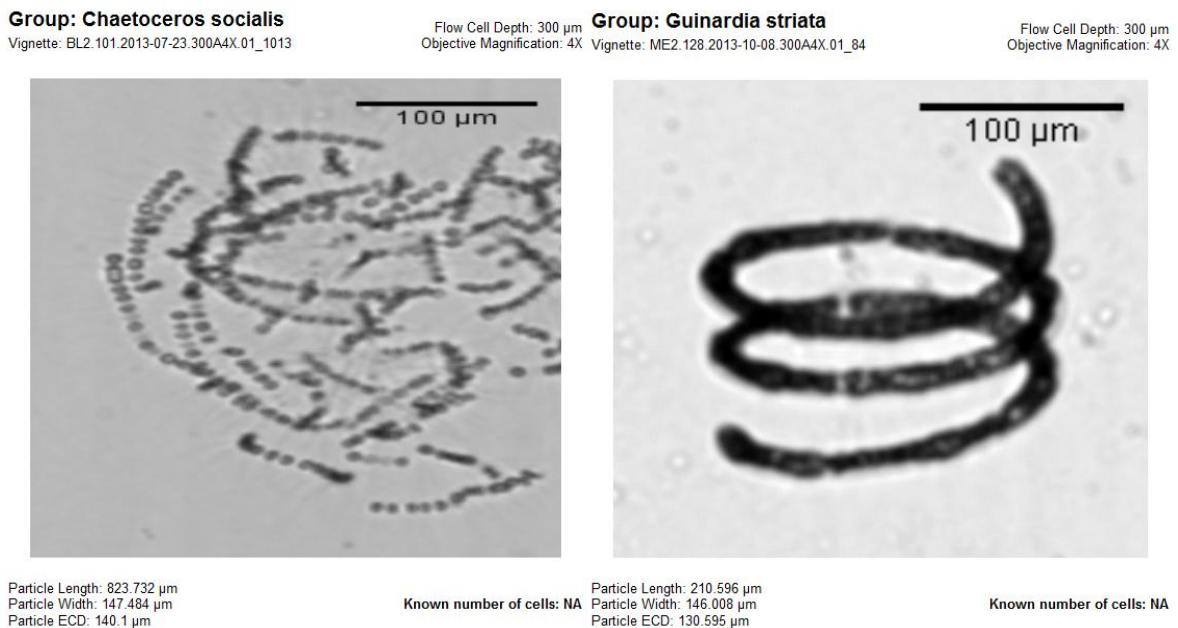


Figure 19. Exemples de colonies (*Chaetoceros socialis* et *Guinardia striata*) dont toutes les cellules ne sont pas discernables.

Comparaison des dénombrements sur 2 sites

Une comparaison sur les comptages de deux laboratoires différents (Nantes et Boulogne-sur-mer) a été réalisée dans le but de choisir la méthode qui sera utilisée pour développer le modèle prédictif et pour déceler d'éventuelles variabilités de comptages. En effet, les comptages n'ayant pas été faits par les mêmes personnes afin de considérer le biais inter-analyste dans une situation « extrême » (taxonomiste expérimenté pour Nantes et stagiaire de M2 non expérimenté pour Boulogne-sur-mer), de variations dans les comptages ont pu être observées. Au total, sur 18 groupes comparables 8 comportent des variations non négligeables.

Ensuite, certains groupes présents dans le module de dénombrement n'ont pas été pris en compte pour la comparaison des performances des différentes méthodes utilisables pour le modèle prédictif. En effet, la qualité des vignettes de certains groupes ne permettait qu'un comptage partiel des cellules ou même une impossibilité de comptage. Ces groupes sont visibles dans le Tableau 4.

Tableau 4. Table représentant l'ensemble des groupes taxonomiques présent dans le module de dénombrement, ainsi que le nombre total de vignettes et le nombre de vignettes comptées pour chaque site (encadré en bleu : les groupes retirés de la comparaison).

Group	Total vignettes	BOULOGNE	NANTES
<i>Odontella</i>	160	160 (100%)	116 (72.50%)
<i>Guinardia flaccida</i>	260	260 (100%)	101 (38.85%)
<i>Dinophysis tripos</i>	36	36 (100%)	36 (100%)
<i>Proboscia Rhizosolenia</i>	265	265 (100%)	...
<i>Dactyliosolen fragilissimus</i>	238	238 (100%)	146 (61.34%)
<i>Guinardia striata</i>	84	76 (90.48%)	58 (69.05%)
<i>Thalassionema</i>	169	169 (100%)	83 (49.11%)
<i>Leptocylindrus</i>	256	256 (100%)	118 (46.10%)
<i>Pseudo-nitzschia</i>	260	260 (100%)	177 (68.08%)
<i>Phaeocystis</i>	264
<i>Chaetoceros socialis</i>	60	...	1 (1.67%)
<i>Chaetoceros</i>	260	135 (51.92%)	62 (23.85%)
<i>Asterionellopsis glacialis</i>	227	227 (100%)	60 (26.43%)
<i>Ditylum brightwellii</i>	261	261 (100%)	157 (60.15%)
<i>Paralia</i>	93	8 (8.60%)	...
<i>Thalassiosira big chaines</i>	99	99 (100%)	92 (92.93%)
<i>Guinardia delicatula</i>	56	56 (100%)	52 (92.86%)
<i>Chaetoceros curvisetus</i>	200	195 (97.50%)	77 (38.50%)
<i>Alexandrium affine</i>	93	93 (100%)	93 (100%)
<i>Bacteriastrum</i>	180	164 (91.11%)	61 (33.89%)
<i>Skeletonema</i>	162	162 (100%)	84 (51.85%)
<i>Thalassiosira small chaines</i>	104	104 (100%)	94 (90.38%)

Les performances des méthodes prédictives sont comparées par validation croisée sur des modèles de régressions linéaires et non linéaires. A l'aide de la validation croisée, il est possible d'évaluer les taux de reconnaissance des différentes méthodes prédictives (Govaerts, 2010). On s'intéressera donc au taux de reconnaissance global du nombre de cellules par colonie logarithmique (%TL) et à l'estimation totale du nombre de cellules (%EstTot).

Ici, 3 méthodes prédictives sont comparées :

- LM : Linear Model (Modèle Linéaire Multivarié)
- LDA : Linear Discriminant Analysis (Analyse Discriminante Linéaire)
- MDA : Mixture Discriminant Analysis (extension de la LDA)

Les résultats pour le dénombrement effectué à Boulogne sont visibles dans le Tableau 5 et ceux de Nantes dans le Tableau 6.

Tableau 5. Pourcentages d'estimation du nombre de cellules par colonie pour l'ensemble des groupes étudiés à Boulogne (en bleu : meilleure méthode par rapport aux 3 utilisées pour le groupe donné).

Group	LM	LDA	MDA
	%TL - %EstTot	%TL - %EstTot	%TL - %EstTot
<i>Odontella</i>	52.69 – 138.43	75.87 – 101.12	84.94 – 100.77
<i>Guinardia flaccida</i>	84.15 – 98.09	89.04 – 98.33	91.23 – 98.99
<i>Dinophysis tripos</i>	60.28 – 153.93	86.67 – 105.65	100.00 – 100.00
<i>Proboscia Rhizosolenia</i>	86.75 – 97.55	88.53 – 95.99	89.89 – 97.12
<i>Dactyliosolen fragilissimus</i>	68.24 – 97.63	77.35 – 101.23	81.47 – 97.40
<i>Guinardia striata</i>	37.76 – 135.75	42.50 – 105.63	74.61 – 96.26
<i>Thalassionema</i>	41.60 – 95.65	58.22 – 97.49	79.70 – 98.47
<i>Leptocylindrus</i>	32.97 – 95.91	35.08 – 98.98	57.50 – 99.35
<i>Pseudo-nitzschia</i>	56.15 – 97.86	56.81 – 98.54	68.54 – 96.81
<i>Phaeocystis</i>	... - - - ...
<i>Chaetoceros socialis</i>	... - - - ...
<i>Chaetoceros</i>	29.41 – 97.17	30.74 – 99.88	57.56 – 100.84
<i>Asterionellopsis glacialis</i>	36.04 – 98.36	45.24 – 100.57	71.54 – 99.60
<i>Ditylum brightwellii</i>	88.70 – 98.39	93.72 – 98.49	95.82 – 98.24
<i>Paralia</i>	... - - - ...
<i>Thalassiosira big chaines</i>	62.83 – 99.66	65.66 – 97.96	84.34 – 99.58
<i>Guinardia delicatula</i>	37.32 – 95.26	48.57 – 98.41	70.00 – 95.54
<i>Chaetoceros curvisetus</i>	41.33 – 97.47	38.87 – 98.62	64.36 – 99.99
<i>Alexandrium affine</i>	82.04 – 99.34	87.31 – 99.59	97.85 – 100.34
<i>Bacteriastrum</i>	43.90 – 98.93	40.91 – 101.37	73.78 – 99.61
<i>Skeletonema</i>	39.44 – 97.86	34.81 – 102.05	54.63 – 100.43
<i>Thalassiosira small chaines</i>	46.73 – 97.33	49.90 – 100.88	69.52 – 100.12

Tableau 6. Pourcentages d'estimation du nombre de cellules par colonie pour l'ensemble des groupes étudiés à Nantes (en bleu : meilleure méthode par rapport aux 3 utilisées pour le groupe donné).

Group	LM	LDA	MDA
	%TL - %EstTot	%TL - %EstTot	%TL - %EstTot
<i>Odontella</i>	52.69 – 138.43	75.87 – 101.12	84.94 – 100.77
<i>Guinardia flaccida</i>	84.15 – 98.09	89.04 – 98.33	91.23 – 98.99
<i>Dinophysis tripos</i>	60.28 – 153.93	86.67 – 105.65	100.00 – 100.00
<i>Proboscia Rhizosolenia</i>	86.75 – 97.55	88.53 – 95.99	89.89 – 97.12
<i>Dactyliosolen fragilissimus</i>	68.24 – 97.63	77.35 – 101.23	81.47 – 97.40
<i>Guinardia striata</i>	37.76 – 135.75	42.50 – 105.63	74.61 – 96.26
<i>Thalassionema</i>	41.60 – 95.65	58.22 – 97.49	79.70 – 98.47
<i>Leptocylindrus</i>	32.97 – 95.91	35.08 – 98.98	57.50 – 99.35
<i>Pseudo-nitzschia</i>	56.15 – 97.86	56.81 – 98.54	68.54 – 96.81
<i>Phaeocystis</i>	... - - - ...
<i>Chaetoceros socialis</i>	... - - - ...
<i>Chaetoceros</i>	29.41 – 97.17	30.74 – 99.88	57.56 – 100.84
<i>Asterionellopsis glacialis</i>	36.04 – 98.36	45.24 – 100.57	71.54 – 99.60
<i>Ditylum brightwellii</i>	88.70 – 98.39	93.72 – 98.49	95.82 – 98.24
<i>Paralia</i>	... - - - ...
<i>Thalassiosira big chaines</i>	62.83 – 99.66	65.66 – 97.96	84.34 – 99.58
<i>Guinardia delicatula</i>	37.32 – 95.26	48.57 – 98.41	70.00 – 95.54
<i>Chaetoceros curvisetus</i>	41.33 – 97.47	38.87 – 98.62	64.36 – 99.99
<i>Alexandrium affine</i>	82.04 – 99.34	87.31 – 99.59	97.85 – 100.34
<i>Bacteriastrium</i>	43.90 – 98.93	40.91 – 101.37	73.78 – 99.61
<i>Skeletonema</i>	39.44 – 97.86	34.81 – 102.05	54.63 – 100.43
<i>Thalassiosira small chaines</i>	46.73 – 97.33	49.90 – 100.88	69.52 – 100.12

Pour les deux sites, la méthode MDA semble fournir en moyenne les meilleurs scores de prédiction. C'est donc cette méthode qui est retenue pour la construction des modèles prédictifs.

Discussion

L'objectif de ce stage était de contribuer à l'optimisation de l'utilisation du système couplé FlowCAM/ZooPhytoImage dans un but d'intégration au REPHY pour la fin de l'année 2015. Néanmoins même si le FlowCAM est dédié au REPHY, il pourrait être utilisé pour d'autres campagnes comme IBTS ou Camanoc. De plus, il s'incère parfaitement dans le cadre de la DCSMM en apportant de nouvelles données pour d'éventuelles prises de décisions.

L'utilité de ce système étant la simplification du travail effectué actuellement par les taxonomistes, il est important de connaître l'efficacité de celui-ci par rapport à la microscopie.

En 2013, Alvarez et al, ont montré des différences de richesses taxonomiques et d'abondances entre l'utilisation du microscope et l'utilisation du FlowCAM/ZooPhytoImage. Pour cela, ils ont étudié une série d'échantillons provenant de la « cantabrian sea » en Espagne et couvrant la période de novembre 2008 à mars 2010. Les résultats de cette étude montrent une surestimation de l'abondance à l'aide du FlowCAM, ainsi qu'une sous-estimation de la richesse taxonomique. Cela peut être dû au fait qu'il existe une perte d'informations lors de la numérisation (soies non visibles, mauvaise orientation de la particule, grossissement trop faible, etc.), empêchant la discrimination de certaines espèces au sein d'un genre par l'analyste lors de la phase de correction d'erreur.

Néanmoins, dans cette étude, il a également été prouvé que même si les résultats différaient, les deux méthodes suivaient les mêmes patrons de variations. De plus, il est aussi nécessaire de prendre en compte les incertitudes des mesures phytoplanctoniques par le biais de la méthode Utermöhl (Soudan et Miossec, 2013). 4 sources de variabilité ont pu être découvertes :

- La préservation des échantillons : en effet, l'utilisation de lugol dénature certaines particules et entraîne un biais avant même de commencer l'analyse. Il est à noter que cette technique est aussi utilisée pour le FlowCAM.
- Le sous-échantillonnage : si les sous-échantillons sont faits à partir d'un échantillon non homogène alors les résultats obtenus ne seront pas représentatifs de celui-ci.
- L'identification taxonomique : c'est la plus grande source d'erreur du fait de la difficulté de l'identification de certains taxons, ainsi que du fait d'avoir plusieurs taxonomistes différents (expérience, niveau de fatigue et de concentration).
- Le comptage : ici, les erreurs sont la conséquence de l'effet de bord lors de la sédimentation dans la cuve.

L'un des atouts de l'utilisation couplée du FlowCAM et du logiciel ZooPhytoImage est sa polyvalence. En effet, il permet de faire des études à long et court termes, ainsi qu'à petites et grandes échelles. De plus, il est possible de numériser et analyser des échantillons vivants et/ou lugolés à l'aide du même set d'apprentissage. La fixation des échantillons au lugol est un avantage considérable car cela permet d'une part de faire de l'échantillonnage à hautes fréquences puisqu'il n'ait pas nécessaire de numériser les échantillons dans la foulée (contrairement aux échantillons non lugolés qui doivent être numérisés sous 24h avant la mort d'un trop grand nombre d'individus). D'autre part, il est possible de faire des échantillonnages même si l'appareil n'est pas à disposition (ceci étant utile lors de campagnes en mer, comme la campagne IBTS par exemple). Néanmoins, l'utilisation du lugol n'est pas sans conséquences. En effet, son utilisation entraîne une dénaturation des particules (éclatement, changement de couleurs, déformations), ceci pouvant donner suite à des différences de résultats entre échantillons vivants et lugolés. Il est donc préférable de n'utiliser que des échantillons vivants ou lugolés afin d'homogénéiser les résultats. Il faut également prendre en compte que l'utilisation du FlowCAM a pour but de faire des analyses quasiment en temps réel. Le but n'étant donc pas d'accumuler des échantillons et de les passer longtemps après leur fixation. En effet, comme son utilisation s'insère dans le programme du REPHY, il est essentiel d'analyser les échantillons rapidement pour ce réseau.

Au vu des différences d'échantillons entre les centres Manche-Mer du Nord et les centres Atlantique, il semble actuellement plus judicieux d'utiliser des training set spécifiques à chaque laboratoire en attendant l'intégration du module d'apprentissage actif. En effet, à Boulogne, un réel travail d'optimisation des groupes débris est nécessaire alors que celui-ci ne l'a pas à Nantes. Il est vrai qu'hormis le groupe « bubble », les performances des groupes débris sont comprises entre 7 et 67%. Ceci nous indiquant un nombre d'erreurs considérables puisque les débris composent la majorité de nos échantillons. De plus, les débris offrent des intervalles de données très larges pour les paramètres utiles à la reconnaissance. Il faudra donc essayer de scinder certains groupes afin de mieux caractériser ces particules. Cela permettra le cas échéant de réduire considérablement le temps de travail pour l'analyse d'un échantillon à Boulogne. Le but étant ici d'éviter les confusions entre les débris et les particules phytoplanctoniques. Par exemple, on pourrait diviser le groupe « dark » en plusieurs sous-groupes dont les vignettes auraient des caractéristiques communes. Les confusions entre les différents groupes débris ne sont pas spécialement importantes. En effet, le but

est que les débris soient reconnus comme tel. Si une erreur de classement intervient au sein des groupes débris cela n'aura aucunes conséquences sur les résultats attendus pour l'analyse des groupes phytoplanctoniques. Ce travail devrait permettre de diminuer le nombre de vignettes de débris à reclasser pendant la correction d'erreur (la correction de la classification des vignettes débris étant la majeure partie de la correction d'erreurs). Ensuite, une refonte des groupes utilisés sera nécessaire. En effet, dans la version 1 du set Manche, le genre *Skeletonema* n'est pas intégré et il faudra obligatoirement qu'il apparaisse dans la deuxième version puisque ce groupe fait partie des espèces d'intérêts dans notre projet. Suite à cela, il sera nécessaire de tester les performances du set pour déceler cette fois ci les confusions entre les groupes d'intérêts dans le but de diminuer les erreurs de classification. Tout ce travail devrait grandement améliorer les performances de l'outil de reconnaissance. Mais si celles-ci ne sont pas encore acceptables nous pourrions implémenter les groupes avec de nouvelles vignettes puisque pour tous, le seuil maximal de 260 vignettes n'est pas atteint (ceci nous laissant une marge de travail importante).

Le fait de posséder des groupes dans le set, qui ne sont pas représentés dans les zones d'échantillonnages pourrait possiblement entrainer des erreurs d'identifications et donc rallonger le temps de correction. Lorsque le module d'apprentissage actif sera disponible, la question d'un set spécialisé à la zone d'étude ne sera plus d'actualité. Ce module pourra « apprendre de ses erreurs » en ajoutant au set les vignettes et leurs mesures, qui n'ont pas été correctement identifiées. Cet apprentissage sera à l'origine d'une régionalisation du set au fur et à mesure des passages d'échantillons.

Les travaux d'optimisation des différents sets d'apprentissage ainsi que la contribution au développement de la version 5 du logiciel ZooPhytoImage ont permis de mettre en exergue certains atouts et inconvénients de l'utilisation du système couplé FlowCAM/ZooPhytoImage. En effet, la numérisation rapide des échantillons permet de diminuer le temps de travail du technicien en charge de l'identification taxonomique des échantillons (Poulton et Martin, 2010). Pour le laboratoire de Boulogne-sur-Mer, la numérisation d'un échantillon de 10 mL dure 12 minutes. Le temps de numérisation est donc très rapide et le technicien peut s'adonner à d'autres tâches pendant celle-ci puisqu'elle est entièrement automatique. Néanmoins cette durée ne prend pas en compte le délai d'analyse des échantillons numérisés (plus précisément, le temps nécessaire à la correction des erreurs faites par l'outil de reconnaissance automatisé). De plus, c'est aussi la partie de l'analyse qui est la plus chronophage. En se concertant avec le laboratoire de Nantes, une variabilité du temps de correction a été mise en avant. Cette variabilité est essentiellement due à la composition des échantillons à traiter mais aussi du technicien responsable de la correction. Pour l'Ifremer Nantes, les échantillons analysés comportent beaucoup moins de vignettes qu'à Boulogne (un maximum de 5000 vignettes pour Nantes contre 160 000 pour Boulogne). Ceci provoque de grandes différences dans le temps d'analyse entre les centres puisque les techniciens du laboratoire de Nantes passent en général 30 minutes à 2h sur un échantillon alors que les techniciens de Boulogne y passent au minimum 2 à 4 heures. Malgré ces inconvénients, l'utilisation de ce système reste tout de même avantageuse puisqu'elle permet de quantifier les erreurs et diminue la fatigue du technicien (une lecture au microscope étant fatigante au vu du temps nécessaire pour une analyse). La composition taxonomique des échantillons a également un rôle important. Il a pu être remarqué précédemment que les échantillons de Boulogne étaient très chargés en débris par rapport à ceux de Nantes. Or, lors d'essais d'analyses, une remarque a été faite sur le fait que la majorité des particules à reclasser était des débris. La question de l'ajout d'un module de fluorescence a donc été posée. Ce module

permettrait d'écarter les débris dès la numérisation mais lors d'une acquisition, plusieurs particules se situent dans le champ de vision de la caméra. Si l'une de ces particules est phytoplanctonique alors toutes les particules (phytoplanctoniques ou non) dans la cuve au moment de l'acquisition seront prises en photo. Néanmoins cette technique pourrait tout de même réduire considérablement le nombre de particules non phytoplanctoniques dans un échantillon. Avant toutes décisions sur l'ajout d'un module de fluorescence, la valeur ajoutée de celui-ci, en fonction de son prix devra être testée dans le but de savoir si ce module serait rentable. Cependant d'autres facteurs entrent en compte, comme la vitesse de correction du technicien (essentiellement caractérisée par l'expérience du technicien) et la fastidiosité de l'exercice. Il est donc préférable que le technicien s'occupant des analyses, ait une formation de taxonomiste afin d'augmenter sa rapidité d'analyse.

Enfin, le problème de la perte d'informations pourrait être résolu à l'aide d'un nouvel objectif (grossissement x10) qui permettrait une acquisition plus détaillée lors de la numérisation. Une autre solution est l'utilisation du FastCAM (actuellement en développement à l'Ifremer de Brest). Cet appareil a un fonctionnement similaire à celui du FlowCAM, mais le FastCAM devrait avoir de meilleures performances d'estimation en général et devrait être aussi plus rapide.

Même si l'étude d'*Alvarez et al* prouve une surestimation de l'abondance avec le FlowCAM, un problème de sous-estimation est à régler. C'est celui de l'estimation du nombre de particules par colonies. En effet, le logiciel prend en compte la colonie comme une seule particule ce qui fait chuter l'estimation de l'abondance pour ces groupes. C'est pourquoi un module permettant l'estimation du nombre de cellules par colonies (en prenant également en compte la taille de la colonie) est nécessaire. Pourtant ce dernier ne règle pas tout. Certains groupes ne sont pas dénombrables (colonies en « 3 dimensions », redondance sur plusieurs vignettes, cellules non visibles, etc.). Cette difficulté n'est donc pas entièrement réglée et il est nécessaire de trouver une solution viable pour les groupes restant car certains sont d'intérêt majeur en Manche orientale (comme *Phaeocystis globosa* par exemple).

L'estimation des biovolumes et des contenus en carbone se fait à l'aide de formules de conversion. Le biovolume est calculé sur la base de mesures faites sur la vignette. Or celle-ci est en 2 dimensions et il est donc impossible de connaître l'épaisseur de la particule. Une méthode de conversion en 3D de la vignette a été développée et pourrait être une solution pour une meilleure estimation des biovolumes. En effet cette méthode semble fournir des résultats plus précis qu'une simple estimation comme le fait actuellement ZooPhytoImage (Moberg et Sosik, 2012). Cependant l'incorporation de cette méthode dans le logiciel serait difficile et demanderait un travail assez conséquent. En effet, la méthode utilisée actuellement permet une insertion simple dans le logiciel tout en gardant une estimation plutôt correcte. Il serait donc judicieux de n'intégrer ce module que si le besoin s'en fait ressentir. En ce qui concerne les contenus en carbone, la formule utilise la relation entre celui-ci et le biovolume. Or, il a été démontré que les facteurs de conversions utilisés dans cette formule sont fortement variables d'une espèce à l'autre (Menden-Deuer et Lessard, 2000). Ceci montre la nécessité d'avoir une multitude de facteurs de conversions. L'optimisation de cette estimation semble donc être compliquée puisque les facteurs de conversion pour les espèces en Manche orientale ne sont pas connus.

Avant de rendre le logiciel disponible au grand public, il sera nécessaire de corriger les bugs et les difficultés rencontrés puisque celui-ci devra être utilisable par n'importe qui simplement à l'aide de la notice, qu'il soit formé ou non à son utilisation. En effet, un certain nombre de

commandes dans le menu ne fonctionnent pas et l'utilisateur est obligé de taper des lignes de commandes dans la console R. De plus, la commande permettant de créer des fichiers zidb n'existe pas. Pourtant ces fichiers sont essentiels. Ils permettent la création de vignettes destinées à être intégrées au set et permettent également l'analyse d'échantillons inconnus. A l'heure actuelle, la seule façon de créer des zidb est d'importer les échantillons à l'aide de l'interface graphique. Or pour ouvrir cette interface il est nécessaire d'avoir un set fonctionnel. Cela engendre un problème considérable pour les nouveaux utilisateurs du logiciel. Ces derniers n'ayant pas de set, ils ne pourront pas créer les zidb nécessaires à la création d'un set et resteront tout simplement bloqués à cette étape. A l'inverse, d'autres commandes sont obsolètes. C'est-à-dire que ces dernières étaient utiles dans les versions précédentes mais n'ont plus aucuns intérêts dans la présente version. Par exemple, il existe une commande permettant de créer des fichiers zid. Ces fichiers avaient la même utilité que les zidb dans les versions antérieures mais ne sont pas utilisables dans la version 5. Il serait donc judicieux de changer la commande « create zid files » en « create zidb files », afin que n'importe qui puisse créer le set de son choix.

Suite à différents tests d'analyses d'échantillons, différents problèmes d'ergonomies ont été décelés dans le module de correction d'erreurs (celui-ci étant le module principalement utilisé par l'analyste). Leur amélioration pourrait réduire de façon considérable le temps passé à la correction d'erreur. Nous avons donc pu remarquer que les colonnes dédiées aux différents groupes dans le module de correction d'erreurs ne peuvent comporter qu'au maximum 200 vignettes. A Nantes le problème ne se pose pas puisqu'ils n'atteignent jamais ce nombre par groupe pendant cette étape mais à Boulogne la chose est différente. Du fait, d'un grand nombre de débris, le maximum de 200 vignettes par colonne est facilement atteint. Ceci pouvant bloquer l'analyste lors de la correction d'erreurs. De plus le reclassement des vignettes dans ce module ne s'avère pas être optimal. Il serait utile de pouvoir déplacer les colonnes des groupes afin de faciliter le transfert des vignettes entre deux colonnes opposées dans le tableau. Il est également impossible de sélectionner plusieurs vignettes à la fois. Or cette fonctionnalité devrait réduire significativement le temps passé par l'analyste sur cette étape. L'ajout d'une barre d'échelle sur les vignettes lors de la phase de correction d'erreur pourrait faciliter grandement les choix de l'analyste pendant cette étape. Il existe également une colonne appelée « Other » permettant de stocker temporairement les vignettes destinées à être reclassées. Néanmoins il est impossible de placer des vignettes dans cette colonne et elle est donc inutilisable pour l'instant. Enfin, en passant le curseur sur une vignette, un zoom de celle-ci devrait apparaître. Or, lors de cet affichage aucun zoom n'est fait, la vignette apparaît bien mais à la même taille que celle dans la colonne. Cette fonctionnalité n'a donc aucune utilité tant qu'elle ne sera pas améliorée.

Une des perspectives à ne pas négliger est le type d'utilisation du FlowCAM. En effet, au vu du nombre de fonctionnalités que possède le système FlowCAM/ZooPhytoImage, il est possible de choisir la façon d'utilisation la plus appropriée aux objectifs du projet. Par exemple, il peut être utilisé pour l'étude de la richesse taxonomique (attention nécessité d'avoir une reconnaissance optimale jusqu'au niveau de l'espèce si possible) ou de la dynamique des communautés phytoplanctoniques (comme la succession des communautés au court du temps). Il est aussi possible d'utiliser le FlowCAM pour un simple dénombrement des particules ou pour déterminer la gamme de taille d'un échantillon (Sieracki *et al*, 1998). Cette méthode fonctionne parfaitement et donne des résultats très satisfaisant. Toutefois, cela exige l'utilisation d'échantillons « purs ». C'est-à-dire que l'échantillon ne doit comporter que les particules d'intérêt car de ce fait il est impossible de séparer les particules

entre elles. Ceci revient donc presque toujours à utiliser des échantillons issus de cultures en laboratoire et dont la composition est totalement connue. Enfin, il est possible de coupler le système FlowCAM/ZooPhytoImage avec d'autres appareils de mesures (cytosence, cytomètre, etc.). Par exemple, en 2015, See *et al* ont couplé l'utilisation du FlowCAM avec celle du Fluoroprobe (celui-ci détermine l'assemblage phytoplanctonique d'un échantillon en analysant sa composition pigmentaire). Cette technique a permis de corriger les erreurs faites par les deux appareils et donc d'avoir une meilleure identification des particules. Il est également possible de faire une étude ciblée sur une seule espèce d'intérêt majeur. Par exemple, en 2006 une étude s'est déroulée sur des échantillons provenant du Texas et de la Floride (Buskey et Hyatt, 2006). Celle-ci consistait à étudier la dynamique d'une espèce responsable d'un phénomène d'eau colorée : *Karenia brevis*. Ces travaux ont également montré une difficulté d'identification entre les cellules de *K. brevis* et d'autres espèces de même taille et de morphologie similaire. Ceci montre donc qu'il est indispensable de travailler sur l'optimisation des confusions entre les groupes d'intérêts. Il est également possible d'imaginer de travailler conjointement avec un modèle mathématique. Il permettrait, en connaissant la dynamique de l'espèce cible, de pouvoir prédire potentiellement l'apparition de blooms à court terme par exemple. Pour *Phaeocystis globosa*, nous pourrions imaginer un couplage de ce modèle avec des données météorologiques permettant de prévoir la formation de mousse suite à la prédiction d'un niveau d'agitation de la mer.

Conclusion

Le système couplé FlowCAM/ZooPhytoImage semble être un outil très utile pour des analyses se déroulant à hautes fréquences. C'est pourquoi il s'insère parfaitement au sein du REPHY puisque ce réseau comporte plus de 300 stations de prélèvement (soit plusieurs milliers d'échantillons à analyser chaque). D'où l'utilité de posséder un outil capable de suivre les fréquences d'échantillonnages. Néanmoins ce système n'est pas parfait. En effet, il reste encore une phase d'optimisation assez conséquente. D'une part le set d'apprentissage nécessite une réelle amélioration ainsi qu'une spécification de son contenu en fonction de la zone affiliée à chaque laboratoire. D'autre part, le logiciel est toujours en version bêta et nécessite encore un travail afin de corriger les bugs et d'améliorer son fonctionnement pour le rendre simple d'utilisation à n'importe qui.

Cependant, les limites de ce système commencent à être atteinte et il est essentiel de pouvoir palier à ce problème. Pour cela, plusieurs solutions ont été envisagées. En effet, l'ajout de modules au FlowCAM, pourrait améliorer ses performances. Par exemple, il serait peut-être avantageux d'utiliser un module de fluorescence pour les laboratoires dont les échantillons sont chargés en débris (pour l'Ifremer c'est le centre de Boulogne). Ce module permettrait d'écarter une partie des débris avant numérisation des échantillons, entraînant aussi une diminution de la place prise par les échantillons sur le disque de stockage. Une autre éventualité serait de changer l'objectif x4 par un objectif x10 mais celui-ci augmenterait considérablement le temps de numérisation (plus d'une heure pour un échantillon). Il est également possible d'utiliser des sets spécialisés à chaque laboratoire sachant que cette option entrainera une variabilité dans les résultats entre les différents centres Ifremer. C'est pourquoi un module d'apprentissage actif va être ajouté au logiciel, ainsi qu'un module de dénombrement des colonies pour améliorer l'estimation de l'abondance de ces espèces.

A la vue des performances du système FlowCAM/ZooPhytoImage, il semble nécessaire de développer une machine permettant une meilleure qualité de numérisation (zoom x10) tout en gardant un minimum de temps de passage de l'échantillon (10 minutes par numérisation). Ces objectifs ouvrent donc la porte au développement d'un nouvel appareil appelé FastCAM. Cet appareil est une version améliorée du FlowCAM et sera couplé avec le logiciel ZooPhytoImage (d'où la nécessité d'une version opérationnelle avant intégration à ce nouvel appareil).

Bibliographie

- Álvarez, Eva, Marta Moyano, Ángel López-Urrutia, Enrique Nogueira, and Renate Scharek. "Routine Determination of Plankton Community Composition and Size Structure : A Comparison between FlowCAM and Light Microscopy." *Journal of Plankton Research*, 2013.
- Breiman, Leo. "Random Forests." *Machine Learning* 45, no. 1 (2001) : 5–32.
- Breton, E., C. Brunet, B. Sautour, and J.-M. Brylinski. "Annual Variations of Phytoplankton Biomass in the Eastern English Channel : Comparison by Pigment Signatures and Microscopic Counts." *Journal of Plankton Research* 22, no. 8 (2000) : 1423–40.
- Buskey, Edward J., and Cammie J. Hyatt. "Use of the FlowCAM for Semi-Automated Recognition and Enumeration of Red Tide Cells (*Karenia Brevis*) in Natural Plankton Samples." *Harmful Algae* 5, no. 6 (2006) : 685–92.
- Denis, Kevin, and Phillippe Grosjean. "Original Method to Optimize Groups in Automatic Plankton Classification, Applied to North Sea Phytoplankton Using Zoo/PhytoImage" in 'Joint Meeting ASLO & NABS. Aquatic Sciences : Global Changes from the Center to the Edge." Sante Fe, USA, 2010.
- Guiselin, Natacha. "Étude de La Dynamique Des Communautés Phytoplanctoniques Par Microscopie et Cytométrie En Flux, En Eaux Côtières de La Manche Orientale." 2010.
- Hernández-Fariñas, Tania, Dominique Soudant, Laurent Barillé, Catherine Belin, Alain Lefebvre, and Cédric Bacher. "Temporal Changes in the Phytoplankton Community along the French Coast of the Eastern English Channel and the Southern Bight of the North Sea." *ICES Journal of Marine Science : Journal Du Conseil* 71, no. 4 (2014) : 821–33.
- Lefebvre, Alain, and Morgane Flesch. "Suivi Régional Des Nutriments Sur Le Littoral Nord-Pas de Calais/Picardie," 2015.
- Lefebvre, Alain, Natacha Guiselin, Frederique Barbet, and Felipe L. Artigas. "Long-Term Hydrological and Phytoplankton Monitoring (1992–2007) of Three Potentially Eutrophic Systems in the Eastern English Channel and the Southern Bight of the North Sea." *ICES Journal of Marine Science : Journal Du Conseil* 68, no. 10 (2011) : 2029–43.
- Lefebvre, Alain, and Claudie Mégret. "Suivi Régional Des Nutriments Sur Le Littoral Nord-Pas de Calais/Picardie," 2014.
- Lepage, Richard. *Reconnaissance D'algues Toxiques Par Vision Artificielle et Réseau de Neurones*. Université du Québec à Rimouski Rimouski, 2004.
- Martin, Corinne S., Andre Carpentier, Sandrine Vaz, Franck Coppin, L. Curet, J.-C. Dauvin, Juliette Delavenne, et al. "The Channel Habitat Atlas for Marine Resource Management (CHARM) : An Aid for Planning and Decision-Making in an Area under Strong Anthropogenic Pressure." *Aquatic Living Resources* 22, no. 04 (2009) : 499–508.
- Menden-Deuer, Susanne, and Evelyn J. Lessard. "Carbon to Volume Relationships for Dinoflagellates, Diatoms, and Other Protist Plankton." *Limnology and Oceanography* 45, no. 3 (2000) : 569–79.
- Moberg, Emily A., and Heidi M. Sosik. "Distance Maps to Estimate Cell Volume from Two-Dimensional Plankton Images." *Limnology and Oceanography: Methods* 10, no. 4 (2012) : 278–88.
- Poulton, Nicole J., and Jennifer L. Martin. "Imaging Flow Cytometry for Quantitative Phytoplankton Analysis-FlowCAM." *Microscopic and Molecular Methods for Quantitative Phytoplankton Analysis*, 2010, 47.
- See, Jason H., Lisa Campbell, Tammi L. Richardson, James L. Pinckney, Rongjun Shen, and Norman L. Guinasso. "Combining New Technologies for Determination of Phytoplankton Community Structure in the Northern Gulf of Mexico 1." *Journal of Phycology* 41, no. 2 (2005) : 305–10.

- Sieracki, Christian K., Michael E. Sieracki, and Charles S. Yentsch. "An Imaging-in-Flow System for Automated Analysis of Marine Microplankton." *Marine Ecology Progress Series* 168 (1998) : 285–96.
- Soudant, Dominique, and Laurence Miossec. "Note Bibliographique : Incertitudes Des Mesures Phytoplanctoniques," 2013.
- Tunin Ley, Alina, and Daniele Maurer. "Mise En Oeuvre Opérationnelle D'un Système Couplé de Numérisation (FlowCAM) et de Traitement D'images (PhytoImage) Pour L'analyse Automatisée, Ou Semi-Automatisée, de La Composition Phytoplanctonique D'échantillons D'eau de Mer-Premières Étapes," 2011.
- Wacquet, Guillaume, Alain Lefebvre, Florent Colas, and Philippe Grosjean. "Finalization of a Pre-Operational Tool FlowCAM/ZooPhytoImage.," 2014.

Annexes

Annexe 1. Tableau récapitulatif des groupes composants le training set Manche-Atlantique V0 et V0.1 ainsi que le nombre de vignettes et leur taux d'erreur (en vert, les taux d'erreurs ayant diminué entre les deux versions ; en rouge, ceux qui ont augmenté et en bleu ceux ayant conservé le même taux).

Classe du groupe	Nom du groupe	Nombre de vignettes	Taux d'erreur V0 (%)	Taux d'erreur V0.1 (%)
Bacillariophyceae	<i>Asterionellopsis glacialis</i>	227	15	14
	<i>Asteroplanus karianus</i>	5	100	
	<i>Bacteriastrum</i>	180	31	34
	<i>Cerataulina</i>	127	39	40
	<i>Chaetoceros</i>	389	53	54
	<i>Chaetoceros curvisetus</i>	195	43	42
	<i>Chaetoceros danicus</i>	120	13	10
	<i>Chaetoceros socialis</i>	60	68	68
	<i>Cylindrotheca</i>	50	46	48
	<i>Hantzschia Nitzschia</i>			
	<i>Dactyliosolen fragilissimus</i>	238	15	14
	<i>Ditylum brightwellii</i>	351	23	23
	<i>Eucampia</i>	3	100	
	<i>Guinardia delicatula</i>	56	95	93
	<i>Guinardia striata</i>	84	10	10
	<i>Guinardia flaccida</i>	335	61	63
	<i>Gyrosigma</i>	281	17	16
	<i>Pleurosigma</i>			
	<i>Helicotheca</i>	7	100	
	<i>Hemiaulus</i>	2	100	
	<i>Lauderia</i>	12	92	
	<i>Leptocylindrus</i>	382	22	23
	<i>Nitzschia longissima</i>	23	83	83
	<i>Odontella</i>	160	30	31
	<i>Paralia</i>	93	35	33
	<i>Proboscia Rhizosolenia</i>	363	26	26
	<i>Pseudo-nitzschia</i>	426	11	12
	<i>Skeletonema</i>	162	20	22
<i>Thalassionema</i>	220	30	34	
<i>Thalassiosira cells</i>	41	44	51	
<i>Thalassiosira chaines</i>	225	48	49	
Dynophyceae	<i>Dinophysis</i>	391	16	15
	<i>Prorocentrum</i>	385	6	6
	<i>Alexandrium</i>	263	13	13
	<i>Neoceratium furca lineatum</i>	206	10	10
	<i>Neoceratium fusus</i>	184	11	12
	<i>Peridinium</i>	178	51	49

Dynophyceae	<i>Protoperidinium</i>			
	<i>Lepidodinium mucous</i>	172	20	19
	<i>Scrippsiella</i> <i>Ensiculifera</i> <i>Pentapharsodinium</i> <i>Bysmatrum</i>	138	22	26
	<i>Gymnodinium</i> <i>Gyrodinium</i> <i>Torodinium</i>	124	62	62
	<i>Alexandrium affine</i>	93	26	24
	<i>Ceratium</i>	54	9	11
	<i>Gonyaulax</i>	48	40	54
	<i>Dissodinium Pyrocystis</i>	38	61	66
	<i>Polykrikos</i>	23	39	35
	<i>Diplopsalis Diplopelta</i> <i>Oblea</i>	6	100	
	<i>Heterocapsa</i>	2	100	
	Autres classes	<i>Phaeocystis</i>	264	33
<i>Dictyocha</i>		106	21	22
<i>Euglenoidea</i>		53	9	8
<i>Noctiluca</i>		16	13	19
<i>Pyramimonas</i>		3	100	
Zooplancton	<i>Ciliophora_big</i>	47	51	49
	<i>Ciliophora_small</i>	200	61	62
	<i>Pisces larva</i>	7	100	
	<i>Tintinnids</i>	40	75	80
	<i>Zooplankton spp</i>	29	31	38
Débris	<i>fiber</i>	337	28	33
	<i>dark</i>	298	58	58
	<i>bubble</i>	287	22	20
	<i>clear</i>	249	40	40
	<i>Black opaque</i>	230	35	31
	<i>Granular</i>	213	46	48
	<i>Empty thecal</i>	87	62	63
	<i>membranous</i>	83	67	71
	<i>aggregates</i>	45	100	98
	<i>mues</i>	41	59	66
	<i>polen</i>	16	44	
	<i>Empty frustule</i>	15	100	

Annexe 2. Tableau récapitulatif des groupes composants le training set Manche-Atlantique V1 ainsi que les taux d'erreurs des versions V0.1 et V1 (en vert, les taux d'erreurs ayant diminués entre les deux versions ; en rouge, ceux qui ont augmenté et en bleu ceux ayant conservé le même taux).

Classe du groupe	Nom du groupe	Nombre de vignettes	Taux d'erreur V1 (%)	Taux d'erreur V0.1 (%)
Bacillariophyceae	<i>Asterionellopsis glacialis</i>	227	14	14
	<i>Asteroplanus karianus</i>	5	 	
	<i>Bacteriastrum</i>	180	31	34
	<i>Cerataulina</i>	127	39	40
	<i>Chaetoceros</i>	260	65	54
	<i>Chaetoceros curvisetus</i>	200	43	42
	<i>Chaetoceros danicus</i>	122	11	10
	<i>Chaetoceros socialis</i>	60	58	68
	<i>Cylindrotheca</i>	73	55	48
	<i>Hantzschia Nitzschia</i>			
	<i>Dactyliosolen fragilissimus</i>	238	13	14
	<i>Ditylum brightwellii</i>	261	26	23
	<i>Eucampia</i>	3	 	
	<i>Guinardia delicatula</i>	56	88	93
	<i>Guinardia striata</i>	84	12	10
	<i>Guinardia flaccida</i>	260	58	63
	<i>Gyrosigma</i>	281	16	16
	<i>Pleurosigma</i>			
	<i>Helicotheca</i>	7	 	
	<i>Hemiaulus</i>	2	 	
	<i>Lauderia</i>	12	 	
	<i>Leptocylindrus</i>	256	23	23
	<i>Nitzschia longissima</i>	23	 	83
	<i>Odontella</i>	160	32	31
	<i>Paralia</i>	93	49	33
	<i>Proboscia Rhizosolenia</i>	265	8	26
	<i>Rhizosolenia setigera</i>	41	24	
	<i>Pseudo-nitzschia</i>	260	24	12
	<i>Skeletonema</i>	162	17	22
	<i>Thalassionema</i>	220	30	34
<i>Thalassiosira cells</i>	41	39	51	
<i>Thalassiosira chaines</i>	225	45	49	
Dynophyceae	<i>Dinophysis</i>	260	15	15
	<i>Dinophysis tripos</i>	36	31	
	<i>Prorocentrum</i>	260	13	6
	<i>Alexandrium</i>	263	8	13
	<i>Neoceratium furca lineatum</i>	206	8	10

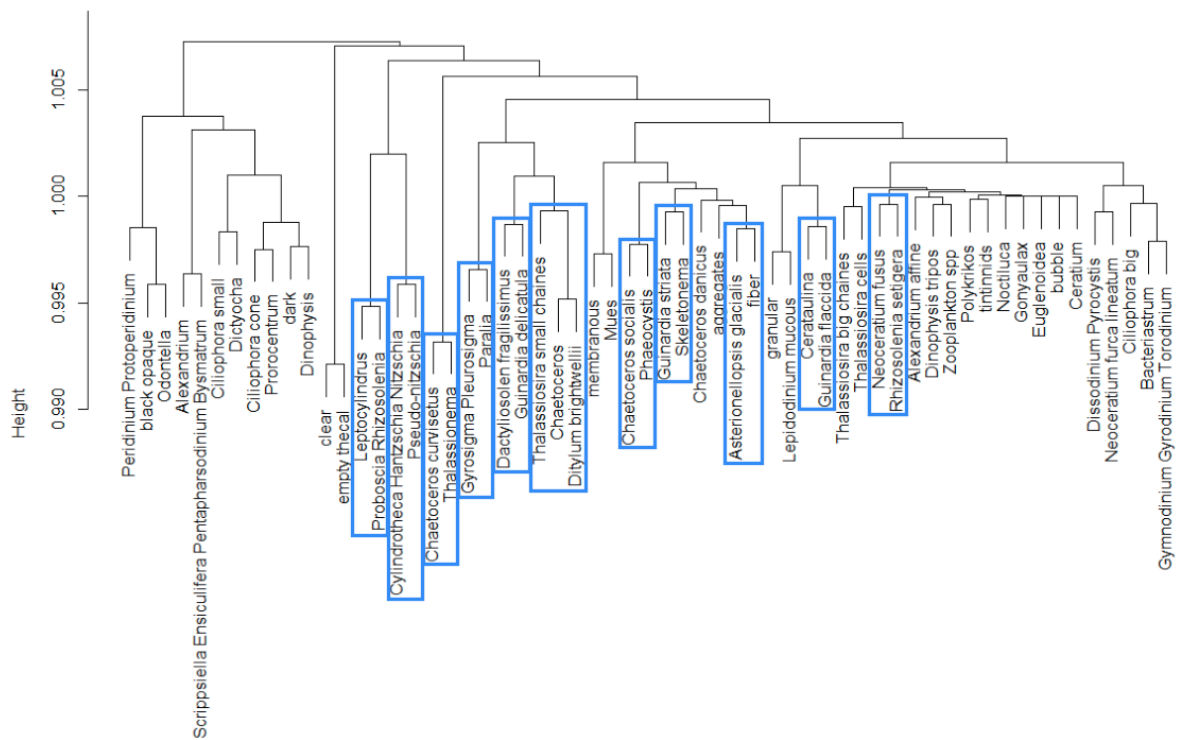
Dynophyceae	<i>Neoceratium fusus</i>	184	11	12
	<i>Peridinium</i> <i>Protoperidinium</i>	178	27	49
	<i>Lepidodinium mucous</i>	172	19	19
	<i>Scrippsiella</i> <i>Ensiculifera</i> <i>Pentapharsodinium</i> <i>Bysmatrum</i>	138	21	26
	<i>Gymnodinium</i> <i>Gyrodinium</i> <i>Torodinium</i>	124	56	62
	<i>Alexandrium affine</i>	93	20	24
	<i>Ceratium</i>	53	11	11
	<i>Gonyaulax</i>	48	38	54
	<i>Dissodinium Pyrocystis</i>	38	61	66
	<i>Polykrikos</i>	23	24	35
	<i>Diplopsalis Diplopelta</i> <i>Oblea</i>	6		
	<i>Heterocapsa</i>	2		
Autres classes	<i>Phaeocystis</i>	264	39	32
	<i>Dictyocha</i>	106	24	22
	<i>Euglenoidea</i>	53	8	8
	<i>Noctiluca</i>	16	24	19
	<i>Pyramimonas</i>	3		
Zooplankton	<i>Ciliophora_big</i>	47	51	49
	<i>Ciliophora_small</i>	200	61	62
	<i>Pisces larva</i>	7		
	<i>Tintinnids</i>	40	73	80
	<i>Zooplankton spp</i>	29	34	38
Débris	<i>fiber</i>	260	25	33
	<i>dark</i>	200	66	58
	<i>bubble</i>	200	5	20
	<i>clear</i>	249	41	40
	<i>Black opaque</i>	230	30	31
	<i>Granular</i>	213	46	48
	<i>Empty thecal</i>	88	69	63
	<i>membranous</i>	83	64	71
	<i>aggregates</i>	45	98	98
	<i>mues</i>	41	61	66
	<i>polen</i>	16		
	<i>Empty frustule</i>	15		

Annexe 3. Tableau récapitulatif des groupes composants le training set Manche-Atlantique V2 ainsi leur taux d'erreurs associés.

Nom du groupe	Nombre de vignettes	Taux d'erreur (%)
<i>Aggregates</i>	45	96
<i>Alexandrium</i>	263	10
<i>Alexandrium affine</i>	93	23
<i>Asterionellopsis glacialis</i>	227	15
<i>Bacteriastrum</i>	180	27
<i>black opaque</i>	230	29
<i>bubble</i>	200	5
<i>Cerataulina</i>	127	38
<i>Ceratium</i>	53	9
<i>Chaetoceros</i>	260	63
<i>Chaetoceros curvisetus</i>	200	42
<i>Chaetoceros danicus</i>	122	8
<i>Chaetoceros socialis</i>	60	63
<i>Ciliophora big</i>	47	53
<i>Ciliophora cone</i>	104	86
<i>Ciliophora small</i>	84	65
<i>clear</i>	249	43
<i>Cylindrotheca Hantzschia Nitzschia</i>	73	56
<i>Dactyliosolen fragilissimus</i>	238	12
<i>dark</i>	200	69
<i>Dictyocha</i>	107	16
<i>Dinophysis</i>	260	15
<i>Dinophysis tripos</i>	36	33
<i>Dissodinium Pyrocystis</i>	38	58
<i>Ditylum brightwellii</i>	261	25
<i>empty thecal</i>	88	59
<i>Euglenoidea</i>	53	8
<i>Fiber</i>	260	29
<i>Gonyaulax</i>	48	35
<i>Granular</i>	213	46
<i>Guinardia delicatula</i>	56	89
<i>Guinardia flaccida</i>	260	13
<i>Guinardia striata</i>	84	58
<i>Gymnodinium Gyrodinium Torodinium</i>	124	58
<i>Gyrosigma Pleurosigma</i>	281	16
<i>Lepidodinium mucous</i>	172	18
<i>Leptocylindrus</i>	256	21
<i>membranous</i>	83	65
<i>Mues</i>	41	61
<i>Neoceratium furca lineatum</i>	206	9
<i>Neoceratium fusus</i>	184	12

<i>Noctiluca</i>	16	13
<i>Odontella</i>	160	28
<i>Paralia</i>	93	35
<i>Peridinium Protoperidinium</i>	178	47
<i>Phaeocystis</i>	264	32
<i>Polykrikos</i>	23	39
<i>Proboscia Rhizosolenia</i>	265	26
<i>Prorocentrum</i>	260	8
<i>Pseudo-nitzschia</i>	260	12
<i>Rhizosolenia setigera</i>	41	34
<i>Scrippsiella Ensiculifera Pentapharsodinium</i> <i>Bysmatrum</i>	138	24
<i>Skeletonema</i>	162	22
<i>Thalassionema</i>	169	36
<i>Thalassiosira big chaines</i>	99	44
<i>Thalassiosira cells</i>	41	39
<i>Thalassiosira small chaines</i>	104	46
<i>tintinnids</i>	40	78
<i>Zooplankton spp</i>	29	34

Annexe 4. Dendrogramme représentant les confusions d'intérêt dans le set Manche V2 (en bleu : les confusions en lien avec les groupes d'intérêt).



Annexe 5. Matrice permettant de visualiser l'intensité des confusions choisies (en bleu : les confusions en rapport avec les groupes d'intérêt).

Actual // Predicted	01	04	07	10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55	58				
inium Protoperidinium 01	98	20	5	12			6	12	1						1			3		9	1		9	
black opaque 02	14	14	1	1	2	4	3								1			4	2	4	2		2	4
Odontella 03		2	1	5						5	1	2			4	8	1		1	1			1	
Alexandrium 04	1	2	2	1																	2	5		
irsodium Bysmatrum 05	2	1	2	2	4	1	1				2				1	14	1		1					
Ciliophora small 06	5	2	3	2	6	1	4	1	6						1									
Dityocha 07			1	2	2	7																		
Ciliophora cone 08	5	5	1	4	9	1	1	7	9		1	1	1			3	1	1					1	7
Prorocentrum 09	1	1		6	3	2	6																	1
dark 10	5	1	1	5	7	9	8	3	2	6	5				12	5	3	1				1		6
Dinophysis 11	1			7	14	4	2	1																1
clear 12				2	4	3	4	2							1	1	14	2	1	4	1	1	5	4
empty thecal 13					1	1	4	2																1
Leptocylindrus 14																								
roboscia Rhizosolenia 15																								
Hantzschia Nitzschia 16																								
Pseudo-nitzschia 17																								
haetoceros curvisetus 18																								
Thalassionema 19																								
rosigma Pleurosigma 20																								
Paralia 21																								
ylitosolen fragilissimus 22																								
Guinardia delicatula 23																								
ssiosira small chaines 24																								
Chaetoceros 25																								
Ditylum brightwellii 26																								
membranous 27																								
Mues 28																								
Chaetoceros socialis 29																								
Phaeocystis 30																								
Guinardia striata 31																								
Skeletonema 32																								
Chaetoceros danicus 33																								
aggregates 34																								
terionellopsis glacialis 35																								
fiber 36																								
granular 37																								
Lepidodinium mucosum 38																								
Cerataulina 39																								
Guinardia flaccida 40																								
lassiosira big chaines 41																								
Thalassiosira cells 42																								
Neoceratium fuscum 43																								
Rhizosolenia setigera 44																								
Alexandrium affine 45																								
Dinophysis tripos 46																								
Zooplankton spp 47																								
Polykrikos 48																								
trintinids 49																								
Noctiluca 50																								
Gonyaulax 51																								
Euglenoidea 52																								
bubble 53																								
Ceratium 54																								
irsodium Pyrocystis 55																								
eratum furca lineatum 56																								
Ciliophora big 57																								
Bacterasterium 58																								
yrodinium Torodinium 59																								

Résumé

Le phytoplancton est un élément essentiel dans les écosystèmes marins. En effet, il est responsable de la fixation de la moitié du carbone sur Terre. Il est également à la base de la régulation de nombreux réseaux trophiques marins. Depuis plusieurs années, des variations dans la dynamique des communautés phytoplanctoniques ainsi que des blooms successifs ont pu être notés.

C'est donc dans ce contexte qu'a pu s'insérer le REPHY qui est un réseau de surveillance et d'observation du phytoplancton et des phycotoxines. Ce réseau est composé de plus de 300 stations de prélèvements, elles même comportant plusieurs points. Afin de pouvoir traiter tous ces échantillons dans les temps et d'effectuer les prélèvements à hautes fréquences, il est nécessaire de trouver de nouvelles méthodes d'analyses. En effet, les méthodes conventionnelles (microscopie) semblent avoir des limites dans cette situation. C'est pourquoi l'utilisation du système combiné FlowCAM/ZooPhytoImage a été développée.

Ce système semble être plus approprié pour l'étude de nombreux échantillons à hautes fréquences. Néanmoins, il est pour l'instant nécessaire d'utiliser un set d'apprentissage spécifique à la région d'utilisation afin de diminuer au mieux les erreurs de classification. Ce logiciel permet également une estimation de l'abondance pour chaque espèce. Cependant pour les espèces coloniales, le logiciel a besoin d'un module de dénombrement des colonies (actuellement en développement). Le logiciel ZooPhytoImage a également besoin d'une optimisation avant ouverture au public car pour l'analyste il est difficile actuellement de corriger les erreurs de l'outil de reconnaissance.

Pour conclure, ce système semble être un bon moyen de substitution à la microscopie. Celui-ci pourrait à terme permettre une diminution du temps de travail pour un échantillon et une augmentation de la fréquence d'analyse. Néanmoins, nous savons que nous atteignons les limites de ce système est c'est pourquoi un nouvel appareil plus performant est en développement : le FastCAM.

Mots clés : FlowCAM, ZooPhytoImage, phytoplancton, bloom, HAB, classification automatique, outil de reconnaissance, estimation de l'abondance, haute fréquence

Abstract

Phytoplankton plays an important role in marine ecosystems. Indeed, it is responsible of half of carbon fixation. It is also the basis of many marine food webs' regulations. For several years, changes in the dynamics of phytoplankton and successive blooms were shown.

REPHY which is an observation and monitoring network of phytoplankton and shellfish toxins, was created in this context. This network has more than 300 sampling sites. To processes all those samples in high frequencies, it is necessary to find new methods of analysis. In fact, conventional methods appear to have limitations in this situation. That's why the use of the combined system FlowCAM/ZooPhytoImage was developed.

This system seems to be more appropriate to study many samples at high frequencies. Nevertheless, it is actually necessary to use a specific training set for the study area in order to have the best recognition. This software also allows an estimation of abundance for each species. However, for colonial species, the software needs a colony counting module (currently in development). The software ZooPhytoImage also needs to be optimized before its distribution because the errors' correction by the analyst is currently difficult.

Finally, this system seems to be a good way of microscopy substitution. This could eventually allow a reduction in working time and an increment of analysis frequency. Nevertheless, we know that we are reaching the limits of this system and that's why a new and more powerful system is in development : FastCAM.

Keywords : FlowCAM, ZooPhytoImage, phytoplankton, bloom, HAB, high frequencies, recognition tools, abundance, estimation, classification

FlowCAM / PhytoImage, système couplé de numérisation et d'analyse d'images de phytoplancton

FORMATION A FLOWCAM / ZOOPHYTOIMAGE V5

CR Mission Formation Nadine Neaud à Arcachon,

Du 31 mars au 3 avril 2015

Objectifs de la mission formation:

- Apporter une mise à niveau aux personnels Ifremer LER / AR pour l'utilisation de l'outil FlowCAM couplé au logiciel ZooPhytoImage version 5
 - perfectionner l'utilisation de l'outil
 - transmettre le plan d'action attendu pour l'activité 2015 de ce projet
-

Plan prévisionnel de la formation

Durée du stage et liste des participants

- 4 jours
- 3 stagiaires :
 - o Danièle Maurer
 - o Myriam Perrière
 - o Claire Meteigner

Dates et lieu de la formation

Laboratoire Ifremer LER Arcachon : du 30/03 après-midi au 03/04 matin.

Pré-requis techniques

Cette formation est basée sur l'utilisation du FlowCAM et de la version 5 du logiciel ZooPhytoImage.

Le laboratoire d'Arcachon est équipé de l'outil d'acquisition d'images FlowCAM.

La version 5 du logiciel ZooPhytoImage est apportée par la formatrice.

Programme de la formation

La formation comprend des présentations théoriques et des travaux pratiques sur le FlowCAM et le logiciel.

Contenu :

Lundi 30 mars après-midi :

Installation, préparation et test du matériel

Mardi 31 mars :

Réunion d'introduction à la formation, présentation de l'état d'avancement du projet, travaux à mener en 2015 et objectifs, présentation de la version 5 du logiciel et plus particulièrement des évolutions, présentation du module de dénombrement des colonies, présentation du disque réseau de travail.

Mercredi 1^{er} avril :

Numérisation d'échantillons au FlowCAM : procédure et contexte à appliquer.

Traitement par le logiciel des fichiers issus des numérisations.

Jeudi 2 avril :

Présentation et exploration du set d'apprentissage Manche-Atlantique 4X

Traitement avec le module de dénombrement des colonies

Vendredi 3 avril matin:

Réunion de clôture de la formation.

Plan d'action 2015

Compte-rendu et bilan de la mission

Matériel

Au laboratoire d'Arcachon, le FlowCAM modèle Bench Top Flowcam US IV acquis en 2008 est équipé du logiciel VisualSpreadSheet 2.2.7 et du so Windows XP.

Avec ce modèle, l'entraînement de l'échantillon dans la cellule est réalisé par une pompe péristaltique.

Pour l'alimentation de l'échantillon, nous avons installé un entonnoir cylindrique en verre et un moteur d'homogénéisation à hélice.



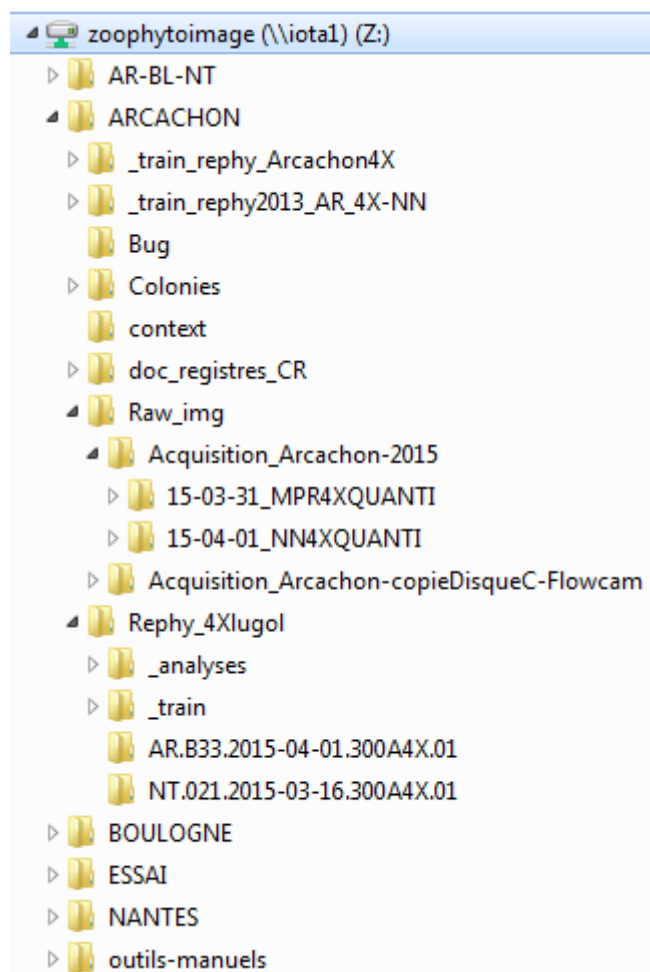
Le système apporte toute satisfaction concernant la qualité des images.

Dans les paramètres d'acquisition, selon les recommandations d'Alina Tunin-Ley, avec l'objectif 4X, le nombre d'images par seconde serait préférentiellement de 7 images/seconde.

Avec ce paramètre, la durée d'acquisition d'un volume imagé de 10 mL est de 33 minutes. La possibilité de numériser à 12 images/seconde (maximum possible avec ce modèle) a été envisagée. L'essai réalisé semble satisfaisant. Les images sont de bonne qualité et la durée de numérisation atteint 18 minutes pour 10 mL. Il est donc proposé de travailler désormais à la vitesse d'acquisition de 12 images/seconde.

Concernant les postes de travail utilisateurs pour le traitement des numérisations par ZooPhytoImage V5, les installations se sont correctement déroulées. Cependant, il s'est avéré que le logiciel ne fonctionnait pas sur l'ordinateur de Myriam du fait de sa vétusté. Il sera prochainement remplacé.

L'environnement informatique, et plus particulièrement, l'organisation des fichiers sur le disque réseau <\\iota1\zoophytoimage> a été présenté et adopté par l'équipe.



Numérisations d'échantillons au FlowCAM

La procédure de numérisation a fait l'objet d'une première démonstration, puis, successivement, chaque utilisatrice a réalisé plusieurs numérisations. Cette étape d'acquisition d'images est maintenant bien maîtrisée par les utilisatrices.

Une fiche mémo résumant la procédure a été fournie.

La procédure de numérisation (objectif 4X) complète et détaillée devra faire l'objet d'une révision pour mise à jour et application au LER/AR (action LER/AR).

Traitement des numérisations avec l'UI de ZooPhytoImage-V5

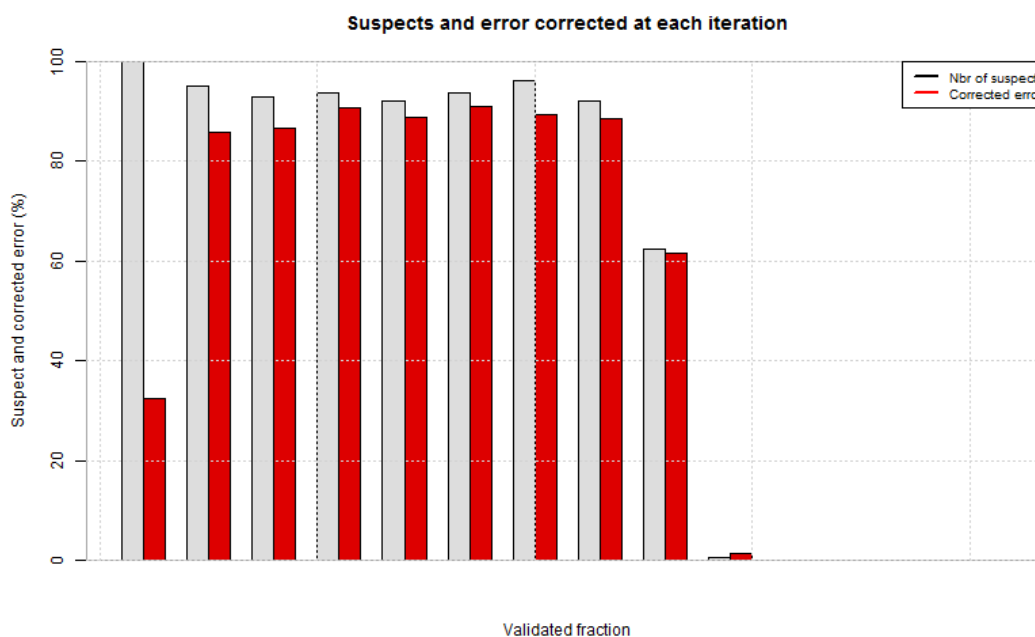
Après installation de la V5 de ZooPhytoImage, une première démonstration de l'UI a été réalisée s'appuyant sur le manuel utilisateur

L'intégration des fichiers des numérisations acquises lors de cette session s'est correctement déroulée à l'exception d'une qui n'avait sans doute pas été correctement réalisée. Lors de cette numérisation, à l'ouverture de Visual Spreadsheet, en session utilisateur, il y avait eu un message proposant le chargement d'un fichier qui avait été ignoré la première fois. En recommençant la numérisation et en acceptant le chargement du fichier proposé, la numérisation obtenue s'est alors bien intégrée lors du traitement par ZooPhytoImage. Ce message n'apparaît jamais en session administrateur.

Après intégration des fichiers, une démonstration d'analyse avec l'UI a été présentée. Une analyse complète d'un échantillon a permis aux utilisateurs d'appréhender l'outil.

Pour cet essai, le diagramme suivant a été obtenu :

AR.B33.2014-06-05.300A4X.01_3/Step10



Cet exercice a entraîné de nombreuses questions et réflexions :

1. Le nombre de vignettes à valider est annoncé à $1/20^{\text{ème}}$ des particules dénombrées. Est-ce toujours le cas ? Y a-t-il des limites ?
2. A l'issue de la première validation (où l'échantillon à valider est tiré au hasard, tous les individus étant a priori suspects), la précision calculée (ici environ 33 %) correspond-elle bien à 33 % d'individus douteux (ici environ 1 500 sur un total de 4 452)? Dans tous les cas, le nombre et la proportion d'individus douteux à chaque étape, pourraient-ils être affichés ?
3. Lors de la 2^{ème} validation et des suivantes, l'échantillon est « choisi principalement parmi les objets suspects » ? Que veut-dire principalement ? Lors des 7 validations suivantes, la proportion d'individus suspects est très élevée, environ 95 %, de même que la proportion (et le nombre) d'individus corrigés (en comparaison de l'exemple donné dans le manuel). Pourquoi ?
4. Lors de notre essai, au total environ 50 % des vignettes ont dû être validées (en 10 passages) soit plus de 2 200. Pourquoi ce nombre est-il beaucoup plus élevé que celui calculé au départ (1 500) ?
5. Le temps nécessaire à la correction d'erreurs jusqu'à obtention d'un pourcentage de suspects acceptable (minimum 10 étapes) est extrêmement long (presque 2 heures pour un opérateur expérimenté et pour un seul des triplicats). Pourra-il être diminué de façon conséquente ?
6. Dans l'ensemble, le prélèvement contient beaucoup de débris (90 %) et peu de particules d'intérêt (phyto). Beaucoup des vignettes à reclasser sont des débris. Cela peut-il être une piste pour des améliorations futures ?
7. Lors du reclassement des vignettes, il faudrait pouvoir en sélectionner facilement plusieurs (voire toute une colonne) ce qui ferait gagner beaucoup de temps.
8. Lorsque l'UI travaille, aucune indication à l'écran n'informe l'utilisateur du déroulement de ce travail (ainsi on ne sait pas si le travail avance ou bien s'il y a eu un problème entraînant l'arrêt du travail).
9. A l'issue de l'analyse, Il serait nécessaire de pouvoir visualiser directement les vignettes et leur classement (douteux, non douteux, classement final). Il est possible de récupérer la liste des items et leur classement, mais l'échantillon doit à nouveau être analysé à partir du menu "Analyse" pour pouvoir visualiser les vignettes et analyser leur classement.
10. Les utilisateurs n'ayant pas forcément participé à l'élaboration du set d'apprentissage ont besoin de visualiser les vignettes du set afin de s'y référer pour les étapes de correction d'erreur. A cette fin, un catalogue d'images du set va être réalisé (action Nadine Neaud).

Module de dénombrement des cellules dans les colonies

L'outil d'aide au dénombrement des cellules dans les colonies a été présenté et une fiche mémo résumant la procédure a été fournie. Cet outil est une première étape afin de permettre l'enrichissement du set d'apprentissage 4X par des informations concernant les nombres de cellules par colonie comptés manuellement par les utilisateurs. Ce travail a démarré à Boulogne et à Nantes sur un extrait des taxons coloniaux du set Manche-Atlantique 4X. Ce même travail est demandé à l'équipe d'Arcachon qui a accepté.

Conclusion et plan d'action

Numérisation : cette étape ne pose aucun problème, le LER/AR s'engage à numériser tous les échantillons de leur labo faisant l'objet d'une analyse de la flore totale (4 échantillons par mois). Chaque numérisation sera répliquée (3 répliqués). Il est envisagé d'écarter les échantillons dont la turbidité excède 2 NFU ; cette proposition est à étudier.

Traitement par ZooPhytoImage : même si l'interface utilisateur a considérablement amélioré l'importation et la mise en forme des données, le module d'analyse, plus particulièrement de correction d'erreur, ne répond pas complètement aux attentes car il est extrêmement chronophage. Par ailleurs, la colonne [other] ne fonctionne pas. Pour ces raisons, le LER/AR propose de ne pas commencer les traitements avant que ces problèmes ne soient résolus.

Comptage des cellules des colonies : cette action ne présente aucun problème de fonctionnement et de disponibilité pour le LER/AR. Quel délai est attendu pour ce travail ?

Set d'apprentissage : à l'occasion du balayage du set ou de l'utilisation du module de comptage de cellules de colonies, quelques vignettes se sont avérées mal classées. Un nouveau contrôle du set doit être réalisé ; il serait souhaitable que ce contrôle soit réalisé par différents membres du COPIL, comme cela avait été évoqué lors de la réunion du 02/12/2014. Ce travail doit être coordonné (action ??).

Par ailleurs, pour les classes d'intérêt qui ne présentent pas encore suffisamment de vignettes (LowAbondances), il est nécessaire de pister l'apparition de ces taxons dans les échantillons et de récupérer les vignettes. Pour cela, le LER/AR dispose de la liste des taxons concernés et indiquera à NMN les numérisations susceptibles d'en contenir. Dans tous les cas, un catalogue des vignettes du set est en préparation (action NMN).

CR de l'audioconférence du 28/05/2015 FlowCAM / ZooPhytoImage – Test de la V5

Ifremer Nantes & Boulogne – Université de Mons

Participants : Alain Lefebvre (Ifremer, LERBL), Remi Cuvelliez (stage M2, Ifremer LER BL), Nadine Neaud-Masson (Ifremer, DYNECO), Guillaume Wacquet (Univ. Mons), Philippe Grosjean (Univ. Mons).

Liste de diffusion

Membres du CoPiL du projet FlowCam/ZooPhytoImage.

Objectifs

Faire un debriefing (i) suite à l'échange de mails au sujet de ZooPhytoImage V5 et aux difficultés rencontrées par les utilisateurs, (ii) sur le CR de la formation à Arcachon, (iii) sur les travaux en cours et à venir.

A cette occasion fixer des échéances et des responsabilités.

Points abordés

Sujets des échanges de messages < 28/05		Audioconf. du 28/05	Décision (qui, quoi, quand)
IU* : Colonne [other] ne fonctionne pas *IU : Interface Utilisateur	PG : hypothèse de la non utilisation de Safari. Réponse Ifr : nous utilisons bien Safari. Hypothèse Ifr : nombre de Classe/colonne trop important. PG va vérifier L : sélection multiples ?	PG : bug à corriger. Autre hypothèse de travail : 2 colonnes other [phyto] et [non phyto] Difficile car code Java Script trop lourd pour un tel dvpt.	Mons, fin juin <i>Hors réunion : possibilité de réorganisation des colonnes dans l'IU selon confusions entre groupes (pour limiter la navigation entre colonnes pendant la validation manuelle).</i> Demande justifiée à engager dans un travail spécifique via un développeur Java Script ou stage Mons. Dans un premier temps, proposition d'une alternative : création de raccourcis clavier/souris.

<p>IU : Correction d'erreur trop chronophage</p>	<p>PG demande une formulation plus précise. Réponse IFR : Pour une acquisition de 2 à 3000 particules : 1 à 2 h. Mais à BL bcp de numérisations à plus de 10 000 particules (contenant 80 à 90% de particules détritiques).</p>	<p>PG : Bien vérifier la configuration Hardware pour avoir un fonctionnement optimal (la réussite de l'installation V5 sur un PC ne garantit pas un fonctionnement optimal) (cf manuel).</p> <p>Chronophage : cf apprentissage actif en cours de dvpt.</p> <p>Le temps de prise en main ne doit pas être considéré dans le temps moyen passé à la validation d'une vignette.</p> <p>Corr stat de l'erreur : on obtient des abondances corrigées par groupe en validant moins, mais problème de ne plus corriger toute l'erreur mais qu'une faible quantité (présence d'erreurs dans les groupes, mais équilibre entre FP et FN) et savoir où s'arrêter.</p> <p>Pb : Corr stat de l'erreur non testée au-delà du nbre de particules par groupe au regard des tests effectués à ce stade (càd pas pour nbre de cell par colonie),</p>	<p>cf manuel</p> <p>Réduction de 35 à 40 % du temps avec ce dvpt. Particulièrement efficace sur les triplicats ou sur des échantillons semblables</p> <p>Présentation graphique à voir pour proposer une aide à la décision si on choisit d'utiliser la correction statistique de l'erreur.</p> <p>Bonne concordance dès 3-4 étapes sur nbre de particules par groupe.</p> <p>Repose la question de l'approche nbre de cellules/colonie ou de l'adaptation de la stratégie (utilisation module corr erreur ou pas en fonction du risque environnementale et donc du temps de réaction à</p>
--	---	---	---

		<p>Créer un film d'une manip de classement pour identifier les phases chronophages et proposer des solutions.</p> <p>Création de raccourci clavier/souris pour accélérer le re-classement via le module de corr d'erreur (en particulier pour [Debris], [Other Phyto] et [Other Non Phyto]).</p> <p>Objectif de passer à moins d'1 heure.</p> <p>Voir pour créer plus de groupes « débris » sur la base de la teinte, de la forme (pas de la taille).</p> <p>A revoir dans contexte optim rephy : coût agent, équipt de tous les LERs, ajout fluo sur fastcam,...</p>	<p>avoir)</p> <p>PG : propose une option supplémentaire : un module pour récupérer les vignettes en fin d'analyse (dossiers et sous-dossiers contenant les prédictions à la manière d'un training set) → possibilité de classer les vignettes de [Other Phyto] et [Other Non Phyto] dans de nouveaux groupes.</p> <p>Action NNM</p> <p>UMons fin juin</p> <p>Concertation BL-NT-AR dans le cadre de l'optimisation des sets en cours</p> <p>Copil interne Ifremer</p>
<p>Pb des débris.</p> <p>Fluorescence pour discrim débris / non débris</p>			
<p>UI : Nb de vignettes par colonne</p>	<p>Nbre max de 200 vignettes par groupe dans l'interface.</p>	<p>Problème identifié par PG (évolution de la méthode)</p>	<p>UMons OK, à corriger pour fin juin.</p>

<p>Menus V5 : Analyse et fonctions permettant les traitements et intégration de nouvelles vignettes dans le set ne fonctionnent pas.</p>	<p>PG : ces défauts seront corrigés d'ici fin d'année ou fin juin ?</p> <p>PG : précise que ce pb avait été signalé par Guillaume qui avait précisé que les fonctions étaient accessibles par ligne de commande R. Objectif Ifr n'est pas d'utiliser des commandes R pour les utilisateurs lambda.</p>	<p>Les menus ne sont pas tout à fait à jour car principe de fonctionnement via des lignes de commande en phase de Beta test.</p>	<p>UMons OK, menu opérationnel pour fin juin.</p>
<p>Comptage cellule /colonie</p>	<p>GW : Phaeocystis et Chaetoceros socialis risquent de poser de gros pb.</p> <p>Ifr : Bien reçu rapport "comparaison comptage BL-NT"</p>	<p>Calcul de biovol des colonies peut se faire tout le temps mais pb de la fiabilité en lien avec ce qui se passe lors de la corr d'erreur (si débris ressemblent aux cellules alors OK, mais si hétérogénéité particules alors pb évaluation biovol).</p>	<p>AR doit livrer son dénombrement ASAP (NNM relance AR).</p> <p>Les nouvelles vignettes de colonies qui seront ajoutées aux sets doivent faire aussi l'objet de ce travail (BL-NT-AR)</p> <p><i>Hors réunion : J'ajouterais qu'il y a des actions identifiées dans le CR de GW suite à ce travail (CR à relire). Notamment les problèmes de sous ou sur-estimation des comptages manuels d'un site à l'autre.</i></p>
<p>CR mission Arcachon</p>	<p>Conclusions et plan d'action :</p> <p>Numérisation systématique : OK sauf pour éch dont TURB > 2FNU</p> <p>Traitement IU ZooPhytoImage : Impossible car trop chronophage (cf. questions et réflexions du CR)</p> <p>Comptage cellule colonie : Oui, quel délai attendu?</p> <p>Set : des corrections de</p>	<p>PG : Traçage des</p>	<p>PG : faisable</p>

	<p>classement à faire + un contrôle par différents membres du COPIIL à organiser/coordonner.</p> <p>Nécessité de création d'un catalogue des vignettes</p>	<p>corrections identifiable grâce à un outil training set 1 vs training set 2 modifié.</p> <p>projet fait par NN en cours de relecture et envoyé ce jour à PG et GW</p>	<p>Orientation du travail sur la base d'un catalogue de vignettes car plus facile par rapport à la sollicitation d'un expert et qui ne permettra pas de connaître les critères de l'expert pour faire les modifications</p> <p>PG : possibilité de créer un catalogue de façon automatique (Mons pour décembre 2015) → vignettes + critères de décision</p>
Travaux réalisés et/ou poursuivis	<p>BL et NT : numérisation systématique des échantillons (AR ?)</p> <p>BL et NT : poursuite optimisation set évalué avec un échantillon test (toujours le même)</p>		<p>Vérifier action AR suite à la formation</p> <p>PG : OK mais avec 1 seul éch on test toujours la même chose, donc en prendre un peu plus pour augmenter la représentativité de ce qui est à traiter comme éch.</p>
proposition	<p>Atelier technique à BL. Le flowcam de Nantes serait transporté à BL. Comparaison des techniques de numérisation. Calibration grâce à des billes calibrées (procédure ?)</p>	<p>Procédure de calibration existe à Mons ? Oui et non – déjà fait avec passage de billes utilisées pour cytométrie en flux avec la procédure de numérisation d'un échantillon + vérification au microscope.</p> <p>Flacons billes disponible à BL.</p> <p>Voir travail sur des cultures pour espèces</p>	<p>Ifr OK, organiser un doodle Ifremer/Mons/LOG : point logiciel + calibration en septembre 2015.</p> <p>Participants : NNM, AL, PhG, GW, FA ?</p>

		toxiques ?	
--	--	------------	--

+ Discussion sur la nécessité de revoir le « discours », la stratégie de mise en œuvre de l'outil FlowCam/ZooPhytoImage par rapport à ce que cet outil est réellement capable de fournir comme information (ouverture de la discussion sur le niveau de précision taxonomique à avoir pour faire une étude que ce soit au microscope comme avec le FlowCam/ZooPhytoImage). L'outil permet d'augmenter la fréquence/Nb d'éch analysés mais on diminue la résolution des identifications. PG rappelle que la plupart du temps, les données REPHY sont agrégées avant d'être exploitées.

PG : regrette de ne pas avoir été destinataire de tous les mails envoyés à Guillaume pour signaler les bugs. La boucle doit comprendre GW-PG-AL-RC-NMN au moins, ce doit être le cas à partir de maintenant. Commentaire PG : Catherine Belin étant responsable scientifique du projet officielle pour le contrat IFREMER-UMONS 2015, il faut qu'elle soit en copie également.

Dans le cadre de la collaboration IFREMER/UMONS, PhG souligne l'importance de l'échange des données acquises (numérisées et traitées) entre les différents partenaires, pour le développement et l'optimisation des méthodes de traitement.

Contrat Ifr/Umons 3 dates clés : fin mars fin juin et décembre (rapports d'avancement + livraison d'une version 6.0 de ZooPhytoImage contenant les modules de dénombrement des colonies, l'apprentissage actif, etc. + correction des bugs).

Livraison fin juin : nouvelle version avec nouveau manuel qu'on installe soit même.

PG rappelle que pour signaler un bug il faut un rapport comprenant un max de détails (copie/film d'écran en plus) et que l'Ifremer doit avoir une personne ressource maîtrisant ZooPhytoImage. AL : à mettre en place lorsque le logiciel sera opérationnel.

+++++

Surveillance du phytoplancton, des phycotoxines et des paramètres physico-chimiques associés

Action « Numérisation du phytoplancton et traitement par ZooPhytoimage »

Responsable scientifique : Alain Lefebvre

Comité de Pilotage :

- Ifremer : A. Lefebvre, C. Belin, N. Neaud-Masson, D. Maurer, F. Colas, J.F. Rolin, M. Repecaud, L. Lampert, R. Siano, M. Lunven, D. Soudant, E. Antajan
 - Université de Mons : Ph. Grosjean
 - Université du Littoral Côte d'Opale : D. Hamad (LISIC), F. Artigas (UMR LOG)
-

Lettre d'information N°1 – Avril 2015

Cette lettre d'information trimestrielle vise à informer le comité de pilotage des travaux en cours ou à venir. Elle peut également proposer des relevés de décisions ou identifier des actions à mener pour mener à bien une tâche dans un délai court, non compatible avec l'organisation semestrielle de la réunion du Comité de Pilotage.

Bilan des contributions à la lettre d'information reçu au 21/04/2015.

- Boulogne-sur-Mer :

* Réunion annuelle du GDR PhycoTox du 31/3 au 2/4, Brest (5 communications orales dont 2 en lien direct avec le projet, en collaboration intra-Ifremer, UMR LOG, Univ. Mons, ULCO/LISIC)

Artigas L.F., Bonato S., Créach V., Didry M., Gomez F., Guiselin N., Hamad D., Hébert P.A., Houliez E., Lefebvre A., Lampert L., Lizon F., Poisson-Caillault E., Prévost E., Rijkeboer M., Thyssen M., Veen A., Rutten T., Wacquet G., 2015. Monitoring of phytoplankton and Harmful Algal Blooms in coastal waters by combining innovative semi-automated tools (scanning flow cytometry & spectral fluorometry). Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Lefebvre A., Neau-Masson N., Maurer D., Wacquet G., Grosjean P., Colas F., Tardivel M., Artigas L.F., Belin C., 2015. Optimization of the monitoring strategy for the French National Phytoplankton and Phycotoxins Network (REPHY) using semi-automated digital images analysis. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Wacquet G., Grosjean P., Hamad D., Lefebvre A., Neaud-Masson N., Colas F., Maurer D., Artigas L.F., 2015. Zoo/PhytoImage : current advances in the semi-automated classification of plankton digital images. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

Lefebvre A., Poisson-Caillault E., Rousseeuw K., 2015. Spatio-temporal dynamics of phytoplankton biomass in the English Channel: high resolution strategy and modeling using unsupervised classification and Hidden Markov Model. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.

*Delegrange A., Lefebvre A., Amara R., Courcot L., Vincent D., 2015. Seasonal variability of *Pseudo-nitzschia* sp. and domoic acid concentrations in the southern bight of the North Sea. Conférence GDR PhycoTox / GIS Cyano, 31 mars – 2 avril 2015, IUEM, Brest.*

* Stage M2 FOGEM (ULCO) de Rémi Cuvelliez

- travail sur le set Manche (test des performances) et identification des besoins d'amélioration
- travail sur le dénombrement des colonies en lien avec G. Wacquet (fichiers transmis à GW pour suite à donner)
- Début du passage des échantillons REPHY/SRN 2014 en mode "classification" via la V5.

- Nantes :

* Poursuite du travail sur le dénombrement des colonies en lien avec G. Wacquet

* Début du passage des échantillons REPHY Bretagne 2015 fournis par LER/BN, LER/BO, LER/MPL/TM et LER/MPL/NT

Points : "Les Hebihens", "Concarneau large", "Loscolo" et "Le croisic (a)"

Ces 4 points sont échantillonnés 2x par mois et font l'objet d'une flore totale

* Formation à l'utilisation de la V5 à Arcachon (action : NNM).

- Brest

L'équipe FastCam est prête à poursuivre les tests avec passage d'échantillons qui devraient leur être fournis par les laboratoires participant au projet (contact : F. Colas pour la fourniture des échantillons).

- ULCO/LISIC

Un sujet de thèse devrait être déposé par D. Hamad. Il concerne la représentation parcimonieuse et l'apprentissage actif pour l'écologie marine.

- Univ. De Mons

Les travaux se poursuivent pour les sujets suivants :

- Dénombrement des cellules par colonies,
- Calculs biomasses/biovolumes et dénombrements dans l'analyse des échantillons Zoo/PhytoImage : intégration des données du point précédent, mais cela nécessite une adaptation des calculs,
- Apprentissage actif : mise au point d'une stratégie optimale sur base de séries d'échantillons rephy,
- Apprentissage actif : intégration de la nouvelle approche dans la V5 de Zoo/PhytoImage,
- Support utilisateur et correction de bugs,
- Collecte des données nouvelles concernant le dénombrement des colonies.

A noter :

- Participation de G. Wacquet à la réunion annuelle du GDR PhycoTox du 31/3 au 2/4, Brest données issues des différents labos.

COPIL de l'Action FlowCAM / ZooPhytoImage

4 novembre 2015 (Visioconférence Arcachon, Boulogne, Brest, Nantes)

Sommaire

Participants.....	2
I) État d'avancement.....	2
I.1. Le FastCAM.....	2
I.2. Le logiciel ZooPhytoImage.....	2
I.3. Bancarisation.....	4
II) Projets en cours et à venir.....	4
II.1. Projet H2020 JERICO-Next.....	4
II.2. CPER MARCO.....	4
II.3 CPER ROEC.....	5
II.4. Groupes « Mesures en continu » et jouvence Navire de recherche « Thalassa ».....	5
II.5. Autres.....	6
III) Livrables pour l'ONEMA.....	6
IV) Révision des objectifs et des stratégies.....	8
V) Comité de Pilotage.....	8

Compte rendu

Coordination : Alain Lefebvre

Contributeurs : Catherine Belin, Nadine Neaud-Masson, Philippe Grosjean, Florent Colas, Danièle Maurer

Participants

COPIL : Alain Lefebvre (Ifremer Boulogne), Philippe Grosjean (Université de Mons), Catherine Belin, Nadine Neaud-Masson (Ifremer Nantes), Danièle Maurer (Ifremer Arcachon), Florent Colas, Morgan Tardivel, Justine Evrard (Ifremer Brest).

Excusés : Felipe Artigas (LOG CNRS ULCO Wimereux), Denis Hamad (ULCO LISIC Calais), Luis Lampert, Raffaele Siano, Michel Lunven, Marie-Madeleine Danielou (Ifremer Brest),

Liste de diffusion

Participants

+ Membres du COPIL absents

+ Jérôme Paillet (ODE.Dir) René Robert (UL.Dir), Chantal Compère, Myriam Rumèbe, Claire Méteigner, Pascale Hébert, Camille Blondel, Cédric Bacher, Martin Plus, Antoine Hugué, Marie Pierre Crassous, tous responsables LERs.

I) État d'avancement

I.1. Le FastCAM

Le FastCAM permet de numériser un échantillon avec un grandissement X10 à une vitesse quinze fois supérieur au FlowCAM. Les calculs d'une cinquantaine de paramètres morphométriques sont alors réalisés par un logiciel développés par l'Ifremer qui fournit les fichiers zid et pid pour une classification avec les logiciels Zoo/phytoimage et PlanktonIdentifier, respectivement.

Pour une évaluation du FastCAM comme outil de numérisation pour la classification, un set d'apprentissage a été constitué. Toutefois, le taux d'erreur est encore de 40 % du fait de la sous-représentation de certaines classes. En effet, le set du REPHY permet une classification avec un taux d'erreur de 20 % mais il est beaucoup plus complet. Une comparaison rigoureuse FastCAM/FlowCAM comme outil de numérisation pour la classification n'est alors possible que par comparaison des images sur des espèces similaires. Ce travail sera mené fin 2015 / début 2016.

Si le FastCAM s'avère un meilleur outil que le FlowCAM, le problème reste la construction à l'échelle industrielle : la difficulté est de trouver un industriel qui ne craindrait pas un procès avec FluidImaging. L'autre solution serait une construction et une maintenance en interne par REM-RDT, mais cette solution nécessite l'aval de la hiérarchie. Ph. Grosjean pense qu'il serait possible de démarcher FluidImaging pour éviter la surenchère de leur part s'ils se sentent concurrencés : F. Colas va voir avec la direction de la valorisation si cette option est envisageable.

I.2. Le logiciel ZooPhytoImage

L'université de Mons a diffusé le 3 novembre 2015 la version 5.4.4. du logiciel ZooPhytoImage accompagné d'un manuel utilisateur rédigé en français.

Il est rappelé que les principaux contributeurs, financeurs et/ou scientifiques, de ce projet sont l'ONEMA, le BELSPO, l'Ifremer, L'université de Mons et l'Université du Littoral Côte d'Opale.

Les principales nouveautés de cette version sont :

- Refonte du code de ZooPhytoImage pour être compatible avec la dernière version de R (3.2.2) et sur toutes les plateformes supportées.

- Améliorations dans l'interface utilisateur, et ajout d'entrées de menu pour toutes les fonctionnalités nécessaires dans le cadre du REPHY.

- Incorporation du comptage de cellules par colonies (outils pour réaliser le training set spécifique de comptage + modélisation du dénombrement des cellules + utilisation dans le calcul des échantillons).

- La correction d'erreur est optimisée et est maintenant intégrée dans la validation des échantillons. Elle permet une diminution très significative du nombre de vignettes qui doivent être validées.

- Une nouvelle présentation graphique de l'état d'avancement de la validation permet de visualiser le travail restant et suggère des points d'arrêt selon deux critères: (1) une analyse satisfaisante de l'échantillon en utilisant la correction d'erreur, ou (2) la validation de la grande majorité des vignettes erronées (voir manuel en français, p. 24-25, par exemple).

- L'apprentissage actif est maintenant incorporé au processus. Lors de la validation d'un échantillon, l'utilisateur peut indiquer quels sont les échantillons validés antérieurement qui peuvent être utilisés pour offrir une amélioration de la classification initiale. La classification est également affinée de manière active à chaque étape de validation d'un même échantillon.

- Dans le cas d'échantillons contenant un très grand nombre de vignettes (> 5000), les fractions à validées sont maintenant contraintes à un nombre maximum d'items (500) (un argument permet de spécifier ce maximum pour pouvoir l'adapter au contexte si nécessaire). Ainsi, les très gros échantillons bénéficient de ce brigade combiné à l'apprentissage actif et à la correction d'erreur statistique pour limiter très, très fortement (jusqu'à un facteur 8-9) le nombre de vignettes qui doivent être validées.

- Un manuel d'utilisation en français et fortement orienté par le traitement IFREMER dans le cadre du REPHY est ajouté au manuel en anglais plus généraliste.

- Une série de bugs est corrigée.

A noter : la demande formulée par N. Neaud-Masson sur la possibilité de déplacer des groupes de particules (au lieu de le faire particule par particule) n'a pas pu être satisfaite, car elle nécessiterait l'intervention d'un spécialiste de JavaScript, or Mons a épuisé son budget pour 2015.

I.3. Bancarisation

Un Appel d'Offre a été lancé par Dyneco/Vigies (E. Gauthier et A. Huguet) afin de permettre d'engager le processus de bancarisation des données issues du système couplé FlowCam/ZooPhytoImage. La société Umanis a été retenue pour faire ce travail. A ce jour, il semble que l'utilisation de niveaux de bancarisation sera nécessaire : un niveau d'information bancarisé dans Quadrigé2 (paramètres intégrés de numérisation et des résultats de la classification) et un niveau d'information détaillé bancarisé dans un système différent (images, protocoles de numérisation et de classification,...).

II) Projets en cours et à venir

L'objectif de cette section est de recenser les projets, les études, ... en lien direct ou indirect avec l'analyse d'images en flux afin de faciliter notamment les prises de contacts, l'optimisation des montages de projets et des demandes de moyens humain et financier.

II.1. Projet H2020 JERICO-Next

Le projet H2020 JERICO-Next (New European eXpertise for coastal observatories) et plus particulièrement les axes 3 et 4 sont dédiés, respectivement, (i) au développement de technologies et de méthodologies avec, d'intérêt particulier pour notre groupe, un projet orienté vers le développement et à la mise en œuvre de plateforme automatisée pour l'observation du phytoplancton et (ii) des cas d'étude dans des zones ateliers (1 zone atelier en Manche Mer du Nord). Il s'agira plus particulièrement de contribuer à l'amélioration des méthodes d'étude de l'eutrophisation, des blooms d'algues nuisibles et de la diversité du plancton. Ce projet s'inscrit dans le cadre des besoins d'amélioration des connaissances en lien avec la mise en œuvre de la DCSMM (2010/56/EU).

Plus d'information via <http://wwz.ifremer.fr/L-institut/Actualites/Reseau-europeen-d-observation-cotiere-JERICO-s-acheve-JERICO-NEXT-prend-la-releve>

II.2. CPER MARCO

Le CPER MARCO (Recherches marines et littorales en Côte d'Opale : des milieux aux ressources, aux usages et à la qualité des produits aquatiques) est un projet structurant multi-laboratoires, multi-organismes associant la mise en place d'instruments et d'outils (enquêtes, indicateurs) pour une approche globale de l'étude du milieu marin, de la ressource et de la qualité des produits aquatiques. Pour répondre à ces enjeux académiques et sociétaux, le projet s'articule autour de 6 axes :

- 1- Observation et évaluation de l'environnement marin (coordination A. Lefebvre, Ifremer et F. Artigas, UMR LOG).
- 2- Structure, fonctionnement et dynamique des écosystèmes
- 3- Productivité et durabilité des ressources halieutiques et aquacoles
- 4- Qualité et sécurité des ressources aquatiques

5- Vulnérabilité et usages des éco-socio-systèmes marins et littoraux

6- Ingénierie marine et littorale

Ce projet fédère :

- 5 partenaires institutionnels : ULCO, Université de Lille 1, CNRS, IFREMER, Anses
- 7 Laboratoires partenaires : LOG, LISIC, TVES, IFREMER-Laboratoire Environnement et Ressources, IFREMER-Laboratoire Comportement des Structures en Mer, IFREMER-Laboratoire Ressources Halieutiques, Laboratoire de sécurité des aliments, site de Boulogne sur Mer, Anses
- Des partenaires associés : PFI Nouvelles Vagues, Parc naturel Marin Estuaires picards et mer d'Opale, Nausicaa, Agence de l'eau Artois Picardie, Institut Viollette.

Le projet est en cours d'instruction et devrait débuter en 2016.

II.3 CPER ROEC

Le CPER ROEC (Réseau d'Observation pour l'Environnement Côtier) vise à étudier l'état environnemental des masses d'eau à long terme avec un focus sur l'oxygène, les nutriments et la diversité du compartiment planctonique. Le contexte générale vise à approfondir les connaissances sur les sujets suivants : évolution climatique, eutrophisation, efflorescences algues toxiques (contact : guillaume.charria@ifremer.fr).

Dans le cadre du CPER ROEC un système de concentration de cellules et d'imagerie en flux automatisé, dérivé du ZooCAM et du FastCAM, doit être développé et installé sur un ferry. Pour éviter les colmatages de la fluide, la profondeur de la cuve sera limitée à 300 μm , impliquant un grandissement maximum X4. Un premier prototype du système d'imagerie a été déployé et testé lors de la campagne IPARO sur les Côtes de la Manche qui a eu lieu en Bretagne Sud et mer d'Iroise en septembre.

II.4. Groupes « Mesures en continu » et jouvence Navire de recherche « Thalassa »

Dans le cadre de la modernisation de Thalassa (travaux prévus au printemps 2017 et cahier des charges à finaliser pour mars 2016), il a été décidé de constituer un groupe de travail concernant les « mesures en route ». Briec Crenan (Ifremer, Service Navires et Equipements) a pour mission d'animer ce groupe de travail (GT). Les objectifs de ce GT sont :

- choix des capteurs à implanter à bord du navire océanographique *Thalassa* afin d'effectuer des mesures en continu,
- définition des contraintes d'intégration, de calibration, de maintenance à prendre en compte
- définition des modes opératoires devant être suivis

Cette réflexion pourra ensuite être étendue dans un second temps aux autres navires océanographiques.

Les systèmes FlowCam et FastCAM ayant été déployés lors de campagnes de prélèvements en mer, le CoPil portera une attention particulière aux propositions de ce groupe quant aux possibilités de mesures pérennes via ces systèmes.

Le lien est également fait avec les besoins du Programme de Surveillance de la DCSMM en terme de besoins de mise en œuvre de ce type de systèmes de mesures (semi) automatisées.

A. Lefebvre, membre de ce GT et pilote scientifique DCSMM (Eutrophisation et Habitats Pélagiques), assurera le lien avec le CoPil.

II.5. Autres

Puisque la convention Ifremer / ONEMA se termine en 2015, le CoPil a envisagé les différentes possibilités de maintien de la dynamique en cours.

L'année 2016 devrait s'avérer être une année de transition lors de laquelle il faudra envisager la préparation d'un certain nombre d'écrits permettant le dépôt de projet d'études et de recherches, de dossiers de demande de financement de post-docs et de thèses.

Des pistes de financements ont été évoquées lors du CoPil : appels à projets Ifremer « Big Data » et « politique de site ». A voir avec la Direction Scientifique de l'Ifremer (contact : P. Farcy).

Il apparaît également nécessaire de formaliser la poursuite des collaborations entre l'Ifremer, l'Université de Mons et l'Université du Littoral Côte d'Opale via une ou plusieurs conventions.

Le CoPil évoque la possibilité d'engager en 2016 des stages de fin d'études dans le contexte des besoins DCSMM et JERICO-Next. C. Belin, F. Colas contacteront, respectivement, J.P. Lecomte et L. Delauney afin d'évoquer les possibilités de financement et d'accueil.

Un entretien spécifique avec la direction du département Ifremer ODE est à programmer (action C. Belin & A. Lefebvre) afin d'envisager les perspectives de cette action.

A propos des versions de ZooPhytoImage : la dernière version livrée à Ifremer est la 5.4-4, elle appartient conjointement à Mons-Ifremer-ULCO, et n'est donc pas publique. Ph. Grosjean propose que la version valide à la fin du projet devienne publique (mais ce n'est pas obligatoire). A noter que la dernière version publique est la 3.1.

III) Livrables pour l'ONEMA

Les différents rapports devront être livrés à l'ONEMA en février 2016.

Les documents attendus sont :

Version évolutive de l'outil opérationnel de numérisation et d'analyse semi-automatique d'images de phytoplancton, utilisant le matériel FlowCAM et le logiciel ZooPhytoImage. Nouvelles perspectives	<i>A faire par Univ. Mons + LOG Wimereux</i>
Mise en œuvre opérationnelle de l'outil FlowCAM / ZooPhytoImage dans le cadre de la surveillance REPHY.	<i>A faire par Ifremer ODE (BL + VIGIES + AR)</i>
Evolution du matériel de numérisation : prototype FastCAM et perspectives	<i>A faire par Ifremer REM RDT</i>

Pour le livrable 1, les documents déjà fournis sont déjà conséquents :

- ZOO/PHYTOIMAGE VERSION 5.4-0. Analyse d'Images de Plancton Assistée par Ordinateur. MANUEL UTILISATEUR. Septembre 2015. L'équipe de développement de Zoolmage : Ph. Grosjean, K. Denis & G. Wacquet : Écologie Numérique des Systèmes Aquatiques, UMONS, Belgique. X. Irigoien, G. Boyra & I. Arregi : AZTI Tecnalia, Espagne. A. Lopez-Urrutia : Centro Oceanográfico de Gijón, IEO, Espagne. M. Sieracki & B. Tupper (FlowCAM plugin). 32 p.
- Projet FlowCAM/ZooPhytoImage. Rapport d'avancement. 30/03/2015. Guillaume WACQUET & Philippe GROSJEAN. 60 p.
- Projet FlowCAM/ZooPhytoImage. Rapport d'avancement. 30/06/2015. Guillaume WACQUET & Philippe GROSJEAN. 19 p.

Pour le livrable 2, les documents déjà fournis sont les suivants :

- Formation à FlowCAM / ZooPhytoImage v5. Compte-rendu mission formation Nadine Neaud à Arcachon, du 31 mars au 3 avril 2015.
- Compte-rendu de l'audioconférence du 28/05/2015. FlowCAM / ZooPhytoImage – Test de la V5. Ifremer Nantes & Boulogne – Université de Mons
- Lettre d'information N°1 – Avril 2015. Alain Lefebvre

Il faudra y ajouter :

- Les descriptifs des numérisations effectuées à Boulogne, Nantes et Arcachon
- Les conclusions de l'étude bancarisation
- Les liens avec la DCSMM

Le livrable 3 est à faire.

IV) Révision des objectifs et des stratégies

Au vu de l'état d'avancement du système de numérisation FastCAM et de la version 5.4.4. du logiciel ZooPhytoImage, il est décidé de revoir les objectifs et les stratégies de l'action FlowCam/ZooPhytoImage aussi bien vis à vis des besoins du REPHY que des besoins des projets d'études et de recherche.

Un brouillon sera mis en circulation au sein du CoPil par A. Lefebvre et C. Belin afin de permettre l'écriture d'un document de référence visant à clarifier les possibilités réelles offertes par le système couplé FlowCam ou FactCAM / ZooPhytoImage en l'état actuel de leurs développements.

V) Comité de Pilotage

Les membres présents du CoPil souhaitent maintenir en l'état le fonctionnement de coordination et d'animation scientifique de l'action FlowCam/ZoophytoImage via le CoPil.

Il est proposé d'envisager le retrait de la participation de M. Repecaud (Ifremer Brest), qui devra cependant être destinataire des comptes-rendus au titre de son implication sur les systèmes instrumentés.

Il est proposé d'ajouter Elvire Antajan (Ifremer Boulogne) et Guillaume Charria (Ifremer Brest) comme membres permanents pour leurs compétences et/ou rôles, respectivement, sur l'analyse d'images pour le zooplancton et pour la coordination du dossier de labellisation Haute Fréquence HOSEA.

Danièle Maurer informe le CoPil de son départ programmé à la retraite en avril 2016. Son remplacement au sein du CoPil devra alors être envisagé afin de permettre la représentation du laboratoire d'Arcachon.