

Mise en œuvre 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

Rapport d'activités de post-doctorat et état d'avancement des travaux (octobre 2011)

Mise en œuvre 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

Rapport d'activités de post-doctorat et état d'avancement des travaux (octobre 2011)



Fiche documentaire

| | |
|--|---|
| Numéro d'identification du rapport : RST/LER/AR/11/002 Diffusion : libre <input checked="" type="checkbox"/> restreinte: <input type="checkbox"/> interdite : <input type="checkbox"/> Validé par : Isabelle Auby ODE-LER-LERAR Benoît Beliaeff RBE-ADMI Catherine Belin ODE-DYNECO-VIGIES Nicolas Chomérat ODE-LER-LERFBN Jean-Claude Cochard ODE-LER Philippe Grosjean ECONUM -UMONS Roger Kantin ODE-LER-LERAR | Date de publication : Octobre 2011 Nombre de pages : 60 + 8 annexes Bibliographie: oui Illustration(s): oui Langue du rapport : Français Rapport définitif <input checked="" type="checkbox"/> |
| Titre et sous-titre du rapport : <p style="text-align: center;"><i>Mise en œuvre 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</i> <i>Rapport d'activités de post-doctorat et état d'avancement des travaux (octobre 2011)</i></p> | |
| Auteur(s) principal(aux) : Alina Tunin-Ley Danièle Maurer Collaborateur(s) : Nadine Neaud Myriame Rumèbe-Perrière Claire Barbier | Organisme / Direction / Service, laboratoire Laboratoire Environnement Ressources d'Arcachon |
| Résumé : Le projet FlowCAM/PhytoImage a pour objectif, à terme, la mise en œuvre opérationnelle d'analyses de la composition phytoplanctonique d'échantillons d'eau de mer grâce à une analyse automatisée, ou semi-automatisée, en utilisant un FlowCAM couplé à une version spécifique du logiciel PhytoImage, développé par l'Université de Mons en Belgique. Ce rapport rend compte des premières phases de la réalisation du projet, effectuées dans le cadre d'un post-doctorat, en constante interaction avec l'équipe de l'UMONS. Les images issues de la numérisation d'échantillons, principalement prélevés dans le bassin d'Arcachon de janvier 2010 à mars 2011, ont permis de constituer des banques d'images des particules du bassin d'Arcachon, principalement de la flore totale phytoplanctonique (matériel frais et fixé au Lugol) mais aussi d'autres catégories de particules non ciblées (zooplancton et particules inertes). Puis un set d'apprentissage remanié (comportant 44 catégories de particules dont 28 phytoplanctoniques) a permis de construire un outil de reconnaissance optimisé dont les performances atteignent un niveau de reconnaissance global de 72% pour l'ensemble des catégories du set. Ce résultat très encourageant n'est pas définitif puisque de nombreuses améliorations peuvent être apportées. Les prochaines étapes à mener pour déterminer sous quelle forme la mise en œuvre de cette méthode pourrait être utilisée en routine pour le suivi des Flores Totales du REPHY du littoral français sont un enrichissement du set d'apprentissage initial par de nouveaux groupes phytoplanctoniques et la résolution des biais quantitatifs mis en évidence sur certains taxons. | |
| Cadre de la recherche : Programme : PJ058 Nouvelles technologies pour la surveillance Projet : A050807 FlowCAM / PhytoImage | Conventions et contrats de collaboration : - IFREMER-ONEMA (2009) (action n°11 : Mise au point de méthodes pour la surveillance du phytoplancton – Volet 2) - Contrat de collaboration IFREMER-UMONS (2008) - Contrat de collaboration IFREMER-UMONS (2010) |

SOMMAIRE

| | |
|---|-----------|
| SOMMAIRE | 7 |
| Remerciements | 11 |
| Contexte | 13 |
| 1. Introduction | 16 |
| 1.1. Présentation générale du système couplé FlowCAM/Phytolmage | 16 |
| 1.1.1. Le FlowCAM..... | 16 |
| 1.1.2. Le logiciel Phytolmage | 17 |
| 1.2. Cahier des charges du post-doctorat | 19 |
| 2. Etapes préliminaires et protocoles de traitement et de numérisation des échantillons | 21 |
| 2.1. Etapes préliminaires..... | 21 |
| 2.1.1. Formation..... | 21 |
| 2.1.2. Intercalibration..... | 21 |
| 2.1.3. Travaux préliminaires..... | 21 |
| 2.2. Protocoles de traitement et de numérisation des échantillons | 22 |
| 3. Les sets d'apprentissage Flore Totale..... | 26 |
| 3.1. Matériel et méthode..... | 26 |
| 3.1.1. Matériel | 26 |
| 3.1.2. Banque d'images | 26 |
| 3.1.3. Set d'apprentissage | 29 |
| 3.2. Résultats | 30 |
| 3.2.1. Banques d'images | 30 |
| 3.2.2. Set d'apprentissage Flore Totale Lugol | 31 |
| 3.2.3. Set d'apprentissage Flore Totale Vivant | 33 |
| 4. Les outils de reconnaissance..... | 35 |
| 4.1. Méthodes d'optimisation et d'évaluation des outils de reconnaissance (UMONS) | 35 |
| 4.1.1. Optimisation logicielle | 35 |
| 4.1.2. Calcul des taux d'erreur | 35 |
| 4.1.3. Matrice de confusion..... | 36 |
| 4.2. Résultats | 36 |
| 4.2.1. Optimisation logicielle (UMONS) | 36 |
| 4.2.2. Statistiques de la classification automatique calculées sur le set d'apprentissage Flore Totale Lugol de mars 2011 | 37 |
| 4.2.3. Statistiques de la classification automatique calculées sur le set d'apprentissage Flore Totale Vivant..... | 40 |
| 5. Evaluation des performances de l'outil de reconnaissance Flore Totale Lugol | 41 |
| 5.1. Performances de classification automatique par l'outil de reconnaissance | 41 |
| 5.1.1. Matériel et méthode | 41 |
| 5.1.2. Résultats | 41 |
| 5.2. Performances de l'estimation automatique de l'abondance par l'outil de reconnaissance | 44 |
| 5.2.1. Matériel et méthode | 44 |
| 5.2.2. Résultats de l'estimation automatique de l'abondance..... | 44 |

| | |
|---|------------|
| 6. Discussion..... | 47 |
| 7. Bilan et perspectives | 51 |
| 7.1. Bilan | 51 |
| 7.2. Perspectives..... | 52 |
| 7.2.1. Enrichissement du set d'apprentissage et création d'un outil de reconnaissance « Atlantique-Manche » | 52 |
| 7.2.2. Nouveaux outils dans Phytolmage..... | 52 |
| 7.2.3. Utilisation de l'assemblage objectif 4X/cellule de flux 300µm..... | 53 |
| 7.2.4. Utilisation du set d'apprentissage 10X sur des images numérisées en 4X | 54 |
| Conclusion | 55 |
| Communications..... | 57 |
| Documents internes Ifremer | 58 |
| Bibliographie..... | 59 |
| Annexes | 61 |
| Annexe 1 – Proposition de sujet de post-doctorat pour une demande de bourse post-doctorale | 63 |
| Annexe 2 – Travail préliminaire sur le genre <i>Dinophysis</i> (Tunin-Ley, 2009) | 75 |
| Annexe 3 – Protocole FlowCAM pour la constitution du set d'apprentissage REPHY (flore totale) (Tunin-Ley, 2009) | 85 |
| Annexe 4 – Proposition de protocole pour l'utilisation du FlowCAM et la constitution d'un set d'apprentissage à partir des échantillons du REPHY (juillet 2009) (Tunin-Ley, 2009)..... | 93 |
| Annexe 5 – Traitement des échantillons avant et pendant leur passage au FlowCAM..... | 109 |
| Annexe 6 – Numérisation au FlowCAM des échantillons du REPHY | 115 |
| Annexe 7 – Illustration des différents groupes du set d'apprentissage Flore Totale Lugol..... | 133 |
| Annexe 8 - Nouvelles fonctionnalités du logiciel Phytolmage (version 2) (UMONS)..... | 159 |

Remerciements

Cette étude a pu être réalisée grâce aux contributions financières de l'ONEMA, du Conseil Régional d'Aquitaine et du FEDER.

Les auteurs tiennent à remercier tout particulièrement Ph. Grosjean, K. Denis du Laboratoire d'Ecologie Numérique des Milieux Aquatiques de l'Université de Mons (Belgique), pour leur aide précieuse et cette collaboration fructueuse.

Merci aussi à B. Sautour et Y. Del Amo, de l'Université de Bordeaux 1 (Laboratoire EPOC) pour leur coopération qui a permis de finaliser ce travail.

Nous remercions enfin toutes les personnes de l'IFREMER qui ont contribué à la réalisation de cette étude :

Catherine Belin, Michel Marchand, Nicolas Chomérat, Elisabeth Nézan ainsi que l'équipe REPHY du LER d'Arcachon : Nadine Neaud, Myriam Rumèbe-Perrière et Claire Barbier.



Contexte

Le REPHY, réseau de surveillance du phytoplancton et des phycotoxines mis en place à l'Ifremer en 1984, a pour but d'observer les communautés phytoplanctoniques des eaux côtières françaises en surveillant plus particulièrement les espèces phytoplanctoniques produisant des toxines. Cette approche globale des populations est fondamentale pour situer dans leur contexte les événements toxiques. Ce réseau a aussi permis de constituer une base de données des communautés phytoplanctoniques sans équivalent en France, potentiellement riche de renseignements sur les impacts du changement climatique ou les pressions anthropiques en milieu côtier. En outre, l'aspect environnemental de cette surveillance permet de répondre aux obligations de la Directive européenne Cadre sur l'Eau (DCE), et aux conventions internationales telles que la convention d'Oslo et de Paris (OSPAR) ou la convention de Barcelone. Ces données sont également essentielles à la réalisation de divers programmes de recherche (en particulier les programmes liés à l'activité conchylicole).

Afin de réaliser ce suivi, des échantillons sont prélevés dans de nombreux sites côtiers répartis sur tout le littoral, permettant l'acquisition à long terme de données sur la biodiversité, la distribution spatiale et l'évolution temporelle du phytoplancton sur les côtes françaises. L'échantillonnage est réalisé toute l'année, à une fréquence variant d'une fois par semaine à une fois par mois. Les dénombrements des taxons phytoplanctoniques¹ sont effectués sur un échantillon d'eau de mer de 10 mL, au microscope photonique inversé selon la méthodologie d'Utermöhl (1958). L'identification se fait au niveau taxinomique le plus bas possible (espèce ou genre), sinon à un rang supérieur (famille, voire ordre ou classe), éventuellement à un niveau intermédiaire (groupe de genres ou d'espèces). Lorsque l'ensemble de la communauté microphytoplanctonique est analysé, les résultats sont mentionnés sous le terme de « Flore Totale » dans le REPHY². Une liste de référence des taxons potentiellement identifiables est disponible dans la base de données Quadrige.

La méthodologie d'Utermöhl (1958) est jusqu'à présent communément utilisée pour la détermination et la quantification des espèces phytoplanctoniques. Elle nécessite une formation approfondie en taxinomie, mais le niveau de compétence des opérateurs s'avérant différent, certains taxons peuvent leur être inconnus ou faire l'objet d'identifications erronées. En outre, l'observateur étant sujet à la fatigue et la déconcentration, des erreurs non régulières et non quantifiables peuvent être commises. Le changement d'observateur peut également biaiser les résultats.

L'optimisation récente des procédés de reconnaissance automatique des particules permet aujourd'hui d'envisager de détecter, de dénombrer et d'identifier automatiquement les cellules phytoplanctoniques, tout en produisant un taux d'erreur quantifiable et reproductible. Cette technique nouvelle ouvre aussi la perspective d'un gain de temps conséquent. A l'Ifremer, les dénombrements au microscope sont actuellement réalisés par des analystes formés à la taxinomie, répartis dans les différents Laboratoires

¹ Il s'agit des espèces dont la taille est supérieure à 20 µm (microphytoplancton), ou qui sont en colonies.

² Par différenciation avec les « Flores Indicatrices Partielles » qui ne concernent que les espèces dont les effectifs sont supérieurs à 100 000 cellules/L et certaines espèces toxiques.

Environnement littoral et Ressources aquacoles (LER). L'automatisation de ce suivi permettrait donc d'alléger le temps alloué aux analyses au microscope, tout en permettant à ce personnel qualifié de réaliser sa tâche avec une pertinence accrue.

C'est dans ce contexte qu'une première étude de faisabilité a été confiée au laboratoire d'Ecologie Numérique des Milieux Aquatiques (Pr. Philippe Grosjean) de l'Université de Mons (Belgique), concernant le logiciel PhytoImage (version 1) développé par ce laboratoire. Cette étude a permis de conclure qu'il était possible d'utiliser le logiciel PhytoImage pour l'automatisation du suivi des communautés phytoplanctoniques, moyennant une évolution de cet outil (Denis et Grosjean, 2006).

Ces résultats prometteurs ont conduit à l'élaboration du projet FlowCAM/PhytoImage dont l'objectif, à terme, est la mise en œuvre opérationnelle d'analyses de la composition phytoplanctonique d'échantillons d'eau de mer grâce à une analyse automatisée, ou semi-automatisée, en utilisant un FlowCAM³ couplé à une version spécifique du logiciel PhytoImage développé par l'Université de Mons. Ce projet a débuté en 2009, en collaboration avec deux partenaires extérieurs, l'ONEMA et l'Université de Mons, avec lesquels les relations sont régies par les documents suivants.

- **La convention IFREMER-ONEMA (2009)** (action n°11 : Mise au point de méthodes pour la surveillance du phytoplancton) comporte deux volets dont le second (Reconnaissance et dénombrement automatisé d'espèces phytoplanctoniques) précise le mode de financement d'un post-doctorat d'une durée de 18 mois, à part égales IFREMER et ONEMA, pour commencer à réaliser l'objectif défini ci-dessus. L'échéance de la convention était fixée mi-2010. Trois rapports ont été remis à l'ONEMA pour rendre compte de l'état d'avancement des travaux (Tunin-Ley, 2009 ; Grosjean et Denis, 2010 ; Tunin-Ley, 2010).
- **Le contrat de collaboration IFREMER-UMONS (2008)** explicite les conditions de la collaboration entre l'IFREMER et l'Université de Mons (Laboratoire d'Ecologie Numérique des Milieux Aquatiques, Pr. Philippe Grosjean). Les engagements de l'UMONS concernent la formation d'agents IFREMER à l'utilisation du FlowCAM, une intercalibration de divers appareils, l'adaptation et l'extension du logiciel PhytoImage pour son utilisation dans le cadre du REPHY (version 2), et enfin le développement et l'optimisation d'un premier outil de reconnaissance du phytoplancton sur la base des images fournies par l'IFREMER. Les échéances étaient fixées fin juin 2010.
- Un nouveau contrat de collaboration (**Contrat de collaboration IFREMER-UMONS, 2010**) a été signé en décembre 2010 pour parachever les travaux entrepris. Il s'agit de la mise au point de la version 3⁴ du logiciel PhytoImage (affinage de l'interface utilisateur et ajout de nouveaux traitements statistiques) et de la fourniture d'un outil de reconnaissance de seconde génération. Les résultats obtenus dans le cadre de la collaboration sont la copropriété de l'IFREMER, de l'UMONS et de la Politique Scientifique de Belgique, à parts égales.

³ Le FlowCAM (Flow Cytometer And Microscope) est le seul matériel de ce type disponible actuellement sur le marché ; il est commercialisé par la société Fluid Imaging (USA).

⁴ Cette version spécifique du logiciel sera fournie sous licence GPL et sera librement utilisable par l'Ifremer sous les termes de cette licence : en particulier le code source complet sera fourni et librement utilisable par l'Ifremer et par ses partenaires.

Le présent rapport rend compte des premières phases de la réalisation du projet, effectuées dans le cadre du post-doctorat en constante interaction avec l'équipe de l'UMONS, et présente un bilan des outils disponibles et de leurs perspectives d'évolution à l'issue de cette étude. Le travail développé par l'équipe belge a fait l'objet d'un rapport spécifique remis à l'ONEMA (Grosjean et Denis, 2010). Seuls leurs principaux résultats seront repris ici. Pour plus de détails concernant les méthodologies, le lecteur pourra se reporter directement à ce document.

Dans l'introduction, une présentation générale des outils utilisés sera faite, car elle est nécessaire à la compréhension des objectifs et des étapes du travail de post-doctorat exposés ensuite.

1. Introduction

1.1. Présentation générale du système couplé FlowCAM/PhytoImage

1.1.1. Le FlowCAM

Le FlowCAM (Flow Cytometer And Microscope) combine deux technologies, la cytométrie en flux d'une part et la microscopie photonique reliée à une caméra numérique d'autre part (Figure 1). Tout d'abord, l'échantillon d'eau de mer à analyser est attiré par une pompe péristaltique dans une cellule passante en verre de dimensions parfaitement connues (chambre de flux), positionnée devant un objectif de microscope lui-même relié à une caméra vidéo. Cette caméra réalise une acquisition d'image des particules selon les paramètres préalablement définis : soit suite à la détection par un faisceau laser, soit à intervalles de temps réguliers. Avec le laser, il est possible de détecter la particule elle-même, ou bien la fluorescence des particules (deux canaux, un pour la chlorophylle, un pour la phycoérythrine). A chaque particule détectée selon les paramètres initialement spécifiés, correspond une photo, à laquelle est associée un ensemble de 21 mesures (surface, diamètre, etc).

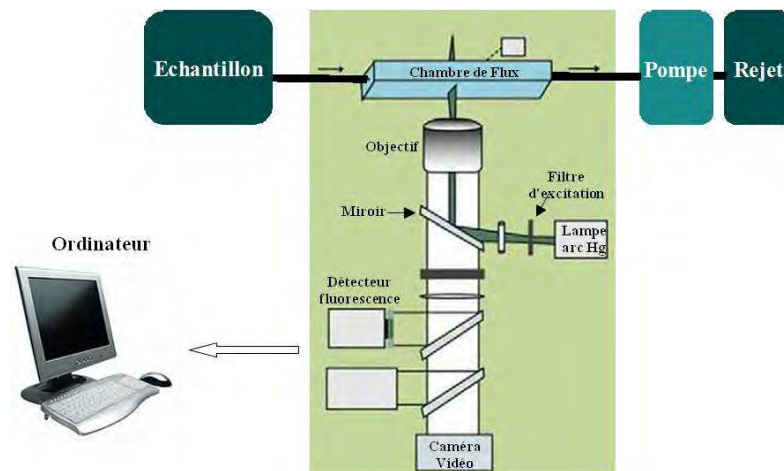


Figure 1. Principe de fonctionnement du FlowCAM.

Le FlowCAM utilisé au LER Arcachon est équipé d'une caméra noir et blanc haute résolution⁵ et de 3 objectifs de grossissement 10X, 4X et 2X, qui s'utilisent respectivement avec des cellules en verre calibrées de profondeur 100 μm pour le premier, 300 μm ou 600 μm pour le deuxième, et 800 μm ou 1mm pour le dernier. Ces assemblages spécifiés par le constructeur garantissent une profondeur de champ idéale pour une netteté optimale des images des particules dans le flux.

⁵ 10224 (H) x 768 (V) FireWire Progressive Scan.

Le logiciel de paramétrage et d'exploitation des données acquises à l'aide du FlowCAM, Visual SpreadSheet, est fourni avec l'appareil. Il permet de définir les paramètres d'acquisition et de mesures des images et propose également un outil de tri par filtres et de classification des particules, qui ne permet cependant pas d'obtenir directement une liste complète d'espèces à un niveau hiérarchique donné.

Les principaux paramètres de réglage du logiciel Visual SpreadSheet à définir avant une numérisation, concernent essentiellement :

- l'illumination de la partie de la cellule calibrée (intensité, homogénéité lumineuse),
- les paramètres vidéos d'acquisition d'images (taille, nombre d'images par seconde, sauvegarde des fichiers de sortie...),
- les paramètres de détection des particules (filtres de taille, critères de teinte...).

1.1.2. Le logiciel Phytolmage

Le logiciel PhytoImage est un logiciel d'analyse d'images et de classification automatique. Se basant sur le principe du « *machine learning* », il permet de réaliser les différentes étapes du processus qui conduit à la classification automatisée, ou semi-automatisée, d'un ensemble d'objets, à partir d'un jeu d'images donné (Figure 2). L'étape préliminaire indispensable est celle de l'apprentissage à la reconnaissance des particules étudiées. Cet apprentissage (à travers une interface) se fait en deux phases.

- La première phase concerne la réalisation d'un set d'apprentissage 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 souhaités⁶.

- La seconde phase de la réalisation d'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. Pour que cet apprentissage soit optimal, plusieurs étapes sont nécessaires :

- choisir le meilleur algorithme possible,
- optimiser les (parfois nombreux) paramètres de cet algorithme sur la base de validations croisées ayant pour objectif de quantifier et de maximiser le taux de reconnaissance global, voire le taux de reconnaissance spécifique d'un ou plusieurs groupes taxinomiques (selon l'objectif visé),
- étudier l'effet du regroupement progressif des taxons en vue d'obtenir un meilleur compromis entre niveau de détail et performances de classification,

⁶ La méthodologie en sera développée dans la suite du rapport (cf. 3.2.).

- enfin, évaluer les performances obtenues, en particulier le taux d'erreur entre différents taxons, la robustesse par rapport à des échantillons nouveaux, le nombre de particules identifiées par unité de temps, etc.⁷.

Une fois l'outil de reconnaissance automatique optimisé, celui-ci peut alors être utilisé sur un jeu d'images nouvelles pour classer automatiquement les particules dans les catégories qui auront été définies manuellement dans le set d'apprentissage initial.

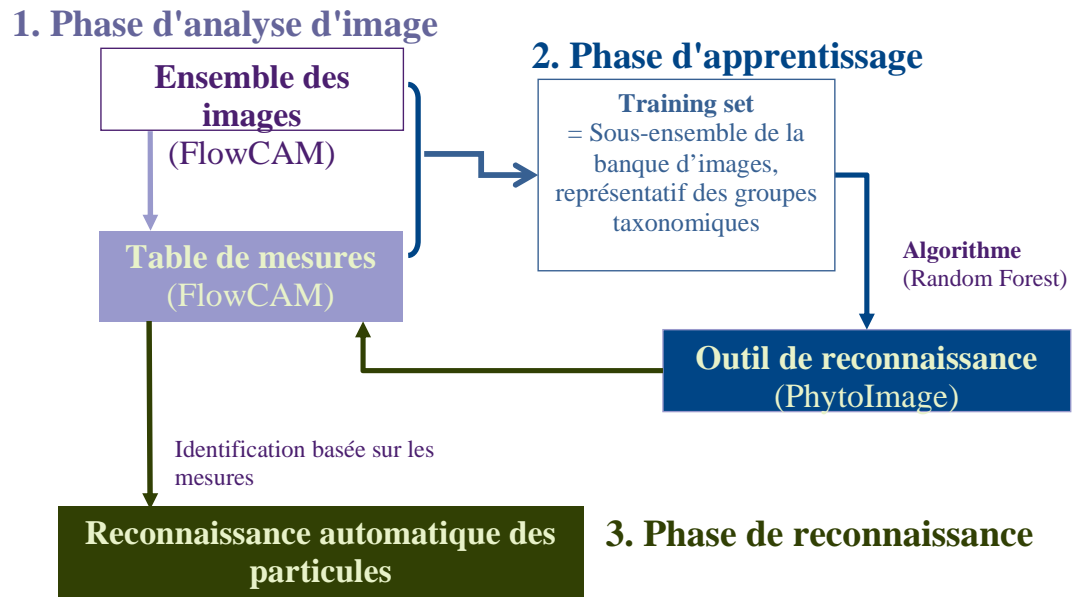


Figure 2. Principe du « *machine learning* » (ici avec l'algorithme « Random Forest ») et des images issues du FlowCAM).

Le logiciel PhytoImage offre la possibilité d'analyser les images acquises par tout type d'outil (microscopie photonique, FlowCAM, ZOOSCAN...). Un plug-in spécifique a été développé pour traiter directement les images issues du FlowCAM et analysées dans un premier temps par le logiciel Visual SpreadSheet. Après exportation des images du FlowCAM, le logiciel PhytoImage mesure 38 paramètres supplémentaires, ce qui permet d'obtenir au final 59 paramètres mesurés pour chaque particule. Par ailleurs, une vignette avec échelle de taille est créée pour chaque particule, afin de faciliter la reconnaissance des particules par un observateur (Figure 3).

⁷ Ainsi, la réalisation d'un outil de reconnaissance optimisé requiert une expérience approfondie en « *machine learning* » (expérience que possède le laboratoire d'Ecologie Numérique des Milieux Aquatiques de l'Université de Mons).

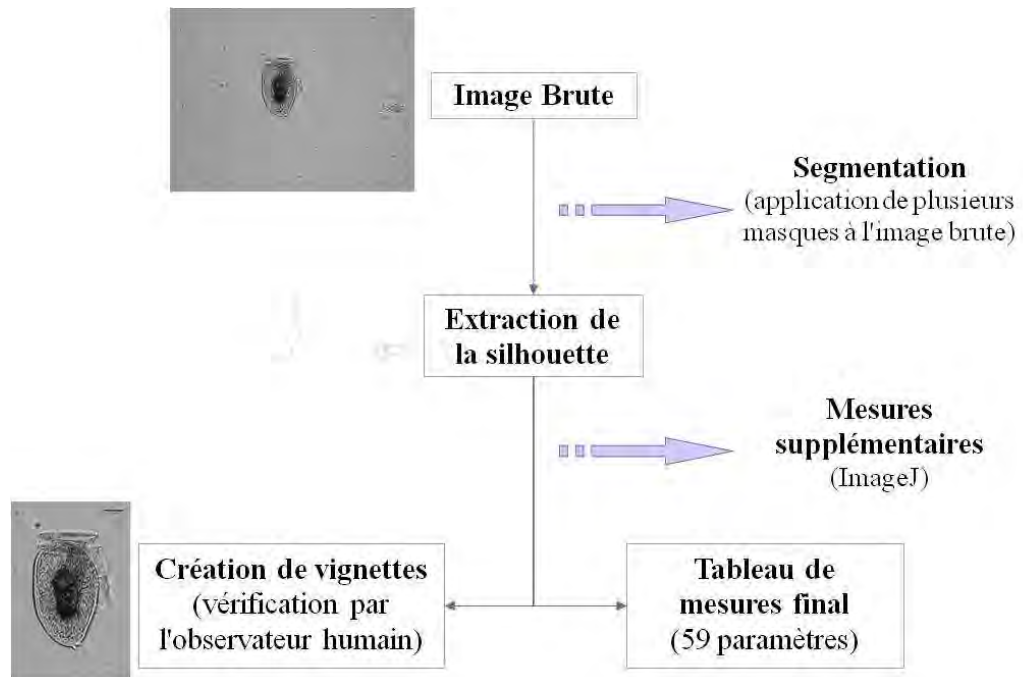


Figure 3. Principe de l'analyse d'images par le logiciel PhytoImage.

1.2. Cahier des charges du post-doctorat

Les premières phases de la réalisation du projet présenté en introduction ont été menées par la post-doctorante Alina TUNIN-LEY, au sein du LER Arcachon où est positionné le FlowCAM⁸. La responsabilité scientifique du post-doctorat a été confiée à Danièle Maurer (LER/Arcachon) et Nicolas Chomérat (LER/Concarneau).

Des problèmes techniques survenus au début du post-doctorat puis fin 2009 ont exigé à deux reprises le renvoi temporaire du FlowCAM chez le constructeur aux Etats-Unis pour réparation. Un retard de 4 mois a ainsi été pris, se répercutant sur le déroulement et les échéances des différentes étapes. Par la suite, une convention a pu être passée avec l'Université de Bordeaux1 pour permettre de mener à leur terme les tâches initialement définies. La reprise de ce travail, pour une durée de 6 mois, n'a pu se faire qu'en octobre 2010, après à une interruption de 3 mois. L'échéance finale a ainsi été reportée à mars 2011.

L'objectif premier du post-doctorat était de déterminer si un niveau de précision taxinomique identique à celui des identifications au microscope par un observateur pouvait être atteint avec la méthode FlowCAM/PhytoImage, sur la flore du bassin d'Arcachon. Il était demandé que ce travail soit réalisé à la fois sur des échantillons fixés (Lugol) et sur des échantillons frais. Dans un deuxième temps, une première évaluation des performances sur le plan quantitatif devait aussi être menée de façon à préparer la suite de la réalisation du projet.

⁸ Le FlowCAM, qui appartient légalement à l'Université de Bordeaux 1, est installé de façon permanente dans les locaux de l'IFREMER selon la convention liant les deux organismes (Convention IFREMER-UB1, 2009).

Les différentes étapes successives à mener au cours de ce travail ont été globalement définies dans la demande de post-doctorat (Annexe 1). Elles sont présentées de façon détaillée ci-dessous.

- Formation et participation à une expérience d'intercalibration de différents FlowCAM.
- Demandes de nouveaux modules dans PhytoImage pour l'adaptation de son utilisation dans le cadre du REPHY.
- Proposition d'une stratégie pour la constitution du set d'apprentissage et rédaction des premiers modes opératoires concernant l'utilisation de l'appareil et la numérisation des échantillons⁹.
- Constitution d'une banque d'images de la flore totale¹⁰ présente dans le bassin d'Arcachon, à partir des prélèvements de phytoplancton effectués dans le cadre du REPHY, au cours d'un cycle annuel, sur des échantillons frais et fixés au Lugol¹¹ ; parallèlement, constitution d'une banque d'images de phytoplancton de Concarneau (fixé au Lugol) à partir d'échantillons considérés comme intéressants du fait de leur richesse spécifique et de leur abondance.
- Réalisation de deux sets d'apprentissage de la flore totale présente dans le bassin d'Arcachon (un set à partir d'images acquises sur du matériel fixé au Lugol et l'autre sur du matériel vivant).
- Contribution au remaniement de ces sets d'apprentissage et à l'optimisation des outils de reconnaissance créés à partir de ces sets (en liaison avec l'UMONS) (cf. 1.1.2.)¹². Evaluation des outils.
- Mise au point et rédaction du mode opératoire final pour le passage des échantillons au FlowCAM. Réalisation d'une première analyse¹³ d'échantillons du REPHY fixés au Lugol recueillis pendant quelques mois avec le dernier outil de reconnaissance et évaluation des performances de l'outil.
- Formation des observateurs phytoplancton du LER d'Arcachon, au passage et traitement des échantillons REPHY au FlowCAM.

Le plan du rapport reprendra le déroulement de ces différentes phases. Les méthodes utilisées, constituant elles-mêmes une partie des résultats de l'étude, seront détaillées à chacune des étapes pour plus de clarté.

⁹ Echéance initiale prévue été 2009, réalisation septembre 2009 du fait d'une première panne sur le FlowCAM..

¹⁰ Au sens défini ci-dessus, note 1.

¹¹ Cette phase devait initialement débiter à l'été 2009 mais a été retardée à septembre 2009 puis à janvier 2010 du fait des deux pannes successives sur le FlowCAM.

¹² Ce travail a été réalisé à l'échéance prévue de juin 2010 sur des sets n'incluant pas un cycle annuel (6 mois de prélèvements). Un nouveau set plus complet a été constitué par la suite en ajoutant des images issues de prélèvements réalisés de juillet 2010 à mars 2011 (soit au final 15 mois de prélèvements). L'outil de reconnaissance correspondant a été évalué.

¹³ Echéance initiale prévue fin juin 2010, réalisation mars 2011 du fait des deux pannes successives du FlowCAM et de l'interruption de contrat de 3 mois.

2. Etapes préliminaires et protocoles de traitement et de numérisation des échantillons

De façon à rendre compte de l'ensemble des activités menées dans le cadre du post-doctorat, une présentation des étapes préliminaires menées est rappelée ici.

2.1. Etapes préliminaires

2.1.1. Formation

La formation à l'utilisation du FlowCAM et du logiciel PhytoImage a été dispensée par l'équipe de l'Université de Mons, par échanges directs lors de trois réunions d'une semaine (en février, avril et juillet 2009), soit à Mons, soit à Arcachon, ainsi que par courriel et par téléphone. Cette formation approfondie a permis une prise en main des outils, nécessitant la maîtrise des nombreux paramètres de réglage, et l'adoption d'habitudes de travail similaires à celles de l'équipe belge.

En outre, la formation s'est poursuivie pendant la phase de mise au point du protocole de numérisation, grâce à la réalisation de travaux-tests (cf. 2.1.3.).

2.1.2. Intercalibration

Une session d'intercalibration entre les deux FlowCAM, celui basé au LER Arcachon et celui de l'université de Mons, était prévue initialement pour s'assurer de la compatibilité de traitement des images acquises par les deux appareils.

Cette semaine d'intercalibration a eu lieu du 27 avril au 1er mai 2009 à Arcachon. Elle a concerné trois appareils, celui du LER Arcachon, et ceux de l'université de Mons et de l'AZTI (Espagne). Cette expérience a permis de démontrer que les FlowCAM étaient compatibles entre eux, c'est-à-dire que les images acquises par différents appareils étaient traitables par le même outil de reconnaissance, du moment qu'une calibration des appareils (en particulier niveaux de gris) et un protocole commun étaient appliqués. Les résultats de cette étude, présentés lors d'un colloque¹⁴ par K. Denis, font l'objet d'une publication qui sera prochainement soumise à *Journal Of Plankton Research*, pour laquelle les collaborateurs IFREMER impliqués dans le projet sont co-auteurs.

2.1.3. Travaux préliminaires

Avant de définir les modes opératoires et la stratégie de constitution des sets d'apprentissage, une prise en main des outils FlowCAM et PhytoImage s'est avérée nécessaire. Des travaux préliminaires ont été menés afin de permettre cet apprentissage. Ils ont été effectués sur des échantillons fixés au Lugol issus de campagnes scientifiques

¹⁴ ICCS 2009, Baton Rouge, Louisiana, U.S.A. May 25-27, 2009.

(immédiatement disponibles car stockés en laboratoire) ou de cultures¹⁵ (Tunin-Ley, 2009).

Les premiers essais ont concerné le genre *Ceratium* (dinoflagellé planctonique). La gamme de taille élevée des espèces de ce genre a permis d'apprendre à utiliser leFlowCAM avec l'objectif 4X et les cellules de flux associées (profondeur 300 µm), qui sont d'un usage moins délicat que celui des cellules de flux (profondeur 100 µm) associées à l'objectif 10X. Par ailleurs, la grande variabilité morphologique de ce genre semblait intéressante pour tester la capacité de discrimination de l'algorithme¹⁶ à un niveau taxinomique détaillé.

Le second essai a concerné le genre *Dinophysis*. En effet, plusieurs espèces appartenant à ce genre, et se développant au large du bassin d'Arcachon de manière saisonnière et récurrente, sont régulièrement détectées en fortes abondances à l'intérieur du bassin. Les réglages propres à l'utilisation de l'objectif 10X et des cellules de flux associées (largeur 100 µm), ont pu être testés.

Ces travaux ont aussi permis une prise en main du logiciel PhytoImage et des modules permettant la constitution de sets d'apprentissage et d'outils de reconnaissance ainsi que l'évaluation des performances de ces outils. Ils ont également permis d'aborder la catégorisation des particules dans le cadre de la constitution du set d'apprentissage, et plus particulièrement la définition des catégories de débris.

Les résultats obtenus, issus des sets d'apprentissage tests, montrent déjà des capacités de discrimination très intéressantes au niveau spécifique pour le genre *Dinophysis* (Annexe 2) et même entre variétés infra-spécifiques pour le genre *Ceratium* (Tunin-Ley, 2009). Ultérieurement, un outil spécifique *Dinophysis*, permettant la discrimination des différentes espèces, pourrait être développé si le besoin en était exprimé.

NB : Les échantillons utilisés n'ayant pas été traités, avant leur numérisation, selon le protocole de conservation fixé ultérieurement (cf. 2.2.), les images qui en étaient issues n'ont pas pu être reprises dans le set d'apprentissage final.

2.2. Protocoles de traitement et de numérisation des échantillons

L'expérience acquise au cours de la réalisation des travaux-tests, a permis la mise au point d'un protocole pour le traitement et la numérisation des échantillons du REPHY **dans le but de constituer un set d'apprentissage** (Annexe 3). Le détail des différentes options testées et analysées, ainsi que les raisons ayant conduit aux choix retenus, sont présentés en Annexe 4. Ce protocole ne prenait cependant pas en compte l'aspect quantitatif des numérisations car l'ensemble des conditions à respecter afin d'assurer une analyse à la fois qualitative et quantitative des particules d'un échantillon ne pouvait être étudié qu'avec l'outil de reconnaissance optimisé.

¹⁵ Echantillons de cultures fournis par le laboratoire EMP/PHYC de Nantes (Véronique Séchet).

¹⁶ L'algorithme « *Random Forest* » a été utilisé car c'était celui qui donnait les meilleurs résultats dans l'étude préliminaire de Denis et Grosjean (2006).

C'est pourquoi la mise au point du mode opératoire final de numérisation des échantillons du REPHY a été effectuée dans la dernière phase de ce travail. Il comprend deux procédures, l'une décrivant le traitement des échantillons phytoplancton avant et pendant leur passage au FlowCAM, l'autre le protocole d'utilisation du FlowCAM pour la numérisation des échantillons du REPHY (Annexes 5 et 6¹⁷).

Les points principaux du mode opératoire défini sont les suivants.

Afin de garantir la compatibilité entre elles des images numérisées par le FlowCAM, il est nécessaire de respecter un protocole précis de traitement et conservation des échantillons. Ainsi une concentration finale en Lugol de 1% (au lieu de 0,1% dans le protocole du REPHY), une conservation des échantillons au réfrigérateur dans des flacons en verre et un délai de passage au FlowCAM inférieur à trois mois ont été retenus afin d'éviter une éventuelle décoloration du Lugol et une dégradation des cellules.

L'assemblage composé de l'objectif de **grossissement 10X** associé à la cellule de **flux calibrée de 100 µm** d'épaisseur a été choisi pour la numérisation des échantillons, afin de bénéficier de la meilleure qualité possible d'image. En effet, l'observation de certains détails (épines, soies, flagelles, etc.) permet l'identification visuelle des taxons par le taxinomiste à un niveau voisin de celle obtenue au microscope en routine. Les plus petits organismes sont également mieux reconnus.

Le **mode *AutoImage***, définissant l'acquisition d'un nombre fixe d'images par intervalle de temps, a été privilégié par rapport au mode *Trigger* (qui utilise le laser pour le déclenchement de l'acquisition d'image), car le premier permet d'avoir une estimation très précise de la concentration cellulaire, contrairement au second (Figures 4a et 4b). En effet, grâce à l'épaisseur connue de la cellule de flux, le volume correspondant au champ de la caméra peut être calculé et ne varie pas. Dès lors, le volume final numérisé correspond au volume d'un champ que multiplie le nombre total d'images acquises (Figure 4a). Ainsi, des numérisations de même durée en mode *AutoImage* correspondent à des volumes imagés identiques. En mode *Trigger*, le temps écoulé entre deux acquisitions d'image est variable, puisque l'acquisition est déclenchée par le passage d'une particule devant le laser en mode *Trigger Scatter* (Figure 4b), et par le passage d'une particule naturellement fluorescente (contenant suffisamment de pigments chlorophylliens) en mode *Trigger Fluorescence*. Pour obtenir une estimation fiable du volume final numérisé, il faut donc connaître très précisément le débit du flux.

¹⁷ Ces documents ont été rédigés selon le modèle de document Qualité du LER Arcachon, ce qui permettra une intégration ultérieure rapide au système Qualité du laboratoire.

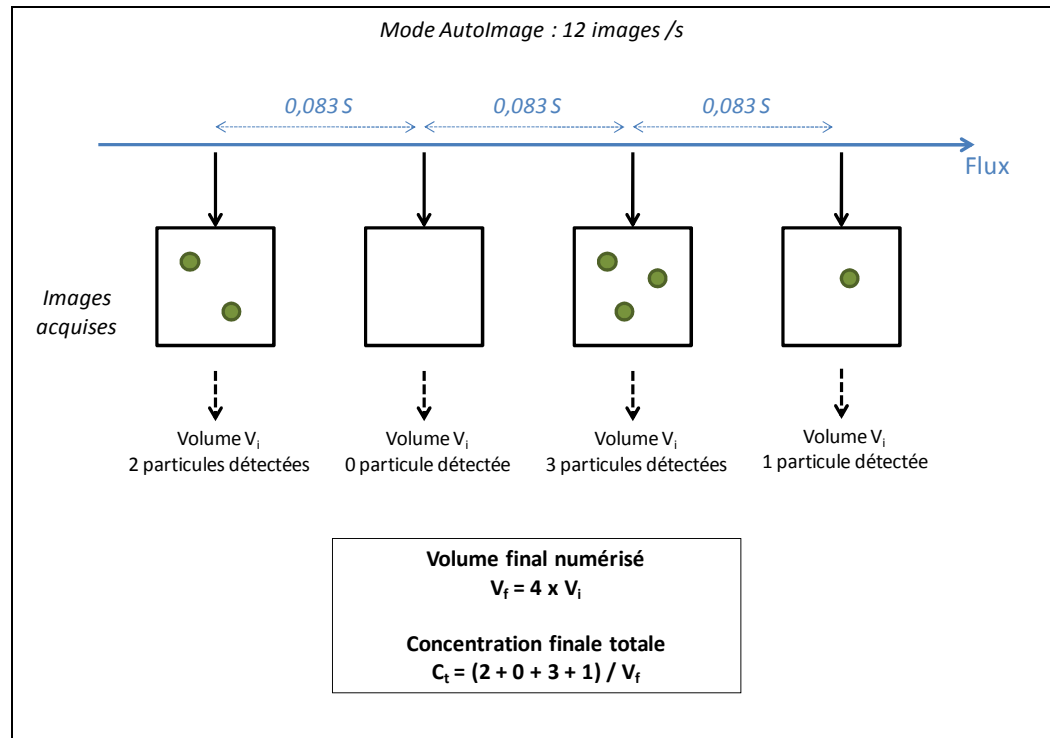


Figure 4a. Schématisation du calcul de la concentration finale en particules en mode *AutoImage*.

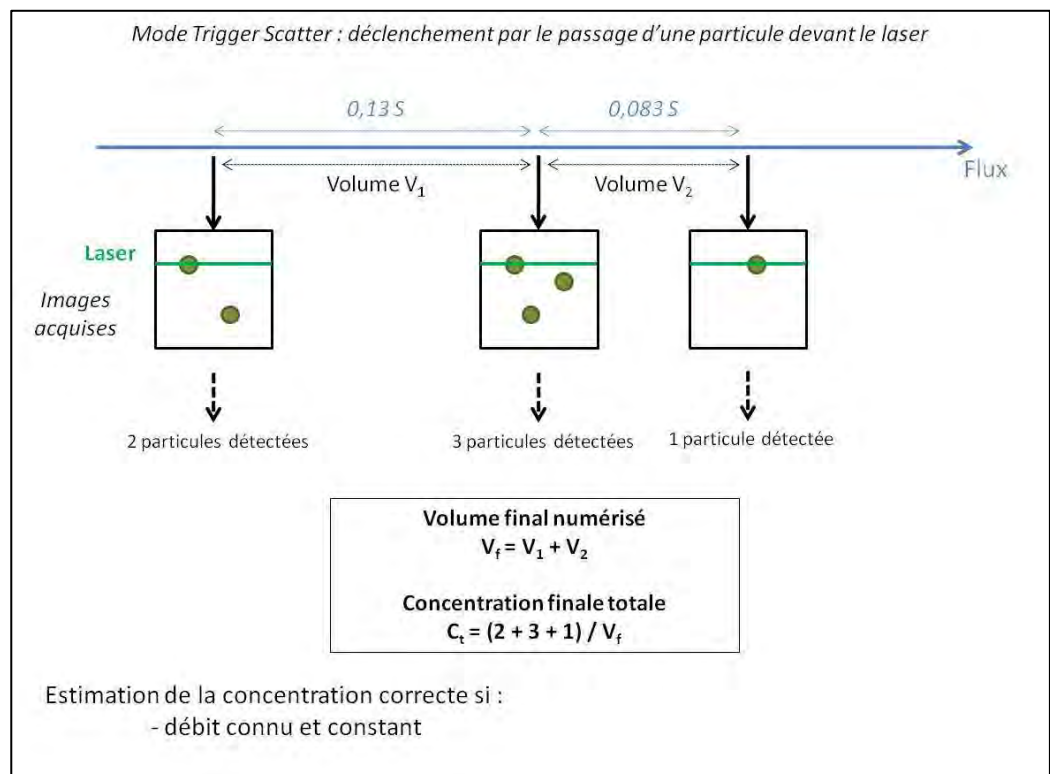


Figure 4b. Schématisation du calcul de la concentration finale en particules en mode *Trigger Scatter*.

La pompe péristaltique intégrée au FlowCAM basé à Arcachon ne permet pas d'obtenir cette précision¹⁸. Dans le cas d'échantillons fixés au Lugol, il n'est pas possible d'utiliser le mode **Trigger en détection de fluorescence** car le Lugol détruit la fluorescence naturelle des cellules chlorophylliennes. La détection de type **Scatter** (toutes les particules dans la gamme de taille qui passent devant le laser déclenchent l'acquisition d'images) pourrait en revanche être utilisée mais avec une pompe de précision.

Un filtre de taille en mode ABD (Area Based Diameter) définissant une plage comprise entre 15 µm et 2000 µm a été appliqué afin de garantir la détection des particules du phytoplancton qui sont habituellement dénombrées par les analystes du REPHY, dont la gamme de taille est comprise entre 20 µm et quelques mm pour les plus grands taxons (colonies de diatomées, *Ceratium* spp., *Noctiluca*, *Phaeocystis*).

Les modalités pratiques de passage de l'échantillon au FlowCAM explicitées en Annexe 6, décrivent en particulier :

- le montage et le démontage des différents composants du FlowCAM (cellules de flux et composants associés, objectifs optiques, systèmes de tubes),
- la manipulation de l'échantillon à numériser,
- les réglages optiques,
- les paramètres d'acquisition à compléter ou vérifier,
- les fonctions du logiciel VisualSpreadsheet à utiliser.

¹⁸ En revanche, l'utilisation d'une pompe additionnelle plus précise, qui pourrait facilement être montée en dérivation, est envisageable.

3. Les sets d'apprentissage Flore Totale

3.1. Matériel et méthode

3.1.1. Matériel

Les échantillons qui ont été numérisés pour la constitution des banques d'images et des sets d'apprentissage, ont été prélevés dans le bassin d'Arcachon entre janvier 2010 et mars 2011 (soit une soixantaine de prélèvements), de manière à caractériser le maximum d'espèces habituellement présentes dans les eaux du Bassin au cours d'un cycle annuel. Ils correspondent aux prélèvements bimensuels à hebdomadaires du REPHY effectués sur les stations Bouée 7 et Teychan. Les échantillons pauvres en espèces ou en abondance n'ont pas toujours été numérisés.

Les échantillons frais ont été numérisés sur une période plus courte (1^{er} semestre 2010), l'interruption du contrat en juin 2010 ayant eu pour conséquence l'arrêt définitif de ces numérisations.

Des échantillons du REPHY fixés au Lugol de l'année 2010, provenant du LER Concarneau, ont été également numérisés. Ils ont été retenus en fonction de leur richesse spécifique et de leur abondance en certains taxons (en particulier de dinoflagellés).

3.1.2. Banque d'images

Une fois l'échantillon numérisé et les images traitées par le logiciel PhytoImage pour obtenir des vignettes, celles-ci font l'objet d'une identification visuelle par un taxinomiste pour constituer une banque d'images. Les vignettes identifiées sont classées manuellement pour former des catégories (ou groupes) soit planctoniques soit de débris ou autres particules.

Dans le cas présent, les catégories phytoplanctoniques des banques d'images ont été constituées de manière à suivre au plus près la liste de taxons utilisée par les analystes REPHY du LER Arcachon. En d'autres termes, la catégorisation des particules a été réalisée de la manière la plus détaillée possible (résolution taxinomique élevée) et la plus exhaustive possible (maximisation du nombre de groupes au détriment du nombre d'images par groupes), comme lors d'un dénombrement au microscope en routine. L'arborescence ainsi constituée contient actuellement 150 groupes taxinomiques (hors débris) dont 137 phytoplanctoniques et 9 zooplanctoniques (Figure 5).

Dans la pratique, les vignettes issues d'une numérisation sont tout d'abord placées dans une catégorie intitulée « non encore classés ». Puis le taxinomiste repère et répartit les vignettes qu'il peut identifier visuellement dans les groupes taxinomiques prédéfinies du REPHY Arcachon. Si le groupe n'existe pas, il est alors créé. Toutes les vignettes ne sont pas placées dans la banque d'images, car elles ne sont pas toutes reconnaissables et/ou retenues (par exemple groupes déjà complets).

| PHYTOPLANKTON | | |
|------------------|--------------------------------|--|
| Diatomées | | |
| Centrales | | |
| | Asterolampraceae | |
| | Asterolampraceae_spp | |
| | Asteromphalus | |
| | Biddulphiaceae | |
| | Cerataulina | |
| | C.pelagica | |
| | Eucampia | |
| | Hemiaulus | |
| | Odontella | |
| | O.mobiliensis | |
| | Odontella_spp | |
| | Triceratium | |
| | Centrales_spp | |
| | Chaetoceraceae | |
| | Bacteriastrum | |
| | Chaetoceros | |
| | Attheya | |
| | C.curvisetus_pseudo | |
| | C.danicus | |
| | C.decipiens | |
| | C.peruvianus | |
| | C.socialis_radians_perpusillus | |
| | Chaetoceros_spp | |
| | Corethron | |
| | Leptocylindraceae | |
| | Leptocylindrus | |
| | L.danicus | |
| | L.minimus | |
| | Lithodesmiaceae | |
| | Ditylum | |
| | D.brightwellii | |
| | Helicotheca | |
| | Lithodesmium | |
| | L.undulatum | |
| | Melosiraceae | |
| | Melosira | |
| | Paralia | |
| | Rhizosoleniaceae | |
| | Dactyliosolen | |
| | D.fragilissimus | |
| | Guinardia | |
| | G.delicatula | |
| | G.flaccida | |
| | G.striata | |
| | Proboscia | |
| | P.alata | |
| | P.alata_indica | |
| | Rhizosolenia | |
| | R.imbricata | |
| | R.robusta | |
| | R.setigera_pungens | |
| | Rhizosoleniaceae_spp | |
| | Striatella | |
| | Thalassiosiraceae | |
| | Lauderia_Schroderella | |
| | Skeletonema_costatum | |
| | Thalassiosira | |
| | T.antartica_gravida | |
| | T.decipiens | |
| | T.levanderi_minima | |
| | Thalassiosira_cells | |
| | Thalassiosira_spp | |
| Pennales | | |
| | Cymatosiraceae | |
| | Fragilariaceae | |
| | Asterionellopsis | |
| | A.glacialis | |
| | Thalassionema | |
| | T.nitzschioides | |
| | Licmophora | |
| | Naviculaceae | |
| | Diploneis | |
| | Haslea | |
| | Meuneria | |
| | M.membranacea | |
| | Naviculaceae_spp | |
| | Pleuro_Gyrosigma | |
| | G.fasciola | |
| | Pleuro_Gyrosigma_spp | |
| | Nitzschiaceae | |
| | Cylindrotheca | |
| | Nitzschia | |
| | N.longissima | |
| | Pseudo_nitzschia | |
| | Pseudo_nitzschia_fine | |
| | Pseudo_nitzschia_spp | |
| | Pennales_spp | |
| | Stauroneis | |
| | Dinoflagellés | |
| | Dinoflagellés cuirassés | |
| | Dinophysales | |
| | Dinophysis | |
| | D.acuminata | |
| | D.acuta | |
| | D.caudata | |
| | D.fortii | |
| | D.tripos | |
| | Dinophysis_spp | |
| | Phalacroma | |
| | P.rapa | |
| | P.rotundata | |
| | Ebriales | |
| | Hermesinum | |
| | Goniomonadales | |
| | Alexandrium_spp | |
| | Gonyaulacales | |
| | Ceratium | |
| | C.candelabrum | |
| | C.furca | |
| | C.fusus | |
| | C.gibberum | |
| | C.kofoidii | |
| | C.lineatum | |
| | C.tripos | |
| | Ceratium_spp | |
| | Gonyaulax | |
| | G.polygramma | |
| | G.spinifera | |
| | Gonyaulax_spp | |
| | Lingulodinium | |
| | Gymnodiniales | |
| | Amphidinium | |
| | Gymnodiniales_spp | |
| | Gyrodinium | |
| | Katodinium | |
| | Polykrikos | |
| | Torodinium | |
| | Noctiluciales | |
| | Pronoctiluca | |
| | Peridinales | |
| | Amphidoma | |
| | A.caudata | |
| | Diplopsalis | |
| | Peridinales_spp | |
| | Protoperidinium | |
| | P.bipes | |
| | P.depressum | |
| | P.diabolus | |
| | Protoperidinium_spp | |
| | Scrippsiella | |
| | Prorocentrales | |
| | P.gracile | |
| | P.micans | |
| | P.minimum | |
| | P.triestinum | |

| PHYTOPLANCTON (suite) | ZOOPLANCTON |
|-------------------------|----------------------|
| Dictyochophycées | Ciliés |
| Euglenophycées | Myrionecta |
| Cryptophycées | M. rubra |
| Prymnesiophycées | Ciliés nus |
| Coccolithophorales | Prostomes |
| Phaeocystis | Tintinnides |
| Raphidophycées | Xystonella |
| Heterosigma akashiwo | Copépodes |
| Chlorophycées | Larves |
| Scenedesmus | Œufs |
| Staurastrum | Foraminifères |
| | Gastéropodes |
| | Radiolaires |
| | Véligères |

Figure 5. Arborescence phytoplancton et zooplancton utilisée pour la banque d'images REPHY Arcachon.

A partir des banques d'images, des sets d'apprentissages peuvent être construits. Ils correspondent à des sous-ensembles de ces banques d'images.

3.1.3. Set d'apprentissage

L'expérience acquise lors de la réalisation des travaux préliminaires (Tunin-Ley, 2009) sur les genres *Dinophysis* et *Ceratium* a permis d'identifier plusieurs points importants à prendre en compte pour l'élaboration d'un set d'apprentissage (Annexe 4).

Les principales conclusions sont les suivantes.

- Le nombre d'images par groupe taxinomique doit être de **100 au minimum**. Si certains groupes (de forme très constante) sont très bien reconnus avec un petit nombre d'images, parfois une douzaine seulement, dans tous les cas, on observe une très nette amélioration des taux de reconnaissance lorsque le nombre d'images par catégorie est supérieur à 100. Cependant, ce seuil peut être plus élevé lorsque la variabilité morphologique des cellules est importante, comme c'est le cas pour une partie des diatomées, notamment pour les espèces coloniales. Les images du set d'apprentissage doivent ainsi représenter de manière significative la variabilité naturelle d'un taxon donné, que celle-ci soit intrinsèque (taille des cellules, état physiologique, forme des extensions, longueur des colonies), ou extrinsèque (position de la cellule par rapport à l'objectif au moment de l'acquisition d'image, netteté de l'image) (Figure 6).

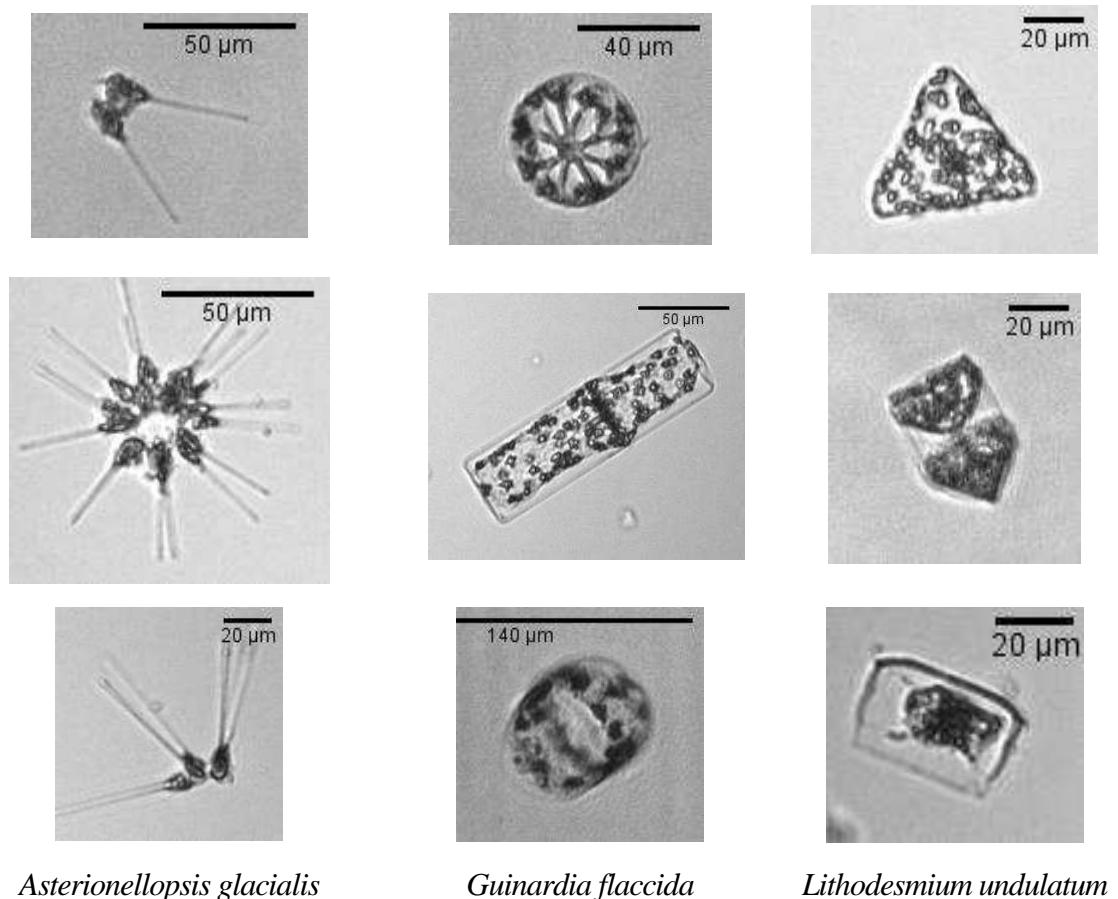


Figure 6. Exemples de la diversité des images du set d'apprentissage pour différentes espèces phytoplanctoniques.

- Idéalement, il faudrait acquérir **un nombre équivalent d'images entre les groupes**. En effet, les groupes dotés du plus grand nombre de vignettes sont aussi les mieux reconnus. 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.

- Il est nécessaire de constituer un set d'apprentissage le plus **détaillé** possible au départ, même si des regroupements sont effectués ultérieurement. Pour les groupes non ciblés¹⁹, il est préférable de créer des catégories assez larges, n'ayant pas forcément de cohérence d'un point de vue taxinomique (cela s'applique particulièrement pour la création d'un outil de reconnaissance spécialisé).

- Les débris sont souvent une source de confusion importante avec les cellules phytoplanctoniques de petite taille. Il est donc primordial de soigneusement définir ces catégories. Une **catégorisation des débris selon leur teinte (niveau de gris)**, même si elle est forcément subjective et imprécise pour l'observateur humain, semble être la plus pertinente (Figure 7). En revanche, une catégorisation selon la forme ou la texture donne rarement de bons résultats. Il peut également s'avérer nécessaire de regrouper avec les débris d'autres particules, comme certains groupes du zooplancton.

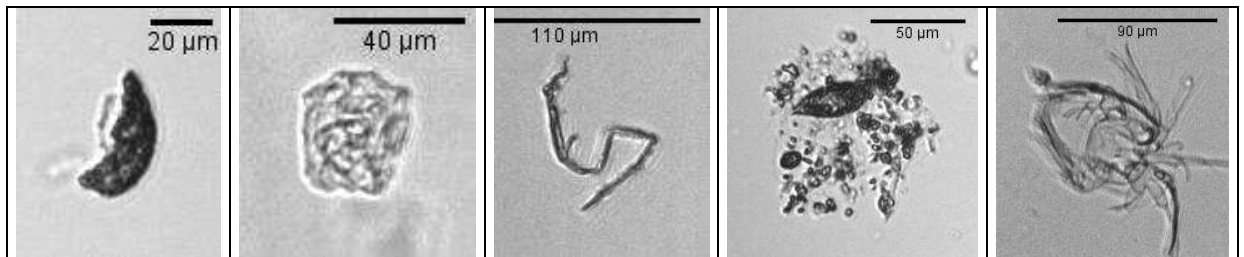


Figure 7. Exemples de débris rencontrés dans les échantillons.

C'est pour ces raisons que, lors du travail d'optimisation des outils de reconnaissance, il a été décidé de ne fournir à l'UMONS que des sets d'apprentissage comportant des groupes représentés par 30 images au minimum, car en dessous de ce seuil, les performances de classification automatique ne sont en général pas probantes. Pour certaines diatomées coloniales posant problème pour la reconnaissance automatique, de 100 à 300 images ont été fournies.

3.2. Résultats

3.2.1. Banques d'images

Trois banques d'images ont été constituées.

La première, réalisée à partir d'échantillons du bassin d'Arcachon fixés au Lugol (dénommée "Flore Totale Lugol"), contient plus de 6 000 vignettes dont plus de 4 000 pour les taxons phytoplanctoniques, classés dans 115 catégories, dont 62 de diatomées et 44 de dinoflagellés.

¹⁹ C'est-à-dire non phytoplanctoniques.

La seconde banque d'images, réalisée à partir d'échantillons frais du bassin d'Arcachon (dénommée "Flore Totale Vivant"), contient plus de 3 300 vignettes dont plus de 2 000 pour les taxons phytoplanctoniques.

La troisième banque d'images, réalisée à partir d'échantillons du REPHY fixés au Lugol provenant du LER Concarneau, comporte moins de groupes phytoplanctoniques (une cinquantaine) mais les échantillons étant souvent dominés par une ou deux espèces, ces groupes présentent un nombre conséquent d'images (au total près de 1 400). Les vignettes de débris ont aussi été classées (848 vignettes dans 11 catégories identiques à celles du bassin d'Arcachon).

3.2.2. Set d'apprentissage Flore Totale Lugol

En juin 2010, un premier set d'apprentissage a été envoyé à l'UMONS pour optimisation. Ce set comprenait un total de 49 groupes. Une quinzaine de groupes ayant été momentanément écartés par manque de vignettes, un set remanié comportant 34 groupes, a été constitué (Tableau 1) (Grosjean et Denis, 2010).

Tableau 1. Liste et effectifs des différents groupes du set d'apprentissage «Lugol» de juin 2010 (Grosjean et Denis, 2010).

| Nom du Taxon | Code training set | nbre de vignettes | Nom du Taxon | Code training set | nbre de vignettes |
|------------------------------------|-------------------------|-------------------|---|--------------------|-------------------|
| Phytoplancton | | | Zooplancton | | |
| Diatomophyceae | | | Larves de crustacés | Larvae_crustaceans | 20 |
| <i>Asterionellopsis glacialis</i> | A. glacialis | 113 | Ciliés nus | Naked_Ciliates | 91 |
| Centrales spp. | Centrales spp. | 100 | Xystonella (Tintinnide) | Xystonella | 125 |
| <i>Chaetoceros</i> spp. | Chaetoceros_spp. | 61 | Nb groupes et vignettes | 3 | 236 |
| <i>Dytilum brightwelli</i> | D. brightwelli | 49 | | | |
| <i>Guinardia delicatula</i> | G. delicatula | 60 | Débris | | |
| <i>Guinardia flaccida</i> | G. flaccida | 151 | Particules noires et opaques | Black_opaque | 102 |
| <i>Leptocylindrus danicus</i> | L. danicus | 116 | Particules fines, allongées et longues | Long_thin | 102 |
| <i>Nitzschia longissima</i> | N. longissima | 52 | Particules détritiques, aspect granuleux, bords irréguliers | Granular | 106 |
| <i>Proboscia alata</i> | P. alata | 37 | Particules transparentes, aspect membraneux | Membranous | 82 |
| <i>Proboscia alata var. indica</i> | P alata_indica | 42 | Particules translucides, plus ou moins claires | Clear | 94 |
| <i>Pseudo-nitzschia</i> spp. | Pseudo-nitzschia | 106 | Fibres foncées ou claires | Fibers | 108 |
| <i>Rhizosolenia imbricata</i> | R. imbricata | 52 | Nb groupes et vignettes | 6 | 594 |
| <i>Thalassiosira decipiens</i> | T. decipiens | 29 | | | |
| <i>Thalassiosira</i> spp. | Thalassiosira_spp_cells | 73 | Autres particules | | |
| <i>Thalassiosira subtilis</i> | T. subtilis | 107 | Pelotes fécales | Fecal_pellets | 65 |
| Nb groupes et vignettes | 15 | 1148 | Grosses bulles occupant tout le champ | Drop | 113 |
| Dinophyceae | | | Agrégats de particules vivantes ou inertes | Aggregates | 53 |
| <i>Ceratium fusus</i> | C. fusus | 61 | Mues de crustacés | Mues | 53 |
| <i>Dinophysis tripos</i> | D. tripos | 51 | Petites bulles d'air | Bubbles | 129 |
| <i>Protoperdinum</i> spp. | Protoperdinum_spp | 45 | Grains de pollen | Pollen | 112 |
| <i>Scrippsiella</i> spp. | Scrippsiella | 26 | Nb groupes et vignettes | 6 | 525 |
| Nb groupes et vignettes | 4 | 183 | | | |

Ce set optimisé a ensuite servi de base à la constitution du set final réalisé dans le cadre du post-doctorat, en intégrant les nouvelles images numérisées entre juillet 2010 et mars 2011. Ce sont les caractéristiques de ce dernier set, arrêté à la date de mars 2011, qui sont présentées ci-dessous. Trois catégories n'ont pas été conservées (par manque de vignettes) et 10 ont été ajoutées conduisant à un set comprenant un total de 44 catégories (30 groupes planctoniques et 14 groupes de particules inertes). Le nombre de vignettes par catégorie a été aussi substantiellement augmenté.

Les groupes phytoplanctoniques cibles sont au nombre de 28 parmi lesquels les diatomées sont largement représentées avec 22 groupes (7 groupes en plus de ceux du set

précédent), tandis que les dinoflagellés comptent 4 groupes. Deux autres classes phytoplanctoniques sont aussi caractérisées (DICTYOPHYCEES et EUGLÉNOPHYCEES). Deux catégories appartenant au zooplancton (larves de Crustacés et Ciliés nus) ont été également intégrées au set d'apprentissage, soit un total de 30 groupes de taxons microplanctoniques (Tableau 2 et Annexe 7).

Tableau 2. Liste des groupes planctoniques représentés dans le set d'apprentissage Flore Totale Lugol de mars 2011.

| Nom du Taxon | Code training set | Nbre de vignettes |
|--|-------------------------|-------------------|
| Phytoplancton | | |
| Diatomophyceae | | |
| <i>Asterionellopsis glacialis</i> | A. glacialis | 181 |
| Centrales spp. | Centrales spp. | 102 |
| <i>Chaetoceros danicus</i> | C. danicus | 23 |
| <i>Chaetoceros decipiens</i> | C. decipiens | 42 |
| <i>Chaetoceros</i> spp. | Chaetoceros_spp. | 114 |
| <i>Dytilum brightwelli</i> | D. brightwelli | 100 |
| <i>Dactyliosolen fragilissimus</i> | D. fragilissimus | 161 |
| <i>Guinardia delicatula</i> | G. delicatula | 81 |
| <i>Guinardia flaccida</i> | G. flaccida | 167 |
| <i>Guinardia striata</i> | G. striata | 204 |
| <i>Leptocylindrus danicus</i> | L. danicus | 254 |
| <i>Lithodesmium undulatum</i> | L. undulatum | 252 |
| <i>Lauderia + Schroderella</i> | Lauderia_Schroderella | 98 |
| <i>Nitzschia longissima + Cylindrotheca closterium</i> | N. longissima | 74 |
| <i>Proboscia alata</i> | P. alata | 47 |
| <i>Proboscia alata var. indica</i> | P alata_indica | 89 |
| <i>Pseudo-nitzschia</i> spp. | Pseudo-nitzschia | 211 |
| <i>Rhizosolenia imbricata+styliformis</i> | R. imbricata | 65 |
| <i>Thalassionema nitzschioides</i> | T. nitzschioides | 72 |
| <i>Thalassiosira subtilis</i> | T. subtilis | 112 |
| <i>Thalassiosira</i> spp. (coloniales) | Thalassiosira_spp | 26 |
| <i>Thalassiosira</i> spp. (solitaires) | Thalassiosira_spp_cells | 196 |
| Nb groupes et nb total de vignettes | 22 | 2671 |
| Dinophyceae | | |
| <i>Ceratium fusus</i> | C. fusus | 143 |
| <i>Ceratium</i> spp. | Ceratium_spp | 21 |
| <i>Dinophysis tripos</i> | D. tripos | 56 |
| <i>Protoperidinium</i> spp. | Protoperidinium_spp | 58 |
| Nb groupes et nb total de vignettes | 4 | 278 |
| Phytoplancton autres | | |
| DICTYOPHYCEAE | DICTYOPHYCEAE | 52 |
| EUGLÉNOPHYCEAE | EUGLÉNOPHYCEAE | 94 |
| Nb groupes et nb total de vignettes | 2 | 146 |
| Zooplancton | | |
| Larves de crustacés | Larvae_crustaceans | 22 |
| Ciliés nus | Naked_Ciliates | 144 |
| Nb groupes et nb total de vignettes | 2 | 166 |

En outre, 14 catégories ne décrivant pas des taxons planctoniques ont été intégrées au set d'apprentissage. En effet, les débris et autres particules inertes représentent environ 80% des particules dans les échantillons du Bassin d'Arcachon. Si ces débris sont ignorés par l'observateur lors d'un dénombrement au microscope, ce n'est pas le cas lors de la numérisation de l'échantillon avec le FlowCAM car ils sont détectés de la même manière que les particules planctoniques ciblées. De plus, au cours du processus de classification automatique, il est impossible d'écarter les particules non ciblées ou de les regrouper dans

une catégorie «poubelle», du fait de la structure même de l'algorithme «*Random Forest*» (cf. 4.2.1). Si une nouvelle particule n'est pas représentée dans le set d'apprentissage, l'outil de reconnaissance va nécessairement la classer, à tort, dans un des groupes existants.

Il est donc essentiel de caractériser ces particules inertes. Les critères retenus portent essentiellement sur : la teinte globale, l'opacité, les contours, la texture et la forme générale (allongée ou non) de la particule. Par ailleurs, les particules inertes se distinguent en deux types : les débris à proprement parler, d'aspect détritique ou minéral, mais dont l'origine reste indéterminée, et les autres particules pour lesquelles la nature est identifiable, comme par exemple les gros agrégats, les grains de pollen ou les bulles d'air (Tableau 3 et Annexe 7).

Tableau 3. Liste des groupes de particules inertes (non ciblées) représentées dans le set d'apprentissage Flore Totale Lugol de mars 2011.

| Type de particule inerte | Code training set | Nbre de vignettes |
|---|-------------------|-------------------|
| Débris | | |
| Particules noires et opaques | Black_opaque | 110 |
| Particules foncées | Dark | 140 |
| Particules fines, allongées et courtes | Short_thin | 103 |
| Particules très fines, allongées et longues | Long_thin | 106 |
| Particules détritiques, aspect granuleux, bords irréguliers | Granular | 313 |
| Particules transparentes, aspect membraneux | Membranous | 107 |
| Particules translucides, plus ou moins claires | Clear | 106 |
| Fibres foncées ou claires | Fibers | 133 |
| Nb groupes et nb total de vignettes | 8 | 1118 |
| Autres particules | | |
| Pelotes fécales | Fecal_pellets | 79 |
| Grosses bulles occupant tout le champ | Drop | 114 |
| Agrégats de particules vivantes ou inertes | Aggregates | 134 |
| Mues de crustacés | Mues | 59 |
| Petites bulles d'air | Bubbles | 135 |
| Grains de pollen | Pollen | 115 |
| Nb groupes et nb total de vignettes | 6 | 636 |

Chacun de ces groupes est illustré par un échantillon d'images du set d'apprentissage dans l'annexe 7.

3.2.3. Set d'apprentissage Flore Totale Vivant

En juin 2010, un set d'apprentissage Flore Totale Vivant a été envoyé à l'UMONS pour optimisation. Ce set comprenait au départ un total de 45 groupes. Après remaniement, 29 groupes ont été conservés (Grosjean et Denis, 2010).

Tableau 4. Liste et effectifs des différents groupes du set d'apprentissage « Vivant » de juin 2010 (Grosjean et Denis, 2010).

| Nom du Taxon | Code training set | nbre de vignettes | Nom du Taxon | Code training set | nbre de vignettes |
|------------------------------------|---------------------|-------------------|---|--------------------|-------------------|
| Phytoplankton | | | Zooplankton | | |
| Diatomophyceae | | | Larves de crustacés | Larvae_crustaceans | 135 |
| <i>Asterionellopsis glacialis</i> | A. glacialis | 121 | Ciliés nus | Naked_Ciliates | 39 |
| Centrales spp. | Centrales spp. | 127 | Xystonella (Tintinide) | Xystonella | 63 |
| <i>Chaetoceros</i> spp. | Chaetoceros_spp. | 110 | Nb groupes et vignettes | 3 | 237 |
| <i>Guinardia flaccida</i> | G. flaccida | 96 | Débris | | |
| <i>Leptocylindrus danicus</i> | L. danicus | 109 | Particules noires et opaques | Black_opaque | 17 |
| <i>Nitzschia longissima</i> | N. longissima | 12 | Particules fines, allongées et longues | Long_thin | 95 |
| <i>Proboscia alata</i> | P. alata | 94 | Particules détritiques, aspect granuleux, bords irréguliers | Granular | 114 |
| <i>Proboscia alata var. indica</i> | P alata_indica | 60 | Particules translucides, plus ou moins claires | Clear | 103 |
| <i>Pseudo-nitzschia</i> spp. | Pseudo-nitzschia | 91 | Fibres foncées ou claires | Fibers | 35 |
| <i>Rhizosolenia imbricata</i> | R. imbricata | 30 | Nb groupes et vignettes | 5 | 364 |
| <i>Thalassiosira subtilis</i> | T. subtilis | 131 | Autres particules | | |
| Nb groupes et vignettes | 11 | 981 | Pelotes fécales | Fecal_pellets | 49 |
| Dinophyceae | | | Grosses bulles occupant tout le champ | Drop | 142 |
| <i>Ceratium fusus</i> | C. fusus | 66 | Agrégats de particules vivantes ou inertes | Aggregates | 99 |
| <i>Dinophysis tripos</i> | D. tripos | 77 | Mues de crustacés | Mues | 113 |
| <i>Protoperidinium</i> spp. | Protoperidinium_spp | 73 | Petites bulles d'air | Bubbles | 110 |
| <i>Scrippsiella</i> spp. | Scrippsiella | 16 | Grains de pollen | Pollen | 123 |
| Nb groupes et vignettes | 4 | 232 | Nb groupes et vignettes | 6 | 636 |

Ces 29 groupes sont communs avec les groupes du premier set « Lugol » présenté dans le tableau 1 ce qui permettra une comparaison des statistiques obtenues (cf. 4.2.3.).

4. Les outils de reconnaissance

4.1. Méthodes d'optimisation et d'évaluation des outils de reconnaissance (UMONS)

Les méthodes d'optimisation ont été exposées dans les deux documents rendant compte du travail de l'UMONS (Denis et Grosjean, 2006 ; Grosjean et Denis, 2010). La présentation ci-dessous en est largement reprise.

4.1.1. Optimisation logicielle

La création de l'outil de reconnaissance²⁰ se faisant à la suite de la lecture du set d'apprentissage par un algorithme d'apprentissage, dit de « *machine learning* », il est nécessaire de déterminer l'algorithme le plus performant. Dans Zoo/PhytoImage, 6 algorithmes sont disponibles : l'analyse discriminante linéaire (LDA), les k plus proches voisins (Knn), le réseau de neurones (Nnet), le « *learning vector quantization* » (LVQ), les arbres de partitionnement (Tree) et « *Random Forest* » (RF).

Les performances de l'algorithme retenu sont ensuite testées afin de définir les paramètres optimaux pour le meilleur compromis "taux de reconnaissance/temps de calcul". Pour l'algorithme « *Random Forest* »²¹ (qui sera retenu comme le plus performant, cf. 4.2.1.), deux paramètres sont modulés : le nombre d'arbres de partitionnement créés dans la forêt d'arbres (qui joue surtout sur la vitesse de traitement) et le nombre de variables échantillonnées aléatoirement à chaque nœud de ces arbres (qui joue sur la performance en terme de reconnaissance automatique). Pour optimiser la reconnaissance par « *Random Forest* », il est nécessaire de définir les meilleures valeurs de chacun de ces deux paramètres.

4.1.2. Calcul des taux d'erreur

Les taux d'erreur, pour l'ensemble du jeu de données et pour chaque groupe, sont calculés par une validation croisée du set d'apprentissage, répétée 10 fois. Plus précisément, 90% du set d'apprentissage sont utilisés pour l'apprentissage proprement dit, et les 10% restants servent de test pour la reconnaissance automatique. Cette opération est réalisée 10 fois, de manière à ce que chaque dixième du set d'apprentissage soit utilisé une fois pour le test. Ces taux d'erreur traduisent donc l'erreur moyenne commise au cours de ces 10 opérations.

²⁰ L'outil créé, nommé *classifier* en anglais, peut être assimilé à une boîte noire capable de déterminer l'appartenance d'une image à un groupe, uniquement sur la base des paramètres mesurés sur celle-ci. L'optimisation de cet objet est donc primordiale pour reconnaître du plancton de manière automatisée.

²¹ Cet algorithme a été présenté pour la première fois en 2001 par Breiman (2001).

4.1.3. Matrice de confusion

La matrice de confusion est une table qui croise les identifications faites visuellement par l'opérateur (en lignes) et les objets classés par l'outil de reconnaissance (en colonnes) (cf. Figure 8). Cette méthode permet de mettre en évidence les sources de confusions entre taxons, en indiquant le nombre d'objets du set d'apprentissage correctement identifiés par l'outil de reconnaissance, situés dans la diagonale du tableau, et les identifications erronées pour chaque taxon situées hors de la diagonale. Les erreurs commises par l'outil de reconnaissance situées dans les lignes du tableau représentent donc les faux-négatifs pour chaque taxon, tandis que les erreurs apparaissant dans les colonnes correspondent aux faux-positifs pour chaque taxon.

Lors de la constitution du set d'apprentissage, nous avons choisi de privilégier la capacité de l'outil de reconnaissance à identifier correctement l'essentiel des particules appartenant à un taxon donné, c'est-à-dire à favoriser le nombre de vrais-positifs (diagonale) au détriment du nombre de faux-négatifs (lignes hors diagonale). Ce faisant, on favorise aussi le nombre de faux-positifs pour chaque groupe (colonnes hors diagonale), c'est-à-dire le nombre de particules classées à tort dans le groupe donné. Cet effet n'est pas un inconvénient car il facilite en fait une vérification visuelle ultérieure de la classification automatique des particules. La matrice de confusion permet de visualiser rapidement ces tendances, et donc d'ajuster plus facilement la proportion d'images dans chaque groupe pour parvenir à cet objectif.

4.2. Résultats

4.2.1. Optimisation logicielle (UMONS)

Cette optimisation a été réalisée par l'UMONS. Les résultats détaillés en sont présentés dans le document Grosjean et Denis (2010) déjà cité. Pour notre application, l'algorithme donnant les meilleurs résultats dans la reconnaissance automatique s'est avéré être « *Random Forest* », même si l'analyse discriminante linéaire a fourni également des taux de reconnaissance intéressants, et avec une vitesse d'identification significativement plus rapide.

L'effet du nombre d'arbres de partitionnement et du nombre de variables utilisés par « *Random Forest* » a ensuite été testé. Il a été montré (Grosjean et Denis, 2010) que l'outil généré à 500 arbres de partitionnement et à 19 variables présentait les meilleures performances en terme de taux de reconnaissance/temps de calcul. Les 19 variables les plus explicatives retenues sont principalement des paramètres métriques dont les variables ajoutées par PhytoImage à celles mesurées par le logiciel du FlowCAM.

Ainsi, il est apparu que, contrairement au zooplancton pour lequel les niveaux de transparence (niveaux de gris) sont très discriminants (Denis, 2005), la reconnaissance automatique du phytoplancton se base avant tout sur les critères de forme et de taille. Cependant, les niveaux de gris (calibrés dans ce travail) ne sont tout de même pas complètement absents des prédictions.

4.2.2. Statistiques de la classification automatique calculées sur le set d'apprentissage Flore Totale Lugol de mars 2011

4.2.2.1. Taux d'erreur

Sur la base du dernier set d'apprentissage Flore Totale Lugol (mars 2011), la validation croisée indique un taux d'erreur global de 28% (Tableau 5). Cela correspond donc à un taux de reconnaissance global de 72%, c'est-à-dire que 72% des images du set d'apprentissage sont correctement identifiés par l'outil de reconnaissance créé.

Il faut de plus souligner que les performances globales prédites sur le set d'apprentissage sont sous-évaluées car elles incluent les erreurs attachées à la discrimination des débris qui ne sont pas des particules d'intérêt.

En ce qui concerne les groupes planctoniques ciblés, les taux d'erreur par groupe sont très variables. Pour 14 des 28 groupes ciblés, ce taux d'erreur est inférieur à 26%. Les groupes théoriquement mal reconnus sont le plus souvent représentés par un nombre restreint d'images. L'ajout d'images dans ces catégories devrait donc améliorer ces résultats.

Tableau 5. Taux d'erreur global et pour chaque groupe du set d'apprentissage Flore Totale Lugol de mars 2011.

| Erreur globale : | | 28% | |
|-----------------------------------|-----------|----------------------------|-----------|
| Erreur par groupe : | % | | |
| Drop | 7 | Long_thin | 27 |
| Pollen | 8 | Thalassiosira_spp | 28 |
| C.fusus | 8 | Fecal_pellets | 29 |
| L.danicus | 9 | D.brightwelli | 31 |
| L.undulatum | 12 | Dark | 35 |
| Thalassiosira_spp_cells | 15 | Aggregates | 36 |
| G.flaccida | 15 | D.fragilissimus | 37 |
| Pseudo-nitzschia | 15 | Naked_Ciliates | 37 |
| Black_opaque | 18 | P.alata_indica | 45 |
| D.tripos | 19 | Fibers | 48 |
| Centrales_spp | 19 | Membranous | 51 |
| Short_thin | 20 | T.nitzschioides | 53 |
| Mues | 20 | Dictyochophyceae | 53 |
| A.glacialis | 20 | G.delicatula | 53 |
| G.striata | 21 | R.imbricata_styli | 58 |
| T.subtilis | 21 | Larvae_crustaceans | 62 |
| Euglenophyceae | 22 | Protoperidinium_spp | 70 |
| Clear | 24 | Chaetoceros_spp | 74 |
| Lauderia_Schroederella | 24 | Ceratium_spp | 75 |
| Bubbles | 25 | P.alata | 85 |
| Granular | 26 | C.decipiens | 93 |
| N.longissima_Cylindrotheca | 26 | C.danicus | 96 |

| | |
|-----------------------------|------------------------------|
| Diatomées | Zooplancton |
| Dinoflagellés | Débris et particules inertes |
| Phytoplancton autres | |

4.2.2.2. Matrice de confusion

4.2.3. Statistiques de la classification automatique calculées sur le set d'apprentissage Flore Totale Vivant

Les statistiques de la classification automatique ont été calculées sur le set optimisé Flore Totale Vivant (arrêté en juin 2010), et présentées dans le rapport remis à l'ONEMA (Grosjean et Denis, 2010). Les auteurs comparent ces statistiques avec celles obtenues sur le set Flore Totale Lugol de la même date et concluent qu'elles s'avèrent très similaires (taux de reconnaissance global de 78,6 % pour le premier et 77,6 % pour le second). Ces taux un peu supérieurs à ceux obtenus sur le set Flore Totale Lugol de mars 2011 s'expliquent par le fait que le nombre de catégories est moins élevé. En effet, plus ce nombre est élevé, plus les erreurs s'ajoutent. Ces résultats montrent que l'augmentation conséquente du nombre de catégories (entre juin 2010 et mars 2011) conduit à une diminution de la précision relativement limitée (4 % -de 77,6 % à 72 %-) ce qui est encourageant pour la suite du travail.

5. Evaluation des performances de l'outil de reconnaissance Flore Totale Lugol

Les indices de performance calculés sur les groupes du set d'apprentissage peuvent se révéler différents de ceux obtenus avec le même outil sur de nouveaux échantillons, par exemple si ces derniers contiennent des particules n'appartenant pas aux groupes définis dans le set (à commencer par ceux qui ont été écartés par manque de vignettes) ou si les proportions relatives des divers groupes sont très variables d'un échantillon à l'autre (Grosjean et Denis, 2010)

C'est pourquoi, il est nécessaire d'évaluer les performances des outils sur de nouveaux échantillons (cf. 1.1.2.) dont on veut déterminer la composition phytoplanctonique. Cette évaluation a été faite avec l'outil de reconnaissance optimisé créé à partir du set d'apprentissage Lugol qui était le plus complet.

5.1. Performances de classification automatique par l'outil de reconnaissance

5.1.1. Matériel et méthode

Sur les 32 prélèvements (fixés au Lugol) réalisés dans le cadre du REPHY entre août 2010 et mars 2011, 14 ont été retenus pour tester les performances de l'outil de reconnaissance sur des assemblages naturels. Il s'agit des prélèvements qui contenaient plusieurs des taxons présents dans le set d'apprentissage, avec une abondance jugée suffisante. Ils ont été numérisés au fur et à mesure de leur collecte, chaque échantillon étant numérisé deux fois pendant 30 min, de manière à obtenir deux séries comparables.

L'outil de reconnaissance a été utilisé pour classer les images de la totalité des 14 échantillons pour obtenir un résultat significatif (soit 3 361 images au total). Au terme de la classification automatique et après vérification visuelle des images de chaque groupe, les taux réels de vrais-positifs (proportion d'images correctement classées dans chaque groupe) et faux-positifs (proportion d'images erronées) ont ensuite été calculés, et ce pour chacune des deux séries.

5.1.2. Résultats

5.1.2.1. Discrimination des groupes ciblés (phytoplancton)

Si les performances de reconnaissance automatique sur les images issues des échantillons testés sont très similaires d'une série à l'autre, ce qui montre la bonne reproductibilité des numérisations et de la reconnaissance automatique, elles s'avèrent variables d'un groupe à l'autre (Figure 10).

Les groupes présentant peu ou pas d'images correspondent à des taxons soit absents, soit très peu abondants dans les 14 échantillons. Il s'agit des groupes *Chaetoceros danicus*, Centrales, *Ceratium* spp., *Dinophysis tripos*, *Proboscia alata*, *Protoperidinium* spp. et *Rhizosolenia imbricata*. Les résultats de la classification automatique traduisent

donc correctement l'occurrence de ces taxons dans les échantillons testés. En outre, il est intéressant de constater que, pour ces groupes, l'outil de reconnaissance identifie également très peu de faux-positifs. Cela suggère que ces groupes seraient potentiellement bien discriminés au sein d'un assemblage phytoplanctonique. Cependant, il faudrait confirmer ces performances en testant des échantillons présentant une forte abondance pour ces taxons spécifiques, afin de s'assurer que l'outil de reconnaissance est capable de les identifier correctement. En effet, lorsqu'un groupe est représenté par un faible nombre d'images dans le set d'apprentissage, l'outil de reconnaissance montre souvent en réponse une faible sensibilité pour ce groupe, c'est-à-dire que la majorité des images ciblées ne sont pas identifiées correctement (= faux-négatifs).

Concernant les autres groupes phytoplanctoniques, on observe une proportion variable de faux-positifs. Ainsi, pour la diatomée *Guinardia striata*, abondante dans le jeu de données testé, la proportion de faux-positifs est très faible en regard du nombre de vrais-positifs. En revanche, pour *Lithodesmium undulatum*, également bien représenté dans les échantillons testés, le pourcentage de faux-positifs est bien supérieur au pourcentage de vrais-positifs.

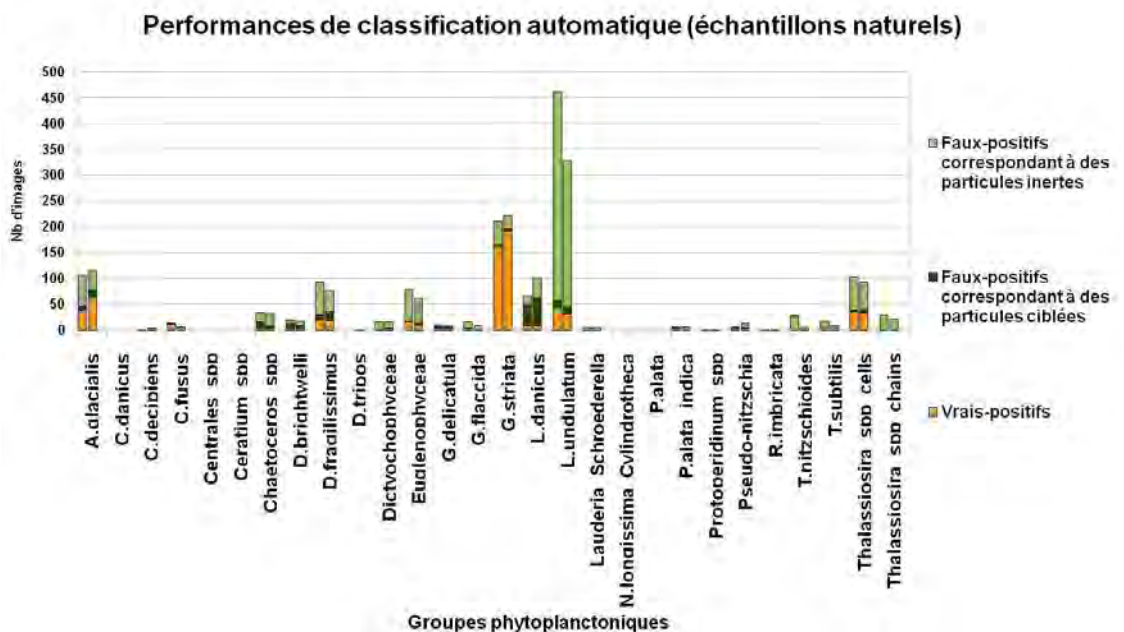


Figure 10. Statistiques de reconnaissance automatique pour les groupes ciblés du phytoplancton, évaluées sur chacune des deux séries de numérisation des 14 échantillons.

Cependant, si le nombre de faux-positifs est assez important pour la majorité des groupes, il est intéressant de constater que parmi ceux-ci, la proportion de faux-positifs appartenant aux groupes ciblés, c'est-à-dire aux taxons phytoplanctoniques, est très faible par comparaison avec le nombre de faux-positifs correspondant aux débris. Cela signifie qu'il y a **peu de confusion entre les groupes ciblés**. Les principales sources de confusion sont donc les particules inertes.

D'autre part, les résultats très similaires entre les deux séries suggèrent une **reproductibilité de l'erreur**, donc une possibilité de correction automatique.

5.1.2.2. Discrimination des groupes non ciblés (particules inertes et débris)

Concernant la reconnaissance des groupes de débris et autres particules, l'outil de reconnaissance se montre très performant, puisque très peu de particules ciblées (phytoplancton) sont classées au sein des débris et autres particules (Figure 11). Cela montre que la **catégorisation des débris est pertinente**, qu'il y ait ou non des confusions entre particules inertes.

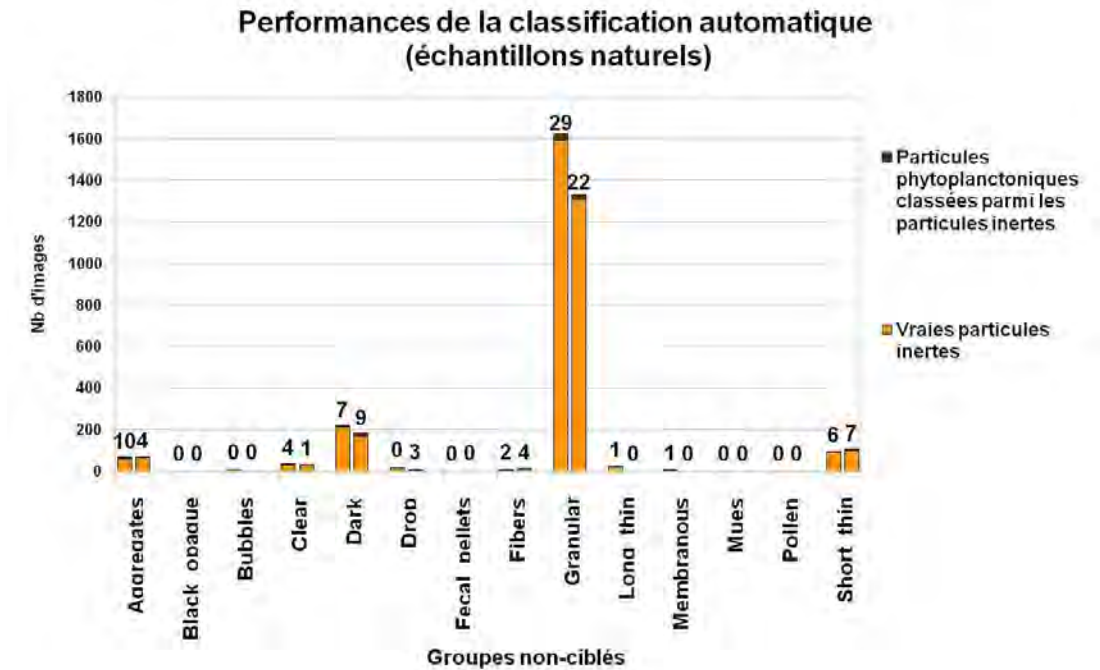


Figure 11. Statistiques de reconnaissance automatique pour les groupes non-ciblés (particules inertes) évaluées sur deux séries identiques de 14 échantillons. Le nombre de faux-positifs appartenant au phytoplancton est affiché pour chaque groupe.

5.1.2.3. Synthèse

La stratégie adoptée pour la constitution et l'optimisation du set d'apprentissage visait à privilégier la sensibilité (*recall*) de l'outil de reconnaissance, c'est-à-dire sa capacité à classer correctement l'essentiel des cellules de chaque taxon. Ce faisant, on tend à favoriser le nombre de faux-positifs dans chaque groupe, ce qui peut diminuer la précision de l'outil c'est-à-dire sa capacité à classer majoritairement dans chaque groupe des vrais-positifs. L'outil de reconnaissance Flore Totale Lugol montre effectivement ces tendances :

- **bonne sensibilité pour la plupart des groupes** ; pour la plupart des taxons ainsi que pour les groupes de particules inertes, la majorité des individus est correctement identifiée par l'outil de reconnaissance ;

- **précision moyenne pour les groupes phytoplanctoniques** ; une proportion élevée de faux-positifs, appartenant essentiellement aux débris, est observée dans la classification automatique pour les groupes phytoplanctoniques ; en revanche, la confusion entre les différents groupes phytoplanctoniques est faible ;

- **précision élevée pour les groupes de particules inertes** ; la proportion de particules phytoplanctoniques ciblées classées à tort dans les débris et autres particules est très faible.

5.2. Performances de l'estimation automatique de l'abondance par l'outil de reconnaissance

5.2.1. Matériel et méthode

Parmi les 14 échantillons REPHY fixés au Lugol utilisés précédemment, seuls 4 échantillons présentaient des abondances élevées et une richesse conséquente en cellules non coloniales, conditions nécessaires pour tester la capacité de l'outil à estimer automatiquement la concentration cellulaire pour chaque taxon du set d'apprentissage. En effet, les cellules agencées en colonies ne sont pas encore identifiées individuellement par le logiciel PhytoImage²². Une colonie est donc considérée comme un objet unique. Il n'est de ce fait pas possible de comparer les comptages cellulaires avec les comptages automatiques dans ces cas précis. Le dénombrement des colonies n'étant pas réalisé dans le cadre du REPHY, cette étude préliminaire s'est focalisée sur les autres taxons.

Les 4 échantillons retenus ont été utilisés afin de comparer l'estimation automatique de l'abondance des taxons phytoplanctoniques fournie par l'outil de reconnaissance avec les résultats des comptages au microscope réalisés dans le cadre du REPHY.

Chaque échantillon a été numérisé trois fois pendant 60 min. Au final, un volume total²³ de 4,35 ml a été numérisé pour chaque échantillon, ce qui offre une précision de comptage comparable à celle résultant d'un comptage manuel. En effet, une particule détectée par le FlowCAM correspond à une concentration cellulaire *in situ* de 230 cellules/L tandis qu'une cellule comptée au microscope (sur la totalité d'une cuve de 10 mL) équivaut à une concentration cellulaire *in situ* de 100 cellules/L (ou plus, si l'aire qui a été observée est inférieure à la surface totale de la cuve) et ce pour chaque taxon.

L'outil de reconnaissance a été utilisé pour classer les images de chaque échantillon, puis les erreurs dans l'estimation automatique de l'abondance ont été évaluées, avant et après une validation visuelle de l'ensemble des images.

5.2.2. Résultats de l'estimation automatique de l'abondance

Afin d'évaluer les performances quantitatives de l'outil de reconnaissance Flore Totale, c'est-à-dire sa capacité à estimer la concentration cellulaire des différents taxons, la différence (moyenne des deux séries) entre les valeurs d'abondance calculées automatiquement par le logiciel PhytoImage et les valeurs d'abondance des comptages au microscope a été représentée pour 3 cas de figures (Figure 12) :

²² Une des améliorations futures du logiciel PhytoImage devrait permettre de résoudre cette question (cf. 7.2..2.2.).

²³ Rappelons qu'il s'agit du volume total effectivement imagé ; en effet, le volume transitant dans le FlowCAM est beaucoup plus important (cf. 2.2.).

- à l'issue de la classification automatique (erreur avant correction),
- après retrait manuel des faux-positifs (erreur après retrait des faux-positifs = validation manuelle partielle),
- après reclassification de l'ensemble des vignettes mal classées (erreur après validation manuelle totale).

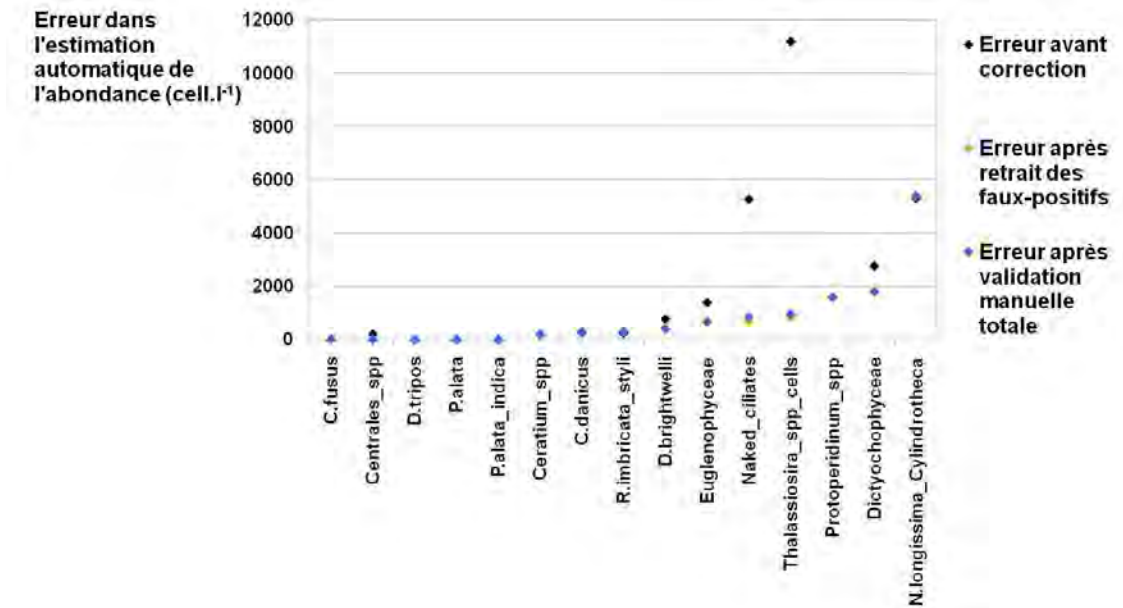


Figure 12. Erreur moyenne dans l'estimation automatique de l'abondance (cell.L⁻¹) par comparaison avec les comptages correspondants au microscope (sur 4 échantillons).

Lorsque l'erreur commise est inférieure à 400 cellules/L, la validation manuelle partielle ou totale de la classification modifie peu les résultats. En revanche, lorsque cette erreur initiale est plus élevée, la validation manuelle améliore sensiblement les résultats d'estimation de l'abondance. Après cette étape, l'erreur dans l'estimation de l'abondance chute en dessous de 2 000 cellules/L pour presque tous les taxons testés, ce qui représente en fait une erreur de classification de 8 images. Ainsi, pour le groupe *Thalassiosira_spp_cells*, cette erreur est diminuée d'un facteur 11.

Parmi les groupes testés, seul le groupe *N. longissima_Cylindrotheca* reste très mal quantifié, même après validation manuelle de la classification. La forme très allongée et la petite taille des diatomées de ce groupe, régulièrement présentes à des concentrations élevées dans les échantillons du bassin d'Arcachon, pourraient expliquer la difficulté à les quantifier de manière précise avec cette méthode automatisée. En effet, cette erreur importante traduit en fait une sous-estimation du nombre de ces diatomées par l'approche automatisée. Il est possible qu'il se produise une perte de ces cellules dans le système de tube au cours de la numérisation de l'échantillon, soit du fait d'une répartition hétérogène des particules dans la cellule de flux (cellules plus petites en marge du flux, donc hors du champ de la caméra centrée sur le flux), soit du fait d'une adhérence de ces diatomées pennées le long des parois des tubes en silicone.

Enfin, la validation manuelle partielle de la classification automatique offre des résultats très similaires à ceux d'une validation manuelle. On peut donc considérer qu'une simple vérification des groupes ciblés (groupes phytoplanctoniques) visant à exclure manuellement les faux-positifs serait suffisante pour obtenir une estimation significative de la concentration cellulaire de ces groupes dans le cadre d'une approche semi-automatique.

6. Discussion

L'outil de reconnaissance « Flore Totale Lugol », au terme d'une longue phase de réflexion et de première mise au point et, fournit des résultats très encourageants en mode semi-automatique.

Ainsi, il permet de discriminer d'ores et déjà 28 groupes phytoplanctoniques à un niveau taxinomique équivalent à celui obtenu en routine par les observateurs sur les échantillons du bassin d'Arcachon Flores Totales du REPHY. La catégorisation pertinente des débris (14 catégories) limite les sources de confusion avec les groupes ciblés. De ce fait, l'essentiel des particules de chaque groupe phytoplanctonique est correctement classé par l'outil de reconnaissance. De plus, les caractéristiques de l'outil de reconnaissance présenté offrent d'intéressantes perspectives pour une approche semi-automatique. En effet, la validation manuelle nécessaire consisterait à vérifier seulement les groupes phytoplanctoniques, puisque la quantité de particules phytoplanctoniques classées dans les débris peut être considérée comme négligeable. En outre, la dominance de débris parmi les faux-positifs classés dans les groupes phytoplanctoniques facilite cette dernière tâche puisque les débris sont très rapidement repérables visuellement. Une validation manuelle constituerait donc une solution tout à fait envisageable, en terme de rapport temps alloué à la validation/précision finale de la classification.

L'estimation de l'abondance des différents taxons par cette méthode automatique pose des problèmes méthodologiques qui restent encore à résoudre. Dans les essais menés, un groupe reste très mal quantifié, malgré la validation manuelle. Cet exemple pose le problème des particules non détectées de manière récurrente au cours de la numérisation. Les particularités morphologiques de certains taxons pourraient expliquer cette absence de détection mais les raisons exactes de cet écueil restent à déterminer.

Les travaux menés par d'autres équipes ont aussi souligné des biais quantitatifs mais qui étaient dus à l'emploi du mode *Trigger* (Tezzo, 2008 ; Campbell et Buskey, 2011). Le mode *AutoImage* que nous avons utilisé ne devrait pas présenter ces inconvénients. En effet, les résultats obtenus dans le cadre du projet AMORE²⁴ montrent que les abondances obtenues, à partir d'échantillons frais de Mer du Nord, par comptage automatisé (FlowCAM/PhytoImage) avec le mode *AutoImage* et par comptage au microscope des mêmes échantillons, sont du même ordre de grandeur (Lancelot *et al.*, 2011).

Ces résultats concernent des numérisations faites sur du matériel vivant. La cause de la sous-estimation observée pour certaines catégories algales dans notre étude²⁵ pourrait donc être recherchée dans la fixation des échantillons. En effet, il a été montré que le Lugol pouvait affecter la taille et la forme de certaines cellules microplanctoniques (Choi et Stoecker, 1989 ; Menden-Deuer *et al.*, 2001), provoquer leur fragmentation ou au contraire favoriser la constitution d'agrégats (Zarauz et Irigoyen, 2008). Ce fixateur pourrait aussi avoir des effets très spécifiques sur la circulation des cellules

²⁴ SSD-Science for a Sustainable Development – Atmosphere and Terrestrial and Marine Ecosystems. Project SD/NS/03.

²⁵ Il s'agit de la première étude faisant état de résultats obtenus à partir d'échantillons **fixés** numérisés au FlowCAM.

phytoplanctoniques dans le système de flux, probablement en rapport avec la forme de chaque espèce, sa flottabilité, son taux de sédimentation naturel, sa capacité d'adhérence, etc, ce qui pourrait en partie expliquer le biais quantitatif mesuré pour certains groupes taxinomiques lors des comptages automatiques avec le FlowCAM.

Il semble difficile d'envisager de ne plus fixer au Lugol les échantillons du REPHY, compte tenu des contraintes qu'imposeraient des numérisations à partir de matériel frais. Il apparaît donc nécessaire d'approfondir l'étude de l'influence de la fixation sur la sous-estimation quantitative par le FlowCAM. Dans ce travail, un outil de reconnaissance sur du matériel vivant a été réalisé, qui montre des statistiques de classification voisines de celles obtenues avec un set constitué à partir de matériel fixé au Lugol (cf. 4.2.3). Cet outil pourrait être complété et utilisé ultérieurement pour comparer les résultats obtenus selon les deux modes, fixation ou vivant, afin de mieux appréhender l'effet de la fixation sur certains taxons.

Les recherches actuelles sur la reconnaissance automatique du phytoplancton restent peu nombreuses. Si l'automatisation du suivi des communautés zooplanctoniques est étudiée depuis de nombreuses années, les défis technologiques et méthodologiques liés à la numérisation de classes de tailles plus petites ont longtemps freiné les recherches sur la reconnaissance automatique en routine du phytoplancton, car il fallait à la fois concevoir l'outil de numérisation proprement dit et le système informatique d'analyse d'images et de classification. Après plusieurs années de mise au point et d'amélioration, le FlowCAM apparaît comme un outil approprié pour imager les particules microplanctoniques qui, combiné au logiciel PhytoImage, permet d'atteindre un niveau de reconnaissance global de 72% du phytoplancton d'un set d'apprentissage pour l'ensemble des catégories du set. Ce résultat est très encourageant pour la suite du projet. En effet, Culverhouse *et al.* (2003), comparant les performances de reconnaissance de 16 taxinomistes et de méthodes automatiques sur des espèces de dinoflagellés toxiques, ont montré que les performances des taxinomistes étaient comprises entre 67 et 95 %, soit en moyenne 72 %, résultat similaire à celui obtenu par la méthode automatique testée dans leur étude. Ces auteurs soulignent la difficulté d'obtenir des résultats consistants d'un échantillon à l'autre, et d'un observateur à l'autre, tout en démontrant que les méthodes automatiques peuvent faire aussi bien que les analystes humains.

Actuellement, seuls les résultats obtenus avec l'Imaging FlowCytobot, développé par des chercheurs du Woods Hole Oceanographic Institution (USA) (cf. Maurer et Tunin-Ley, 2011), offrent un niveau de précision comparable aux travaux menés ici. L'Imaging FlowCytobot est un système submersible d'imagerie en flux, dont le principe de fonctionnement est comparable à celui du FlowCAM (combinaison de la cytométrie en flux et de la microscopie photonique), mais qui s'avère être un appareil beaucoup plus complexe que le FlowCAM (Olsen et Sosik, 2007). Le processus de classification automatique également basé sur le principe du « *machine learning* » développé pour le FlowCytobot est aussi plus lourd. Cet équipement plus complexe, plus difficile à mettre en œuvre et probablement plus coûteux offre des résultats très intéressants (Sosik et Olson, 2007 ; Campbell *et al.*, 2010). L'outil de reconnaissance mis au point par cette

équipe permet de discriminer 22 groupes, parmi lesquels 17 groupes de diatomées (essentiellement décrites au niveau du genre), 4 autres groupes planctoniques (dinoflagellés, ciliés, nanoflagellés, cellules indéterminées) et un seul groupe caractérisant les détritiques. Le taux de reconnaissance global est de 88% et varie entre 68% et 99% pour chaque groupe. L'outil de reconnaissance Flore Totale Lugol permet donc une reconnaissance taxinomique plus détaillée et plus large (30 groupes planctoniques majoritairement définis au niveau spécifique) que l'outil issu du FlowCytobot, ce qui explique la moindre précision obtenue (taux global de reconnaissance de 72% calculé sur le set d'apprentissage). En revanche, l'estimation automatique de l'abondance des différents groupes phytoplanctoniques semble assez précise avec le FlowCytobot, comparée aux comptages manuels. En effet, la numérisation d'échantillons non fixés permet d'éviter certains biais évoqués ci-dessus. Par ailleurs, le système de flux du FlowCytobot, très proche du cytomètre en flux classique, permet le passage des cellules une à une, en file, ce qui évite probablement la perte de particules dans le système de flux mais implique un débit nécessairement très lent.

Les résultats remarquables obtenus avec l'Imaging FlowCytobot ont été rendus possibles grâce à un investissement financier très élevé, une importante équipe et plusieurs années de recherche et d'innovations technologiques. L'approche novatrice du couplage entre le FlowCAM et le logiciel PhytoImage nous permettent d'envisager des résultats comparables, avec un investissement beaucoup moins lourd.

Les méthodes automatisées présentent de nombreux avantages par rapport à la méthode traditionnelle, c'est-à-dire le dénombrement au microscope photonique par un opérateur (Sutor, 2011).

- En premier lieu, elles permettent de quantifier l'erreur associée au dénombrement, et cette erreur étant souvent reproductible, elle est susceptible d'être corrigée. L'erreur commise par un observateur humain pour la même tâche est très difficilement quantifiable, et est rarement prise en compte.
- Par ailleurs, la numérisation des données phytoplanctoniques permet un stockage peu encombrant et quasiment illimité de ces données, en offrant à tout moment la possibilité d'une vérification (contrôle de la qualité) ou d'une nouvelle analyse (par exemple avec un outil de reconnaissance plus performant obtenu ultérieurement).
- De plus, l'analyse d'images amène un nombre considérable d'informations nouvelles, paramètres de taille et de forme, qui sont très difficiles à extraire en routine au cours d'un suivi classique du phytoplancton et qui sont très intéressantes pour toutes les études écologiques.
- Enfin il ne faut pas oublier les perspectives de durée d'analyse et de coût moindres.

L'intérêt de ces méthodes a conduit à la constitution du groupe de travail international SCOR 130²⁶ Identification Visuelle Automatique du Plancton, ayant pour objectifs d'encourager une coopération internationale sur ce thème, d'évaluer les limites de résolution taxonomiques, de lister les différentes pratiques, d'établir une méthode de comparaison et de développer des méthodes « open source ». Notre travail s'inscrit précisément dans ces finalités.

Pour clore ce rapport, après un bilan des livrables fournis dans le cadre de ce post-doctorat et de la collaboration avec l'UMONS, plusieurs améliorations méthodologiques et informatiques sont proposées afin d'augmenter les performances actuelles et corriger les biais observés.

²⁶ Automatic Plankton Visual Identification, Scientific Committee on Oceanic Research, International Council for Science.



7. Bilan et perspectives

7.1. Bilan

A l'issue du post-doctorat et du travail réalisé par l'UMONS, les livrables suivants ont été fournis :

- la version 2 du logiciel PhytoImage incluant de nouvelles fonctionnalités (UMONS) (Annexe 8),
- 3 banques d'images, deux provenant du bassin d'Arcachon à partir d'échantillons fixés au Lugol (6 000 vignettes) et frais (3 300 vignettes), et une de Concarneau sur du matériel fixé (1 400 vignettes),
- 2 sets d'apprentissage optimisés de la flore totale du bassin d'Arcachon, l'un portant sur la flore fixée au Lugol, l'autre sur la flore vivante,
- 2 outils de reconnaissance optimisés construits à partir des deux sets précédents et les taux de reconnaissance correspondants,
- une première évaluation des performances de l'outil de reconnaissance Flore Totale Lugol sur de nouveaux échantillons du milieu naturel,
- deux modes opératoires détaillés permettant à un nouvel utilisateur formé au FlowCAM de numériser des échantillons REPHY pour alimenter les banques d'images et, à terme, dénombrer en routine les flores phytoplanctoniques (Annexes 5 et 6).

Ces résultats ont fait l'objet de 5 communications orales dont deux dans des colloques internationaux (cf. Communications).

Une publication des résultats concernant le set d'apprentissage Flore Totale Lugol et l'outil de reconnaissance associé optimisé, en cours de rédaction, doit être soumise à l'automne prochain à une revue internationale.

Enfin, la formation à l'utilisation du FlowCAM et de son logiciel VisualSpreadsheet des 4 personnes du LER Arcachon impliquées dans le REPHY a été assurée en mars 2011²⁷ (formation appuyée sur les protocoles écrits détaillés - Annexes 5 et 6). Il s'agit de M. Rumèbe-Perrière, N. Neaud-Masson, C. Barbier et D. Maurer. Cet apprentissage a aussi concerné la première phase du traitement par PhytoImage (création des vignettes) des images acquises par le FlowCAM. Ces formations ont été intégrées au plan de formation de l'Ifremer pour 2011. Ce personnel pourra assurer ultérieurement le transfert de compétences vers d'autres laboratoires Ifremer.

²⁷ Lors d'un CDD de l'Université de Bordeaux 1, au laboratoire EPOC.

7.2. Perspectives

7.2.1. Enrichissement du set d'apprentissage et création d'un outil de reconnaissance « Atlantique-Manche »

Au commencement de cette étude, il avait été choisi de travailler uniquement sur la flore du bassin d'Arcachon. L'expérience acquise permet de penser que la création d'un set et d'un outil de reconnaissance dédiés aux communautés phytoplanctoniques de l'ensemble des côtes atlantiques et de la Manche, est tout à fait envisageable. Les banques d'images constituées au cours de ce travail contiennent déjà de nombreuses vignettes qui pourront servir à enrichir le set d'apprentissage actuel, en y incluant les vignettes provenant des prélèvements de Concarneau et en y ajoutant celles obtenues lors de la numérisation de nouveaux échantillons REPHY d'Arcachon et d'autres LERs (action poursuivie par le personnel d'Arcachon depuis avril 2011).

Ceci permettra de compléter les groupes du set d'apprentissage pour lesquels le nombre d'images est encore trop faible, concernant notamment les taxons de dinoflagellés, et de créer de nouveaux groupes au fur et à mesure que les espèces rares ou plus occasionnelles apparaîtront dans ces nouveaux échantillons. Cet enrichissement pourrait se faire assez rapidement compte tenu du matériel déjà engrangé. L'optimisation de l'outil de reconnaissance créé à partir de ce nouveau set d'apprentissage « Atlantique-Manche » (version 3) sera assuré par l'UMONS dans le cadre de la collaboration entre l'IFREMER et l'UMONS citée en début de rapport (2010).

Il convient aussi de poursuivre les travaux sur l'aspect quantitatif du comptage automatisé et sur la résolution des biais mis en évidence. Ensuite, il sera nécessaire de :

- réévaluer **les performances de reconnaissance automatique** de l'outil, en calculant manuellement les taux de vrais-positifs et de faux-positifs sur de nouveaux échantillons (les taux d'erreur annoncés à partir de la validation croisée ne reflétant pas les performances de reconnaissance automatique sur un nouvel échantillon) ; cette étape pourra être automatisée grâce au logiciel PhytoImage ;
- **valider les performances** de l'outil final concernant **l'estimation automatique (ou semi-automatique) de l'abondance**, en comparant les résultats avec les dénombrements REPHY.

7.2.2. Nouveaux outils dans PhytoImage

7.2.2.1. Correction d'erreur

Parmi les pistes actuellement explorées, la correction d'erreur, ou plus précisément la modélisation et l'intégration des biais systématiques mesurables, et donc corrigibles, apparaît comme une technique complémentaire prometteuse.

Par ailleurs, plusieurs fonctions nouvelles doivent être ajoutées au logiciel afin de faciliter la manipulation de celui-ci et la validation manuelle des classifications automatiques générées. Notamment, la méthode de validation partielle des individus sélectionnés sur la base de leur probabilité de reconnaissance incorrecte, présentée par K. Denis lors du dernier congrès ASLO en février 2011, doit permettre de diminuer

considérablement le temps alloué à cette vérification. Grâce au calcul d'une probabilité d'appartenance à un groupe, Ph. Grosjean et K. Denis ont en effet démontré qu'il était possible d'améliorer significativement les performances de reconnaissance automatique en ne reclassant qu'une faible proportion des images considérées comme les plus douteuses par l'algorithme. Cette approche particulière offre donc un compromis idéal entre la méthode automatique et la validation manuelle totale, tout en garantissant des performances de reconnaissance identiques ou supérieures. En plus de l'emploi des probabilités *a posteriori* retournées par la méthode de classification elle-même, l'utilisation de critères bio-écologiques (par exemple : est-il plausible de rencontrer cette espèce-là à cet endroit-là et à ce moment-là de l'année?) peut également être envisagée pour pointer automatiquement des identifications suspectes.

7.2.2.2. Espèces coloniales

Rappelons que les cytomètres de flux basés sur l'analyse d'image et utilisés en écologie marine, ne disposent pas encore de systèmes qui permettent de dénombrer automatiquement les cellules à l'intérieur de colonies (Olson et Sosik, 2007). Une première étude menée à l'UMONS (Govaerts, 2010) montre que l'information permettant de quantifier le nombre de cellules d'une colonie phytoplanctonique est présente dans les variables mesurées automatiquement sur les vignettes par PhytoImage, et que, moyennant la calibration de modèles prédictifs relativement simples comme la régression linéaire ou l'analyse discriminante linéaire, il serait possible de compter de manière automatisée les cellules de la majorité des taxons phytoplanctoniques rencontrés. Les espèces coloniales étant très fréquentes dans les échantillons, c'est un travail qui devra être mené en priorité si l'on veut comparer les résultats obtenus en semi-automatique avec ceux des dénombrements REPHY.

7.2.3. Utilisation de l'assemblage objectif 4X/cellule de flux 300µm

Un nouveau set d'apprentissage pourrait être constitué à partir de la numérisation des échantillons du REPHY à l'aide de l'objectif 4X (en plus de celle à l'objectif 10x). Bien que la résolution des images soit fortement réduite, cet assemblage permet de numériser les échantillons beaucoup plus rapidement (grâce à la cellule de flux correspondante qui est plus épaisse). Ceci présente trois avantages :

- constitution des catégories du set d'apprentissage plus rapide,
- détection plus rapide de taxons rares,
- numérisation d'un plus grand volume d'échantillon offrant une meilleure précision de comptage (meilleure estimation de l'abondance pour chaque groupe),
- meilleur flux (ou circulation ou passage) des cellules les plus grandes.

Ce nouveau set d'apprentissage proposerait une identification à un niveau taxinomique moins élevé que le précédent, du fait de la perte d'information inhérente à la résolution plus faible des images. Les taxons les plus petits seraient difficilement identifiés, donc probablement absents du set d'apprentissage. Cependant, ce set

d'apprentissage pourrait offrir un outil intéressant venant compléter le premier, notamment pour quantifier précisément les espèces les plus grandes. La numérisation double d'un échantillon sur les deux assemblages apporterait donc deux informations complémentaires balayant l'ensemble des classes de taille du microphytoplancton.

7.2.4. Utilisation du set d'apprentissage 10X sur des images numérisées en 4X

Une alternative intéressante pour diminuer le temps d'analyse des échantillons tout en conservant une précision taxinomique élevée va être étudiée dès septembre 2011 par Ph. Grosjean. Elle consiste à dégrader les images du set d'apprentissage acquises avec l'objectif 10X de manière à obtenir une qualité et un format d'image comparable à celles issues d'une numérisation avec l'objectif 4X. Cela permettrait d'utiliser le set d'apprentissage actuel pour classer automatiquement les particules d'échantillons numérisés avec l'objectif 4X, en bénéficiant ainsi d'un temps d'analyse bien plus rapide, tout en conservant la finesse des informations taxinomiques contenues dans le set d'apprentissage 10X. Si cette approche montre des résultats probants, elle constituera une solution idéale en terme de rapport de gain entre performances de reconnaissance et temps d'analyse.

Conclusion

Les résultats présentés dans ce rapport établissent que l'approche combinée de deux technologies, le FlowCAM pour la numérisation des particules, et le logiciel PhytoImage (développé par l'UMONS) pour la classification automatique, offre des perspectives très prometteuses pour la reconnaissance et le dénombrement automatiques, ou semi-automatiques, des espèces phytoplanctoniques. Nous avons ainsi montré qu'il était possible, à partir d'un set d'apprentissage provisoire (comportant 44 catégories dont 28 phytoplanctoniques), d'atteindre un niveau de reconnaissance global de 72% pour l'ensemble des catégories du set. Ce résultat très encourageant n'est pas définitif, puisque de nombreuses améliorations peuvent être apportées, tant sur le plan méthodologique que sur le plan informatique. Ainsi, une approche semi-automatique, incluant une validation manuelle partielle des images classées automatiquement, permettrait d'augmenter sensiblement les performances de reconnaissance.

Par ailleurs, si la numérisation d'échantillons REPHY du bassin d'Arcachon et d'autres sites était poursuivie, le set d'apprentissage initial pourrait s'enrichir relativement rapidement de nouveaux groupes, permettant d'espérer discriminer automatiquement la majorité des taxons reconnus actuellement par les analystes du REPHY, avec le minimum de perte de résolution taxinomique.

Ainsi, au terme du projet, il est envisageable de créer un outil de reconnaissance du phytoplancton pour la côte Manche-Atlantique, dont les performances devront être évaluées. Il est possible que des performances voisines de celles d'un comptage en routine au microscope puissent être atteintes. Cependant, il convient de poursuivre les travaux sur l'aspect quantitatif du comptage automatisé et sur la résolution des biais mis en évidence. C'est la prochaine étape à mener pour déterminer sous quelle forme la mise en œuvre de cette méthode pourrait être utilisée en routine pour le suivi des Flores Totales du REPHY de l'ensemble du littoral français.

Les méthodes automatisées présentent de nombreux atouts qui s'avèrent incontournables dans un proche avenir. En premier lieu, elles permettent de quantifier l'erreur associée à la mesure ce qui paraît indispensable pour toute accréditation des résultats. En effet, l'erreur commise par un observateur humain pour la même tâche n'est jamais évaluée et serait extrêmement difficile, voire impossible, à chiffrer. Par ailleurs, la numérisation des données phytoplanctoniques permet un stockage peu encombrant et quasiment illimité de ces données, en offrant à tout moment la possibilité d'une vérification (contrôle de la qualité) ou d'une nouvelle analyse (par exemple avec un outil de reconnaissance plus performant obtenu ultérieurement). De plus, l'analyse d'images amène un nombre considérable d'informations nouvelles, paramètres de taille et de forme, qui sont très difficiles à extraire en routine au cours d'un suivi classique du phytoplancton et qui sont très intéressantes pour toutes les études écologiques. Enfin il ne faut pas oublier les perspectives de durée d'analyse et de coût moindres.

Pour ces différentes raisons, il paraît essentiel de poursuivre les travaux engagés lors de ce post-doctorat, en privilégiant les pistes énoncées ci-dessus pour améliorer les performances. Les interruptions successives du contrat de l'unique personne formée à l'IFREMER pour mener à bien le projet ont occasionné une perte de temps importante. Il

serait nécessaire de consacrer à ce projet des moyens en personnel suffisants, pour qu'il puisse aboutir dans des délais raisonnables, et pérennes pour en assurer la continuité.

Dès juin 2010, ces préoccupations avaient été communiquées au CORS lors de la réunion qui s'est tenue à Arcachon. Le CORS avait fait la recommandation suivante dans son compte rendu de réunion :

« Le CORS s'inquiète de la pérennité à court et moyen terme des compétences nécessaires à la poursuite du développement de la reconnaissance automatisée du phytoplancton. Il recommande de revoir avec la DRH le point de blocage concernant la poursuite en CDD à partir de janvier 2011, d'Alina Tunin-Ley au regard de l'importance du projet et de la rareté de sa compétence en France. »

Plus tard, lors de la réunion Outils Identification et Quantification du Plancton qui s'est tenue en janvier 2011 à Boulogne (Belin *et al.*, 2011), l'attention avait aussi été portée sur ce problème :

*« En conclusion, il est indispensable de maintenir à Ifremer un plein temps sur l'amélioration des sets d'apprentissage (et la constitution de nouveaux sets), la participation à l'évolution du logiciel par la description de spécifications techniques, les actions de formation et de soutien aux autres utilisateurs. En outre les évolutions du matériel et du logiciel sont encore nombreuses, et il est nécessaire de coordonner ces évolutions dans les années qui viennent. Ce plein temps, assuré à ce jour par une personne en CDD hébergée hors Ifremer, et dont le contrat se termine à court terme, doit se transformer en un plein temps pérenne en interne Ifremer, dans le cadre du pôle expertise phytoplancton dont les missions doivent être élargies. **Faute de cela, le projet serait en grand danger.** »*

Communications

Journées REPHY (2008, IFREMER Nantes), restitution ASCOBAR (2008, Talence) et CORS (2010, Arcachon)

Tunin-Ley, A. Automatisation du suivi des communautés planctoniques : Utilisation du FlowCAM couplé au logiciel PhytoImage. Présentation orale.

Colloque ICCS (International Conference on Computational Science) 2009 (Baton Rouge, Louisiana, U.S.A.)

Denis K., Tunin-Ley A., Fernandes J. A., Maurer D., Parent J.-Y., Belin C., Irogoien X., Grosjean Ph. (2009). FlowCAM/PhytoImage intercalibration exercise. Présentation orale.

Meeting ASLO 2011 (San Juan, Porto Rico)

Tunin-Ley A., Maurer D., Denis K., Belin, C., Grosjean, P. Coupling optical detection by FlowCAM to automatic classification by the software Zoo/PhytoImage for automation of phytoplankton community identification. *ASLO 2011 Aquatic Science Meeting*, San Juan, Puerto Rico, 2011, February 13-28. Présentation orale.

Denis K., Tunin-Ley A., Maurer, D., Grosjean, P. Statistical error correction of Zoo/PhytoImage identification by partial manual validation of suspect particles. *ASLO 2011 Aquatic Science Meeting*, San Juan, Puerto Rico, 2011, February 13-28. Présentation orale.

Maurer D., Tunin-Ley A., Denis K., Barbier C., Pouvreau, S., Grosjean P. Semi-automated identification of planktonic larvae of the oyster *Crassostrea gigas* using FlowCAM coupled to the image analysis software Zoo/PhytoImage. *ASLO 2011 Aquatic Science Meeting*, San Juan, Puerto Rico, 2011, February 13-28. Présentation orale.

Congrès de la Société Phycologique de France 2011 (Villefranche-sur-Mer)

Tunin-Ley A., Denis K., Maurer D., Chomérat N., Belin, C., Grosjean P. Vers une automatisation du suivi des communautés phytoplanctoniques : utilisation du FlowCAM couplé au logiciel PhytoImage. Congrès annuel de la Société phycologique de France, Villefranche-sur-Mer, 2011, Avril 4-5. Présentation orale.

Documents internes Ifremer

Belin C., Rolin J.F., Haberkorn H., Maurer D. (2011). Compte rendu de la Réunion Outils Identification et Quantification du Plancton, Boulogne, 11 janvier 2011 : 9 p.

Convention IFREMER-ONEMA (2009). Fiche action n°11. Mise au point de méthodes pour la surveillance du phytoplancton : 2 p + annexe 5 p. Université de Mons Hainaut (Belgique) pour le développement d'outils logiciels couplés au FlowCAM : analyse automatisée d'échantillons de phytoplancton, septembre 2008, 5 p. + annexe technique 4 p.

Contrat de collaboration IFREMER-UMONS (2008). Collaboration entre l'IFREMER et l'Université de Mons Hainaut (Belgique) pour le développement d'outils logiciels couplés au FlowCAM : analyse automatisée d'échantillons de phytoplancton. Lettre-contrat n° 08/5210090 : 5 p. + annexe technique 4 p.

Contrat de collaboration IFREMER-UMONS (2010). Collaboration entre l'IFREMER et l'Université de Mons Hainaut (Belgique) pour l'évolution d'outils logiciels couplés au FlowCAM : amélioration de l'analyse automatisée d'échantillons de phytoplancton, décembre 2010. Lettre-contrat n° 10/5210681 : 7 p. + annexe technique 6 p.

Grosjean P., Denis K. (2010). Zoo/PhytoImage. Optimisation du set d'apprentissage REPHY. Rapport d'activités Convention Ifremer/ONEMA. Action n° 11 - Mise au point de méthodes pour la surveillance du phytoplancton, septembre 2010 : 44 p.

Lefebvre A., Cugier P. (2010). Comité d'Orientation de la Recherche et de la Surveillance CORS. Journée Thématique : Evolution à long terme des habitats, 1 et 2 juin 2010, Arcachon. Rapport : 25 p.

Maurer D., Tunin-Ley A. (2011). Rapport de mission IFREMER Colloque ASLO (Advancing the Science of Limnology and Oceanography), février 2011, San Juan (Porto Rico) : 5 p.

Tunin-Ley A. (2009). Numérisation des espèces phytoplanctoniques par FlowCAM et reconnaissance de ces espèces par analyse d'images par le logiciel PhytoImage. Rapport d'activités Convention Ifremer/ONEMA. Action n° 11 : Mise au point de méthodes pour la surveillance du phytoplancton, novembre 2009 : 111 p (incluant 6 annexes).

Tunin-Ley A. (2010). Numérisation des espèces phytoplanctoniques par FlowCAM et reconnaissance de ces espèces par analyse d'images par le logiciel PhytoImage. Rapport d'activités Convention Ifremer/ONEMA. Action n° 11 : Mise au point de méthodes pour la surveillance du phytoplancton, septembre 2010 : 5 p.

Bibliographie

- Breiman L.B. (2001). Random forests. *Machine learning*, 45 (1): 5-32.
- Campbell J.R., Buskey E.J. (2011). Imaging technology and microplankton monitoring in the Mission-Aransas National Estuarine Research Reserve. *ASLO 2011 Aquatic Science Meeting*, San Juan, Puerto Rico, 2011, February 13-28.
- Campbell L., Olson R.J., Sosik H.M., Abraham A., Henrichs D.W., Hyatt C.J., Buskey E.J. (2010). First harmful Dinophysis (Dinophyceae, Dinophysales) bloom in the U.S. is revealed by automated imaging flow cytometry. *Journal of Phycology*, 46: 66–75.
- Choi J.W. Stoecker D.K. (1989). Effects of Fixation on Cell Volume of Marine Planktonic Protozoa. *Applied and Environmental Microbiology*, 55: 1761-1765.
- Culverhouse P.F., Williams R., Reguera B., Herry V., Gonzalez-Gil S. (2003). Do expert make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, 247: 17-25.
- Denis K. (2005). Etude de la diversité du plancton au large du récif de Tuléar (Madagascar) par analyse d'image. Mémoire de fin d'études Licence en Sciences Biologiques, Laboratoire d'Ecologie Numérique des Milieux Aquatiques, Université de Mons Hainaut : 122 p + Annexes.
- Denis K., Grosjean P. (2006). Reconnaissance automatique du phytoplancton par analyse d'images numériques (PhytoImage). Première approche en utilisant la banque d'images IFREMER. Rapport Université de Mons Hainaut. Laboratoire d'Ecologie Numérique des Milieux Aquatiques : 88 p.
- Govaerts P. (2010). Comptage automatique du nombre de cellules par colonies de phytoplancton de la Mer du Nord à l'aide du FlowCAM et de PhytoImage. Mémoire de fin d'étude présenté en vue de l'obtention du diplôme du Grade de Maître en Biologie : Finalité Approfondie. Université de Mons Hainaut, UMONS – Service d'Écologie Numérique des Écosystèmes Aquatiques : 101 p.
- Lancelot C., Rousseau V., Lacroix G., Denis K., Gypens N., Grosjean P., Van Nieuwenhove K., Desmit J.Y., Terseleer Lillo N., Ruddick K., Delbare D. (2011). Combined effect of changing hydroclimate and human activity on coastal ecosystem health – AMORE III. Final report. Brussels : Belgian Science Policy 2011 – 52 p. (Research Program Science for a sustainable Development).
- Menden-Deuer S., Lessard E.J. Satterberg J. (2001). Effect of preservation on dinoflagellate and diatom cell volume and consequences for carbon biomass predictions. *Marine Ecology Progress Series*, 222: 41-50.
- Olson R.J., Sosik H.M. (2007). A submersible imaging-in-flow instrument to analyze nano and microplankton: Imaging FlowCytobot. *Limnology and Oceanography: Methods* 5, 195–203.
- Sosik H.M., Olson R.J. (2007). Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography: Methods* 5, 204-216.
- Sutor M. (2011). Application of imaging instrument and semi-automated classification techniques for plankton analysis: an overview. *ASLO 2011 Aquatic Science Meeting*, San Juan, Puerto Rico, 2011, February 13-28.

Tezzo X. (2008). Identification automatisée du phytoplankton de la baie sud de la mer du Nord par imagerie numérique en cytométrie en flux (FlowCAM) associée à des techniques de classification supervisée. Mémoire de fin d'études en vue de l'obtention du Grade de Bioingénieur : sciences agronomiques. Année académique 2007-2008 Université Libre de Bruxelles : 89 p.

Utermöhl H. (1958). Zur vervollkommung der quantitativen phytolankton-methodik. *Mitteilungen der Internationalen Vereinigung für Limnologie*, 9: 1-38.

Zarauz L., Irigoien X. (2008). Effects of Lugol's fixation on the size structure of natural nano-microplankton samples, analysed by means of an automatic counting method. *Journal of Plankton Research*, 30 (11): 1297-1303.

Annexes

Annexe 1 – Proposition de sujet de post-doctorat pour une demande de bourse post-doctorale

1. FICHE SYNTHETIQUE

- Titre du sujet de post-doctorat proposé

Mise en œuvre opérationnelle d'analyses de la composition phytoplanctonique d'échantillons d'eau de mer grâce à une analyse automatisée ou semi-automatisée, en utilisant un FlowCAM couplé à une version spécifique du logiciel PhytoImage, développé par l'Université de Mons-Hainaut.

- Responsables scientifiques Ifremer

LER / Arcachon : Danièle Maurer / encadrant du laboratoire d'accueil
LER / Concarneau : Nicolas Chomérat / expert taxinomiste

- Programme Ifremer de rattachement

SE3L

- Projet contractualisé de rattachement

Nouvelles technologies

- Unité (laboratoire ou service) et département d'accueil

Ifremer / LER / AR, Arcachon

- Co-financement envisagé

- Résumé et mots-clés en français

Le réseau français de surveillance du Phytoplancton REPHY acquiert depuis vingt ans des séries de données sur les populations phytoplanctoniques sur de nombreux points du littoral. La méthode d'identification et de dénombrement utilisée étant très fastidieuse, il est nécessaire d'en améliorer largement les performances. L'objectif du post doctorat est la mise en œuvre opérationnelle d'un système couplé de numérisation et de traitement d'images de phytoplancton : après numérisation par un FlowCAM, le traitement des images sera réalisé avec une version spécifique du logiciel PhytoImage développé par l'Université de Mons Hainaut (Pr Philippe Grosjean). La première étape du travail consiste en une prise en main du FlowCAM (en lien avec l'Université de Mons Hainaut) et en la rédaction d'un cahier des charges des objectifs à atteindre et des contraintes à respecter. La deuxième phase est la constitution d'un set d'apprentissage, banque d'images numérisées et représentatives des taxons rencontrés dans les échantillons du REPHY, pour permettre à l'Université de finaliser une version optimisée de PhytoImage. Une troisième étape est l'analyse des résultats acquis en temps réel avec cette version optimisée, ainsi que la mise en œuvre opérationnelle proprement dite (rédaction de procédures, formation des observateurs phytoplancton). La collaboration de qualité avec l'Université de Mons Hainaut, ainsi que le caractère innovant (l'utilisation d'un tel outil dans un programme de surveillance national serait une première) sont des points forts de ce projet.

- Titre, résumé et mots-clés en anglais

Title

Operational implementation of automated or semi-automated analysis of phytoplankton composition in sea water samples, with FlowCAM coupled to a specific version of PhytoImage software, developed by Mons-Hainaut University (Belgium).

Abstract

The French phytoplankton monitoring network (REPHY) has acquired phytoplankton times series for twenty years on numerous sampling stations along the coast. The method used for

identification and counting is very time-consuming, and it is necessary to improve it. The objective of this post doctorat is the implementation of a coupled system of digitisation and phytoplankton picture analysis : after digitisation by a FlowCAM, the picture analysis will be performed with a specific version of PhytoImage software developed by Mons-Hainaut University (Pr Philippe Grosjean). The first step of the work consists in training to FlowCAM use (with the University), and in writing requirements (objectives to reach, constraints). A second step is to build a training set of phytoplankton pictures, which are representative of taxa encountered in REPHY samples, in order to allow the University to finalize a new version of PhytoImage. A third step is the analysis of the results acquired in real time with this new version, and the operational implementation (procedures writing, training for phytoplankton technicians). The premium collaboration with Mons-Hainaut University, as well as the innovative aspect (the use of such a tool in a monitoring network would be very new) are strong points of this project.

- Profil de candidature souhaité

Les compétences requises sont principalement des compétences pointues en taxinomie du phytoplancton marin, ainsi qu'une motivation pour le traitement d'images.

A sharp competence in marine phytoplankton taxonomy is required, as well as motivation for picture analysis.

2. PROGRAMME DE RECHERCHE DETAILLE

2.1. Programme Ifremer de rattachement

SE3L (Surveillance et Evaluation de l'Etat des Eaux Littorales)

2.2. Exposé du projet

Contexte

Le Réseau de Surveillance du Phytoplancton et des Phycotoxines (REPHY), créé en 1984 par l'Ifremer, couvre l'ensemble du littoral français métropolitain. Ses objectifs sont à la fois environnementaux et sanitaires :

- la connaissance de la biomasse, de l'abondance et de la composition du phytoplancton marin des eaux côtières et lagunaires, qui recouvre notamment celle de la distribution spatio-temporelle des différentes espèces phytoplanctoniques, le recensement des efflorescences exceptionnelles telles que les développements d'espèces toxiques ou nuisibles susceptibles d'affecter l'écosystème,
- la détection et le suivi des espèces phytoplanctoniques productrices de toxines susceptibles de s'accumuler dans les produits marins de consommation, et la recherche de ces toxines dans les mollusques bivalves présents dans les zones de production ou dans les gisements naturels.

L'aspect environnemental de cette surveillance permet de répondre en particulier aux obligations de la Directive européenne Cadre sur l'Eau (DCE), et aux conventions internationales telles que la convention d'Oslo et de Paris (OSPAR) ou la convention de Barcelone. Ces données sont également de plus en plus utilisées pour des programmes de recherche.

La surveillance du phytoplancton est organisée dans le REPHY de sorte qu'elle puisse répondre à des questions diverses. En particulier, l'acquisition sur une trentaine de points du littoral, de séries temporelles de données comprenant la totalité des taxons phytoplanctoniques présents, permet d'établir des liens avec les problèmes liés à l'eutrophisation, au changement climatique, ou à une dégradation de l'écosystème, de calculer des indicateurs pour une estimation de la qualité de l'eau d'un point de vue diversité biologique, de détecter des nouvelles espèces, ou de revenir sur les observations historiques d'espèces rares.



L'acquisition sur un nombre conséquent de points supplémentaires, de séries de données sur les espèces qui « bloom », permet de façon complémentaire de calculer des indicateurs pour une estimation de la qualité de l'eau d'un point de vue abondance et composition, et de suivre l'évolution des blooms, espèce par espèce.

Ces deux types d'observations phytoplanctoniques sont actuellement effectués sur un total d'environ 130 points de prélèvement du REPHY, répartis sur tout le littoral. L'échantillonnage est réalisé toute l'année, à une fréquence d'une fois par quinzaine ou d'une fois par mois. Les dénombrements des taxons phytoplanctoniques pouvant être identifiés dans les conditions d'observation, sont effectués sur un échantillon d'eau de mer de 10 ml, au microscope optique inversé par la méthode d'Utermöhl (1958). L'identification doit se faire au plus précis (espèce ou genre) si possible, sinon à un niveau taxinomique supérieur (genre, famille, voire ordre ou classe), éventuellement à un niveau intermédiaire (groupe de genres ou d'espèces). Une liste de référence des taxons potentiellement identifiables est utilisée pour la saisie dans la base de données Quadriga.

Problématique et justification du sujet

Cette méthode d'identification et de dénombrement du phytoplancton étant très consommatrice de temps, il a depuis longtemps été envisagé d'utiliser des méthodes permettant d'améliorer les performances, soit en terme de rapidité (le même nombre d'échantillons en un temps plus réduit), soit en terme d'efficacité (plus d'échantillons dans le même temps, ce qui permettrait d'augmenter la précision des dénombrements). Les méthodes les plus prometteuses à court terme sont celles qui consistent à traiter des images numérisées, car elles sont actuellement les seules potentiellement capables de fournir une information sur l'ensemble des taxons phytoplanctoniques, à l'instar de ce qui se fait dans le cadre du REPHY.

Ce sujet de post doctorat s'inscrit donc dans cette problématique, puisqu'il a pour objectif la mise en œuvre opérationnelle d'un système couplé de numérisation et de traitement d'images de phytoplancton, afin de pouvoir analyser la composition phytoplanctonique d'échantillons fixés (lugolés) et frais. La numérisation des images de phytoplancton sera assurée par un FlowCAM, couplé à une version spécifique et adaptée pour ces besoins, du logiciel PhytoImage, afin de permettre l'analyse automatisée ou semi-automatisée des échantillons.

Collaboration avec un laboratoire extérieur

Ce travail se déroulera dans le cadre d'une convention de collaboration (2008-2009) entre l'Ifremer et l'Université de Mons Hainaut en Belgique (Pr Philippe Grosjean), qui a développé un logiciel de traitement des images de plancton (PhytoImage : <http://www.sciviews.org/zooimage/>). Une étude préliminaire menée en 2006 par l'Université sur la base de microphotographies de phytoplancton fournies par Ifremer, avait conduit à des résultats suffisamment intéressants pour que l'adaptation et l'extension du logiciel PhytoImage puissent être envisagées, afin de fournir un outil final utilisable dans le cadre du REPHY. Cette version spécifique du logiciel sera fournie sous licence GPL et sera librement utilisable par l'Ifremer sous les termes de cette licence : en particulier le code source complet sera fourni et librement utilisable par l'Ifremer et par ses partenaires.

Par ailleurs, le matériel FlowCAM, co-financé par l'Université de Bordeaux 1 (UMR-EPOC) pour 60% de son coût, sera bien positionné dans les locaux d'Ifremer Arcachon.

Compétences requises

L'adaptation du logiciel étant assurée par l'Université de Mons Hainaut, la compétence requise pour ce travail est essentiellement une compétence taxinomique : il s'agit en effet de fournir un maximum d'informations sur les taxons à identifier, afin que le logiciel puisse faire son apprentissage dans un premier temps, et être utilisé dans un deuxième temps de la façon la plus optimale possible. Pour cet aspect le post doctorant pourra compter sur le personnel observateur phytoplancton du laboratoire accueillant, et sera encadré par un expert sur la taxinomie du phytoplancton, appartenant au laboratoire Ifremer de Concarneau.



Par ailleurs, les informations étant acquises grâce à la numérisation d'images effectuées avec le matériel FlowCAM, le post doctorant devra se former à l'utilisation de cet appareil, afin d'en maîtriser à terme tous les aspects, et pouvoir devenir lui même un formateur potentiel pour d'autres observateurs phytoplancton du REPHY.

Approches méthodologiques

Le sujet sera mené en plusieurs étapes.

1. Formation et intercalibration

La numérisation des images de phytoplancton sera effectuée avec un matériel identique (FlowCAM) à Ifremer et à l'Université de Mons-Hainaut. Le FlowCAM (<http://www.fluidimaging.com>) est un appareil produit par la société Fluid Imaging qui numérise des images de particules phytoplanctoniques en cytométrie de flux (l'échantillon passe dans une cellule de mesure où des capteurs détectent son passage et déclenchent la prise de photographies à l'aide d'une microcaméra numérique). De nombreux paramètres doivent être réglés et maîtrisés, ce qui implique un apprentissage qui sera assuré par l'Université. Des habitudes de travail similaires devant être adoptées par les deux équipes de l'Ifremer et de l'Université, une intercalibration des deux appareils sera réalisée simultanément, à l'aide d'échantillons de phytoplancton d'origines différentes.

Cette phase du travail comprendra également la rédaction d'un cahier des charges décrivant les objectifs à atteindre en terme de performance (niveau de d'identification, pourcentage de reconnaissance minimal, etc) et les contraintes à respecter (par ex : traitement d'échantillons lugolés).

2. Vers l'adaptation de PhytoImage pour son utilisation dans le cadre du REPHY : constitution d'un set d'apprentissage.

Dans une deuxième phase, un set d'apprentissage sera constitué : en effet PhytoImage n'est pleinement utilisable qu'après la réalisation d'un apprentissage à la reconnaissance des particules étudiées. Le set d'apprentissage consistera donc en une banque d'images issues de la numérisation par le FlowCAM, et représentatives des particules rencontrées dans les échantillons à analyser ultérieurement. Ces images seront identifiées à l'œil et classées dans autant de dossiers ou sous-dossiers que nécessaires pour représenter les niveaux d'identification de plus en plus détaillés souhaités. Pour effectuer un apprentissage valable à la reconnaissance d'un groupe, il faut au minimum 50 images de diverses particules de ces groupes (si possible, issues d'échantillons différents). Lorsque l'abondance des particules le permet, un ensemble de 100 à 200 images est idéal par groupe. Ces images seront accompagnées des données de mensurations.

Ce set d'apprentissage fourni à l'Université de Mons Hainaut lui permettra de finaliser l'outil de reconnaissance spécifique aux besoins de l'Ifremer et adapté aux contraintes décrites dans le cahier des charges.

3. Vers un outil de reconnaissance du phytoplancton spécifique aux images acquises par le FlowCAM dans le cadre du REPHY

Pendant la phase de finalisation de l'outil, le post doctorant assurera les échanges avec l'Université de Mons Hainaut (demandes d'images supplémentaires, questions diverses, etc). Il organisera également une ou des sessions de formation pour des observateurs phytoplancton afin de préparer une mise en œuvre ultérieure étendue d'autres laboratoires Ifremer. Des procédures d'utilisation spécifiques au fonctionnement de l'outil dans le cadre du REPHY seront également rédigées.

Après fourniture par l'Université de la version de PhytoImage optimisée, la suite du travail consistera en l'analyse des résultats acquis en temps réel sur les échantillons du REPHY pendant quelques mois, afin de prévoir une nouvelle évolution du logiciel en fonction des besoins identifiés à l'utilisation (par ex spécialisation d'outils de reconnaissance différents en fonction de la zone géographique, de la saison, ou de besoins particuliers)

Originalité et caractère innovant des recherches

Ce type de méthode d'analyse des échantillons de phytoplancton n'est actuellement utilisée en routine dans aucun autre programme de surveillance du phytoplancton en Europe, ni dans le monde. La mise en œuvre opérationnelle d'un outil tel que celui là dans le cadre du REPHY serait donc une première. L'originalité du travail et la coopération de qualité avec l'Université de Mons-Hainaut, pourront conduire à une publication et/ou communication.

Résultats attendus

L'objectif d'une future utilisation de cet outil d'analyse des échantillons de phytoplancton dans le cadre d'un réseau de surveillance implique que tout soit fait pour que le logiciel réponde le mieux possible aux contraintes et aux exigences qui auront été décrites auparavant : la constitution du set d'apprentissage est donc une étape cruciale. Par ailleurs, les résultats attendus au terme du post doctorat sont également une mise en œuvre effective de l'outil dans le laboratoire accueillant (Arcachon), comprenant une prise en main et une appropriation par les observateurs phytoplancton de ce laboratoire.

Applications possibles

L'expérience et le savoir-faire acquis pourront facilement être valorisés, dans le cadre de programmes de surveillance ou de recherches impliquant l'acquisition de nombreuses données phytoplancton, en France (Universités, Insu stations marines, etc), ou à l'étranger (programmes de surveillance de type DCE par exemple).

3. ENCADREMENT SCIENTIFIQUE

3.1. Proposant du sujet / Premier co-responsable scientifique du post-doctorat

- Nom, prénom

Danièle Maurer

- Fonction et spécialité à l'Ifremer

Cadre de recherche

Ecologie générale, écosystème conchylicole, phytoplancton

- Diplômes (Doctorat d'Etat, HDR)

Ingénieur agronome ENSA Montpellier (1975)

Doctorat en Océanologie Biologique (1978) Université Aix-Marseille II; "Phytoplancton et pollution : la lagune Ebrié (Côte d'Ivoire) et le secteur de Cortiou (Marseille)".

- Centre Ifremer de

Nantes (Station rattachée d'Arcachon)

- Département

Laboratoire Environnement et Ressources (LER)

- Unité (laboratoire ou service)

LER Arcachon

- Tél. : 05 57 72 29 89 **E-mail** : dmaurer@ifremer.fr

Expérience d'encadrements d'étudiants

- Nom des post-doctorants dirigés antérieurement (bourses Ifremer et autres) :

Co-encadrement en cours (mai 2006–mai 2008) avec Benoît Sautour (Université de Bordeaux I) de Béatrice Bec, post doctorante de l'Université de Bordeaux I :

"Identification et répartition spatio-temporelle du phytoplancton potentiellement toxique dans le Bassin d'Arcachon et sur le proche plateau continental".

- Situation professionnelle actuelle des doctorants et post-doctorants encadrés (si connue) :

Nom et signature du co-responsable scientifique

Danièle Maurer



3.2. Proposant du sujet / Deuxième co-responsable scientifique du post-doctorat :

- Nom, prénom

Nicolas Chomérat

- Fonction et spécialité à l'Ifremer

C I, taxinomiste des micro-algues

- Diplômes (Doctorat d'Etat, HDR)

Doctorat en Sciences de l'Environnement (spécialité Écologie, Phycologie), obtenu en novembre 2005 à l'Université Aix-Marseille III

- Centre Ifremer de

Brest (Station rattachée de Concarneau)

- Département

Laboratoire Environnement et Ressources (LER)

- Unité (laboratoire ou service)

LER Finistère Bretagne Nord (site de Concarneau)

- Tél. : 02 98 97 43 38 **E-mail :** nicolas.chomerat@ifremer.fr

Expérience d'encadrements d'étudiants

- Nom des post-doctorants dirigés antérieurement (bourses Ifremer et autres) :

Aucune expérience de ce type pour l'instant

- Situation professionnelle actuelle des doctorants et post-doctorants encadrés (si connue) :

Nom et signature du co-responsable scientifique

Nicolas Chomérat



3.3. Avis concertés des responsables d'accueil (unité et département Ifremer)

- **Intitulé de l'Unité (laboratoire ou service) :** Ifremer / LER / AR - Arcachon
- **Nom du responsable du laboratoire ou du service :** Jean Paul Dreno
- **Nom de la secrétaire :** Florence Trut
- **Centre Ifremer de :** Nantes
- **Tél. :** **E-mail :** dopler.ar@ifremer.fr
- **Tél. (secrétariat) :** 05 57.72.29.80
- **Intitulé du Département :** Ifremer / DOP / LER
- **Nom du Responsable de Département :** Luc Drévès
- **Nom de la secrétaire :** Marie-Pierre Le Bris
- **Centre Ifremer de :** Brest
- **Tél. :** **E-mail :** dopler@ifremer.fr
- **Tél. (secrétariat) :** 02 98 22 44 66
- **Expérience acquise par l'Unité d'accueil :**
Très bonne expertise en matière de taxonomie du phytoplancton dans le cadre du réseau REPHY.
- **Moyens nécessaires et disponibles pour la réalisation du travail :**
En collaboration avec LER/FBN, encadrement du post-doc et fournitures d'échantillons.
- **Période d'accueil envisagée :**
à partir d'octobre - novembre 2008
- **Avis concertés des Responsables d'Unité et du Département d'accueil :**
Le candidat sera accueilli au LER/AR-Arcachon
Son encadrement sera assuré en collaboration avec les spécialistes du phytoplancton du LER/FBN-Concarneau.
Avis très favorable sur un projet finalisé très attendu par les LER, leur permettant une optimisation quantitative et qualitative de l'observation du phytoplancton littoral.

Nom et signature du responsable de l'unité (lab. ou service)et/ou du département d'accueil

Le Chef du LER-AR
Jean Paul DRENO



Le Responsable du département
Luc DRÉVÈS



3.3. Avis concertés du responsable du programme et de la Direction des Programmes et de la Coordination des Projets (DPCP)

Cette proposition de post doctorat est intéressante à plusieurs titres : elle répond à une demande formulée depuis longtemps dans le cadre du REPHY, pour alléger les tâches demandées aux observateurs du phytoplancton, elle correspond à un intérêt actuellement accru pour ce type de technique dans le monde des chercheurs travaillant en écologie du phytoplancton, enfin elle s'inscrit dans le cadre d'une collaboration avec une équipe reconnue de l'Université de Mons-Hainaut, collaboration qui s'est déjà révélée très fructueuse.

L'importance, maintenant internationalement reconnue, des séries temporelles sur le phytoplancton qui ont été acquises par le REPHY depuis vingt ans, n'est plus à justifier : ces séries sont utilisées aussi bien pour répondre à des demandes de type réglementaire –par exemple les indicateurs de qualité de la Directive Cadre sur l'Eau, ou des questions des administrations sanitaires-, que pour une utilisation dans le cadre de programmes de recherche –les demandes de données sont en augmentation constante en interne Ifremer, mais aussi de plus en plus en externe Ifremer-.

Les techniques de numérisation des images, en elles-mêmes prometteuses pour un certain nombre de domaines, ne sont cependant convaincantes pour le phytoplancton qu'avec un matériel adapté, et un traitement des images capable de répondre aux très fortes contraintes posées par ces organismes (certaines cellules de phytoplancton sont en effet particulièrement difficiles à discriminer). La combinaison d'un matériel (FlowCAM) qui a fait ses preuves dans la communauté internationale, et d'un logiciel (PhytoImage) qui a déjà montré sa capacité à évoluer, est une garantie de succès.

L'investissement d'Ifremer se doit d'être à la hauteur, afin de faire avancer ce projet qui stagne depuis presque deux ans faute de moyens humains : un post doctorat est donc tout à fait adapté, pour assurer la complémentarité avec l'équipe de l'Université de Mons-Hainaut.



Michel MARCHAND

Responsable du Programme SE3L

(Surveillance et Evaluation de l'Etat des Eaux Littorales)

Patrick VINCENT

Directeur DPCP

(Programmes et Coordination des Projets)

Annexe 2 – Travail préliminaire sur le genre *Dinophysis* (Tunin-Ley, 2009)



Avant propos :

Les sets d'apprentissage et outils de reconnaissance sont présentés ici à titre d'exemples. Dans tous les cas, le nombre d'images n'est pas suffisant pour représenter significativement les groupes taxonomiques, sauf quelques exceptions. C'est pourquoi je n'ai pas analysé trop en détail ces résultats. Pour les mêmes raisons, je n'ai pas trop expliqué les étapes de l'analyse d'images, ni les choix taxonomiques.

Ces exemples permettent cependant d'illustrer la manière dont fonctionne la reconnaissance automatique avec le logiciel Zoo/PhytoImage, et de donner quelques pistes de réflexions pour la constitution du set d'apprentissage final destiné à l'analyse de la flore totale des échantillons du REPHY. Différents sets d'apprentissage ont été créés, l'un focalisé sur le genre *Dinophysis*, basé sur les échantillons de la campagne Arcadino de mai 2008, l'autre concernant le genre *Ceratium*, à partir d'échantillons méditerranéens prélevés au filet (vide de maille 53 µm). Les sets d'apprentissage ont été constitués dans le but de créer des outils de reconnaissance spécialisés, contrairement au set d'apprentissage final qui sera un outil de reconnaissance généraliste.

Pour chaque étude sont présentés :

- le nombre d'image par taxon,
- le taux d'erreur global ainsi que les taux d'erreur spécifiques de chaque taxon,
- la matrice de confusion,
- les statistiques concernant les prédictions erronées (faux-négatifs et faux-positifs).

Training set test *Dinophysis*

--> Paramètres d'expérimentation

| | |
|--------------------|----------------------------|
| Prélèvement | filtration 2 l sur 20µm |
| Lugol | oui |
| Assemblage | 10X/100µm |
| Vitesse | Purge |
| Mode | Autolmage |

--> Training set détaillé

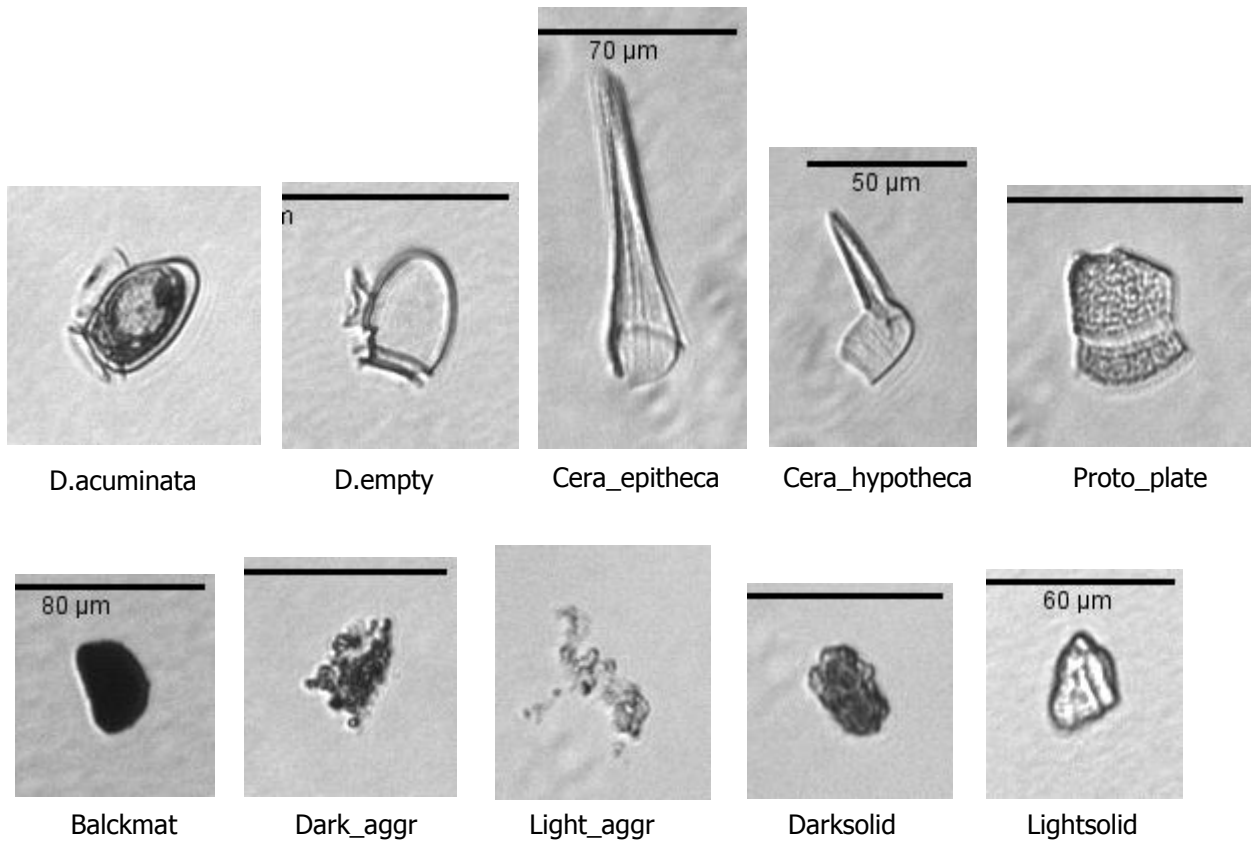
Deux catégories de *Dinophysis* ont pu être créées (pour lesquelles le nombre d'images est suffisant) : *D. acuminata* et les cellules de *Dinophysis* vides (D.empty), nombreuses dans les échantillons.

Les autres espèces du phytoplancton pour lesquelles les images étaient assez nombreuses ont été identifiées au niveau taxonomique le plus précis.

Les débris ont été caractérisés par un maximum de catégories, en fonction des particules largement représentées dans les échantillons. Ont été ainsi distingués :

- les particules noires opaques ("blackmat")
- les agrégats foncés aux contours flous et irréguliers ("Dark_aggr")
- les agrégats translucides aux contours flous et irréguliers ("Light_aggr")
- les particules denses foncées ("Darksolid")
- les particules denses translucides ("Lightsolid")

- les fibres ("fiber")
- les plaques thécales de *Protoperdinium* ("Proto_plate") et de *Ceratium* ("Cera_epitheca" et "Cera_hypotheca"), abondantes dans les échantillons.



Classification stats (nombre d'images par groupe) :

| | | |
|------------------|----|----------------------|
| Blackmat | 9 | Débris |
| C.furca | 19 | Dinophysis |
| C.kofoidii | 12 | Diatomées |
| Cera_epitheca | 14 | Autres dinoflagellés |
| Cera_hypotheca | 15 | |
| Chaetoceros | 42 | |
| D.acuminata | 72 | |
| D.empty | 43 | |
| Dark_aggr | 49 | |
| Darksolid | 40 | |
| Fiber | 15 | |
| L.danicus | 56 | |
| Light_aggr | 18 | |
| Lightsolid | 85 | |
| Proto_plate | 60 | |
| Pseudonitzschia | 18 | |
| Rhizosoleniaceae | 85 | |

k-fold cross validation error estimation (k = 10): 29.6%

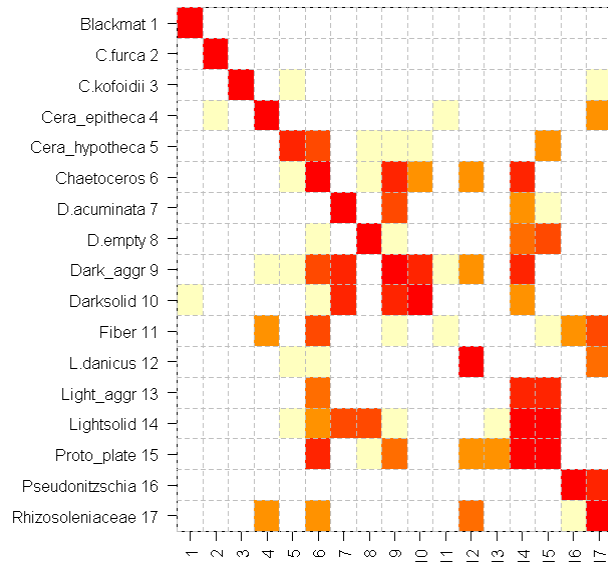
Error per class:

| | Error (%) |
|------------------|-----------|
| Blackmat | 0 |
| C.furca | 0 |
| L.danicus | 8.93 |
| Rhizosoleniaceae | 9.41 |
| D.acuminata | 9.72 |
| C.kofoidii | 16.67 |
| D.empty | 20.93 |
| Lightsolid | 25.88 |
| Cera_epitheca | 28.57 |
| Pseudonitzschia | 38.89 |
| Darksolid | 40 |
| Chaetoceros | 40.48 |
| Proto_plate | 45 |
| Dark_aggr | 57.14 |
| Cera_hypotheca | 60 |
| Fiber | 93.33 |
| Light_aggr | 100 |

Matrice de confusion :

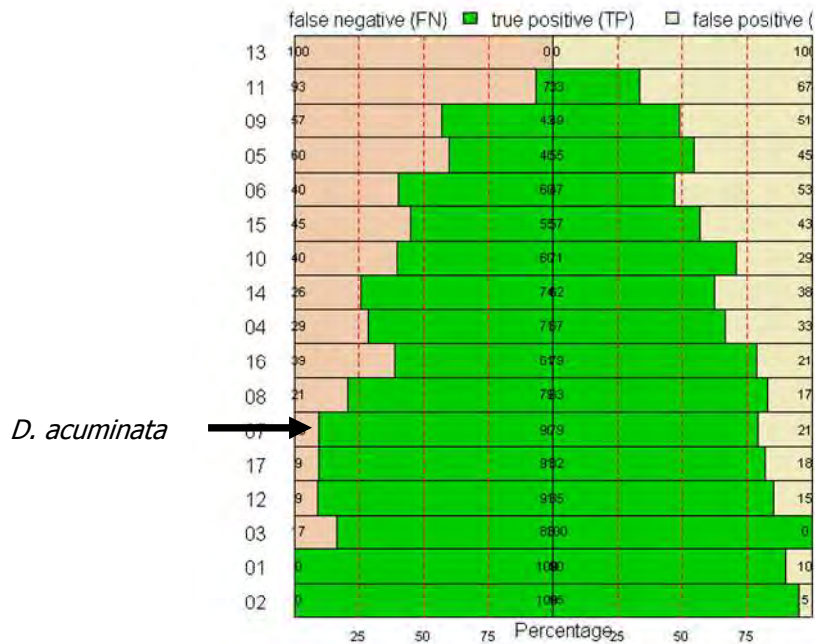
Pour rappel, la matrice de confusion permet de mettre en évidence les sources de confusion entre taxons, en indiquant le nombre d'objets correctement identifiés par l'algorithme, et les identifications erronées pour chaque taxon. En ligne sont classés les groupes identifiés manuellement, et en colonne les groupe prédits par l'outil de reconnaissance. Les nombres qui constituent la diagonale de la matrice correspondent donc aux identifications correctes de l'outil de reconnaissance. Les chiffres en dehors de cette diagonale correspondent aux erreurs d'identification de l'outil de reconnaissance, avec en ligne les faux-négatifs pour chaque taxon, et en colonne les faux-positifs.

| Classes | predicted | | | | | | | | | | | | | | | | |
|---------------------|-----------|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 Blackmat | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 C.furca | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 C.kofoidii | 0 | 0 | 10 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 Cera_epitheca | 0 | 1 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 5 Cera_hypotheca | 0 | 0 | 0 | 0 | 6 | 4 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| 6 Chaetoceros | 0 | 0 | 0 | 0 | 1 | 25 | 0 | 1 | 5 | 2 | 0 | 2 | 0 | 6 | 0 | 0 | 0 |
| 7 D.acuminata | 0 | 0 | 0 | 0 | 0 | 0 | 65 | 0 | 4 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 |
| 8 D.empty | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 34 | 1 | 0 | 0 | 0 | 0 | 3 | 4 | 0 | 0 |
| 9 Dark_aggr | 0 | 0 | 0 | 1 | 1 | 4 | 7 | 0 | 21 | 7 | 1 | 2 | 0 | 5 | 0 | 0 | 0 |
| 10 Darksolid | 1 | 0 | 0 | 0 | 0 | 1 | 6 | 0 | 6 | 24 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 11 Fiber | 0 | 0 | 0 | 2 | 0 | 4 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 4 |
| 12 L.danicus | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 51 | 0 | 0 | 0 | 0 | 3 |
| 13 Light_aggr | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 8 | 0 | 0 |
| 14 Lightsolid | 0 | 0 | 0 | 0 | 1 | 2 | 4 | 4 | 1 | 0 | 0 | 0 | 1 | 63 | 9 | 0 | 0 |
| 15 Proto_plate | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 1 | 3 | 0 | 0 | 2 | 2 | 13 | 33 | 0 | 0 |
| 16 Pseudonitzschia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 7 |
| 17 Rhizosoleniaceae | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 77 |



Faux positifs et faux négatifs :

Rappel : Pour un taxon A, les faux-négatifs sont les images de A qui sont identifiées comme appartenant à d'autres taxons, tandis que les faux-positifs sont des particules n'appartenant pas au taxon A mais identifiées comme telles.



Pour *D. acuminata*, il y a 10% de faux-négatifs.

Les *Dinophysis* vides sont moins bien identifiés (confondus avec *D. acuminata*, les thèques de *Ceratium* et les agrégats clairs).

Deux groupes sont très mal identifiés : les fibres et les agrégats translucides.

--> Training set simplifié

Des modifications ont été apportées au set d'apprentissage précédent, notamment au niveau des groupes qui étaient largement confondus :

- les débris foncés ont été regroupés dans un même groupe indépendamment de leur aspect
- les débris translucides ont été regroupés, y compris les plaques thécales de *Proto-peridinium*
- *Pseudo-nitzschia* et Rhizosoleniaceae ont été regroupées
- les fibres ont été enlevées du set d'apprentissage, puisqu'elles ne sont jamais identifiées comme appartenant aux catégories du genre *Dinophysis*

Classification stats (nombre d'images par groupe) :

| | | |
|------------------|-----|----------------------|
| C.furca | 19 | Débris |
| C.kofoidii | 12 | Dinophysis |
| Cera_epitheca | 14 | Diatomées |
| Cera_hypotheca | 15 | Autres dinoflagellés |
| Chaetoceros | 42 | |
| D.acuminata | 72 | |
| D.empty | 43 | |
| Dark_particle | 97 | |
| L.danicus | 56 | |
| Light_particle | 163 | |
| Rhizosol_Pseudon | 103 | |

k-fold cross validation error estimation (k = 10): 18.08%

| Error per class: | Error (%) | Error per class: | Error (%) |
|------------------|-----------|------------------|-----------|
| C.furca | 0 | Blackmat | 0 |
| Rhizosol_Pseudon | 4.85 | C.furca | 0 |
| Light_particle | 9.2 | L.danicus | 8.93 |
| L.danicus | 10.71 | Rhizosoleniaceae | 9.41 |
| D.acuminata | 15.28 | D.acuminata | 9.72 |
| C.kofoidii | 16.67 | C.kofoidii | 16.67 |
| D.empty | 25.58 | D.empty | 20.93 |
| Cera_epitheca | 28.57 | Lightsolid | 25.88 |
| Dark_particle | 29.9 | Cera_epitheca | 28.57 |
| Chaetoceros | 50 | Pseudonitzschia | 38.89 |
| Cera_hypotheca | 73.33 | Darksolid | 40 |
| | | Chaetoceros | 40.48 |
| | | Proto_plate | 45 |
| | | Dark_aggr | 57.14 |
| | | Cera_hypotheca | 60 |
| | | Fiber | 93.33 |
| | | Light_aggr | 100 |

Le taux d'erreur global a diminué, mais les taux d'erreur pour les deux catégories de *Dinophysis* a augmenté (car le nombre d'images dans les groupes prêtant à confusion a augmenté du fait des regroupements).

Pour diminuer ce taux spécifique, il faut augmenter le nombre d'images pour *Dinophysis*.

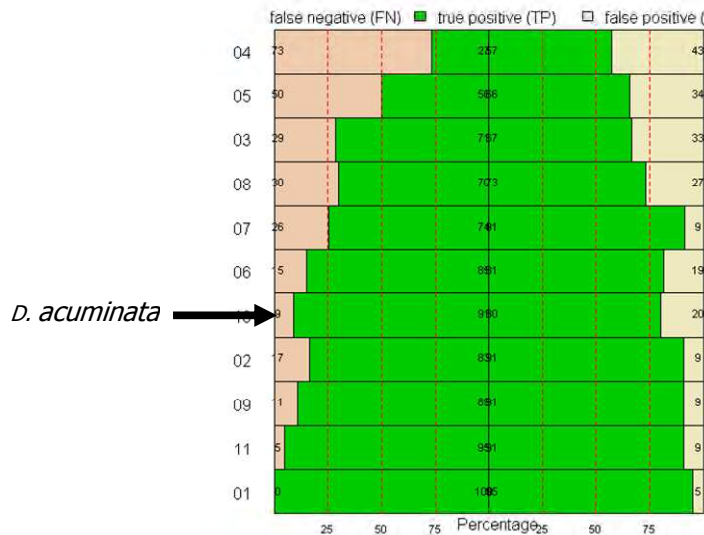
En revanche le taux d'erreur a diminué pour d'autres groupes : les Rhizosoleniaceae et *Pseudo-nitzschia*, les particules translucides, les particules sombres.

Matrice de confusion :

| Classes | predicted | | | | | | | | | | |
|----------------------|-----------|----|----|---|----|----|----|----|----|-----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 C.furca | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 C.kofoidii | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 Cera_epitheca | 1 | 0 | 10 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 4 Cera_hypotheca | 0 | 0 | 1 | 4 | 3 | 0 | 2 | 1 | 0 | 4 | 0 |
| 5 Chaetoceros | 0 | 0 | 0 | 1 | 21 | 0 | 0 | 8 | 1 | 9 | 2 |
| 6 <i>D.acuminata</i> | 0 | 0 | 0 | 0 | 0 | 61 | 0 | 9 | 0 | 2 | 0 |
| 7 D.empty | 0 | 0 | 0 | 0 | 1 | 0 | 32 | 1 | 0 | 9 | 0 |
| 8 Dark_particle | 0 | 0 | 2 | 0 | 1 | 12 | 0 | 68 | 2 | 12 | 0 |
| 9 L.danicus | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 50 | 0 | 5 |
| 10 Light_particle | 0 | 0 | 0 | 1 | 4 | 2 | 1 | 6 | 1 | 148 | 0 |
| 11 Rhizosol_Pseudon | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 98 |

D. acuminata est majoritairement confondu avec les particules translucides et les particules sombres.

Faux positifs et faux négatifs :



Le nombre de faux négatifs pour *D. acuminata* est passé de 10 à 15%.

--> **Améliorations à viser :**

Le nombre d'images représentatives de *D. acuminata* est encore insuffisant pour rendre compte pertinemment de la variabilité de l'espèce, et en comparaison avec le nombre d'images des groupes qui prêtent à confusion.

Le set d'apprentissage doit être retravaillé pour diminuer les faux-négatifs de *D. acuminata*, dans le but de créer un outil de reconnaissance spécifique de *Dinophysis*, en mode semi-automatique. Dans cette optique, les catégories de débris, souvent sources d'une importante confusion avec les taxons d'intérêt, doivent être soigneusement définies.

Il reste à acquérir des images supplémentaires pour les autres espèces de *Dinophysis*, afin d'évaluer l'efficacité de reconnaissance de l'outil au niveau intra-générique.

**Annexe 3 – Protocole FlowCAM pour la constitution du set d'apprentissage
REPHY (flore totale) (Tunin-Ley, 2009)**

Les indications suivantes concernent uniquement le FlowCAM hébergé au LER d'Arcachon, dans le cadre du projet de reconnaissance automatique des échantillons REPHY. Le protocole et les réglages proposés ne sont pas forcément applicables à un autre FlowCAM ou à un autre projet.

1. Préparation de l'échantillon

- Fixation et conservation

Le prélèvement et la fixation des échantillons suit le protocole établi pour le REPHY, soit l'utilisation de Lugol acide (préparation du Lugol acide détaillée dans le document **arC10mo01v00**) pour la conservation des échantillons (1 ml de Lugol pour 1 litre d'échantillon). Les échantillons ne sont habituellement pas gardés sur le long terme, lorsqu'ils ne présentent pas d'intérêt floristique particulier.

Pour l'usage du FlowCAM, il est nécessaire de passer l'échantillon **dans les 15 jours suivant le prélèvement**, pour éviter une trop grande incidence de la décoloration sur la reconnaissance automatique.

Pour une conservation plus longue, il est important de réaliser un sous-échantillon qui sera conservé à l'obscurité et au frais, dans un récipient en **verre**.

- Pré-filtration

Afin de limiter le colmatage éventuel des cellules et de diminuer le nombre de groupes non ciblés dans le set d'apprentissage, une pré-filtration systématique des échantillons s'impose. Une **pré-filtration sur filtre de vide de maille de 1 mm** doit être appliquée à chaque échantillon.

- Concentration / Dilution

Dans le cas où l'échantillon à analyser présente une très forte concentration en particules, il est nécessaire de le diluer dans de l'eau de mer filtrée avant le passage dans le FlowCAM, pour éviter d'une part le colmatage des cellules de flux et d'autre part une possible saturation de l'ordinateur au moment de l'acquisition d'images.

De la même manière, si l'échantillon est très faiblement concentré, il est envisageable de le concentrer sur filtre ou par sédimentation.

Dans les deux cas, il est important d'indiquer dans le tableau des métadonnées (fichier Excel) le coefficient de concentration ou de dilution appliqué.

2. Réglages du FlowCAM

- Assemblage objectif / cellule

Le set d'apprentissage REPHY sera réalisé uniquement en utilisant l'assemblage **objectif 10X/cellule 100 µm**. Les éventuelles espèces dont la taille et la forme ne permettraient pas le passage dans la cellule de 100 µm sont a priori peu nombreuses (*Noctiluca* essentiellement). Les espèces trop grandes pour être numérisées avec un tel dispositif seront dénombrées manuellement, uniquement dans le cas où elles seraient présentes en concentrations élevées.

- Réglage de l'illumination

L'illumination est réglée à l'aide de l'histogramme de niveaux de gris. Le bon réglage correspond au mode de l'histogramme centré vers les niveaux de gris 180 et un étalement d'environ 50 niveaux de gris maximum.

Il faut vérifier la répartition des niveaux de gris après chaque nouveau montage (option "Raw Pixel Intensity").

Si nécessaire, la LED peut être délicatement repositionnée à l'aide des trois vis pour obtenir la distribution de niveaux de gris adéquate. Une fois que la courbe est obtenue, l'intensité lumineuse doit être ajustée à l'aide des fonctions "gain" et "flash duration" pour placer le mode de l'histogramme à 180.

- Mode de détection

L'analyse des échantillons REPHY doit être réalisée en **mode *AutoImage***. Le calcul de l'abondance n'étant pas utile pour la constitution du set d'apprentissage, il n'est pas nécessaire de définir précisément la valeur du nombre d'images prises par seconde (fonction "Flash delay", pouvant varier de 1 à 12 images par seconde). Les éventuels doublons étant aisément identifiables, donc faciles à écarter au cours de la classification manuelle, il est conseillé de régler la paramètre "**Flash delay**" **entre 9 et 12 images par seconde**, pour permettre l'acquisition d'un maximum d'images.

- Filtre logiciel de taille des particules

Le filtre de taille à appliquer dans Visual Spreadsheet pour la numérisation des échantillons REPHY se base sur des mesures en **mode ABD (Area Based Diameter)**. La gamme de taille du filtre doit être définie entre **15 et 2000 µm**. En mode ABD, une limite inférieure de 15 µm permet de numériser les particules dont la longueur est supérieure à 19 µm environ, ce qui correspond au seuil de taille appliqué par les analystes phytoplancton du réseau REPHY.

Une limite supérieure de 2000 µm en mode ABD permet de numériser l'ensemble du phytoplancton, y compris les plus grosses cellules.

- Capture d'images

L'ensemble des configurations possibles de particules pour un même taxon doit être représenté dans le set d'apprentissage. Il est donc important de numériser toutes les particules, y compris celles qui sont floues ou qui sont coupées sur les bords de l'image.

Pour la **zone acceptable de numérisation** des particules, les dimensions suivantes doivent être indiquées : **0 - 1023- 0 -767**. Les particules tronquées par le bord du champ de la caméra seront ainsi acceptées, et pourront être intégrées au set d'apprentissage.

- Mise au point

Ce réglage doit être effectué après tout nouveau montage de la cellule du FlowCAM, avant la numérisation des échantillons. Après avoir réalisé la mise au point vers le plan médian de la cellule de flux, à mi-chemin entre les deux parois en verre, il est nécessaire d'affiner le réglage en utilisant un petit volume d'échantillon à analyser. A l'aide de la vis micrométrique, la mise au point est ajustée en utilisant le mode "**AutoImage - No Save**", qui permet de visualiser en temps réel les images numérisées, sans les sauvegarder. Le réglage est satisfaisant lorsque le nombre de particules nettes est maximum (au jugé, par observation visuelle des collages).

Important : la mise au point ne doit surtout pas être modifiée au cours d'une numérisation. Si les images sont floues, il est préférable d'abandonner la numérisation en cours et de recommencer les réglages.

2. Numérisation de l'échantillon

- Calibration des niveaux de gris

Avant chaque numérisation d'échantillon, il est nécessaire d'effectuer une calibration des niveaux de gris, à l'aide des **filtres gris neutres ND2, ND4, ND8, et du filtre noir opaque**.

Au cours d'une même numérisation, alors que la cellule de flux est remplie d'eau de mer filtrée, les filtres gris neutres sont successivement positionnés devant la source lumineuse et laissés en place pendant 3-4 secondes. Le fichier correspondant à la calibration doit porter le même nom que celui de l'échantillon, le suffixe "_NDx" étant ajouté à la fin. Ces images seront utiles pour calibrer les niveaux de gris des images en densité optique.

- Volume d'échantillon à numériser

Pour la phase de constitution du set d'apprentissage, le volume est indiqué à titre informatif, puisque l'abondance n'a pas besoin d'être connue. Le volume d'échantillon à introduire dépend de la concentration initiale de l'échantillon et des taxons ciblés, et reste donc fonction des objectifs de l'utilisateur.

- Dilution systématique pour pallier la sédimentation des particules

Afin de numériser l'ensemble des particules, y compris celles qui sédimentent le plus vite, il suffit d'introduire dans la cellule de flux un volume connu d'eau de mer filtrée, puis d'ajouter dans la continuité l'échantillon à analyser et de lancer l'acquisition d'images immédiatement après. Cette méthode assure que toutes les particules de l'échantillon ont la même probabilité d'être détectées. Pour l'assemblage objectif 10X/cellule 100 µm, le volume à introduire avant l'échantillon est de **500 µl**.

- Début de l'acquisition

En premier lieu, les 500 µl d'eau de mer filtrée sont introduits dans l'entonnoir puis pompés jusqu'à remplir complètement la cellule de flux, en actionnant la pompe péristaltique. En utilisant le mode "Setup and Focus", l'utilisateur s'assure alors que le champ de la caméra est net de tout débris ou défaut et qu'aucune bulle ne vient gêner la calibration du fond. Si des bulles sont observées, il convient de les éliminer soit en pinçant plusieurs fois le tube de sortie, soit en actionnant la pompe en mode "Reverse". L'acquisition est lancée dès que l'échantillon a été introduit dans l'entonnoir à la suite des 500 µl d'eau de mer filtrée.

- Fin de l'acquisition

L'acquisition est stoppée lorsque la fin de l'échantillon (bulle) atteint l'entrée de la cellule de flux, de manière à numériser les quelques cellules qui tendent à rester en surface.

La durée de l'acquisition (en secondes) doit être reportée dans le fichier de métadonnées (feuille Excel).

Annexe 4 – Proposition de protocole pour l'utilisation du FlowCAM et la constitution d'un set d'apprentissage à partir des échantillons du REPHY (juillet 2009) (Tunin-Ley, 2009)

. 1 Utilisation du FlowCAM

- . 1 Fixation des échantillons
- . 2 Assemblages objectif/cellule
- . 3 Réglage de l'illumination
- . 4 Modes de détection
- . 5 Filtres de taille
- . 6 Paramètres de capture d'images
- . 7 Focalisation des particules
- . 8 Calibration des niveaux de gris
- . 9 Numérisation d'un échantillon
- . 10 Volume à numériser
- . 11 Temps de passage d'un échantillon
- . 12 Rinçage des cellules

. 2 Utilisation du logiciel Zoo/PhytoImage

- . 1 Constitution du set d'apprentissage
- . 2 Demandes spécifiques de l'Ifremer concernant le logiciel Zoo/PhytoImage

1. Utilisation du FlowCAM

1.1. Fixation des échantillons

L'utilisation du Lugol pour fixer les échantillons du REPHY reste probablement la plus appropriée, et ne semble pas gênante pour la numérisation des images et la reconnaissance des particules. Étant données les importantes modifications qu'un changement de protocole entraînerait pour les LER de l'Ifremer, il paraît plus judicieux de conserver cette méthode de fixation. Pour les échantillons servant à réaliser le set d'apprentissage, une analyse dans les 15 jours suivant l'échantillonnage paraît raisonnable.

Pour une conservation à plus long terme, des sous-échantillons devraient être conservés dans des flacons en verre afin de limiter la décoloration due au plastique des flacons habituellement utilisés.

1.2. Assemblages objectif/cellule

- Objectif 10X / cellule 100 µm

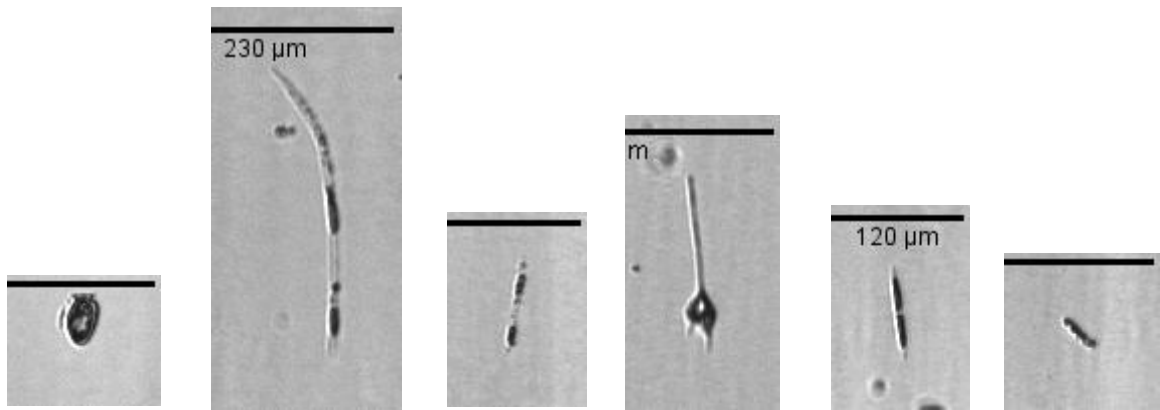
Cet assemblage permet d'obtenir les détails les plus fins, et la meilleure qualité d'image. Lorsque la mise au point est correctement réalisée, il est possible de discerner nettement les organes et les structures fines, telles que les soies ou les processus des diatomées, souvent nécessaires pour une identification au niveau spécifique. Les plus petits organismes sont également mieux détectés.

- Protocole

L'utilisation d'un objectif de grossissement moindre (4X) associé à une cellule de plus grande profondeur peut s'avérer utile de manière occasionnelle dans le cas où certaines espèces de grande taille sont présentes en abondance, mais ne semble pas appropriée pour l'analyse en routine des échantillons du REPHY, car la résolution des images n'est pas suffisante pour identifier manuellement la majorité des espèces présentes à Arcachon. En effet, les échantillons de REPHY sont dominés par des espèces de taille relativement petite.

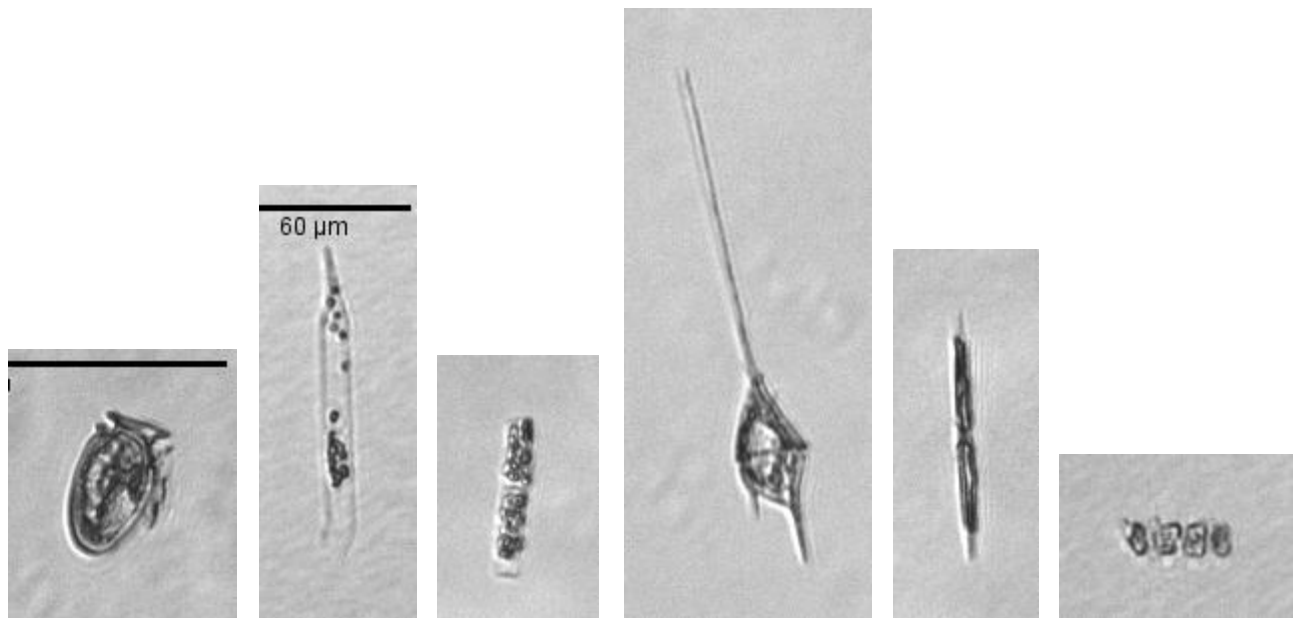
D'autre part, les tests ont montré que de nombreuses cellules dont la taille dépasse 100 µm (en terme de longueur maximale) passent à travers la cellule 100 µm sans problème apparent. La cellule de 100 µm mesurant 2 mm de largeur, la plupart des cellules parviennent à passer dans la cellule, car le flux tend à les positionner de telle sorte que la plus grande longueur soit orientée dans le sens du flux. L'assemblage objectif 10X/cellule 100 µm paraît donc le meilleur choix pour analyser les échantillons du REPHY avec le FlowCAM.

Images acquises avec l'objectif 4X :



Dinophysis acuminata *Proboscia alata* *Leptocylindrus* *Ceratium kofoidii* *Pseudo-nitzschia* *Chaetoceros*

Images acquises avec l'objectif 10X :



Dinophysis acuminata *Proboscia alata* *Leptocylindrus* *Ceratium kofoidii* *Pseudo-nitzschia* *Chaetoceros*

Cependant, l'utilisation préférentielle de l'assemblage objectif 10X/cellule 100 µm pose le problème des cellules de très grande taille et de certaines espèces coloniales (notamment les diatomées formant de longues chaînes ou présentant des soies très longues). En grand nombre, ces espèces ont tendance à s'accumuler à l'entrée de la cellule de flux, retenant probablement avec elles une partie des cellules des autres espèces, ce qui peut entraîner un biais dans l'estimation de l'abondance des espèces, dans le cas de l'analyse en routine des échantillons. Dans cette situation, il convient soit de diluer suffisamment l'échantillon afin de limiter le colmatage à l'entrée de la cellule de flux, soit d'utiliser une cellule plus profonde pour numériser les cellules de grande taille. Cette dernière alternative suppose l'élaboration de 2 sets d'apprentissage distincts correspondant aux deux assemblages utilisés, pour lesquels auront été préalablement définies les listes des espèces numérisées respectivement par chaque type d'assemblage. De plus, cela induit un doublement du temps consacré à la constitution du set

d'apprentissage, depuis l'étape de la numérisation de l'échantillon jusqu'à l'identification manuelle des images acquises.

Dans un premier temps, il paraît plus pertinent de choisir la dilution pour passer les échantillons riches en espèces coloniales ou de grandes tailles. Si cette méthode s'avère insuffisante au fur et à mesure que la composition spécifique des échantillons du REPHY évolue, il faudrait envisager de débiter un second set d'apprentissage ne concernant que les espèces de très grande taille qui ne peuvent pas passer par la cellule de 100 µm et/ou qui entraînent un colmatage à l'entrée de la cellule lorsque les concentrations sont fortes (espèces à définir ultérieurement au cas par cas ?).

Dans tous les cas, si la présence de grandes cellules ou de grandes colonies dans l'échantillon gêne la numérisation des espèces de taille plus petite, il sera nécessaire de réaliser une filtration pour permettre l'utilisation de la cellule de 100 µm.

1.3. Réglage de l'illumination

Cette étape est primordiale pour assurer une numérisation correcte et exploitable des particules, mais s'avère délicate avec le FlowCAM présent au LER d'Arcachon. En effet, le dispositif d'illumination, constitué d'une LED dont le centrage est réalisé à l'aide de trois vis, est particulièrement fragile, et nécessite une certaine prise en main. Notamment, les vis étant très courtes, il faut veiller à ne pas trop dévisser pour ne pas sortir les vis de leur pas. Les trois vis exercent une pression sur un anneau de centrage. La position de la LED dépend donc de la pression simultanée exercée par les 3 vis, et non pas de l'enfoncement de la vis dans le pas de vis. Le réglage est donc assez délicat, et doit être vérifié à chaque fois qu'une cellule de flux est installée, même s'il s'agit de la même cellule de flux qui a été démontrée et remontée. De même, deux cellules de même profondeur présentent des défauts variables au niveau des parois de verre qui influent différemment sur la diffusion de la source lumineuse. En pratique, il est donc difficile d'obtenir avec exactitude une distribution similaire des niveaux de gris entre deux montages.

- *Protocole*

Pour obtenir une illumination optimale du champ, il est préférable d'ajuster la position de la LED en jouant sur la pression des vis avec une faible illumination (utiliser les paramètres de gain et de durée de flash) en se basant sur la courbe de distribution des niveaux de gris (option "Raw Pixel Intensity"). Ainsi qu'il est expliqué dans le manuel d'utilisation du FlowCAM, la portion de la courbe correspondant au fond de l'image doit être la plus régulière et la plus étroite possible. Ensuite, il faut augmenter le gain et la durée du flash de manière à obtenir une illumination d'intensité moyenne supérieure à 180, et inférieure à 200 pour éviter une surexposition. Idéalement, l'illumination du fond doit s'étendre sur une cinquantaine de niveaux de gris environ, afin de laisser une large plage de niveaux de gris pour la caractérisation des particules d'intérêt. Cependant, ces réglages varient d'un montage à l'autre en fonction de la cellule utilisée ; il n'est donc pas possible de définir strictement les valeurs.

1.4. Modes de détection

- *AutoImage*

Le mode *AutoImage* (numérisation des particules à intervalle de temps régulier, défini par l'utilisateur) est applicable aux échantillons du REPHY, lorsque la concentration en phytoplancton est suffisante. Lorsque l'échantillon est faiblement concentré, ce mode ne permet pas de numériser efficacement l'ensemble des espèces, car les espèces les plus rares ont peu

de chances d'être numérisées. D'autre part, ce mode permet d'avoir une bonne estimation de l'abondance des cellules.

- *Trigger (scatter)*

Lorsque l'échantillon est faiblement concentré, le mode *Trigger* utilisant la détection par diffusion des particules (*scatter*) s'avère le plus pertinent. Il permet de détecter toutes les particules passant devant le laser dans la gamme de tailles spécifiée par l'utilisateur. Les espèces rares, très faiblement représentées dans l'échantillon, ont davantage de chance d'être détectées, ce qui permet de rendre compte plus précisément de la composition spécifique.

- *Trigger (fluorescence)*

Le mode de détection par fluorescence n'est a priori pas utilisable dans le cadre de l'analyse des échantillons du REPHY, puisque les échantillons sont habituellement fixés avec du lugol, qui interfère avec la fluorescence naturelle des pigments chlorophylliens. En outre, de nombreux genres notamment parmi les dinoflagellés sont mixotrophes ou complètement hétérotrophes, et ne sont donc pas reconnus efficacement en mode fluorescence. De la même manière, les organismes zooplanctoniques prédateurs du phytoplancton sont détectés lorsqu'ils contiennent une quantité suffisante de pigments chlorophylliens. Une étude discriminant les organismes hétérotrophes des espèces photosynthétiques ne semble donc pas évidente.

- *Protocole*

Dans le cadre de l'analyse des échantillons du REPHY, les modes *AutoImage* et *Scatter* sont à privilégier, l'utilisation de chacun des modes dépendant de la concentration du phytoplancton et du type d'étude. Pendant la phase de constitution du set d'apprentissage, les deux modes peuvent être utilisés. Lorsque l'échantillon est suffisamment concentré, le mode *AutoImage* doit être préféré, car il permet d'obtenir la meilleure précision en terme d'abondance cellulaire. Si l'on cherche à acquérir spécifiquement des images d'espèces plus rares dans l'échantillon, le mode *Scatter* peut s'avérer plus approprié. Pour l'analyse d'échantillons en routine, c'est le mode *AutoImage* qui est le plus approprié car il permet d'obtenir la meilleure estimation de l'abondance.

1.5. Filtres de taille

Seules les cellules du microphytoplancton dont la taille est supérieure à 20 μm sont comptabilisées par les analystes REPHY. Il est donc prévu d'appliquer un filtre de taille pour la numérisation des particules par le FlowCAM, correspondant aux gammes de taille respectées par les observateurs humains. Cependant, pour numériser effectivement les cellules supérieures à 20 μm , il faut appliquer un filtre avec un minimum de détection inférieur à 20 μm , car cette mesure correspond au diamètre d'un disque de surface équivalente à l'aire de la particule (en mode ABD). En pratique, des cellules présentant des parties très transparentes, parfois mal détectées, ou bien mal positionnées par rapport çà l'objectif, apparaissent de moindre taille qu'en réalité. Après plusieurs tests, il semble qu'une limite inférieure de **15 μm** soit appropriée pour détecter l'ensemble des espèces communément considérées comme appartenant au microphytoplancton.

La limite supérieure est fixée à **2000 μm** , certaines espèces dépassant facilement le mm (plusieurs espèces de *Ceratium* notamment).

Par la suite, il serait intéressant de pouvoir appliquer un deuxième filtre de taille, avant l'analyse d'image par ImageJ, de manière à ne conserver que les données concernant les particules de taille supérieure à 20 μm , peut-être en utilisant un autre paramètre que l'ABD,

comme le plus grand diamètre par exemple (voir plus loin dans la partie "Demandes spécifiques concernant le logiciel PhytoImage").

1.6. Paramètres de capture d'images

Les valeurs des paramètres de capture d'images fournies par Ph. Grosjean et K. Denis ont été utilisées pour les tests, et semblent convenir pour l'acquisition des images à partir des échantillons du REPHY et des échantillons des missions Arcadino.

Avec de fortes concentrations en particules, l'ordinateur du FlowCAM ne semble pas saturer au cours de l'acquisition, même lorsque le nombre d'images par seconde est défini au maximum (12 images par seconde). En revanche, en augmentant la valeur de ce paramètre, on observe aussi davantage de doublons (même particule numérisée deux fois), même s'ils restent peu nombreux si l'on considère le nombre total d'images. Ce paramètre peut donc être modifié au cas par cas pour la constitution du set d'apprentissage (les doublons peuvent être manuellement écartés), mais il faudra définir une valeur seuil pour l'analyse d'échantillons en routine.

D'autre part, en ce qui concerne la zone acceptable de numérisation des particules (c'est-à-dire si l'on accepte ou non les particules tronquées par le bord du champ), il semble plus logique de ne conserver que les particules entières dans un premier temps, puisque seules les images des cellules complètes seront intégrées au set d'apprentissage. Soit les dimensions suivantes : 1 - 1022 - 1 - 766.

1.7. Focalisation des particules

Afin d'optimiser le nombre de particules correctement focalisées, il est nécessaire de réaliser la mise au point au centre de la cellule de flux, en repérant l'une après l'autre les parois de la cellule puis la seconde (grâce aux aspérités naturelles du verre), puis en positionnant la vis micrométrique à mi-distance des deux parois. Cependant l'absence de repères sur la molette de la vis micrométrique et surtout le fait que la vis force davantage dans un sens que dans l'autre rendent cette opération relativement peu précise sur l'appareil qui équipe le LER d'Arcachon. Il est donc préférable de **réaliser la mise au point au début de chaque expérience**, en utilisant le mode *AutoImage (No Save)* ou *Scatter (No Save)* avec un petit volume d'échantillon à analyser, ce qui permet de visualiser directement la netteté des images acquises et de rectifier la position de la vis micrométrique si besoin.

D'autre part, selon leur taille, les cellules ne se positionnent pas au même endroit dans le flux, les plus grosses cellules étant plus nombreuses au centre du flux, alors que les particules de petite taille ont tendance à passer plus près des parois. Pour la constitution du set d'apprentissage, il est donc nécessaire **de réaliser plusieurs passages pour un même échantillon** en faisant varier la mise au point, afin d'obtenir une numérisation correcte des particules pour l'ensemble des gammes de taille considérées.

En aucun cas, la mise au point ne doit être modifiée au cours d'une numérisation, car les imperfections de l'arrière-plan, qui sont normalement considérées comme faisant partie du fond lors de sa calibration, risquent alors d'être détectées comme des particules d'intérêt, ou d'être ensuite prises en compte dans la segmentation des objets, rendant les images inutilisables pour le set d'apprentissage.

1.8. Calibration des niveaux de gris

Avant toute numérisation d'échantillon, il est nécessaire de procéder à une calibration des niveaux de gris, à l'aide de filtres gris neutres, de densité 2x, 4x, 8x et d'un filtre opaque noir (appliquer les filtres l'un après l'autre devant la source lumineuse pendant 4 secondes environ)

1.9. Numérisation d'un échantillon

Selon les espèces, la sédimentation des particules dans la cellule de flux s'avère très différente. Les grosses cellules et les particules sphériques ont tendance à sédimenter en premier, alors que d'autres cellules, possédant de grandes expansions ou des réserves lipidiques, sédimenteront beaucoup plus lentement, comme c'est le cas de *Dinophysis acuminata* qui s'accumule à la surface de l'échantillon (zone de tension superficielle). Pour numériser l'ensemble des espèces présentes dans l'échantillon, il est donc nécessaire de faire passer l'échantillon en entier dans la cellule de flux. Il faut cependant éviter la formation des bulles qui sont détectées aussitôt comme des particules, d'où la numérisation d'un très grand nombre d'images qui "polluent" le set d'acquisition.

- Protocole

Pour pallier ces problèmes, il est possible de passer avant l'échantillon **un volume connu d'eau de mer filtrée**, correspondant à peu près à la longueur de tube séparant la cellule de l'entonnoir, de façon à remplir la cellule en entier tout en s'assurant d'éliminer les bulles éventuellement formées. Le volume d'échantillon à analyser est ensuite introduit dans l'entonnoir en continuité avec l'eau de mer filtrée, et l'acquisition lancée tout de suite après. Les cellules sédimentant le plus vite sont ainsi détectées sans risque d'omission. L'échantillon passe en entier dans la cellule, et la numérisation est stoppée manuellement dès que la fin de l'échantillon atteint l'entrée de la cellule de flux, permettant ainsi de détecter les dernières particules susceptibles d'être retenues en surface à cause de la tension superficielle ou de leurs caractéristiques particulières.

L'utilisation d'un additif permettant d'augmenter la viscosité de l'eau pour homogénéiser la vitesse de sédimentation des différentes particules n'est pas conseillée, car cela pourrait entraîner une accumulation non désirée de ces particules le long des parois, biaisant l'estimation de l'abondance, et boucher éventuellement la cellule.

La méthode proposée exige de prendre en compte le volume d'eau de mer filtrée initialement introduit dans la cellule dans le calcul du **taux de dilution**, qui se combine au taux de dilution éventuellement appliqué à l'échantillon lui-même (dans le cas de concentrations élevées en particules par exemple). Bien que cette étape soit contraignante, elle paraît indispensable pour éviter la perte des cellules qui sédimentent le plus vite.

Volume d'eau de mer filtrée à introduire avant l'échantillon

| | |
|----------------|--------|
| Cellule 100 µm | 500 µl |
| Cellule 300 µm | 500 µl |
| Cellule 600 µm | 1 ml |

Pour limiter la sédimentation des particules, il faudrait envisager un dispositif d'homogénéisation automatique (agitation délicate) applicable au niveau de l'entonnoir.

1.10. Volume à numériser

Le volume d'échantillon à numériser est fonction de nombreux paramètres :

- **concentration en particules** : si l'échantillon est trop riche en particules, il y a un risque de colmatage de la cellule de flux ; il faut donc soit diluer l'échantillon, soit réaliser une pré-filtration, soit passer plusieurs volumes faibles d'un même échantillon et rincer entre-temps la cellule pour éliminer les agrégats formés.

- **espèces ciblées** : il se peut que les espèces ciblées soient présentes en faible concentration dans un échantillon par ailleurs riche en particules. Puisqu'il est impossible alors de concentrer, il peut être nécessaire de passer un grand volume d'échantillon pour acquérir suffisamment d'images de l'espèce d'intérêt.

- **type d'étude** : création d'un set d'apprentissage ou analyse en routine.

Lors de la constitution du set d'apprentissage, le but est d'acquérir un nombre d'images représentatif de la variété morphologique des différentes espèces. La notion d'abondance n'intervient donc pas ici, et le volume d'échantillon à numériser dépend des paramètres cités plus haut.

Pour l'analyse en routine des échantillons, il sera nécessaire de définir un volume minimal représentant significativement la communauté phytoplanctonique naturelle en termes de composition spécifique et d'abondances, en prenant en compte les contraintes de temps d'analyse.

En outre, les entonnoirs fournis avec le FlowCAM ont une contenance limitée (environ 8 ml pour l'entonnoir qui s'adapte aux cellules de 100 μm et 300 μm). Il faudra peut-être envisager d'utiliser un autre dispositif pour augmenter le volume analysable en une fois, notamment pour l'analyse en routine.

Le volume d'échantillon passé dans le FlowCAM doit donc être défini au cas par cas.

1.11. Temps de passage d'un échantillon

La contrainte de temps au niveau de l'acquisition d'images par le FlowCAM doit être précisément évaluée, car si le FlowCAM est présenté comme un outil très avantageux pour l'identification automatique du phytoplancton, le gain de temps ne paraît pas si évident. En effet, le débit de la pompe avec l'assemblage objectif 10X/cellule 100 μm est beaucoup plus lent que pour l'assemblage 4X/300 μm . Il faut environ 1h30 pour passer 5 ml d'échantillon dans la cellule de 100 μm . Si l'on augmente le diamètre du tube de sortie, il est possible d'augmenter le débit de la pompe, mais les images alors acquises avec l'objectif 10X sont presque toutes floues. Au grossissement 10X, la vitesse maximale (Purge) correspondant au montage tubes+cellule associés, fourni par le constructeur, est la plus appropriée. L'utilisation de cet objectif lors de la constitution du set d'apprentissage à partir des échantillons du REPHY implique donc de prendre en compte le **temps d'analyse relativement élevé**.

| | | | |
|------------------------------------|---------------|---------------|---------------|
| Assemblage | 10X/100µm | 4X/300µm | 4X/600µm |
| Vitesse | Purge | Purge | Purge |
| Running time (min) | 5 | 5 | 5 |
| Eau | Distillée | Mer Filtrée | Distillée |
| Densité | 1 | 1.02 | 1 |
| Masse en sortie (g) | 0.34 | 6.54 | 6.72 |
| Volume pompé (ml) | 0.34 | 6.38 | 6.72 |
| Débit (ml.min⁻¹) | 0.0672 | 1.2768 | 1.3447 |

Temps de passage pour 5 ml d'échantillon

| | |
|-----------|--------|
| 10X/100µm | 74 min |
| 4X/300µm | 4 min |
| 4X/600µm | 4 min |

1.12. Rinçage des cellules

Le rinçage des cellules de flux est primordial, pour éliminer les particules collées aux parois et éviter la formation de traces, qui sont susceptibles de gêner la détection et la segmentation des particules lors des expériences suivantes. La présence de traces sur les parois augmente aussi le risque de segmentation "aberrante" lié au mouvement de la cellule de flux pendant la numérisation de l'échantillon. Le séchage est tout aussi important, et doit être effectué systématiquement au moyen d'une soufflerie (une poire de pulvérisation par exemple).

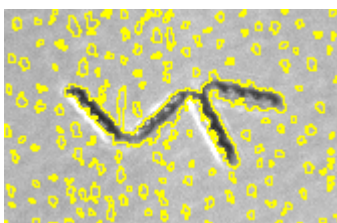
2. Utilisation du logiciel Zoo/PhytoImage

2.1. Constitution du set d'apprentissage

- Analyse d'images avec ImageJ

Une fois l'analyse d'images terminée, il est important de vérifier dans le répertoire contenant les silhouettes (répertoire "_outlines") si la segmentation des particules est correcte. En effet, malgré le soin apporté à la mise au point, il arrive que les imperfections du fond soit détectées en plus de la particule d'intérêt, biaisant ainsi les mesures finales. Ce phénomène peut se produire notamment lorsque le FlowCAM bouge au cours de la numérisation (coup involontaire, vibration extérieure), lors du passage de bulles dans la cellule de flux, ou encore lorsque la cellule de flux bouge elle-même à cause de la pompe péristaltique qui a tendance à tirer sur les tubes connecteurs (le support de la cellule de flux est en effet un peu lâche). Dans ce genre de cas, l'ensemble des images acquises est le plus souvent inutilisable.

Exemple d'une mauvaise segmentation suite au mouvement de la cellule :



- Classification manuelle des particules

Deux sets d'apprentissage provisoires ont été réalisés pour tester les différents paramètres évoqués plus haut. Le premier set d'apprentissage développé à partir des échantillons de la campagne Arcadino de mai 2008, permet d'évaluer l'efficacité d'un outil de reconnaissance spécifique des espèces du genre *Dinophysis*. Le deuxième a été réalisé à partir d'échantillons de Méditerranée (Baie de Villefranche-sur-mer) pour identifier plus spécifiquement les espèces et taxons infraspécifiques du genre *Ceratium*. Pour l'heure, la quantité d'images acquises n'est pas suffisante pour représenter de manière significative la variabilité des taxons ciblés, mais permet d'identifier quelques points à améliorer pour la constitution du futur set d'apprentissage généraliste à développer pour le REPHY.

Les résultats préliminaires de ces sets d'apprentissage sont présentés dans un document annexe. A noter, ce sont des résultats *très* préliminaires, puisque le nombre d'images par groupe est acceptable (pour que l'algorithme fonctionne) mais pas significativement représentatif de chaque groupe. Ces résultats permettent de comprendre comment réagit l'algorithme en réponse aux modifications apportées au set d'apprentissage, et de mettre en évidence les difficultés futures pour le set d'apprentissage généraliste.

Les études préliminaires réalisées sur les genres *Dinophysis* et *Ceratium* ont permis d'identifier plusieurs points importants à prendre en compte pour l'élaboration d'un set d'apprentissage et la création d'un outil de reconnaissance associé.

- au vu des tests effectués, il apparaît que le nombre d'images par groupe taxonomique doit être de **100 minimum**. Si certains groupes sont très bien reconnus avec un petit nombre d'images, parfois une douzaine seulement, dans tous les cas, on observe une très nette amélioration des taux de reconnaissance lorsque le nombre d'images par catégorie est supérieur à 100.

- idéalement, il faudrait acquérir **un nombre équivalent d'images entre les groupes**. Les groupes les mieux représentés sont aussi les mieux reconnus, mais cela induit davantage de faux-négatifs et faux-positifs dans les autres groupes. 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.

- il est nécessaire de constituer pour chaque étude un set d'apprentissage qui soit le plus **détaillé** possible, y compris dans l'optique de créer un outil de reconnaissance spécialisé. Bien que des regroupements seront probablement à envisager dans un deuxième temps, c'est le seul moyen de mettre en évidence les particules prêtant à confusion et éventuellement les critères qui sont sources de confusions. Cependant, il faut noter que le taux d'erreur global diminue fortement après avoir effectué ces regroupements. Pour les groupes ne présentant pas d'intérêt particulier dans le cadre de l'étude menée, il est donc préférable de créer dès le départ des catégories assez larges, n'ayant pas forcément de cohérence d'un point de vue taxonomique. Cela s'applique particulièrement pour la création d'un outil de reconnaissance spécialisé.

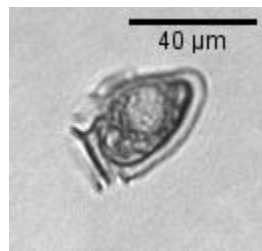
- la catégorisation des débris peut s'avérer problématique. Il est visuellement très difficile de définir des catégories de débris homogènes, sachant que pour une apparence similaire les particules peuvent avoir une origine différente. Ces débris sont souvent une source de confusion importante avec les cellules phytoplanctoniques de petites taille. Il est donc primordial de soigneusement définir ces catégories. Une **catégorisation des débris selon leur teinte (niveau de gris)**, même si elle est forcément subjective et imprécise pour l'observateur humain, semble être la plus pertinente. En revanche, une catégorisation selon la forme ou la texture donne rarement de bons résultats. Il peut également s'avérer nécessaire de regrouper parmi les débris d'autres particules, comme certains groupes du zooplancton, sans qu'il y ait de pertinence taxonomique.

2.2. Demandes spécifiques de l'Ifremer concernant le logiciel Zoo/PhytoImage

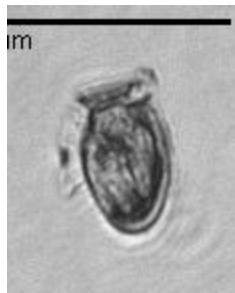
- *Problème d'incrustation de l'échelle de taille*

La première demande concerne la résolution du problème d'incrustation de l'échelle de taille sur les vignettes. Avec la version du logiciel actuellement utilisée, l'échelle dépasse la largeur de l'image ou n'est pas du tout ajoutée, pour une majorité des vignettes. Il n'est donc pas possible dans ces cas-là de connaître la taille des particules.

Exemple d'échelle correctement intégrée à la vignette :



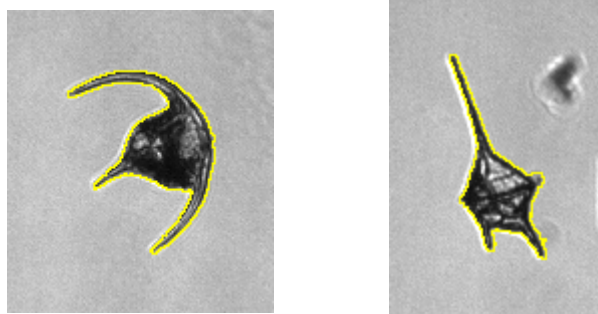
Exemples d'échelle mal, ou non incrustée sur la vignette :



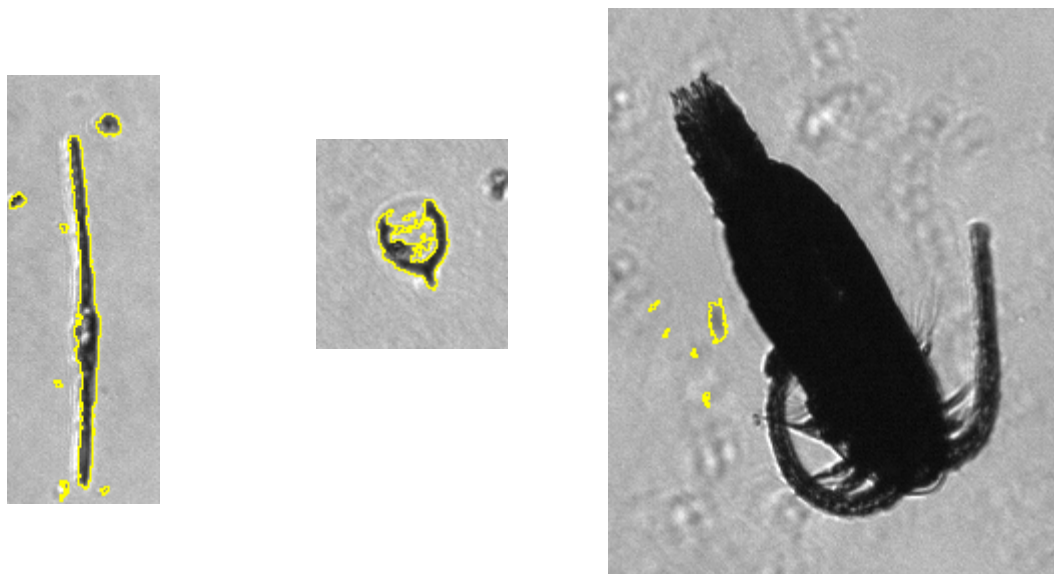
- *Visualisation des silhouettes*

Il serait très utile de pouvoir visualiser rapidement la silhouette (et le masque ?) associée à chaque image au moment de la constitution du set d'apprentissage, pour s'assurer que la segmentation a été correctement effectuée. En effet, outre le problème lié au mouvement de la cellule de flux évoqué plus haut, il se peut que la segmentation ne prenne pas en compte toutes les structures de la cellule, celles-ci étant trop transparentes pour être détectées. Inversement, certains débris de petite taille, proches de la particule d'intérêt, peuvent être comprises à tort dans la segmentation. Hors il est indispensable de choisir les meilleures images pour décrire chaque taxon d'un set d'apprentissage, et il n'est pas toujours aisé de savoir si l'ensemble de la particule considérée a bien été détecté.

Exemples de segmentation correcte :



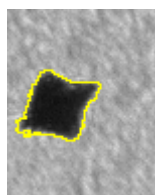
Exemples de mauvaise segmentation :



Idéalement, il faudrait pouvoir consulter la silhouette associée directement à partir de la vignette.

- Problèmes de segmentation

Avec la version actuellement utilisée, certaines particules n'apparaissent pas au centre de la vignette. Cela pourrait-il suggérer un défaut dans le processus même de segmentation ?



- Association "virtuelle" des images aux catégories du set d'apprentissage

Ainsi qu'il avait été évoqué lors de la réunion de Février 2009, il serait intéressant si les images utilisées pour chaque set d'apprentissage n'étaient pas physiquement déplacées dans chaque répertoire, mais plutôt associées "virtuellement" à chaque catégorie, de telle sorte que plusieurs personnes puissent travailler sur un même set d'apprentissage sans modifier le travail

réalisé par les autres. Cela faciliterait aussi les modifications mineures apportées à un set d'apprentissage, en évitant de surcharger les disques par la copie de plusieurs variantes d'un même set d'apprentissage.

- Association de données aux vignettes

Pour contourner le problème des colonies qui sont pour l'instant détectées comme des objets uniques, une des solutions consiste à pouvoir entrer des valeurs associées à chaque image du set d'apprentissage, comme par exemple le nombre de cellules dénombrées sur l'image, qui constituent la colonie. Cette solution serait appliquée en attendant qu'une méthode soit développée pour détecter directement les cellules d'une colonie au moment de l'analyse d'image.

- Introduction du taux de dilution résultant de l'ajout d'eau de mer filtrée

Comme expliqué précédemment, l'ajout d'un faible volume d'eau de mer filtrée avant de passer l'échantillon dans le FlowCAM permet d'optimiser la détection des particules qui sédimentent le plus vite. Serait-il possible d'ajouter ce paramètre à ceux qui sont rentrés dans les champs du tableau décrivant chaque expérience, pour un calcul automatique du taux de dilution final ?

- Définition du taux de dilution

Il faudrait changer le codage du taux de dilution, ainsi qu'il avait été déjà discuté. La dilution est actuellement exprimée en pourcentage, sachant qu'aucune dilution correspond à un pourcentage de 100. Il serait plus logique de coder l'absence de dilution à 1, de façon à pouvoir introduire soit un facteur de dilution, soit un facteur de concentration, car ce cas risque de se présenter lorsque les espèces ciblées seront en trop faible abondance.

- Filtre de taille au niveau de l'analyse d'image

Ainsi qu'il a été expliqué plus haut, le filtre de taille appliqué par le FlowCAM (en mode ABD) ne permet pas de trier précisément les cellules dont la longueur est supérieure à 20 μm . Il serait très utile de pouvoir appliquer un filtre de taille (sous forme de question) avant que l'analyse de taille par ImageJ ne débute, de façon à ce que seules les particules d'intérêt (pour le REPHY, celles de taille supérieure à 20 μm) soient analysées et prises en compte par la suite. Cette mesure pourrait s'effectuer sur la plus grande longueur de la cellule considérée, de la même manière qu'un observateur mesure la longueur d'une cellule au microscope. Cela permettrait également d'alléger significativement le processus de calcul ainsi que le stockage d'images non utilisées.

- Affichage des paramètres retenus par l'algorithme

Si cette option n'est pas déjà disponible avec la version actuelle du logiciel, est-ce qu'il serait possible de connaître les paramètres des particules que l'algorithme a retenus pour discriminer les différents taxons ? En effet, il est intéressant d'accéder à cette information, pour affiner les critères qui servent visuellement pour la classification manuelle des particules. Cela peut s'avérer tout particulièrement utile pour la définition des catégories de débris, qui jusqu'ici reste basée sur des critères très subjectifs.

Annexe 5 – Traitement des échantillons avant et pendant leur passage au FlowCAM

1. Objet et domaine d'application

Ce document décrit la procédure de traitement des échantillons phytoplancton avant et pendant leur passage au FlowCAM pour numérisation. Il s'agit soit des échantillons REPHY Arcachon pleine-mer soit d'autres échantillons provenant de tous les LERs dont les images serviront à enrichir les sets d'apprentissage utilisés par PhytoImage.

2. Opérations préliminaires

2.1. Récolte d'échantillons

2.1.1. Echantillons demandés aux autres LERs

Les échantillons intéressants pour le projet au vu des résultats de comptage réalisés, doivent :

- soit présenter une relative richesse phytoplanctonique (par exemple à partir d'au moins 20 000 cellules/L) et une population diversifiée ou non,
- soit correspondre à un bloom mono ou pluri spécifique,
- soit présenter des dinoflagellés même en relativement faible abondance (à partir de 1 000 /L par exemple pour les espèces les plus grosses) (sachant qu'il y a un déficit important en images de dinoflagellés pour le moment).

Les échantillons répondant à ces critères sont destinés à un passage qualitatif au FlowCAM (cf. 6.2.) c'est-à-dire dont les images serviront à enrichir les sets d'apprentissage utilisés par PhytoImage.

Les échantillons trop pauvres ne valent pas la peine d'être retenus.

Protocole

Les échantillons devront être expédiés au LER/AR au plus tard 1 mois après le prélèvement, le protocole de numérisation au FlowCAM imposant de ne pas dépasser 3 mois de délai. Les échantillons de 250 mL, fixés au Lugol à 1 %, placés dans un flacon en verre brun seront, en attendant l'envoi, stockés au réfrigérateur. Une photocopie de la feuille de comptage sera jointe à l'échantillon.

Solution de Lugol utilisée : 100g KI + 50 g I₂ (iode bisublimée) + 100 mL acide acétique glaciale qsp 1000 mL à l'eau distillée.

2.1.2. Echantillons récoltés par le LER/AR

Tous les échantillons faisant l'objet d'une analyse de la flore totale dans le cadre du REPHY sont à réserver pour un passage quantitatif (cf. 6.1.) au FlowCAM, soit tous les échantillons des stations "Bouée7" et "Teychan bis" récoltés lors des tournées hydrologiques du programme ARCHYD de pleine mer.

Tous les autres échantillons répondant aux critères intéressants cités ci dessus doivent aussi être réservés pour un passage qualitatif.

2.2. Réception/préparation des échantillons

2.2.1. Echantillons fournis par les autres LERs

Le contenant doit être en verre brun, correctement identifié et sa coloration indiquer qu'il y a bien eu ajout de Lugol dans les bonnes concentrations.

Enregistrer l'échantillon dans le registre du FlowCAM 2011 : échantillons reçus.

Enregistrer le type de numération prévu pour cet échantillon dans le registre du FlowCAM 2011 : numérisations.

Conserver l'échantillon au réfrigérateur n° 09/06 en attente du passage au FlowCAM.

2.2.2. Echantillons récoltés par le LER/AR

Dès l'arrivée des échantillons destinés à l'analyse d'une flore totale, procéder au traitement suivant :

- a. homogénéiser l'échantillon de 1 L par 30 retournements doux,
- b. remplir rapidement un flacon en verre de 250 mL,
- c. y ajouter 2,25 mL de Lugol (concentration finale 1 %).
- d. enregistrer cet échantillon dans le registre du FlowCAM 2011. Identifier le flacon et le stocker au réfrigérateur n° 09/06.
- e. les 750 mL restant demeurent dans le flacon d'origine pour comptage au microscope.

Pour les autres échantillons observés par le LER/AR et faisant l'objet de flores partielles indicatrices, attendre les résultats des analyses et mettre en œuvre les étapes "a" à "d" pour les échantillons sélectionnés.

3. Numérisation au FlowCAM

La procédure détaillée de numérisation des échantillons au FlowCAM est décrite dans le document « Numérisation au FlowCAM des échantillons du REPHY ».

Lors des passages à l'objectif 4x, selon la charge en phytoplancton, si des saturations se produisent, modifier le nombre d'images prises par seconde. S'il y a des colmatages, envisager éventuellement une dilution de l'échantillon.

3.1. Protocole quantitatif

Pour le passage à l'objectif 10x en cellule de 100 µm ou 4x en cellule de 300 µm :

- a. Homogénéisation de l'échantillon par retournement doux 30 fois
- b. Remplissage complet (environ 50 mL) de l'entonnoir
- c. Réalisation de 2 passages d'1 heure à l'objectif 10x en cellule de 100 µm (le niveau d'eau de l'échantillon arrive à la limite de l'hélice d'homogénéisation) ou de 2 passages de 10 minutes à l'objectif 4x en cellule de 300 µm
- d. Vidange complète (entonnoir + circuit)
- e. Mettre de nouveau en œuvre les étapes a) à d) de façon à réaliser deux autres passages

Récupérer éventuellement la poubelle pour du qualitatif ultérieur

3.2. Protocole qualitatif

Procéder librement en réalisant le maximum de passages possibles.

Annexe 6 – Numérisation au FlowCAM des échantillons du REPHY

1. Objet et domaine d'application

Protocole d'utilisation du FlowCAM pour la numérisation des échantillons du REPHY (flore totale)

2. Préparation de l'échantillon

Cf. le document « Préparation et traitement des échantillons avant passage au FlowCAM ».



3. Montage des éléments du FlowCAM

3.1. Assemblage Objectif / cellule

L'objectif **10x** doit être utilisé avec une cellule de flux calibrée de **100 µm** et le **collimateur associé**.
L'objectif **4x** doit être associé à une cellule de flux calibrée de **300 µm sans le collimateur**.

3.2. Mise en place du système optique

Objectifs et collimateur se vissent dans le sens des aiguilles d'une montre et se dévissent dans le sens inverse des aiguilles d'une montre.

L'objectif est vissé jusqu'à la butée dans son emplacement sur la gauche du bloc central noir. Si l'objectif utilisé est le 10X, il faut également visser le collimateur associé dans son emplacement à la droite du bloc central. Les autres objectifs ne nécessitent pas l'ajout du collimateur.

Si besoin, penser à déplacer le bloc central vers la gauche ou la droite grâce à la vis supérieure pour permettre un montage plus aisée de l'objectif et du collimateur.

3.3. Montage de la cellule de flux

Pour les cellules de flux calibrées de 100 μm et 300 μm , le même porte-cellule, étiqueté « QCFC 300 », est utilisé.



Dévisser la bague supérieure du porte-cellule.



Après avoir ôté délicatement la protection plastique de la cellule de flux, poser celle-ci sur le porte-cellule en positionnant la partie en verre dans l'encoche prévue à cet effet, le tube le plus court dépassant du côté de la vis.



Pour la cellule de 300 μm , l'anneau noir situé à l'intérieur du porte-cellule (vissé dans la bague inférieure) doit être vissé de manière à ne pas dépasser de l'encoche. Au contraire, pour la cellule de 100 μm , il faut dévisser cet anneau pour qu'il dépasse de l'encoche d'environ 0,5 mm. La bague supérieure est ensuite revissée pour immobiliser la cellule de flux dans son support.



Il faut veiller à ne pas serrer trop fort la bague supérieure pour éviter d'écraser la cellule en verre ; cependant, il faut serrer suffisamment pour que la cellule ne bouge pas dans son support.



L'ensemble porte-cellule + cellule de flux est ensuite inséré sur le bloc central, la vis et le tube court dirigés vers le haut. Pour cette étape, procéder comme suit :

- insérer tout d'abord le tube long dans l'ouverture du bas,
- baisser la main pour permettre de passer le tube court dans l'ouverture du haut entre les deux clapets de caoutchouc noir,
- insérer horizontalement le porte-cellule sur son pivot et serrer la vis pour immobiliser l'ensemble.

Pendant le montage, veiller à ne pas effectuer de torsion aux jonctions de la cellule avec les tubes de silicone. Faire également attention à ne pas rayer les lentilles des objectifs et du collimateur.

Passage **4x** : cellule **300µm**, enlever le collimateur et changer l'objectif.

- Ajout des tubes d'entrée et de sortie



Poser l'entonnoir sur son support et connecter l'embout au tube supérieur de la cellule de flux. Relier ensuite l'embout du connecteur de la pompe avec le tube inférieur de la cellule de flux.
 Pour la cellule de **100 µm**, utiliser le connecteur de pompe avec les **bagues blanches**.
 Pour la cellule de **300 µm**, utiliser le connecteur de pompe avec les **bagues rouges**.



Fixer les deux bagues du connecteur de pompe dans les encoches en métal prévues : la première bague se fixe sur l'encoche du bas (« in »), la deuxième bague sur l'encoche du haut (« out »). Mettre le pot de récupération du liquide en sortie



3.4. Réglages du FlowCAM

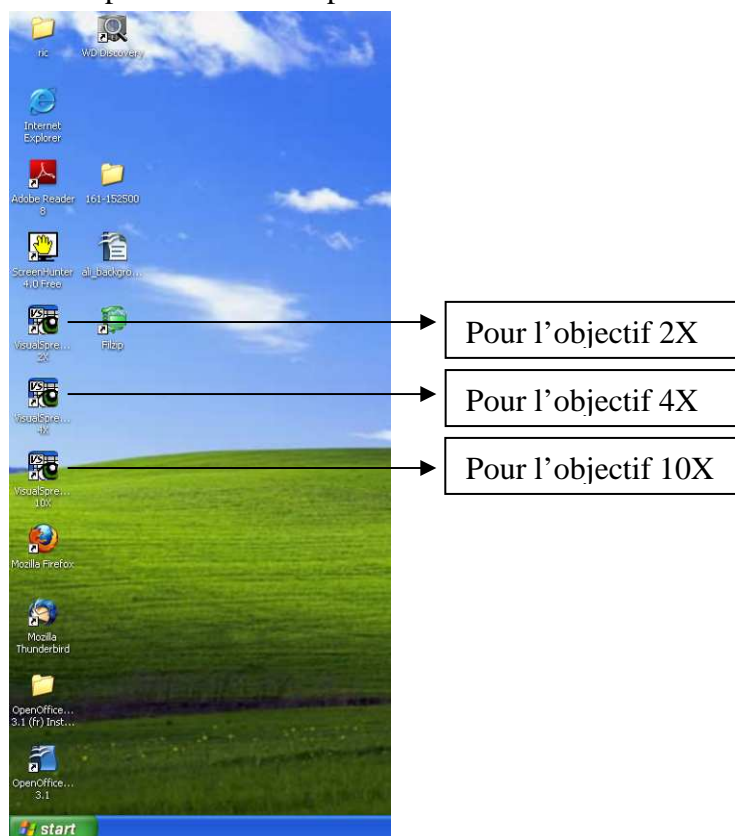
- Démarrage du FlowCAM

Allumer le FlowCAM à l'aide du commutateur situé à l'arrière de la machine (en haut à gauche). Allumer l'écran.

Après avoir verrouillé le pavé numérique du clavier (touche NumLock en haut à gauche du pavé numérique), entrer le mot de passe :

flowcam507

En fonction de l'objectif choisi pour la numérisation, lancer le programme de VisualSpreadsheet correspondant.



- Utilisation de la pompe péristaltique

L'interrupteur de gauche permet d'attirer le liquide dans la cellule (position bas), de la refluer (position haut) ou de l'éteindre (position médiane).

L'interrupteur de droite permet de régler le débit du plus rapide au plus lent :

En haut, la vitesse («prime-purge »)

En centre, la vitesse (« slow»)

En bas, La vitesse (« fast »)

Pour les vitesses « slow » et « fast », la molette graduée permet de moduler la vitesse de 1 à 10.

La vitesse « prime-purge » est la plus rapide

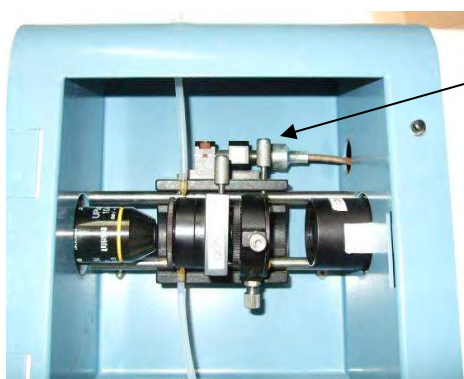
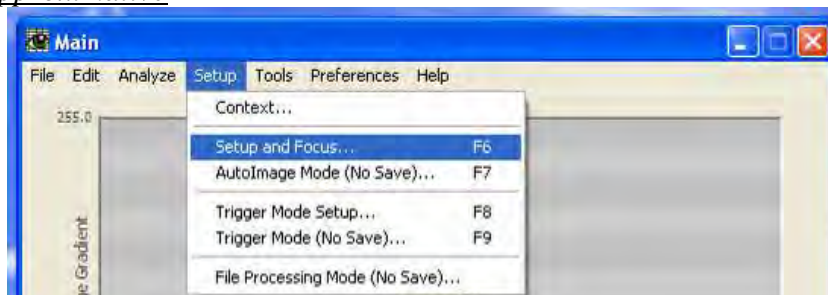


- Introduction de l'échantillon

Remplir l'entonnoir avec l'échantillon lugolé, actionner l'agitation douce (minimum de rotation) et actionner la pompe. Stopper la pompe lorsque le liquide a rempli complètement la cellule de flux.

- Mise au point approximative

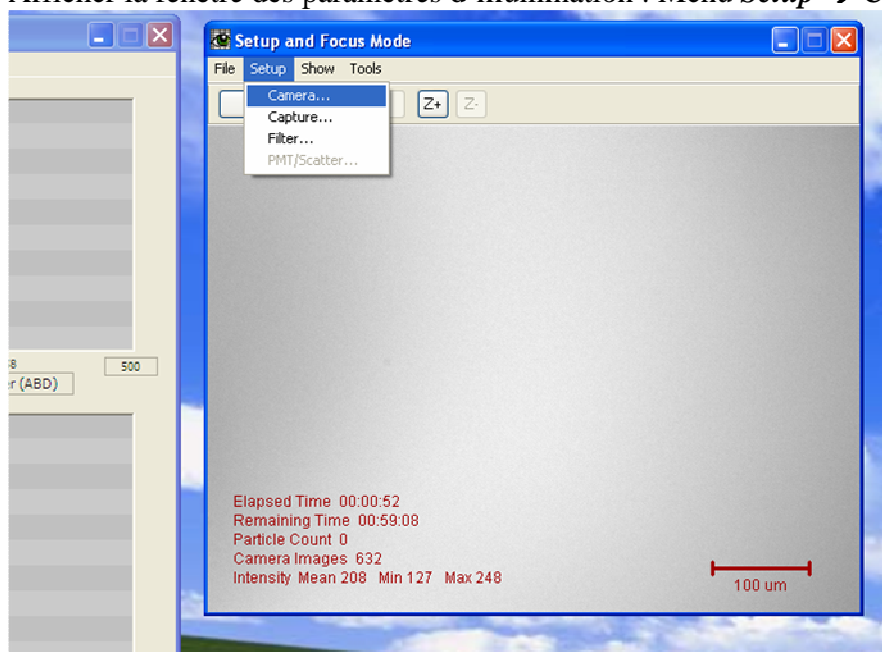
Afficher la fenêtre de la caméra
Menu **Setup**
→ **Setup and Focus**



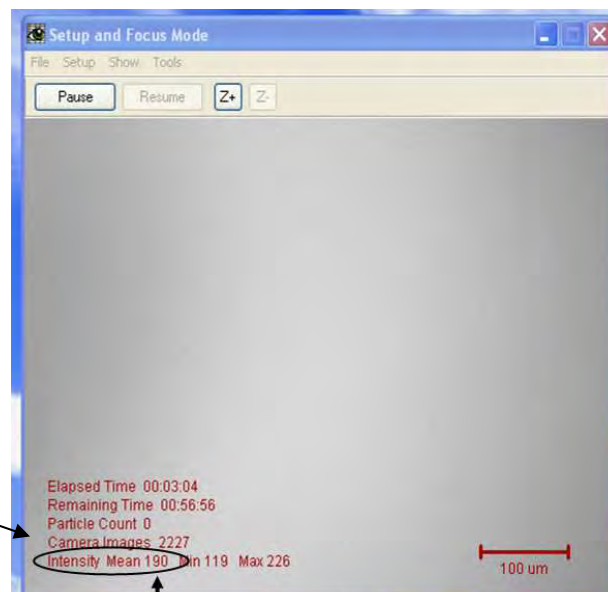
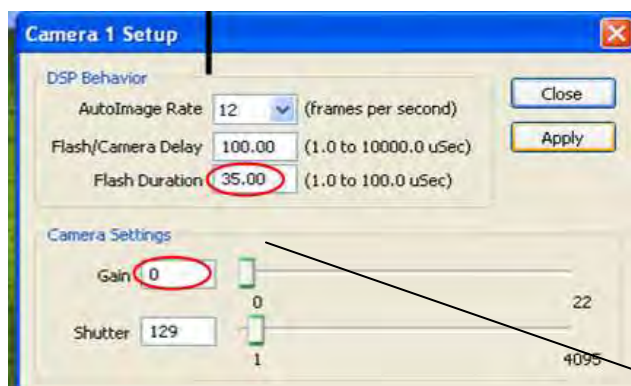
Après avoir desserré la vis supérieure déplacer le bloc central le long des rails de manière à visualiser une des parois de la cellule de flux sur la fenêtre de la caméra.

- Réglage de l'illumination

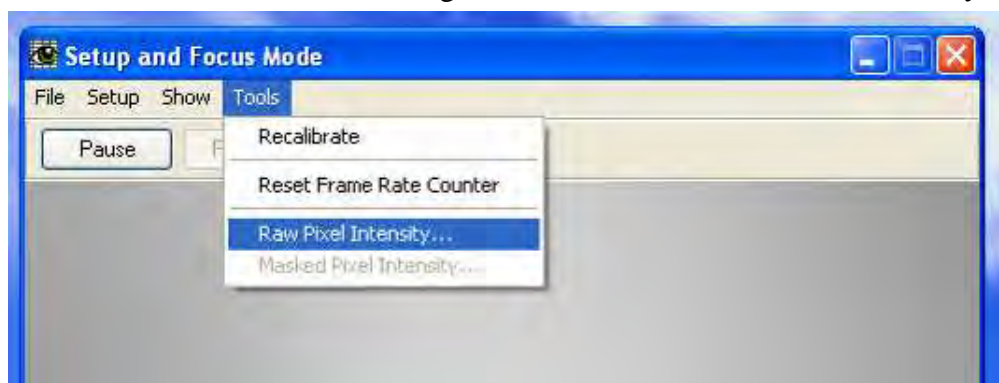
Afficher la fenêtre des paramètres d'illumination : Menu **Setup** → **Camera**



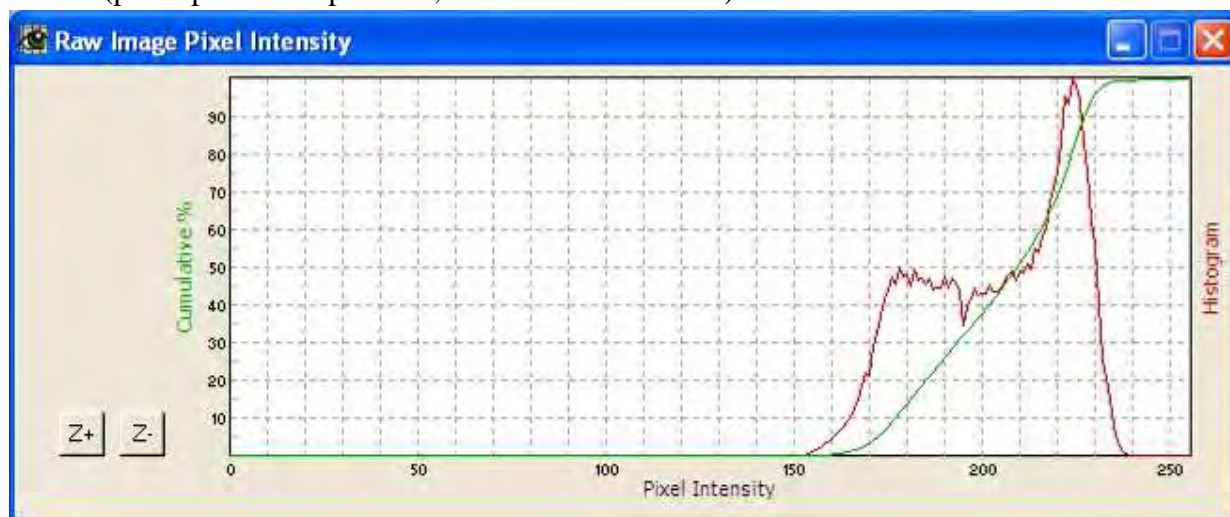
Modifier les paramètres « Flash duration » et « Gain » pour obtenir une intensité moyenne comprise entre 180 et 200.



Afficher la courbe des niveaux de gris : Menu **Tools** → **Raw Pixel Intensity**



Vérifier régulièrement grâce à cette option que la courbe des niveaux de gris présente une forme correcte (pic le plus étroit possible, situé entre 150 et 250).

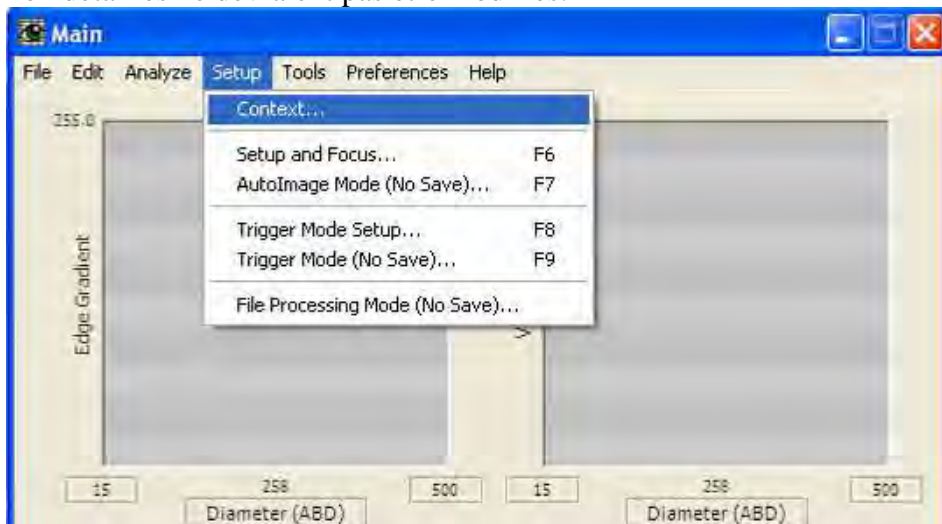


Avec le 4x et la cellule de 300 µm régler le centrage du spot (source lumineuse) grâce aux clés alen et aux vis situées sur la paroi droite de l'appareil. Ceci afin d'obtenir la meilleurs courbe possible.

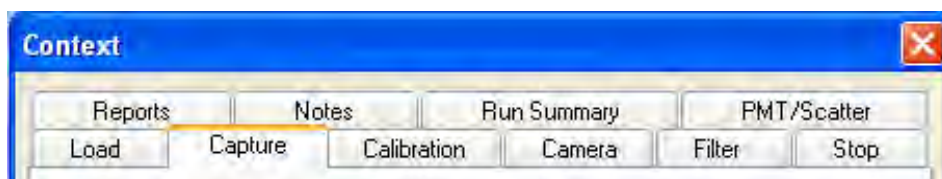
3.5. Réglages des paramètres d'acquisition

Ouvrir la fenêtre « Context » : Menu *Setup* → *Context*

Les paramètres détaillés plus bas sont ceux susceptibles d'être modifiés si nécessaire. Les paramètres non détaillés ne devraient pas être modifiés.

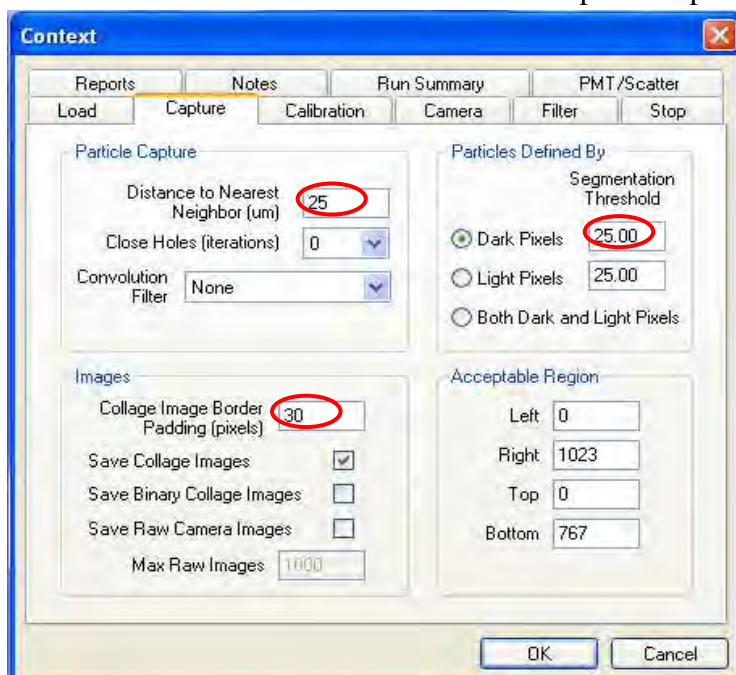


Dans la fenêtre « Contexte » qui s'affiche, plusieurs onglets sont accessibles :



- Onglet « *Capture* »

Introduire ou vérifier les valeurs suivantes pour les paramètres entourés.



Distance to Nearest Neighbor

Distance au plus proche voisin ; seuil minimal en μm entre deux particules distinctes. En dessous de cette distance, deux particules sont détectées comme une.

Collage Image Border Padding

Nombre de pixels détournés autour de la particule

Segmentation Threshold – Dark Pixel

Nombre minimum de niveaux de gris plus foncés que le fond pour détecter une particule

- Onglet « Camera »

Régler le nombre d'images par seconde désiré (en général 12 images/seconde pour l'objectif 10X, 7 images/secondes pour l'objectif 4X).

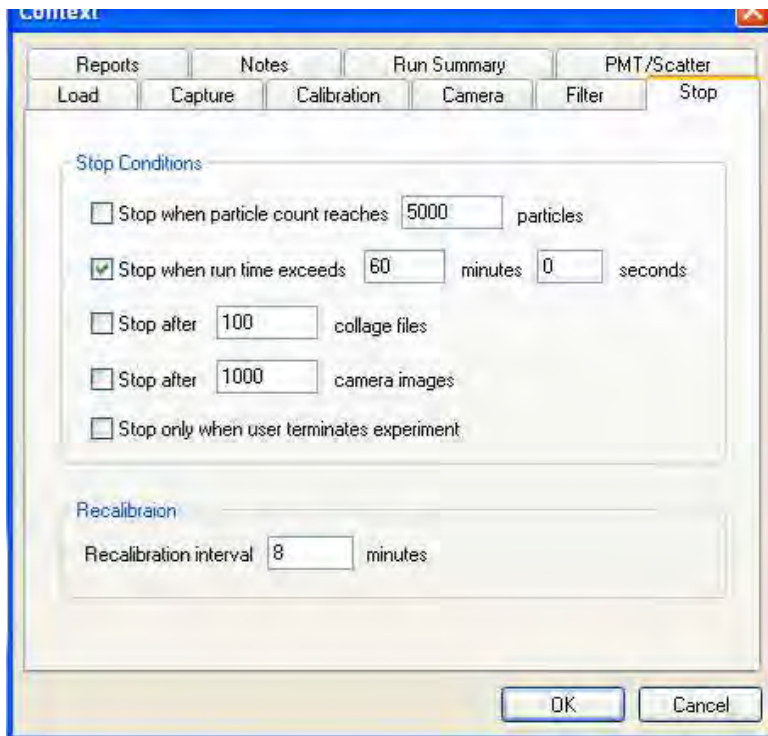


Introduire ou vérifier les valeurs suivantes pour les paramètres entourés. Utiliser le mode ABD pour définir la gamme de taille des particules qui seront détectées



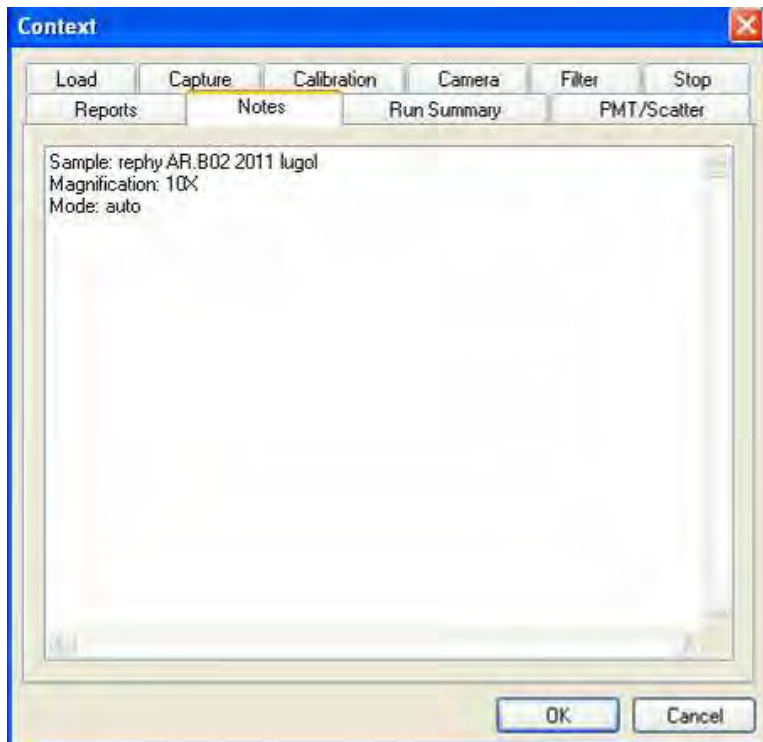
- Onglet « Stop »

Définir le mode d'arrêt de la numérisation. La 2^e option, qui définit un temps de numérisation (par exemple 60 ou 30 min), paraît la plus adaptée pour le moment.



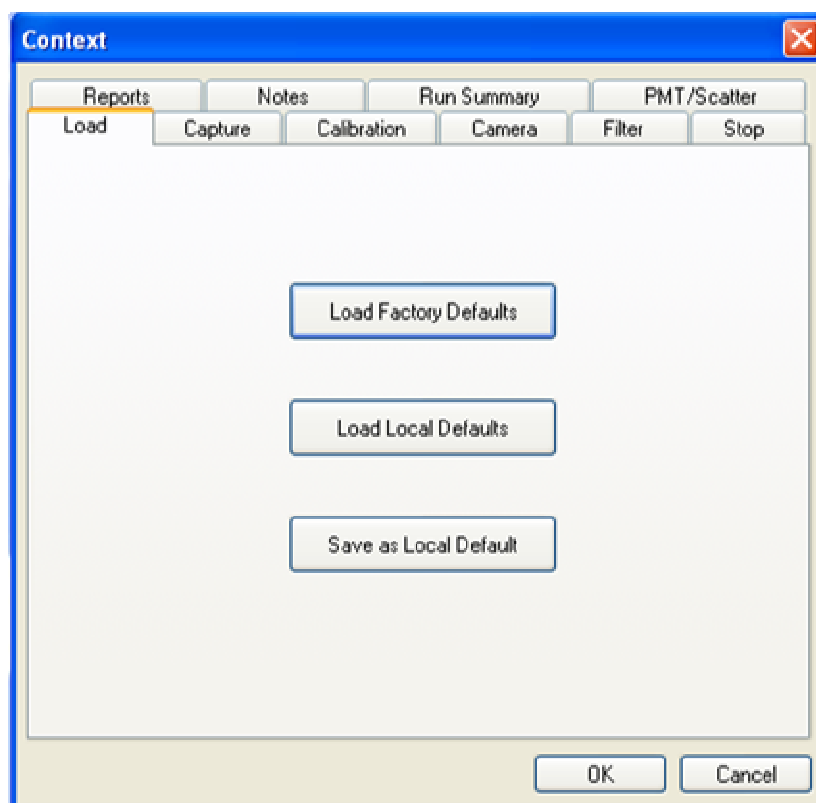
- Onglet « Notes »

Vérifier et rajouter si nécessaires les informations importantes concernant la numérisation.



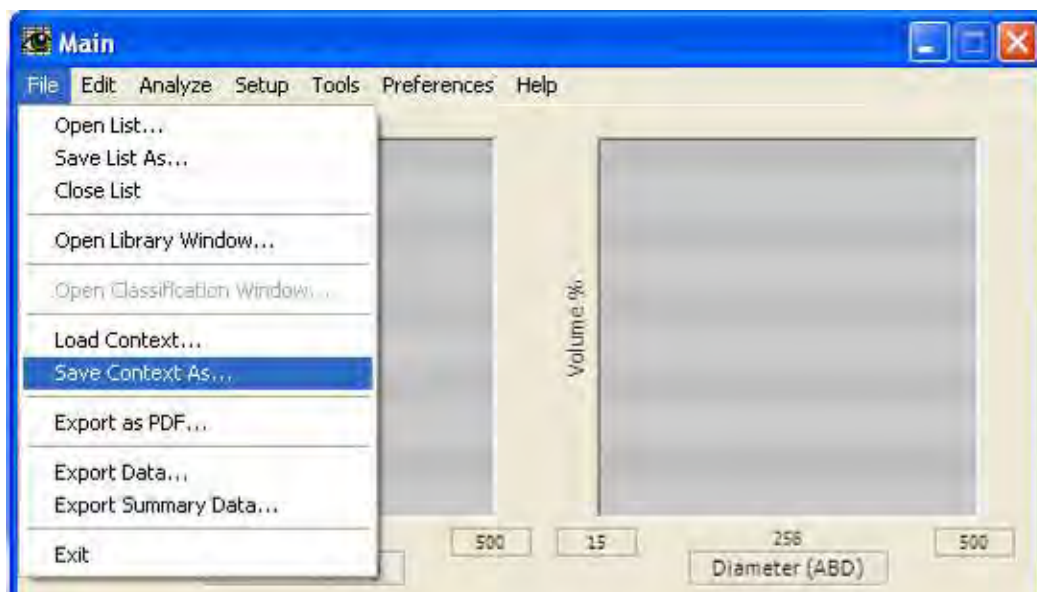
- Onglet « Load »

Il permet de sauvegarder les valeurs des paramètres en 1 clic, et de les ouvrir aussi rapidement



Cliquer sur « Save as Local Defaults » pour enregistrer les valeurs de paramètres en cours.
 Cliquer sur « Load Local Defaults » pour ouvrir les valeurs de paramètres précédemment enregistrés.

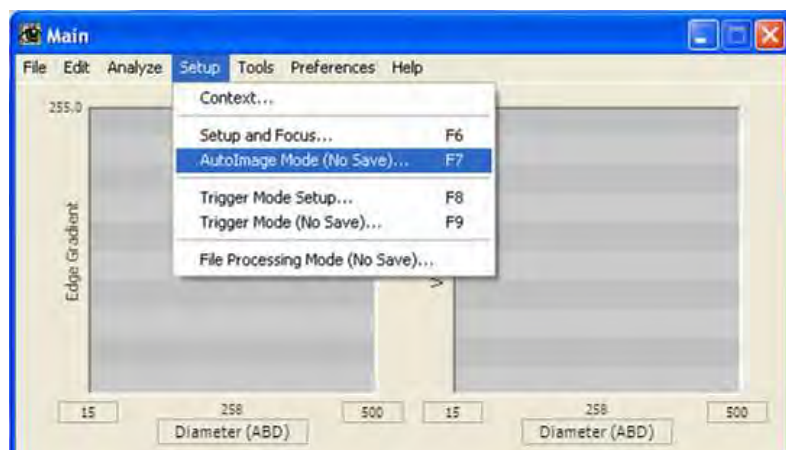
Il est également possible d'utiliser les deux fonctions situées dans le menu déroulant de la fenêtre principale, notamment pour sauvegarder plusieurs modèles de réglages des paramètres, selon l'étude concernée.



3.6. Numérisation de l'échantillon

- Mise au point fine

Démarrer le mode AutoImage sans sauvegarde : Menu **Setup** → **AutoImage Mode (no Save)**.
Démarrer la pompe péristaltique.



Ajuster la mise au point à l'aide de la vis micrométrique (mollette blanche située à gauche du panneau frontal de la machine), de manière à ce que les images des particules détectées en temps réel par le FlowCAM deviennent les plus nettes possible. Fermer ensuite les fenêtres de collages et de la caméra

- Création d'un répertoire du jour

Ouvre le Disque dur C:\Data\REPHY\Acquisition

- **Créer un Dossier** portant le **date du jour** suivi des **initiales de l'analyste** (Ex : 2011-03-15-NN)

Dans ce dossier :

- **recopier un fichier Excel d'un autre dossier (ATTENTION : choisir un dossier d'une numérisation identique 4x ou 10x), le nommer avec la date du jour suivi de Rephy et du lieu de provenance** (Ex : 2011-03-15.Rephy.AR)

ATTENTION : ne pas oublier de modifier tous les champs concernant l'échantillon et la date de numérisation, ce fichier sera important pour le traitement des images avec phytoimage.

- **Créer un sous-dossier « grey »** dans lequel sera stockée la calibration des niveaux de gris

▪ Dans ce dossier apparaîtront les numérisations, les nommer de la manière suivante :

- Pour une numérisation au 10X avec une cellule de 100µm
XX.Ynn.2011-03-15.100A10X_01
- Pour une numérisation au 4X avec une cellule de 300µm
XX.Ynn.2011-03-15.300A4X_01

Avec :
XX initial de la provenance
AR : éch du Bassin d'Arcachon
HO : éch du Lac d'Hossegor
DCE : éch côte basque et landes
ATL : éch d'atlantique
ou initiales du labo fournisseur
Ynn le numéro de l'échantillon

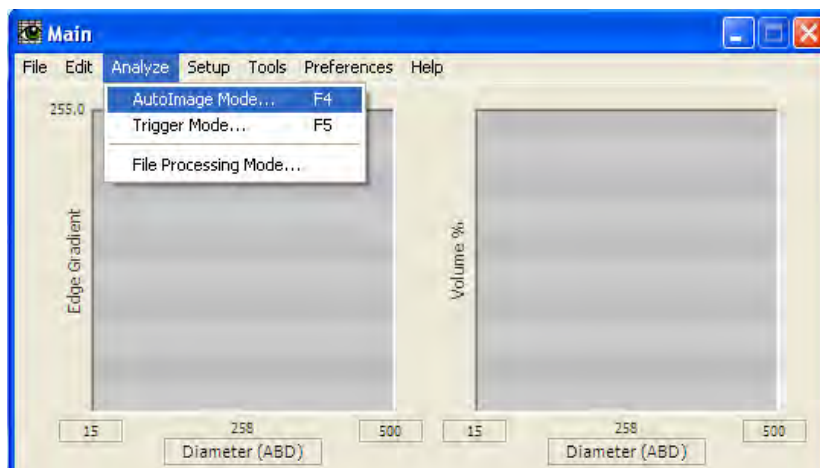
Ex : AR.B02.2011-03-15.100A10X_01

- Calibration des niveaux de gris

Avant chaque numérisation d'échantillon, il est nécessaire d'effectuer une calibration des niveaux de gris, à l'aide des **filtres gris neutres ND2, ND4, ND8, et du filtre noir opaque**.



Ces images seront utiles pour calibrer les niveaux de gris des images en densité optique. Lancer une numérisation en mode AutoImage : Menu *Analyse* → *AutoImage Mode*, sans actionner la pompe péristaltique.



Cliquer sur « Ok » et sauvegarder le fichier de calibration des gris dans le sous-répertoire **<grey>** du répertoire de la date de numérisation.

Le fichier correspondant à la calibration doit porter le même nom que celui de l'échantillon, le suffixe "**_NDx**" étant ajouté à la fin (ex : AR.B02.2011-03-15.100A10X_01_NDx).

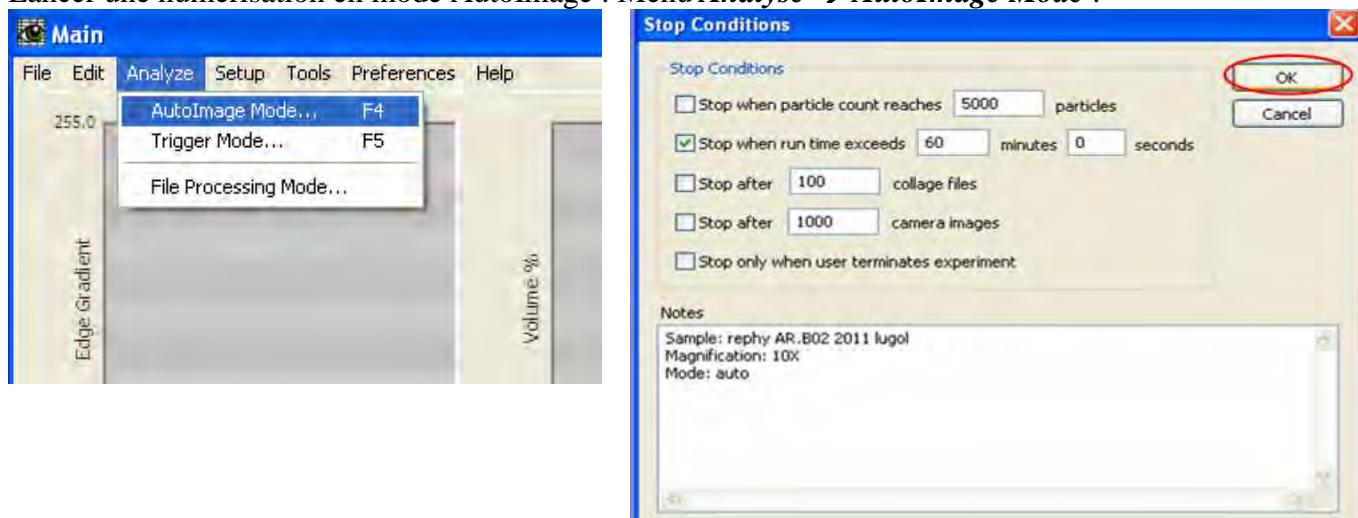
Save ↵, puis, une fois la calibration du fond terminée, placer successivement **entre le collimateur (ou la LED) et le porte-cellule** les filtres gris neutres dans l'ordre ND2, ND4 et ND8, puis le filtre opaque noir pendant 1 à 2 secondes.



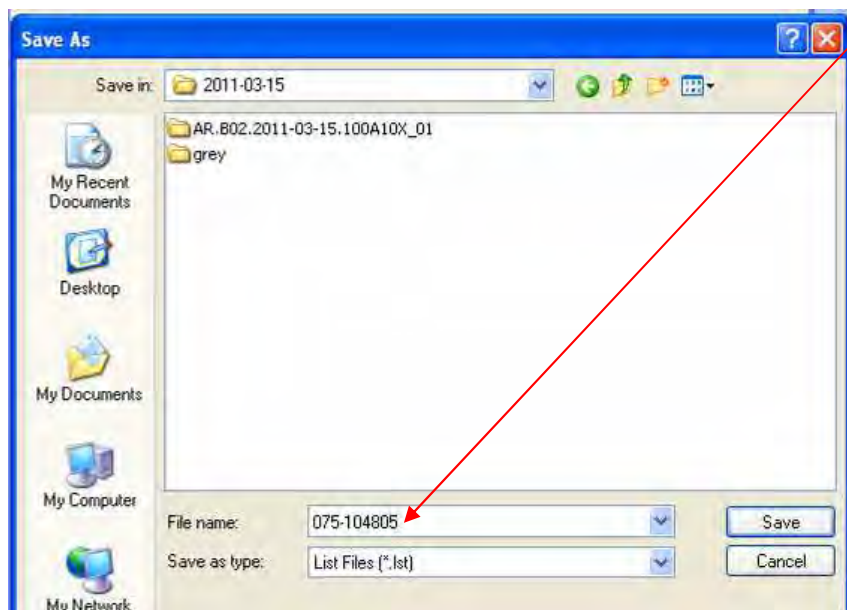
Fermer la fenêtre de la caméra une fois l'opération terminée.

- Acquisition des images de l'échantillon

Ne pas oublier de **remplir le fichier Excel associé** qui liste les échantillons numérisés dans la journée.
Lancer une numérisation en mode AutoImage : Menu *Analyse* → *AutoImage Mode* .



Cliquer sur « **Ok** » après avoir modifié si nécessaire le champ des notes ou le mode d'arrêt de la numérisation.



Nommer le fichier de numérisation : nom copier dans le fichier excel du jour de numérisation

(ex : AR.B02.2011-03-15.100A10X_01)

L'enregistrer **dans le répertoire de la date de numérisation**, attention pas dans le <grey>.

Cliquer sur **Save**, un nouveau répertoire se crée automatiquement et porte le nom collé dans « File name ».

Lancer aussitôt la pompe péristaltique Vitesse :
"slow" en 10x
"fast" en 4x

Laisser les fenêtres ouvertes jusqu'à la fin de l'acquisition lorsqu'elle est automatique (intervalle de temps, nombre de particules max, ou autre option de fin choisie).

Sinon, mettre fin à l'acquisition au moment choisi en cliquant sur Pause dans la fenêtre de la caméra, puis fermer cette fenêtre.

En fin d'acquisition, stopper la pompe.

3.7. Démontage et nettoyage de la cellule

- Rinçage de la cellule

A la fin de la journée de numérisation, ou si un nouvel échantillon doit être analysé, il faut procéder au rinçage de la cellule :

- Pipeter le reste d'échantillon dans l'entonnoir
- Démontez le tube connecteur de la pompe péristaltique
- Laissez s'écouler le reste de liquide par gravité
- Baissez le porte entonnoir, déconnectez l'entonnoir
- Sortez l'entonnoir, le rincez à l'eau distillée puis à l'éthanol à 95%, mettez à sécher
- déconnectez le tube connecteur de la pompe du tube inférieur de la cellule, et le re-connectez à la place au tube supérieur
- à l'aide des pissettes d'eau distillée dont l'embout est introduit directement dans le tube inférieur, de la cellule de flux, procédez à un 1^{er} rinçage à l'eau distillée, évacuez le liquide avec la poire, faites un 2^e rinçage avec la pissette d'alcool, évacuez le liquide avec la poire, et terminez par un dernier rinçage à l'eau distillée

Pour un nettoyage plus efficace, affichez la fenêtre de la caméra Menu *Setup* → *Setup and Focus*

Pour une durée courte entre deux utilisations (quelques jours) laissez la cellule de flux en place, remplie d'eau distillée en circuit fermé.

Pour une durée plus longue, ôtez la cellule après un dernier rinçage à l'alcool, la placez dans son étui de protection et la stockez à l'abri de la poussière. Laissez un objectif en place afin d'éviter une accumulation de poussière au niveau de la caméra, et fermez le portillon.

4. Sauvegarde des fichiers

Une fois la séquence d'analyse terminée :

Ouvrez le disque (y :) entrez le mot de passe :

Admin

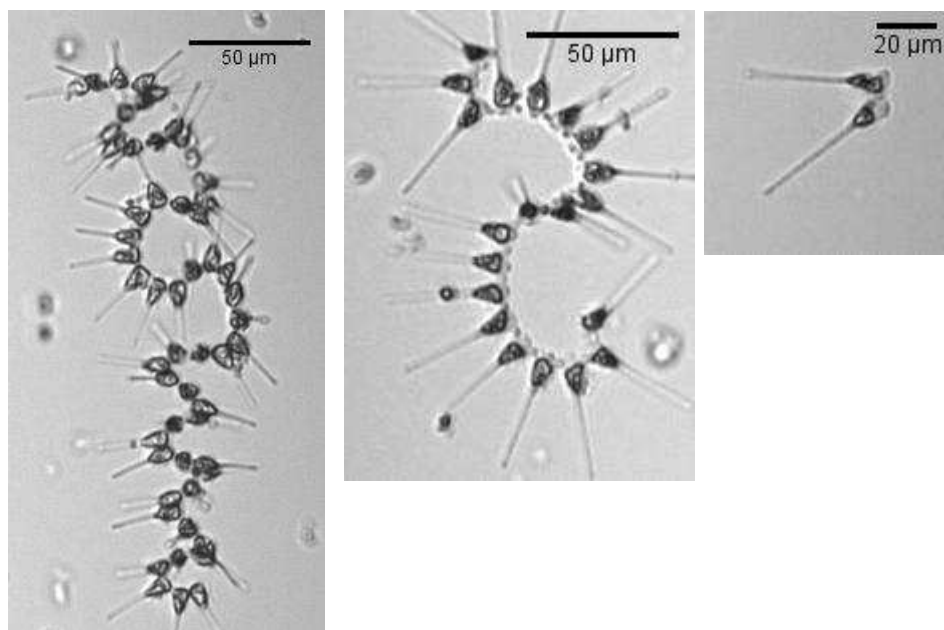
Sauvegardez les fichiers de la journée sous y:\FlowCAM Acquisition\Data\REPHY\Ifremer_2011.

Annexe 7 – Illustration des différents groupes du set d'apprentissage Flore Totale Lugol

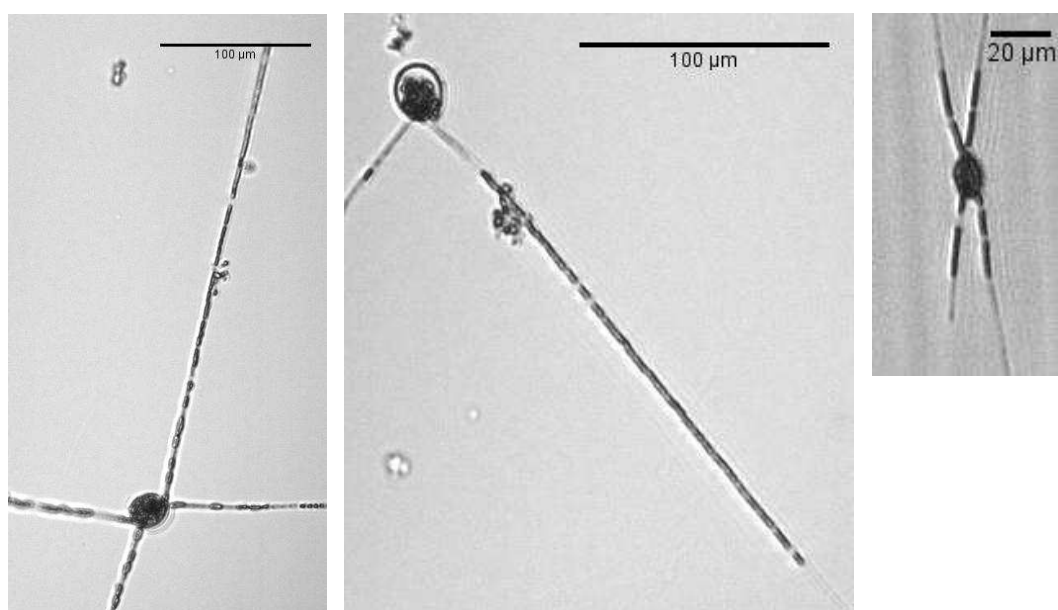
Phytoplancton

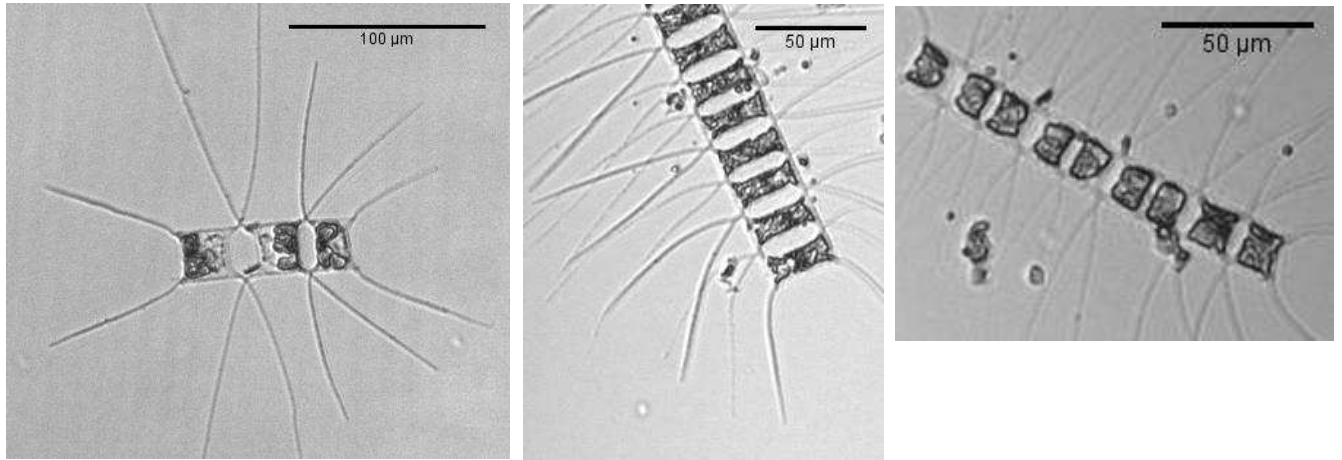
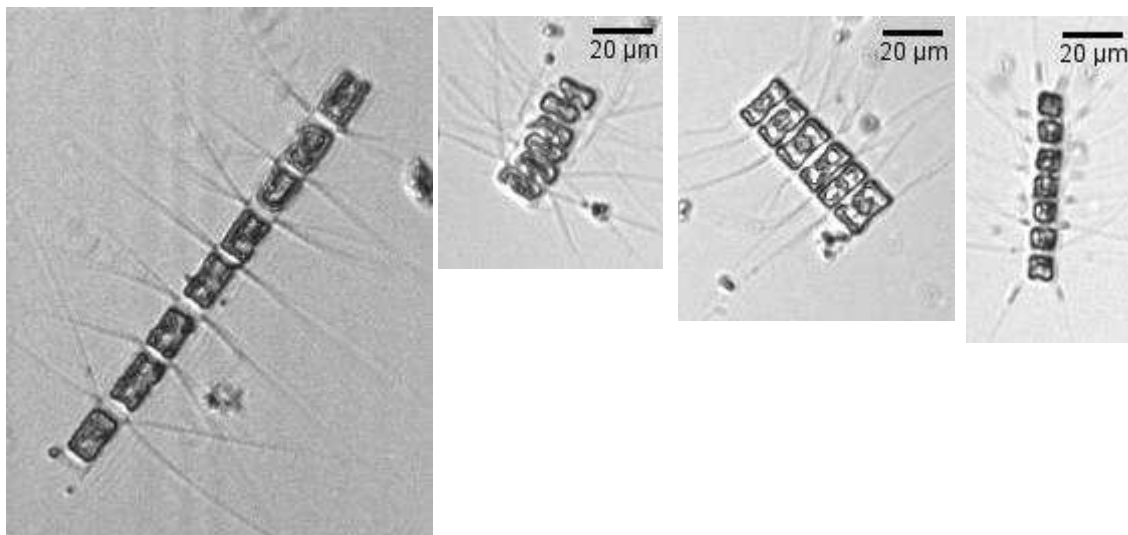
Bacillariophyceae (diatomées)

Asterionellopsis glacialis

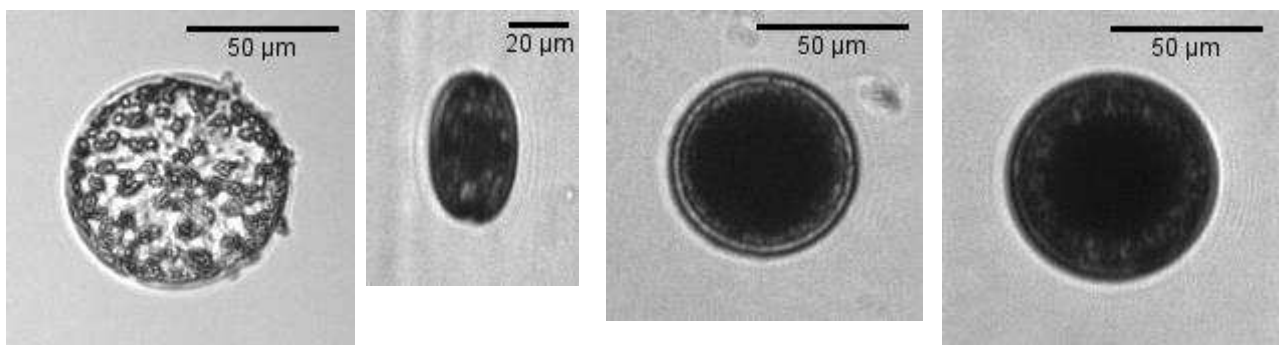


Chaetoceros danicus

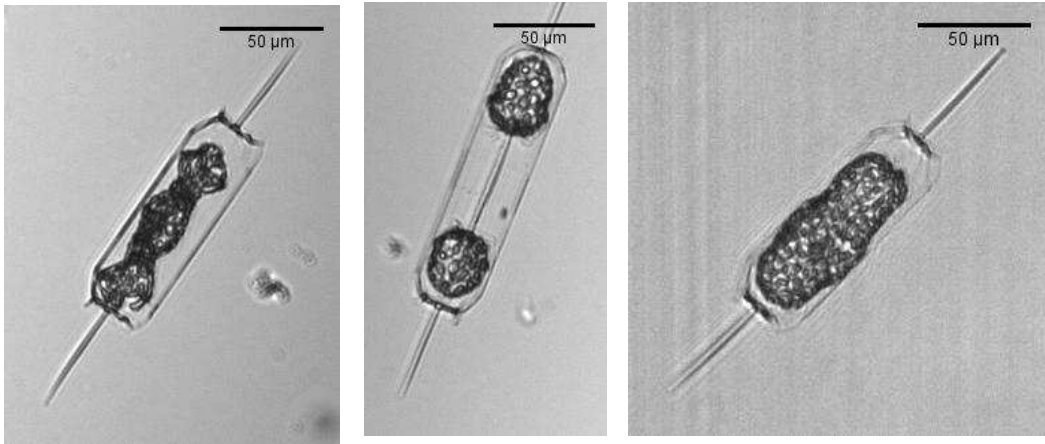


Chaetoceros decipiens*Chaetoceros* spp.

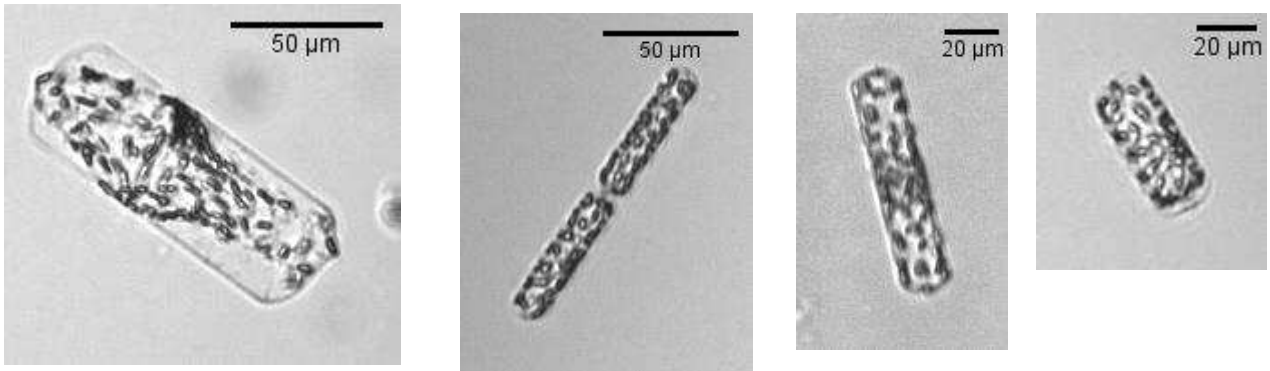
Centrales spp.



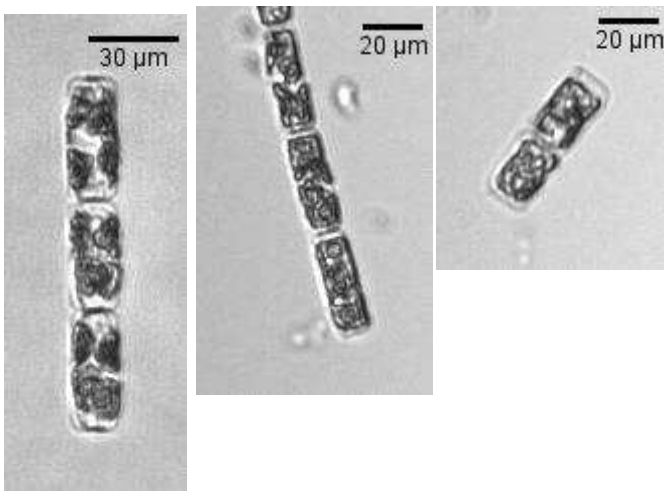
Dytilum brightwellii

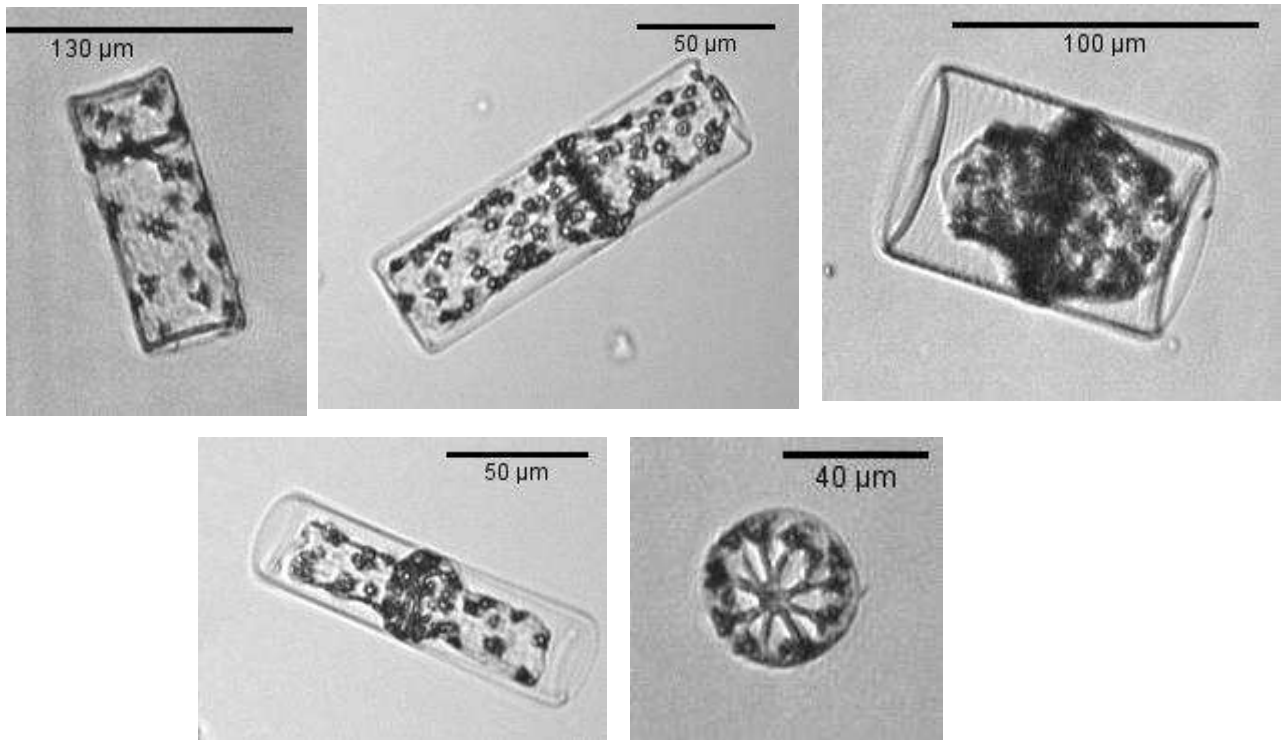
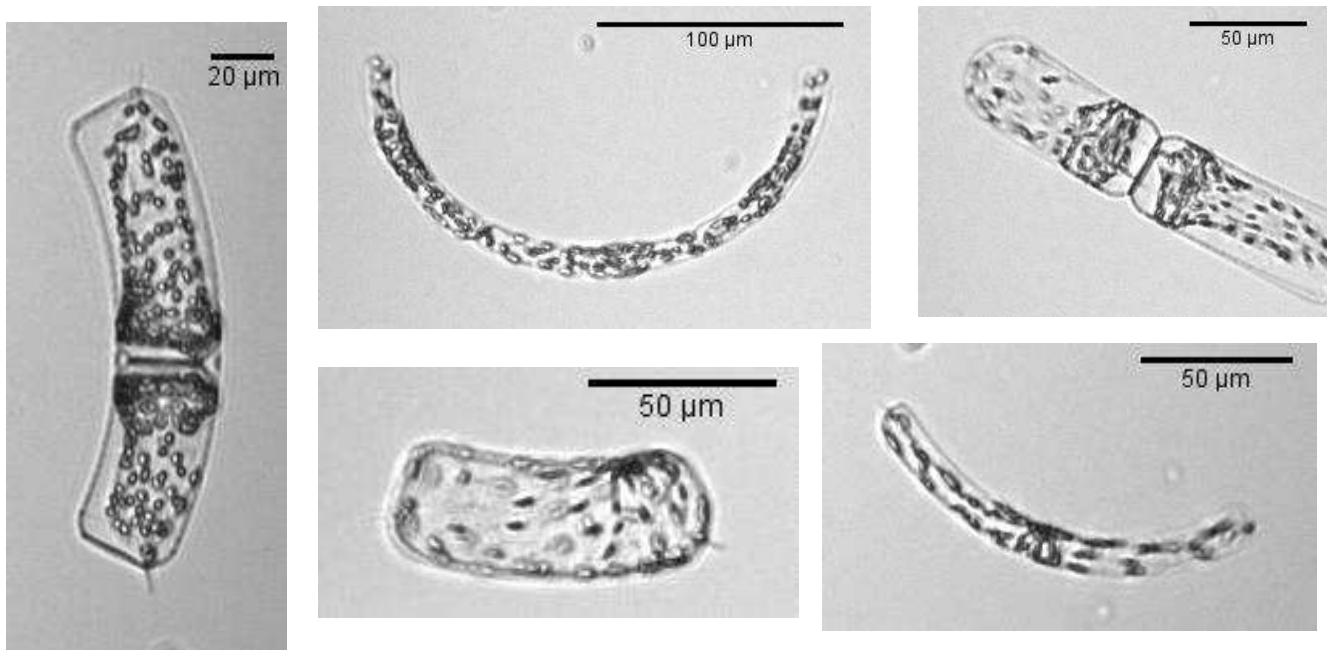


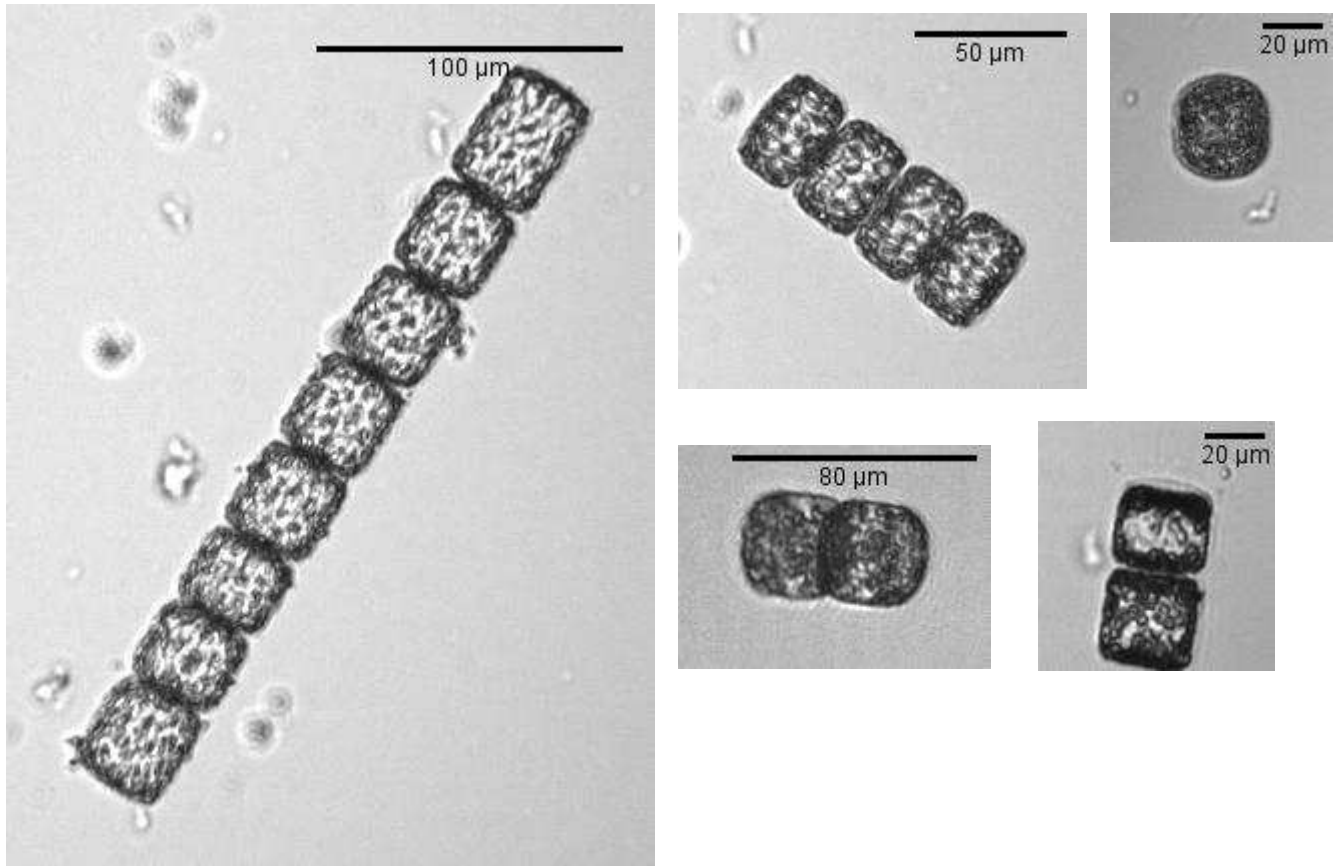
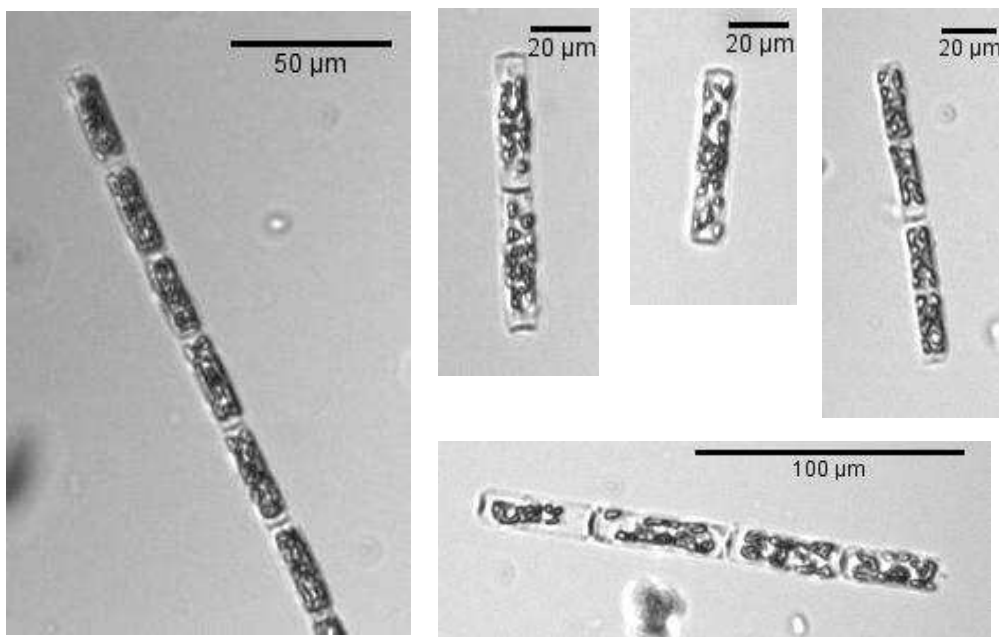
Dactyliosolen fragilissimus

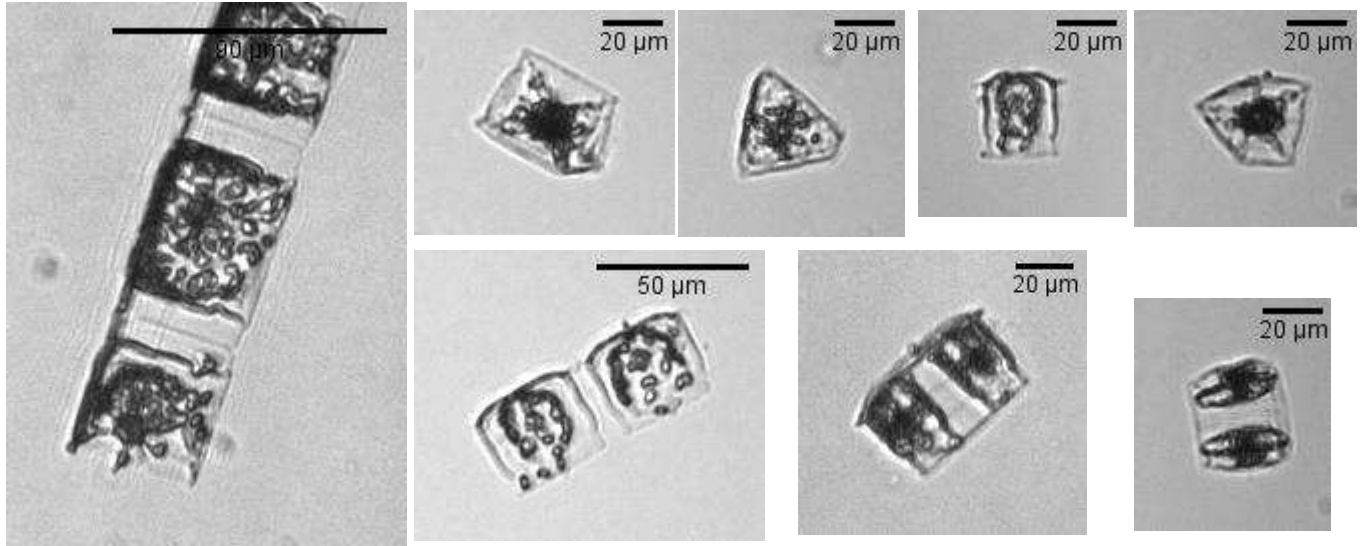
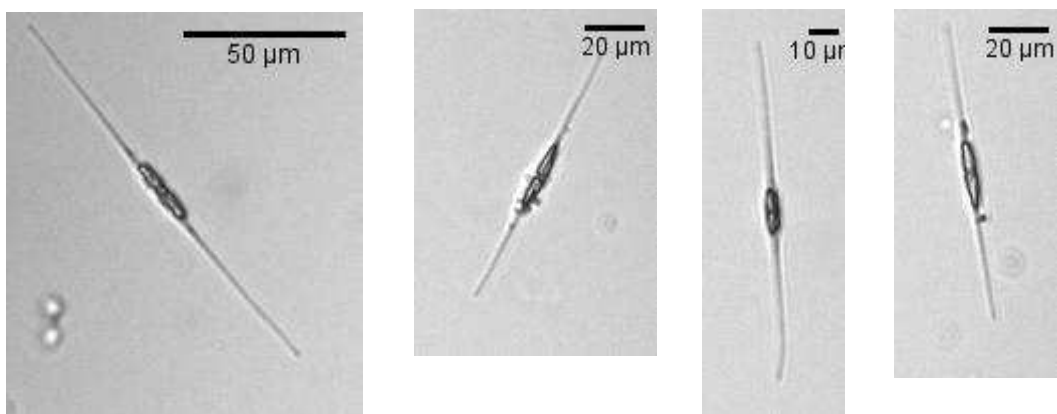


Guinardia delicatula

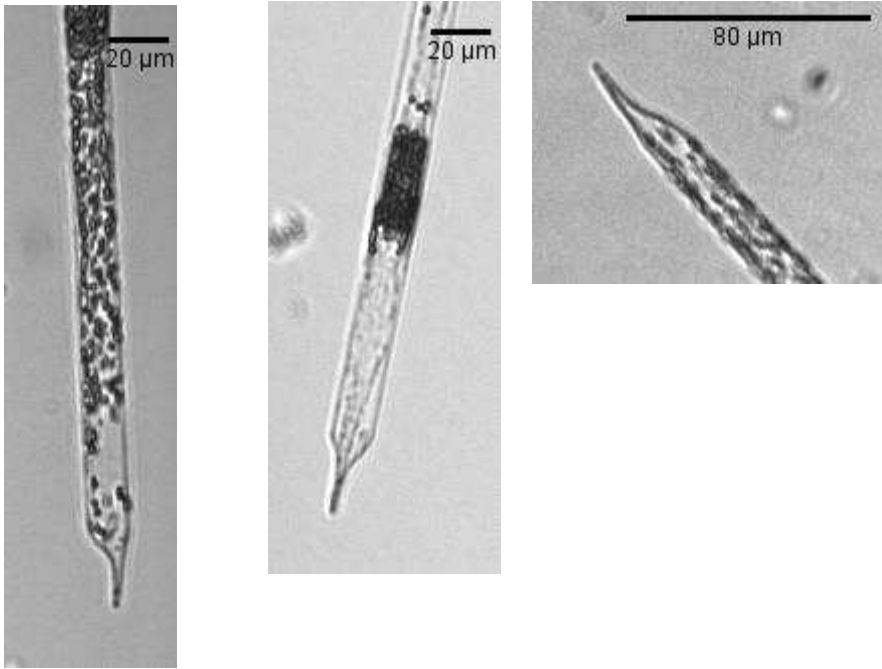


Guinardia flaccida*Guinardia striata*

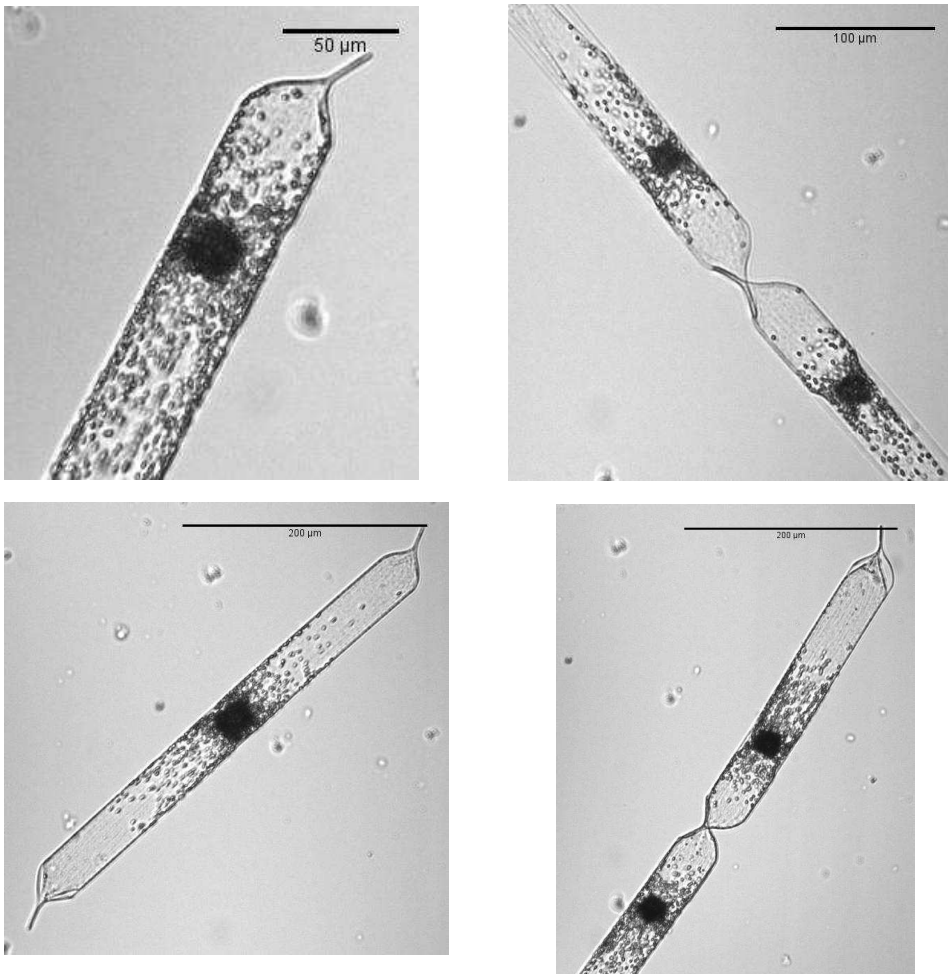
Lauderia + Schroederella*Leptocylindrus danicus*

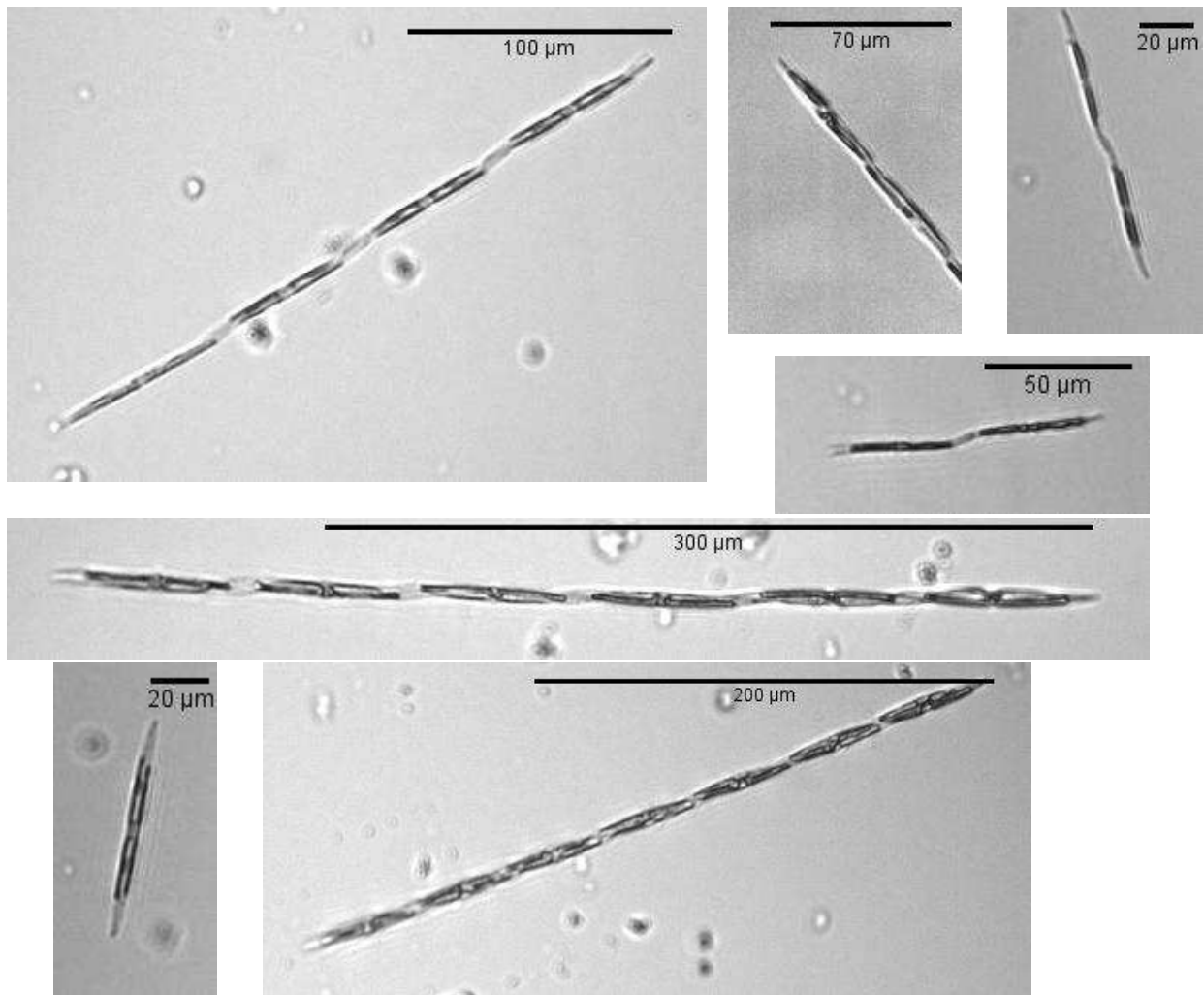
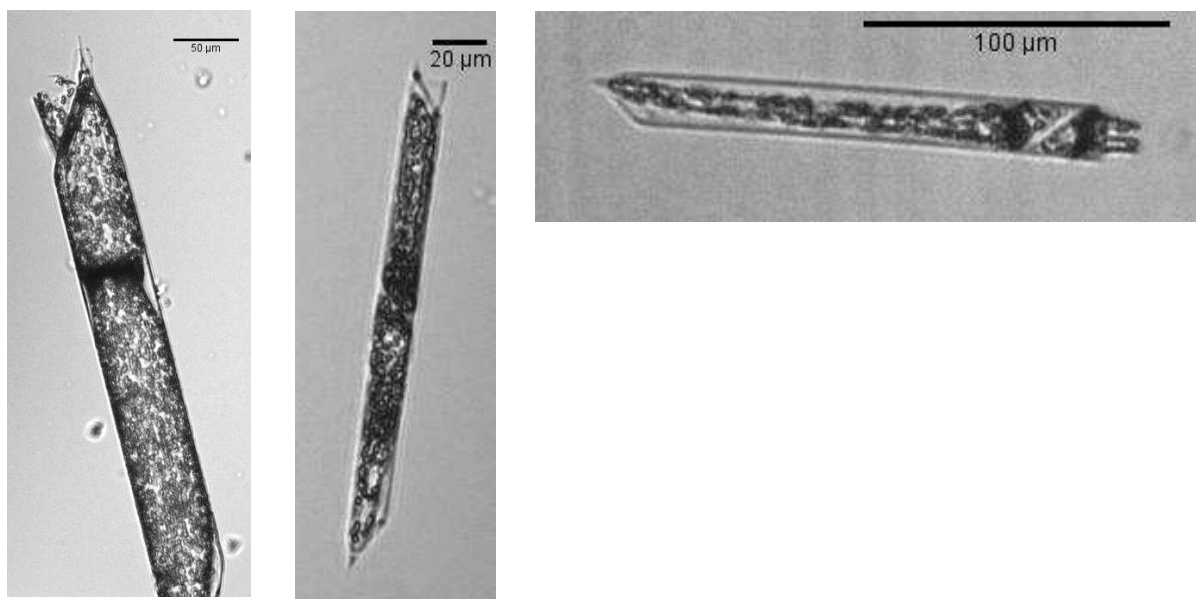
Lithodesmium undulatum*Nitzschia longissima + Cylindrotheca closterium*

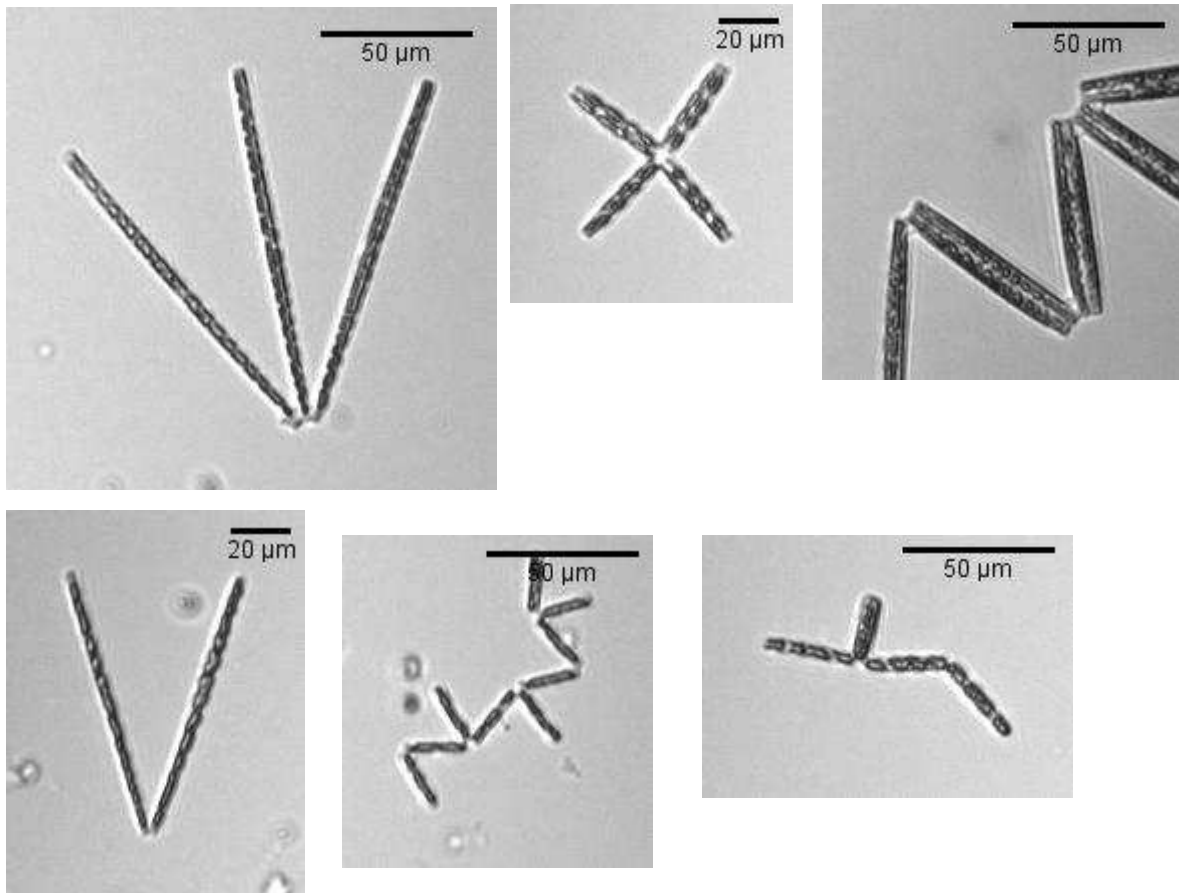
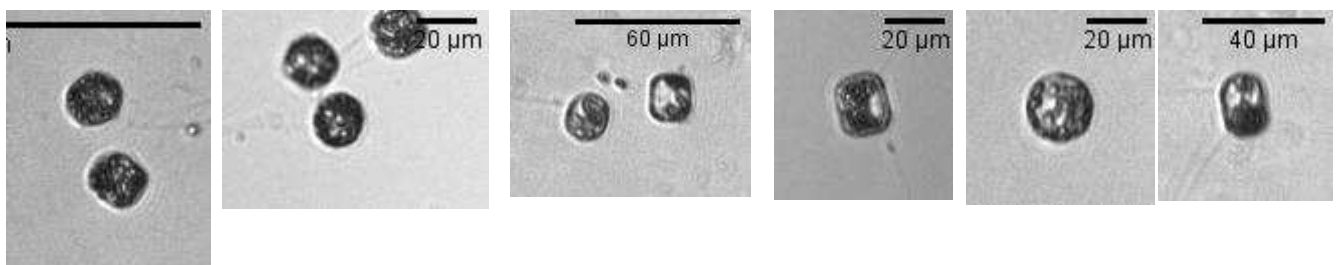
Proboscia alata

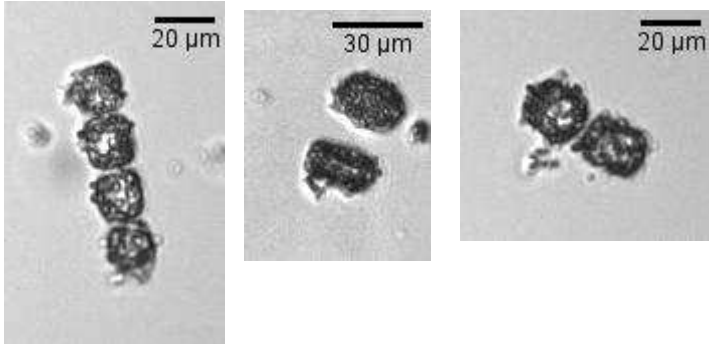
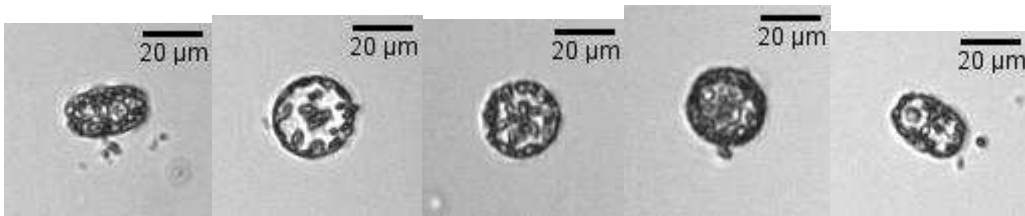
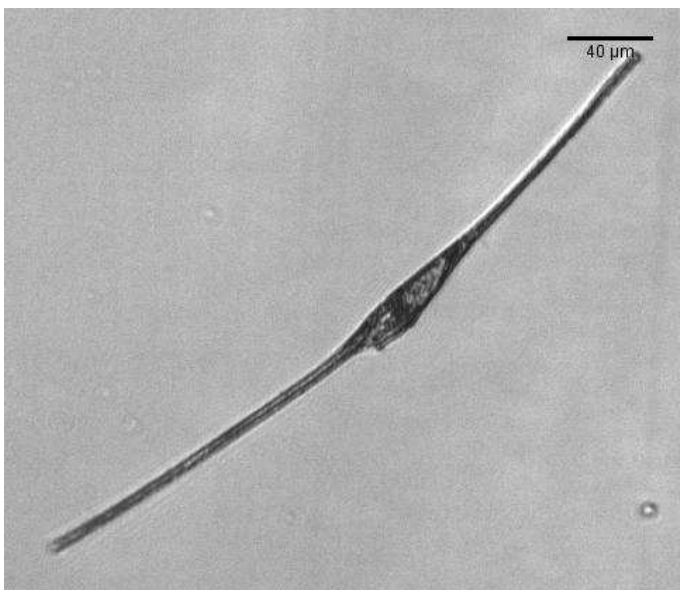


Proboscia alata var. *indica*

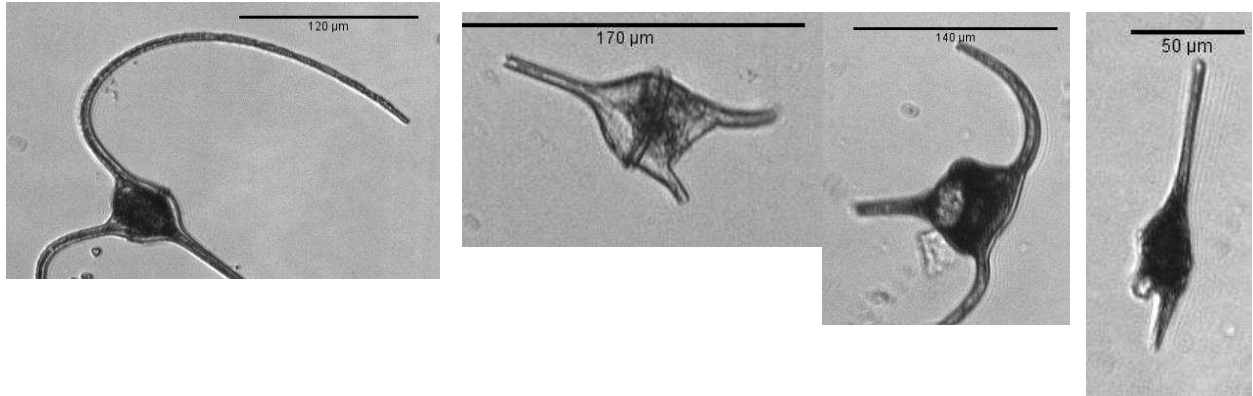


Pseudo-nitzschia spp.*Rhizosolenia imbricata* + *styliformis*

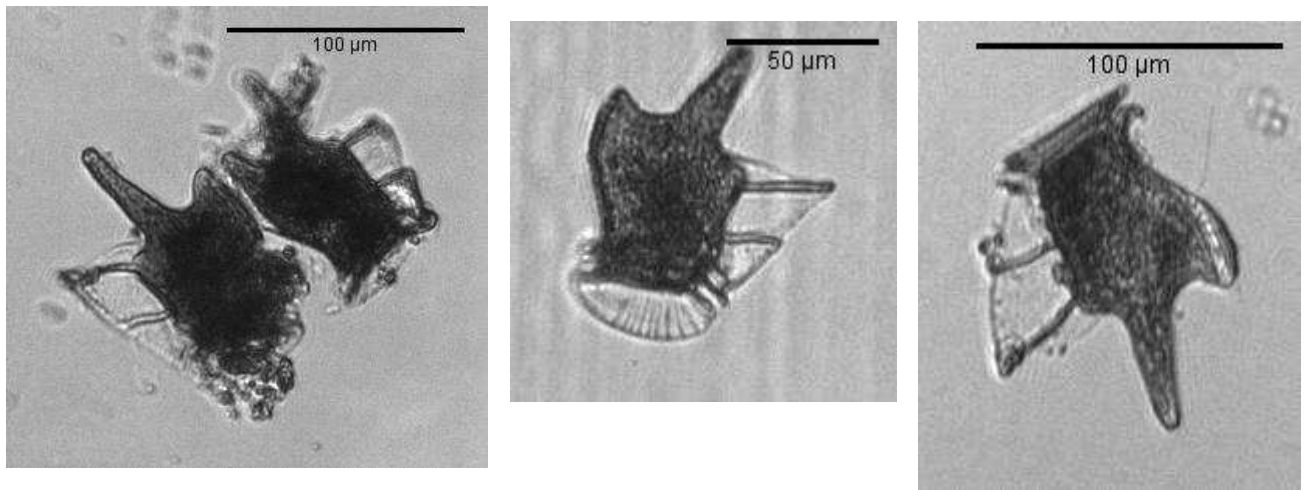
Thalassionema*Thalassiosira subtilis*

Thalassiosira spp. (espèces coloniales)*Thalassiosira* spp. (espèces solitaires)**Dinophyceae (dinoflagellés)***Ceratium fusus*

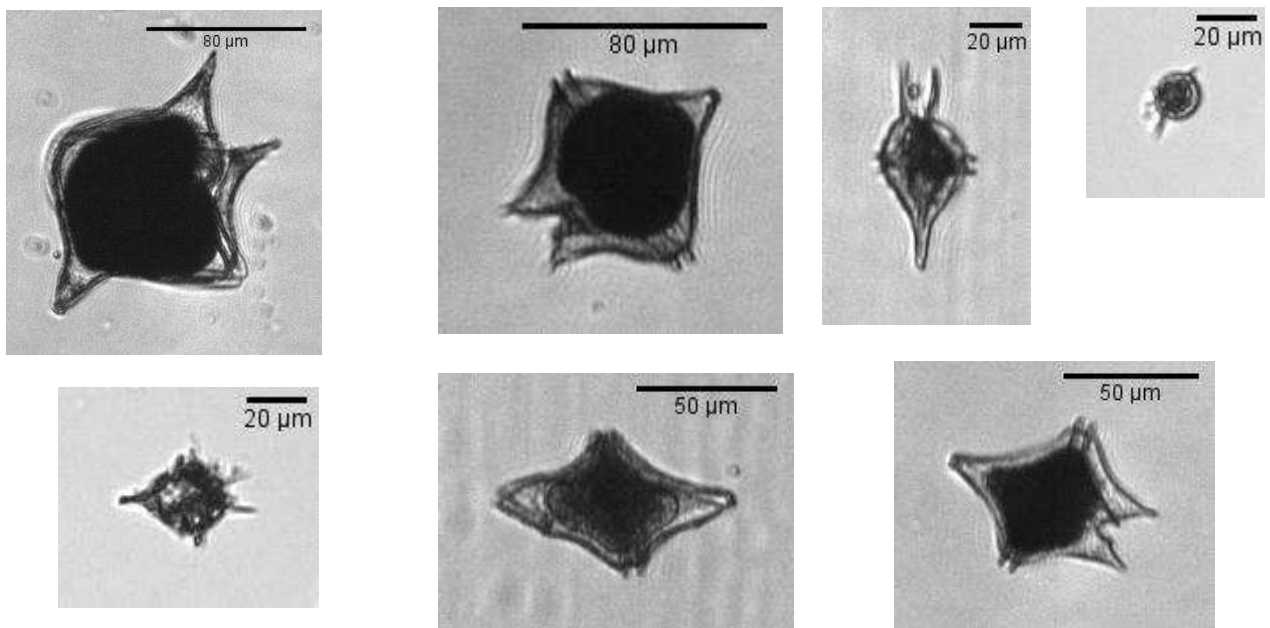
Ceratium spp.



Dinophysis tripos

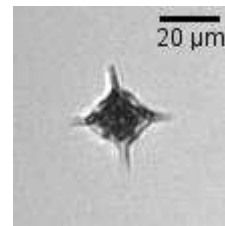
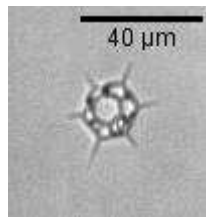
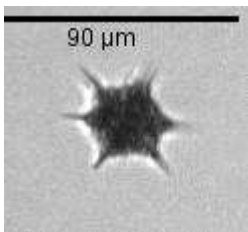
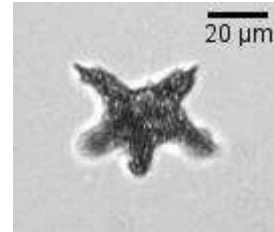
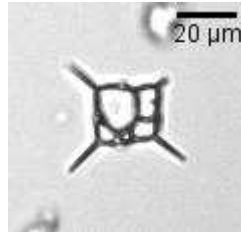
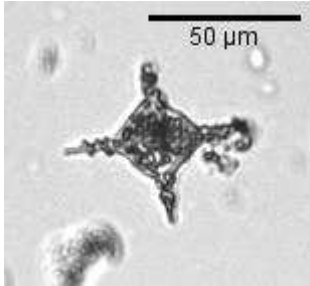


Protoperidinium spp.

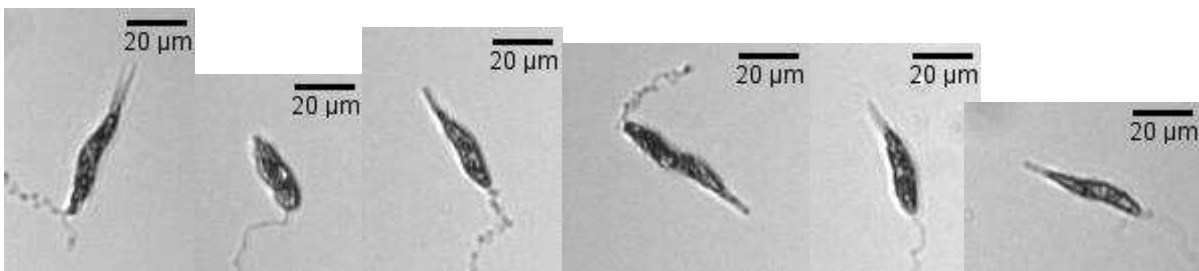


Autres groupes phytoplanctoniques

Dictyochophyceae

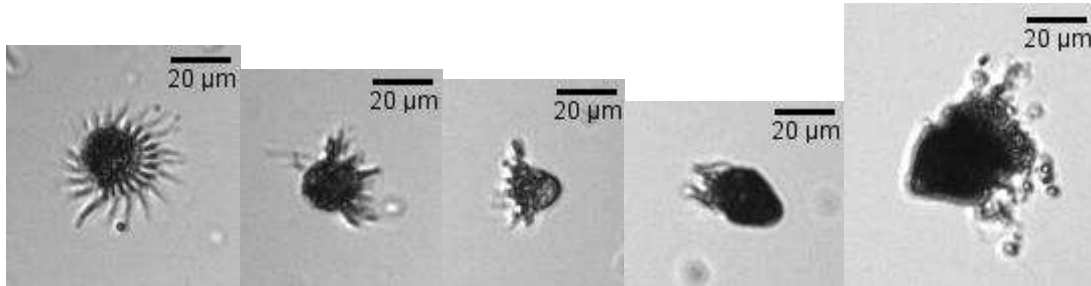


Euglenophyceae (Euglenida dans Quadrige2)

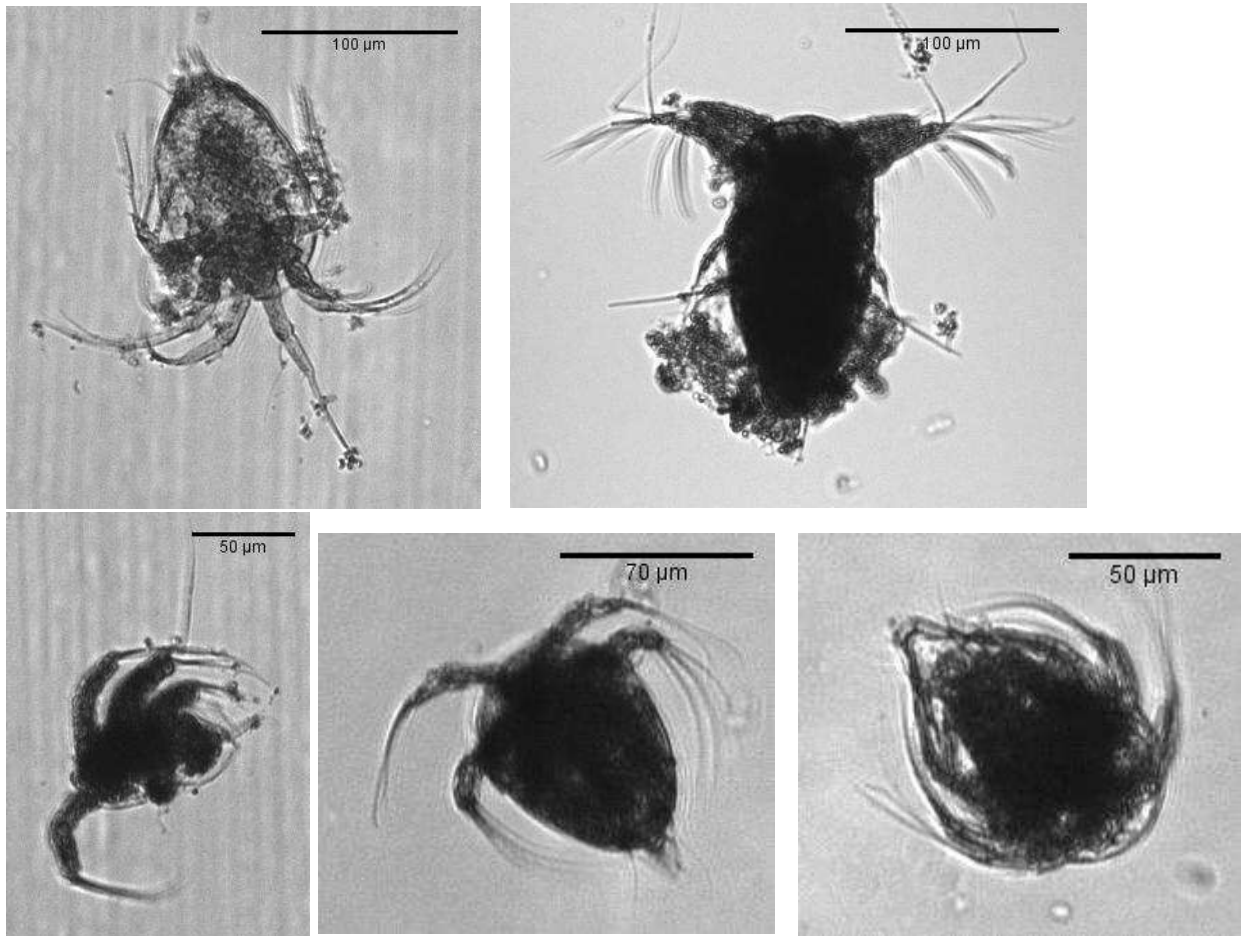


Zooplancton

Ciliés nus



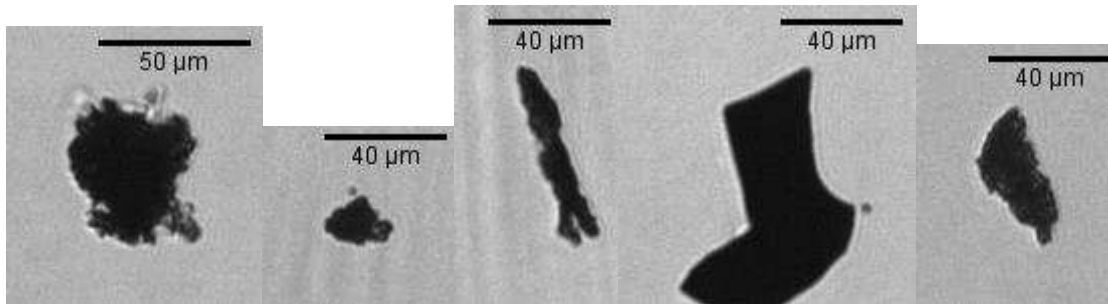
Larves de crustacés



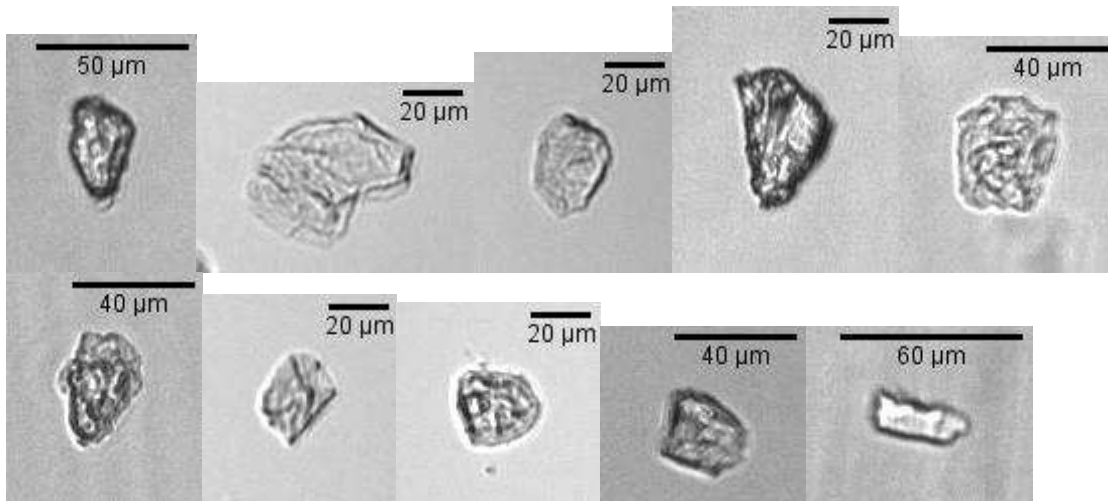
Particules inertes

Débris

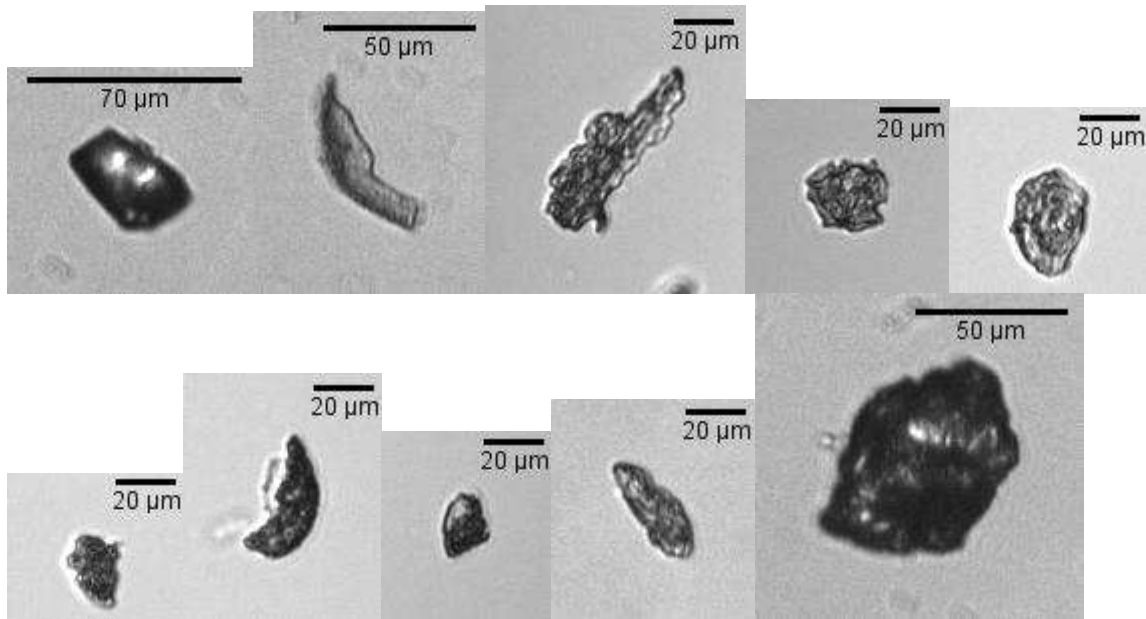
Particules noires opaques (Black_opaque)



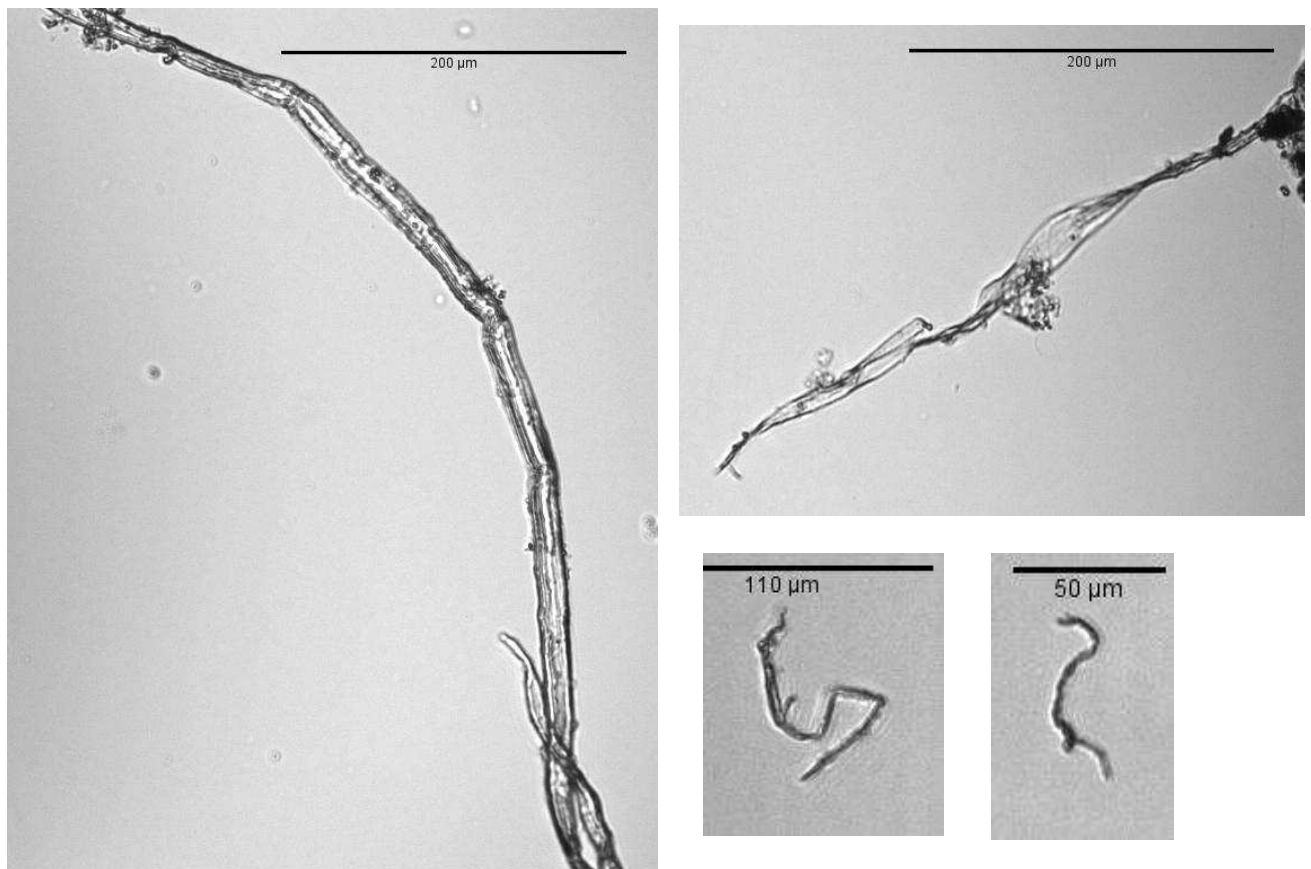
Particules claires translucides (Clear)

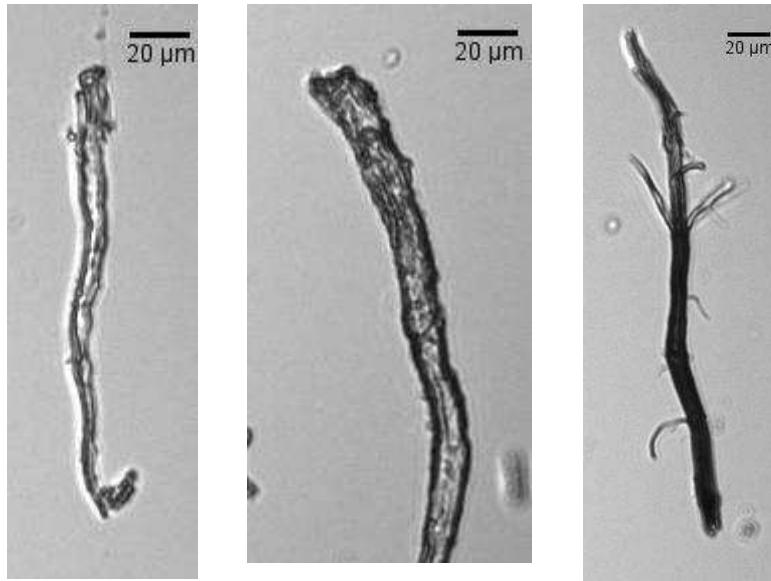


Particules foncées (Dark)

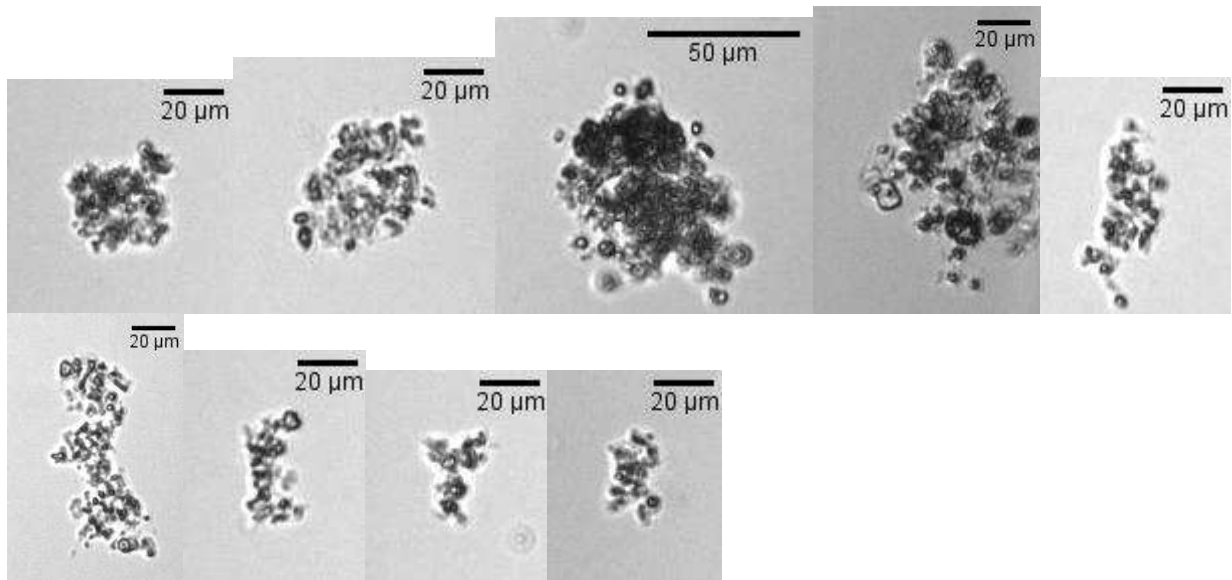


Fibres (Fibers)

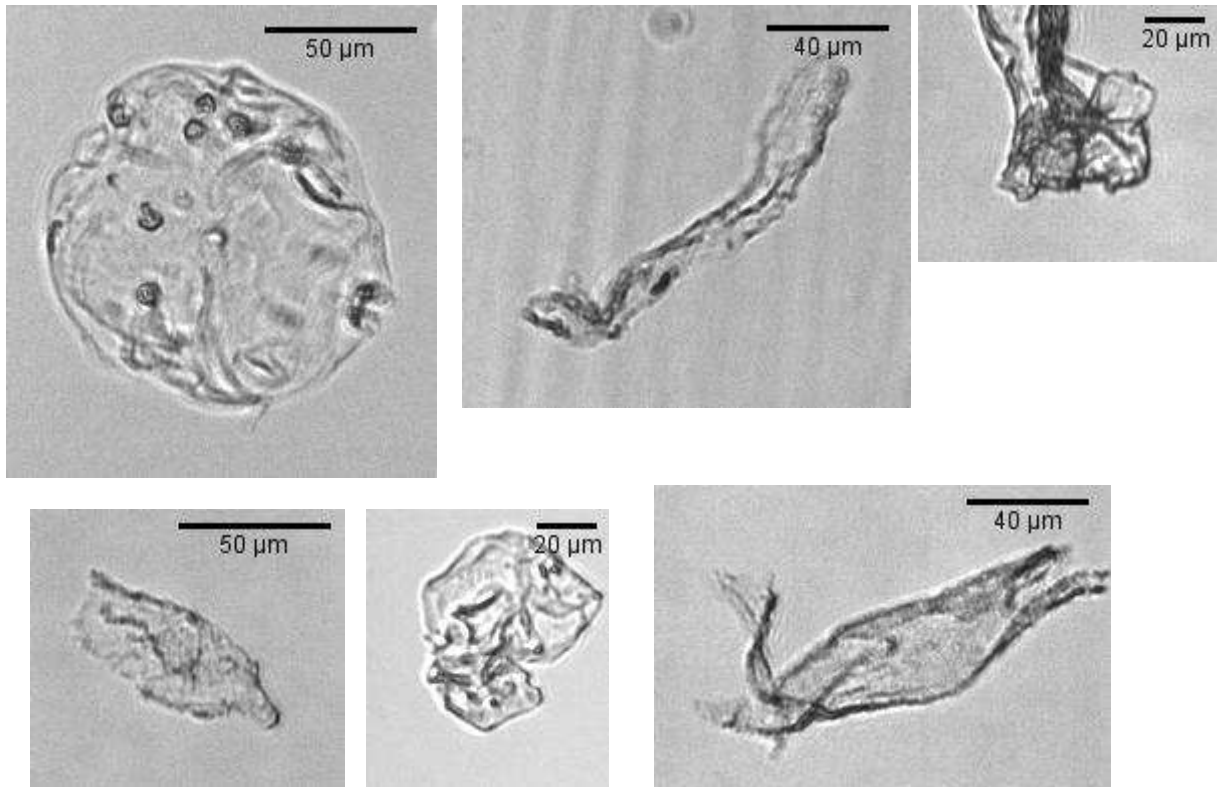




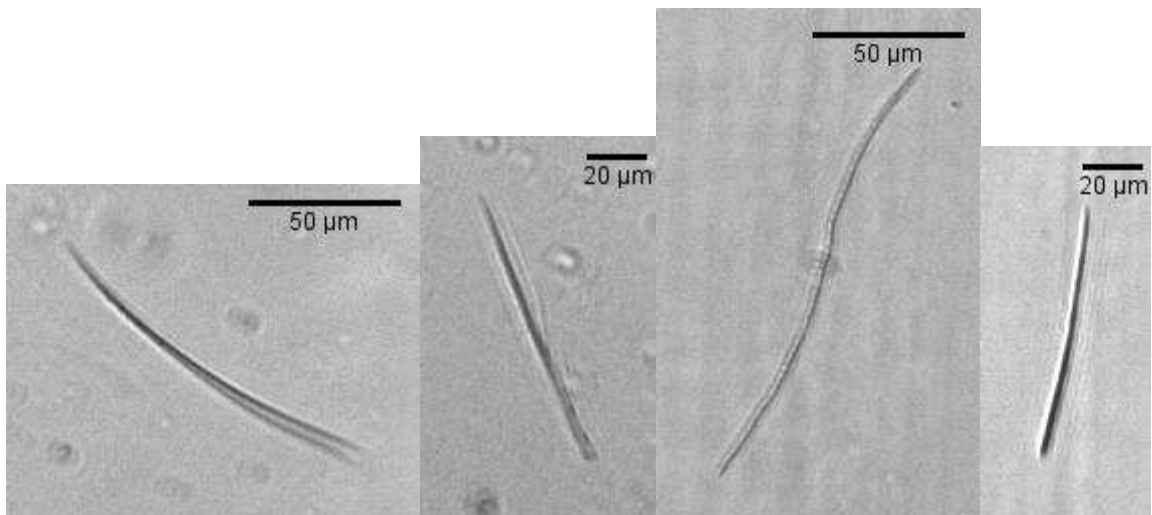
Particules détritiques granuleuses (Granular)



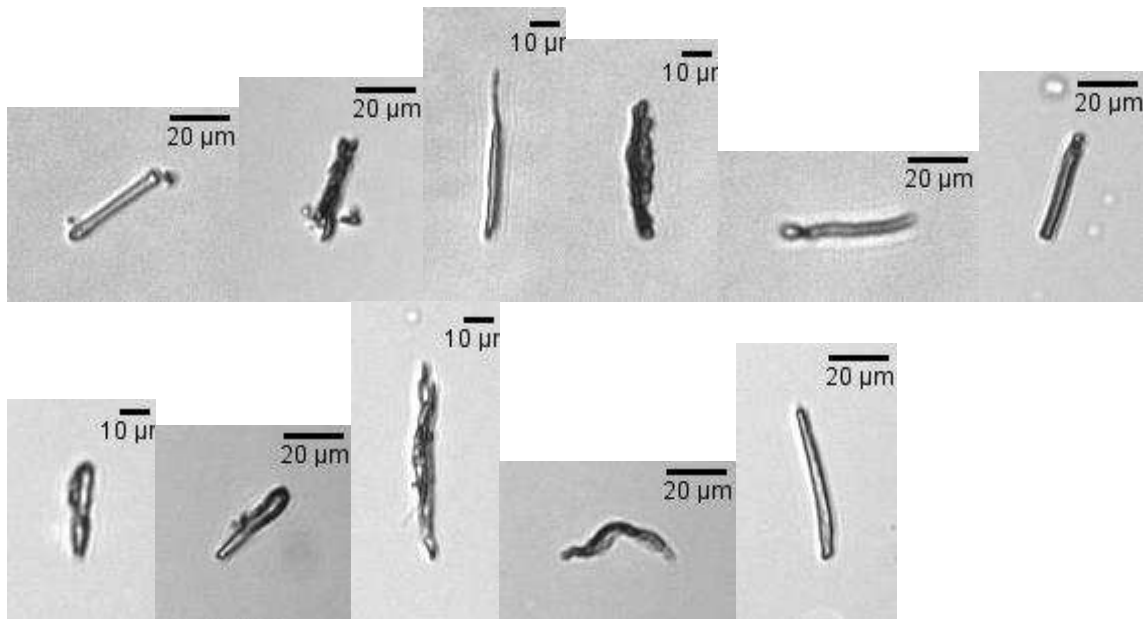
Particules détritiques membraneuses (Membranous)



Particules allongées longues (Long_thin)

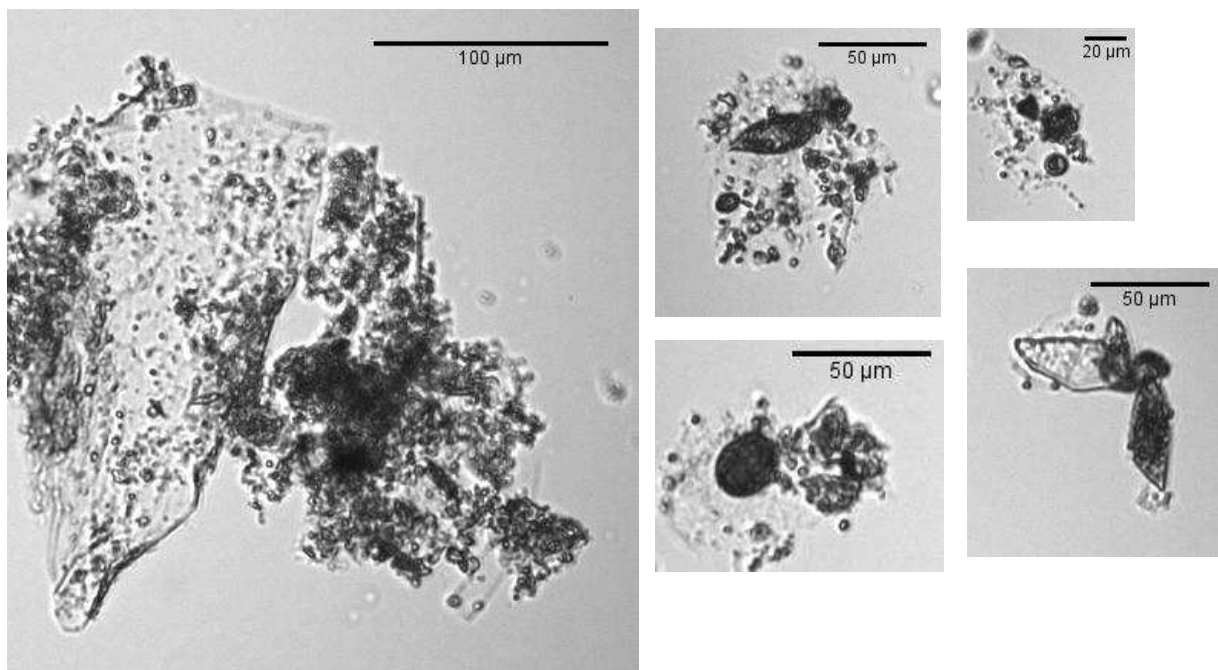


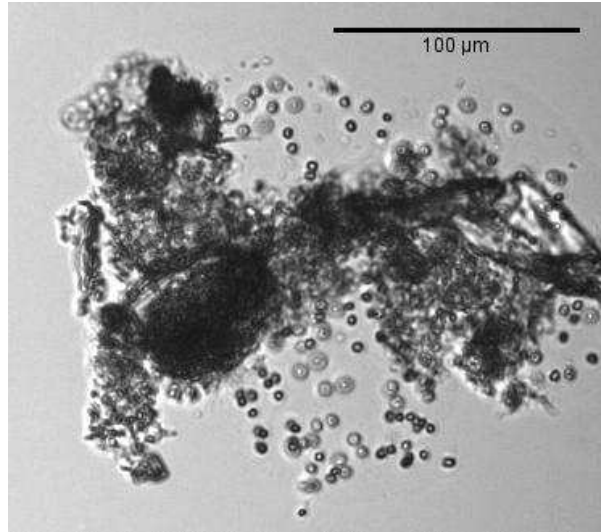
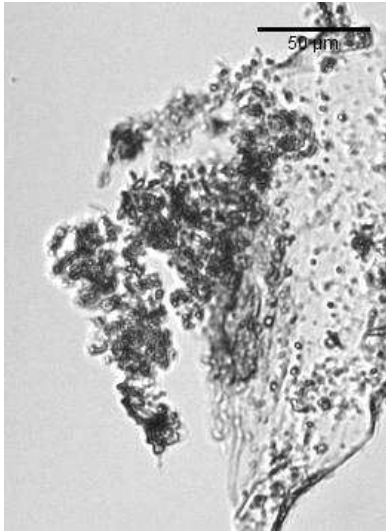
Particules allongées courtes (Short_thin)



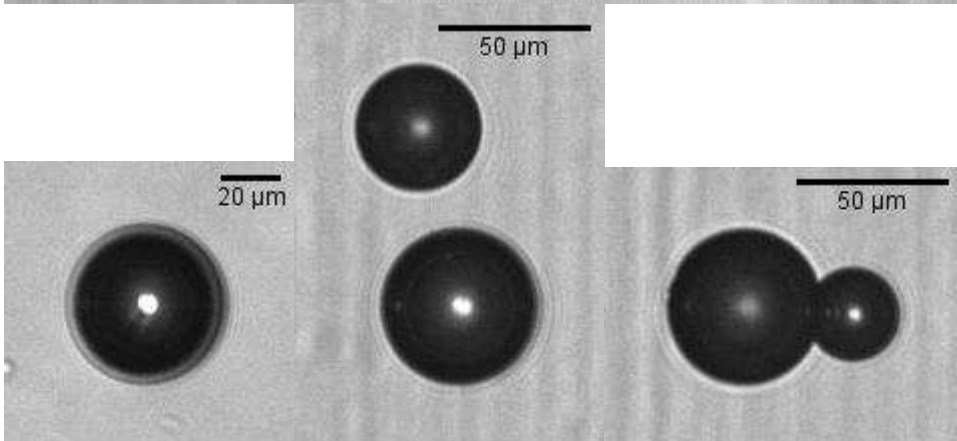
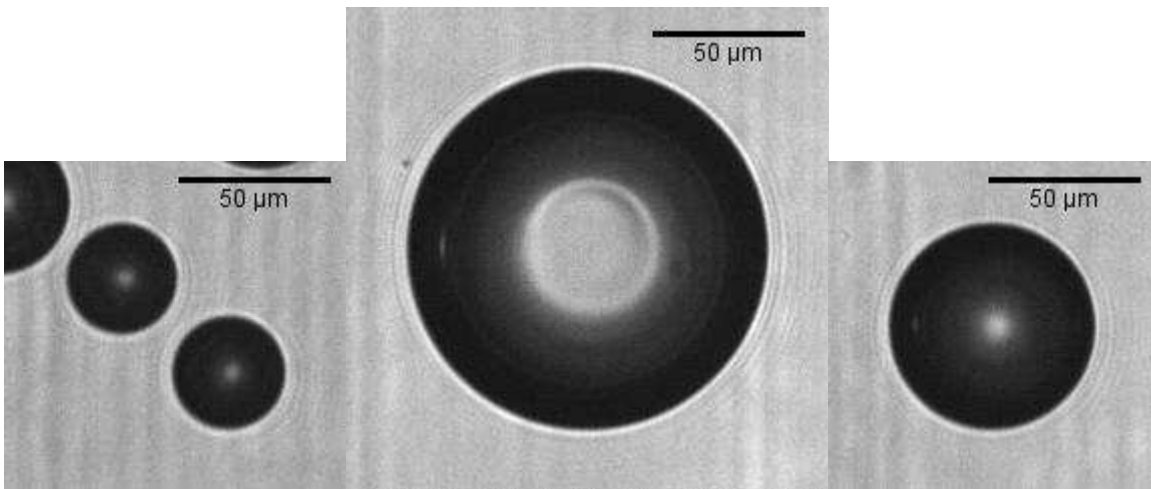
Autres particules d'origine connue

Agrégats (Aggregates)

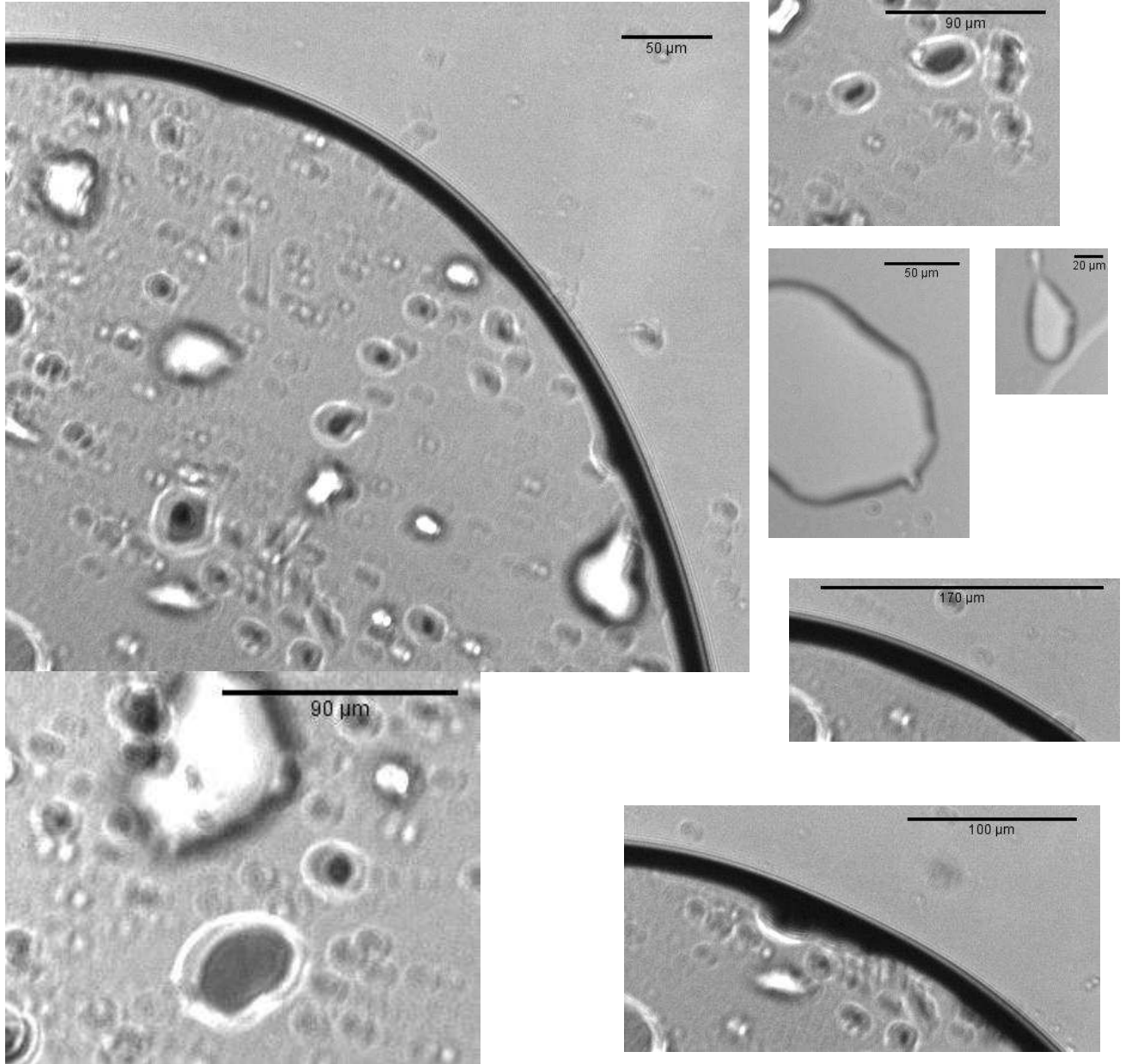




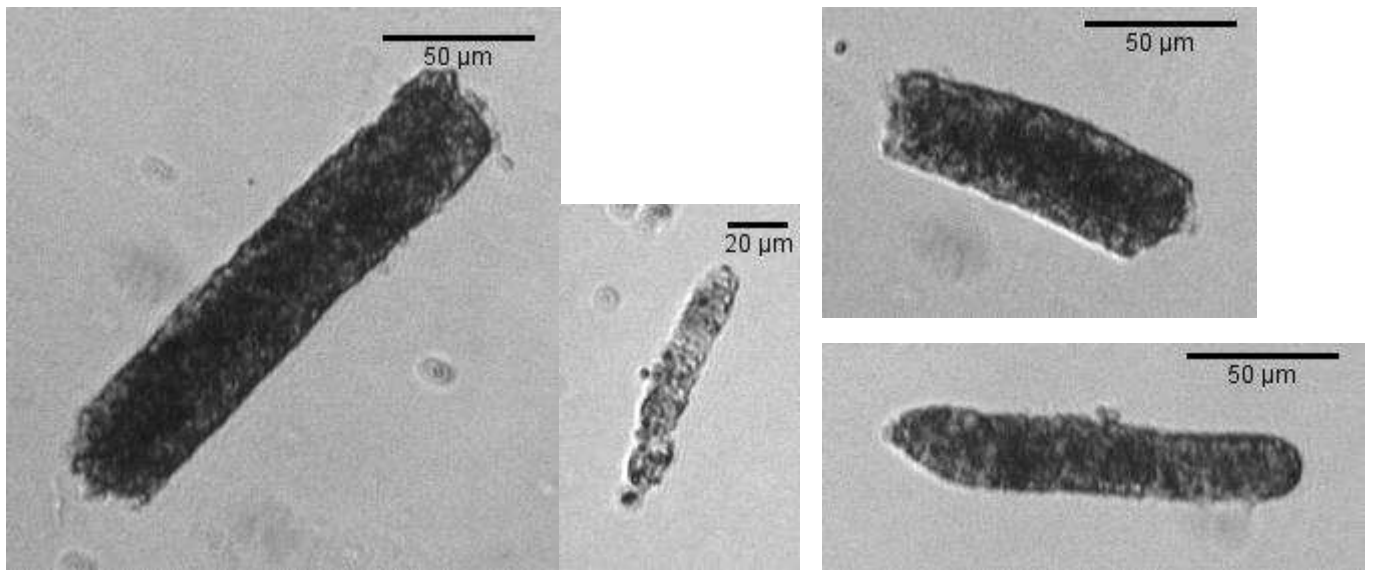
Petites bulles (Bubbles)



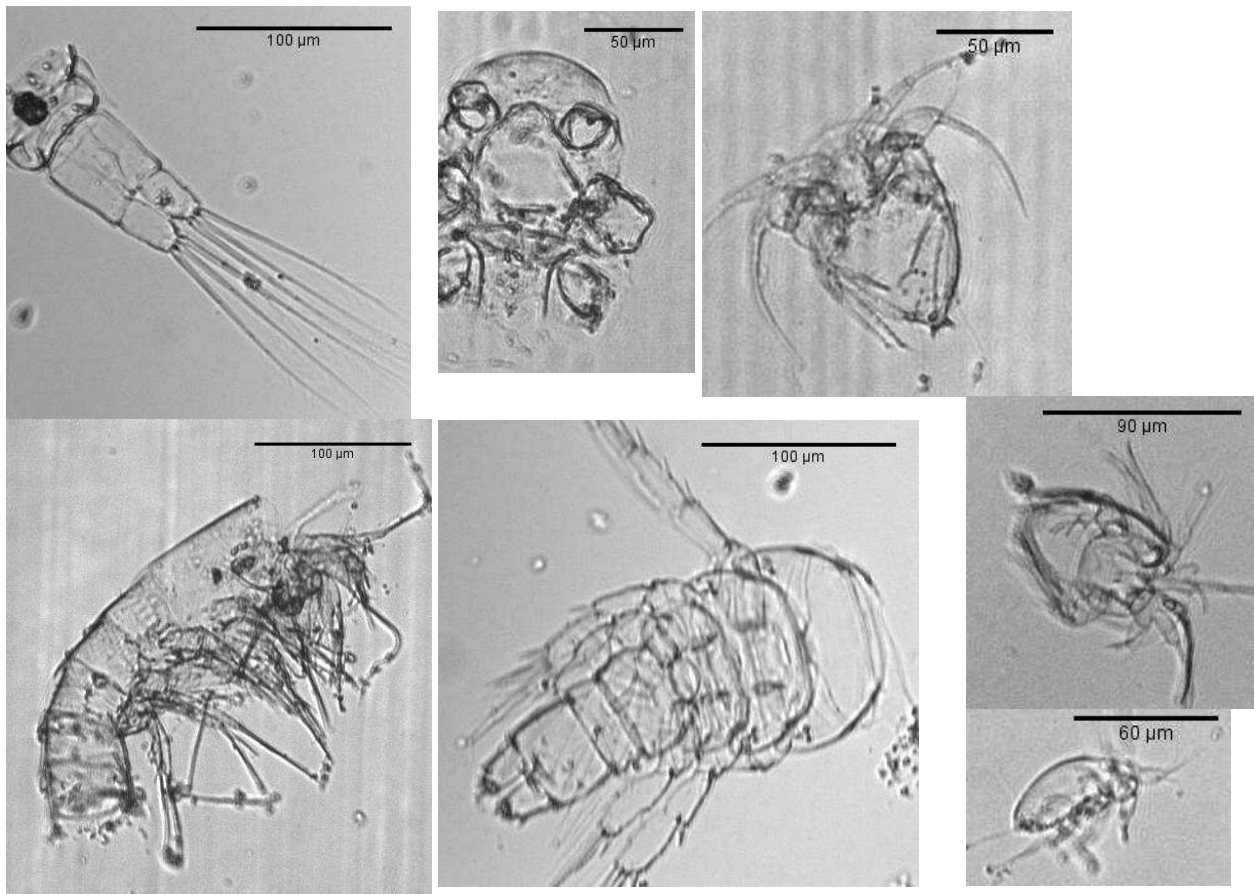
Grandes bulles (Drop)



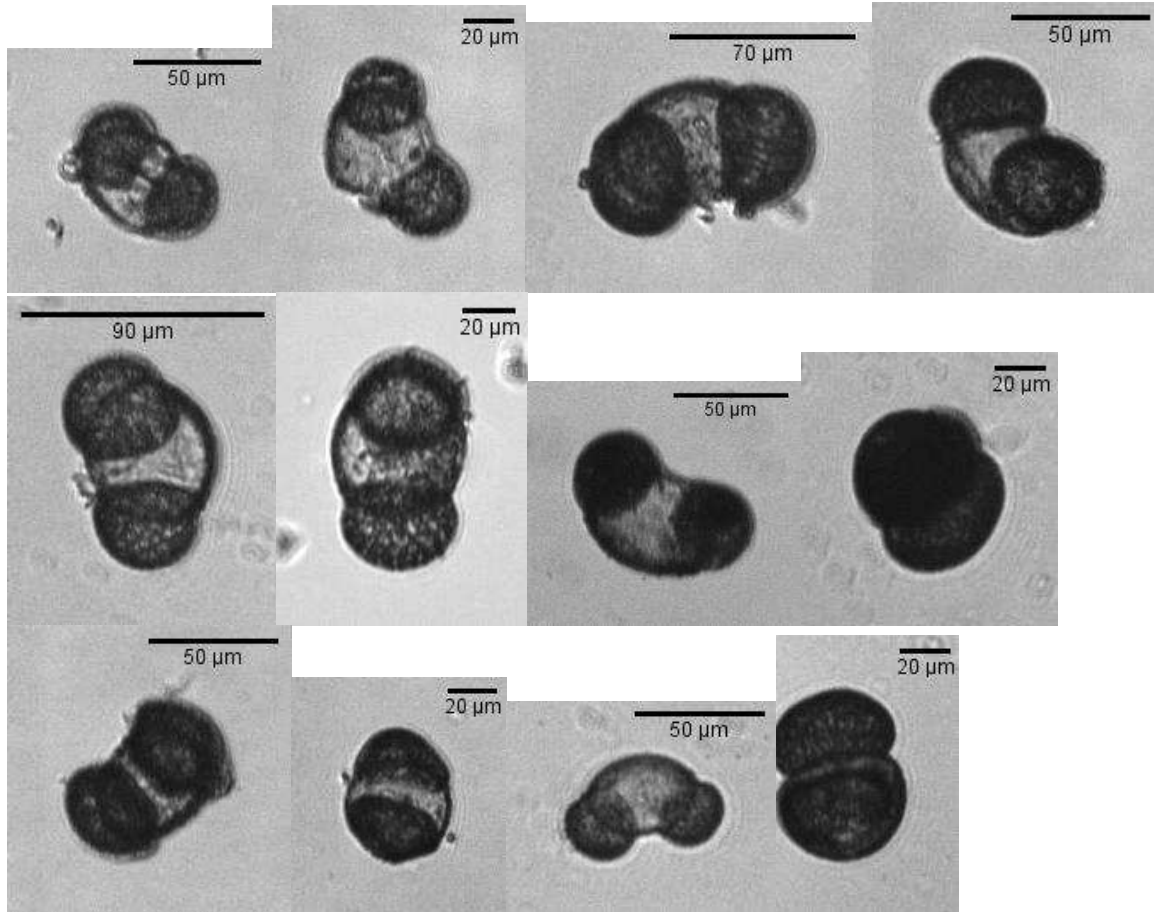
Pelotes fécales (Fecal_pellets)



Mues de crustacés (Mues)



Pollen



Annexe 8 - Nouvelles fonctionnalités du logiciel PhytoImage (version 2) (UMONS)

Dans le but de faire évoluer le logiciel PhytoImage pour répondre aux objectifs de l'IFREMER, plusieurs modules ont été intégrés à la version disponible en accès libre. Ces modules, qui offrent un usage plus dynamique des capacités du logiciel et augmentent la précision d'analyse en routine, sont développés par l'équipe EcoNum de l'UMONS, en collaboration. Ils sont la copropriété des deux organismes et de la Politique Scientifique de Belgique, et doivent être distribués dans le respect de la licence GPL 2²⁸.

Les nouvelles options du logiciel sont les suivantes.

- Application possible d'un filtre avec choix de la variable pour la classification automatique. Cette option offre ainsi la possibilité de choisir les particules à classer selon la taille, par exemple celles dont la taille est supérieure à 20 µm (limite inférieure théorique du microplancton, utilisée pour le comptage des flores REPHY), ou selon les différents paramètres mesurés par le FlowCAM.
- Possibilité de sélectionner les variables à utiliser pour la phase d'apprentissage. Cette option permet d'utiliser pour la prédiction, des variables autres que celles utilisées par défaut.
- Possibilité d'ajouter de manière automatisée de nouvelles vignettes à classer au set d'apprentissage existant.
- Méthode de classification semi-automatique permettant de classer dans des dossiers et sous dossier les vignettes d'après la prédiction faite par un outil de reconnaissance.
- Méthode de validation manuelle des prédictions automatiques : elle permet de corriger les erreurs faites par un outil de reconnaissance et d'utiliser cette correction dans l'analyse des échantillons.
- Méthode de classification de vignettes après un filtrage par un paramètre défini par l'utilisateur du logiciel.
- Nouvelles représentations graphiques (utilisées dans ce rapport) des performances de l'outil de reconnaissance qui facilitent l'analyse et la validation de ceux-ci.
- Nouvelles variables calculées sur les particules pour quantifier la forme de celles-ci.

²⁸ Rappelons que les résultats obtenus dans le cadre des conventions entre l'IFREMER et l'UMONS sont également la copropriété des deux organismes.