

2013 - **Domaine** Outils pour la surveillance environnementale - **Action 9** FlowCAM Phytolmage

# Finalisation d'un outil pré- opérationnel FlowCAM / ZooPhytolmage

**Action9 - FlowCAM ZooPhytolmage. Livrable n° 3**

**Rapport final**


**Guillaume WACQUET (Ifremer)**

**Alain LEFEBVRE (Ifremer)**

**Florent COLAS (Ifremer)**

**Philippe GROSJEAN (Université de Mons)**

Février 2014, mis à jour en Juillet 2014

 Ifremer

 **UMONS**  
Université de Mons

## **AUTEURS**

**Guillaume WACQUET (Ifremer)** - guillaume.wacquet@ifremer.fr

**Alain LEFEBVRE (Ifremer)** - alain.lefebvre@ifremer.fr

**Florent COLAS (Ifremer)** - florent.colas@ifremer.fr

**Philippe GROSJEAN (Université de Mons)** - philippe.grosjean@umons.ac.be

## **CORRESPONDANTS**

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

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

## **AUTRES CONTRIBUTEURS**

**Bertrand FOREST (Ifremer)**

**Morgan TARDIVEL (Ifremer)**

**Marie-Pierre CRASSOUS (Ifremer)**

**Liliane FIANT (Ifremer)**

**Droits d'usage** : libre accès

**Niveau géographique** : national

**Couverture géographique** : nationale

**Niveau de lecture** : experts

## RESUME

Aujourd'hui, la méthode la plus utilisée pour l'analyse des échantillons dans le cadre du REPHY est celle de la microscopie. Cependant, cette méthode souffre de nombreux inconvénients : niveaux de compétence différents selon les opérateurs, taxons pouvant faire l'objet d'identifications erronées, observateur sujet à la fatigue et à la déconcentration, erreurs non quantifiables. C'est pourquoi, différents axes d'évolution ont été proposés pour adapter le FlowCAM couplé à ZooPhytoImage aux exigences du REPHY qui sont la justesse et la répétabilité de la mesure. L'objectif de ces nouveaux outils est d'offrir un gain de temps tangible aux observateurs REPHY par rapport aux observations au microscope optique.

Le temps d'analyse étant un point clef de ce système, des études ont été menées afin de pouvoir réaliser une acquisition rapide. En ce sens, afin que la mesure soit représentative du milieu, au moins 2000 particules doivent être comptées. Pour réduire le temps d'analyse, nous choisissons donc de pré-concentrer les échantillons lorsque la concentration cellulaire est inférieure à 105 cellules/L. Parallèlement à cela, la répétabilité et la justesse de la mesure sont conditionnées par les réglages optiques du système. C'est pourquoi, un actionneur automatique permettant d'améliorer la répétabilité du positionnement de la cellule de flux par rapport à l'objectif a été développé et installé sur le FlowCAM.

Pour effectuer des analyses quantitatives des échantillons, il est également nécessaire de procéder à différents réglages liés au système fluidique du FlowCAM. Des expérimentations sur le sens du débit et sur la viscosité des échantillons ont été réalisées. Il en résulte qu'une pré-filtration systématique des échantillons s'impose pour certaines analyses. Cependant, une pré-filtration de l'échantillon ne permet pas (ou plus) l'analyse de flores totales. Cette solution est donc dépendante des objectifs fixés par l'utilisateur (flores partielles, flores indicatrices, etc.) et doit être appliquée au cas par cas.

Les comparaisons quantitatives des abondances et concentrations obtenues par lectures au microscope et par FlowCAM/ZooPhytoImage ont ensuite été réalisées pour des échantillons REPHY 2013. Un volume de 10 mL a été numérisé pour chacun de ces échantillons, 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 100 cellules/L, ce qui est équivalent à un comptage au microscope inversé. Les moyennes des taux de reconnaissance des espèces sont d'environ 69% pour les échantillons lugolés et 67.50% pour les échantillons vivants. Ces indices de performances sont donc encourageants. Cependant, de tels scores sont trop faibles pour donner des classifications fiables en routine. Il faut donc valider les données, grâce à une étape de vérification manuelle des classements effectués par l'ordinateur. Cette opération est accélérée par le fait que l'ordinateur a effectué un classement correct à 65-75%, par rapport à une classification purement manuelle des vignettes. Le module de correction détermine une probabilité d'être suspect pour chaque particule, et seules les particules les plus suspectes sont à valider. Cette approche particulière offre un compromis idéal entre la méthode automatique et la validation manuelle totale, tout en garantissant des performances de reconnaissance similaires ou meilleures dans un laps de temps acceptable.

Un problème important concerne le stockage et la bancarisation des données issues du système FlowCAM/ZooPhytoImage. En effet, la gestion des données numériques associées à une image ainsi que le problème de la flexibilité du système avec plusieurs protocoles pour un jeu de données, peuvent rendre *a priori* complexe la bancarisation des données. L'intégration des données sous Quadrigé<sup>2</sup> passe donc par la rédaction d'un cahier des charges afin de définir les exigences du système.

Le système couplé FlowCAM / ZooPhytoImage devient un outil véritablement opérationnel en 2014. Totalement adapté aux observations du phytoplancton réalisées dans le cadre du réseau d'observation REPHY, il permettra de mieux répondre aux sollicitations présentes et futures concernant l'évaluation de la qualité des eaux littorales et marines dans le cadre des exigences européennes, telles que la DCE et la DCSMM. Un des bénéfices immédiats sera par

exemple pour l'acquisition des données nécessaires au calcul de l'indice abondance composant l'indicateur phytoplancton pour la DCE en Manche- Atlantique, indice qui est basé sur la proportion de taxons du micro-phytoplancton présents en quantité importante dans un échantillon.

### **MOTS CLES (THEMATIQUE ET GEOGRAPHIQUE)**

Optimisation technologique; système optique; protocole de numérisation; analyse quantitative; correction d'erreur.

## TITLE

Finalization of a pre-operational tool FlowCAM/ZooPhytoImage.

## ABSTRACT

Today, the most widely used method for the analysis of samples within the REPHY network, is the microscopy. However, this method has several drawbacks: different skill level depending on the analyst, taxa may be misidentified, analyst subject to fatigue and loss of concentration, non-quantifiable errors. That is why different axis of development have been proposed to adapt the FlowCAM and ZooPhytoImage to the requirements of REPHY network which are the accuracy and repeatability of the measurement. The purpose of these new tools being to provide a tangible gain of time for analysts compared to the observations with optical microscope.

The analysis time being a key point of this system, studies were conducted in order to achieve a fast acquisition. To have a representative measurement of the environment, 2,000 particles must be counted. To reduce the analysis time, we choose to pre-concentrate the samples when the cell concentration is less than 105 cells/L. At the same time as this, the repeatability and accuracy of measurement are determined by the optical system settings. This is why, an automatic actuator allowing to improve the repeatability of the flow cell positioning with respect to the lens, has been installed on the FlowCAM.

To perform quantitative analyzes of the samples, it is also necessary to make various settings associated to the fluid system of FlowCAM. Experiments on the flow direction and on the viscosity of the samples were performed. The results show that a systematic pre-filtration of the samples is required for some analyzes. However, a pre-filtration of the sample does not allow analysis of total flora. This solution is therefore dependent on the objectives set by the user (partial flora, indicator flora, etc.) and must be applied on a case by case basis.

The quantitative comparisons of abundances and concentrations obtained by the microscope and by FlowCAM/ZooPhytoImage were then performed for samples of REPHY 2013. A volume of 10 mL was digitized for each of these samples, which offers a counting accuracy comparable to the one resulting from manual counting. Indeed, one particle detected by FlowCAM corresponds to in situ cell concentration of 100 cells/L, which is equivalent to a count with the inverted microscope. The averages of species recognition rates are approximately 69% for the samples fixed with lugol's solution and 67.50% for the live samples. These performance indices are encouraging. However, such scores are too low to provide reliable classifications routinely. It is therefore necessary to validate the data through a manual verification step of predictions done by the computer. This process is accelerated by the fact that the computer has made a correct classification of 65-75 %, with respect to a manual classification of thumbnails. This module determines a probability of being suspect for each particle, and only the most suspect particles are validated. This particular approach offers an ideal compromise between the automatic method and the total manual validation, while guaranteeing similar or better performance of recognition within an acceptable time.

An important problem lies in the storage and banking of data from the FlowCAM/ZooPhytoImage system. Indeed, the management of digital data associated with an image and the problem of the flexibility of the system with multiple protocols for a dataset, can be complex. Therefore, the integration of data in Quadrige <sup>2</sup> requires the writing of specifications to define system requirements.

The FlowCAM / ZooPhytoImage is becoming an operational tool in 2014. Completely adapted to the phytoplankton observations performed in the context of the French monitoring network REPHY, it will allow answering more accurately to the questions of WFD and MSFD concerning the evaluation of marine water quality. For instance, the first benefit will be for the acquisition of data necessary to the calculation of abundance index, part of the phytoplankton index for WFD in Channel and Atlantic water bodies : as a matter of fact, this

index is based on the proportion of micro-phytoplankton taxa which are very abundant in a water sample.

### **KEY WORDS (THEMATIC AND GEOGRAPHICAL AREA)**

Technological optimization; optical system; digitization protocol; quantitative analysis; error correction.

## SYNTHESE POUR L'ACTION OPERATIONNELLE

Ce travail vise à développer des outils innovants permettant de faciliter en l'automatisant l'identification de taxons phytoplanctoniques (au niveau le plus fin possible) et leur dénombrement et de les rendre ainsi plus aisément utilisables en routine tout en limitant les biais dus aux opérateurs. Il s'agit d'un système couplant la numérisation des cellules phytoplanctoniques par analyse d'images (FlowCAM) et un logiciel de reconnaissance des images numérisées (PhytoImage). En 2013, trois actions sont menées :

- PhytoImage : améliorations logicielles (interface, mais aussi pour la classification des images) et édition d'un guide d'utilisation (livrable 1)
- PhytoImage : amélioration de la reconnaissance et de la classification des espèces (livrable 2)
- FlowCAM : amélioration du fonctionnement pour rendre l'utilisation en routine plus fiable et rapide (préparation et conservation des échantillons, réglages optiques, réglages de la circulation des flux).

Par ailleurs, afin d'améliorer les capacités de l'ensemble FlowCAM/PhytoImage, deux actions complémentaires ont été engagées :

- la recherche d'attributs comme la texture des particules et les adaptations de FlowCAM/PhytoImage permettant de les détecter
- la comparaison des observations obtenues avec FlowCAM/PhytoImage et le cytomètre de flux (Cytosense).

Ces travaux constituent le livrable 4.

Aujourd'hui, plus de 3000 échantillons d'eau sont examinés annuellement pour des lectures de phytoplancton dans le cadre du REPHY et des réseaux régionaux pilotés par l'IFREMER. Les données obtenues (identification et dénombrement par taxon, genre ou espèce) seront utilisées en particulier pour répondre aux exigences de la DCE, puisqu'elles permettront de calculer les indices de l'élément de qualité phytoplancton, qui fait partie des éléments biologiques obligatoires pour l'évaluation de la qualité écologique des masses d'eau.

La méthode la plus utilisée à ce jour est celle de la microscopie. Les identifications des taxons phytoplanctoniques sont effectuées sur 10 mL d'eau de mer fixée au Lugol, au microscope inversé selon la méthode d'Utermöhl (décantation de l'échantillon dans une cuve). L'identification se fait au niveau le plus bas possible (espèce ou genre) sinon au rang supérieur (famille voire ordre ou classe). Cette méthode nécessite une formation approfondie en taxonomie mais le niveau de compétence des opérateurs étant différent, certains taxons peuvent leur être inconnus ou faire l'objet d'identifications erronées. De plus, 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.

De ce fait, des nouveaux systèmes sont mis en place pour l'identification automatique des espèces de phytoplancton contenues dans les échantillons d'eau de mer. L'IFREMER mène des actions pour intégrer ces équipements au REPHY. Suite au travail de post-doctorat de Tunin-Ley (LER Arcachon puis Université de Bordeaux), différents axes d'évolution ont été proposés pour adapter le FlowCAM couplé à ZooPhytoImage aux exigences du REPHY qui sont la justesse et la répétabilité de la mesure (Tunin-Ley et Maurer, 2011). A cela s'ajoute la rapidité : l'objectif de ces nouveaux outils étant d'offrir un gain de temps tangible aux observateurs REPHY par rapport aux observations au microscope optique.

En effet, le temps d'analyse d'un échantillon est un point clef de ce système. En effet, pour l'analyse du phytoplancton deux grossissements sont possibles : 4X et 10X. L'utilisation d'un objectif de grossissement faible tel que le 4X (qui correspond à un grossissement global de 40X) associé à une cellule de grande profondeur (300 ou 600  $\mu\text{m}$ ) s'avère utile dans le cas

où certaines espèces de grande taille sont présentes en abondance. Cependant, dans ce cas, la résolution des images n'est pas toujours suffisante pour identifier manuellement certaines espèces présentes dans l'échantillon. Pour pallier ce problème, l'utilisation de l'assemblage optique 10X/100 $\mu$ m (grossissement global de 100X) peut être utilisé. Cet assemblage permet d'obtenir une meilleure description morphologique grâce aux détails plus fins, et une 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. Néanmoins, le temps d'acquisition s'avère être bien plus long pour le second. Un gain considérable serait de pouvoir réaliser une acquisition au grandissement 10X avec un temps d'analyse plus court, comparable au 4X. Des pistes d'évolution du système sont étudiées dans ce sens. En effet, le volume à numériser doit être choisi pour que la mesure soit représentative du milieu. Historiquement, les analyses du REPHY reposent sur l'observation de 10 mL au microscope optique. La transposition de ce protocole au FlowCAM peut se faire en considérant que le volume à numériser doit être de 10 mL. Cependant, l'analyse au 10X est alors très longue. Par ailleurs, l'approche FlowCAM/ZooPhytoImage impose une contrainte supplémentaire : l'analyse d'un échantillon est considérée statistiquement représentative si au moins 2000 particules sont comptées et analysées. Pour réduire les analyses à 20 min, nous choisissons donc de pré-concentrer les échantillons lorsque la concentration cellulaire est inférieure à 105 cellules/L.

Parallèlement à cela, la répétabilité et la justesse de la mesure sont conditionnées par les réglages optiques du système. Sur le dispositif original, l'illumination produite par la source lumineuse (LED) souffre d'un manque d'homogénéité sur l'ensemble du champ et est très dépendante de son orientation. Deux nouvelles configurations ont alors été testées, essentiellement pour le grossissement 10X. La première consiste à ajouter un dépoli devant la LED et déplacer le collimateur. La seconde consiste à injecter la lumière issue de la LED dans une fibre optique et d'utiliser l'embout de cette fibre comme source. Cependant, en raison de l'encombrement de cette dernière, la première configuration est préférable et envisageable au regard des LED disponibles sur le marché et à moindre coût. De plus, la mise au point de l'échantillon est réalisée à l'aide d'une vis micrométrique manuelle réglant le positionnement de la cuve par rapport à l'objectif. Ce procédé montre des limites car le microscope a une profondeur de champ de plusieurs dizaines de microns au regard du système d'illumination. Il n'est donc pas possible de positionner la cuve de manière répétable à plusieurs dizaines de micron près. C'est pourquoi, l'utilisation d'un actionneur automatique peut permettre d'améliorer la répétabilité du positionnement. Actuellement, ce système, qui peut s'apparenter à un système d'autofocus a été installé sur le FlowCAM et permet un positionnement plus aisé de l'échantillon.

Pour effectuer des analyses quantitatives des échantillons, il est nécessaire de procéder à différents réglages liés au système fluide du FlowCAM, et en particulier le compromis vitesse de flux / temps de numérisation / efficacité doit être optimisé. Dans un système d'imagerie de flux tel que le FlowCAM, le nombre de particules comptées dans un volume donné n'est pas toujours équivalent à la concentration de l'échantillon. Cela s'explique principalement par le fait que le volume imagé ne correspond pas toujours au volume passé dans la chambre de flux. En effet, en mode *AutoImage*, qui définit l'acquisition d'un nombre fixe d'images par intervalle de temps, il est possible de calculer le pourcentage de volume réellement imagé (nommé "efficacité") et d'estimer la distance entre l'acquisition de deux images :

- si cette distance est trop faible, le risque de "multi-imaging" augmente (une même particule sera imagée plusieurs fois). Dans ce cas, il est probable que le comptage soit sur-estimé,
- si cette distance est trop importante, le risque de perte de particules augmente. Dans ce cas, il est probable que le comptage soit sous-estimé.



L'utilisation de l'objectif 10X pose le problème des particules fixées de très grande taille et de certaines espèces coloniales (notamment les diatomées formant de longues chaînes ou présentant des soies très longues). En effet, dans un flux descendant, et bien que le flux d'entraînement tende à positionner certaines particules de telle sorte que la plus grande longueur soit orientée dans le sens du flux et permette ainsi de faire passer des particules longues et fines (par exemple, *Proboscia alata*) ou des chaînes (par exemple, *Pseudonitzschia*), la faible profondeur de la cellule de flux associée (100µm) entraîne l'accumulation de certaines espèces à l'entrée de la cellule de flux, retenant probablement avec elles une partie des particules 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. C'est pourquoi des expérimentations sur le sens du débit et sur la viscosité des échantillons ont été réalisées. Il en résulte qu'une pré-filtration systématique des échantillons sur filtre de vide de maille s'impose pour des analyses au 10X. Cependant, une pré-filtration de l'échantillon ne permet pas (ou plus) l'analyse de flores totales. Cette solution est donc dépendante des objectifs fixés par l'utilisateur (flores partielles, flores indicatrices, etc.) et doit être appliquée au cas par cas.

De plus, 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 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. Les comparaisons quantitatives des abondances et concentrations obtenues par lectures au microscope et par le système couplé FlowCAM/ZooPhytoImage ont donc été réalisées pour des échantillons naturels lugolés et vivants, parmi les échantillons REPHY 2013. Ces échantillons ont été choisis car ils présentent des abondances intéressantes et une diversité conséquente en cellules non coloniales (ou facilement dénombrables dans les colonies). Un volume de 10 mL a été numérisé pour chacun de ces échantillons, 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 100 cellules/L, ce qui est équivalent à un comptage au microscope inversé. Cependant, pour ce dernier, un facteur multiplicateur peut également être affecté à certains taxons si l'aire qui a été observée est inférieure à la surface totale de la cuve (par exemple, si le comptage a été effectué sur un diamètre, etc.). C'est pourquoi, il est possible d'obtenir des biais pour un comptage manuel.

Pour nos expérimentations, les moyennes des taux de reconnaissance des espèces sont d'environ 69% pour les échantillons lugolés et 67.50% pour les échantillons vivants. Ces indices de performance représentent donc des scores de reconnaissance encourageants. Cependant, de tels scores sont trop faibles pour donner des classifications fiables en routine. Il faut donc valider les données, grâce aux options nouvellement ajoutées au logiciel (étape de vérification manuelle des classements effectués par l'ordinateur). Cette opération est accélérée par le fait que l'ordinateur a effectué un classement correct à 65-75%, par rapport à une classification purement manuelle des vignettes. Les études menées par l'Université de Mons (Grosjean et Denis, 2010; Dereume-Hancart, 2013) ouvrent la voie vers des techniques plus pointues de reconnaissance automatisée permettant une estimation moins biaisée de l'abondance sans avoir à valider manuellement la totalité des vignettes. En effet, deux choix étaient possibles au départ : une validation manuelle totale de l'échantillon (comme pour le zooplancton traité avec le ZooSCAN) ou une validation partielle. Pour le phytoplancton, le choix s'est porté sur la deuxième solution. Le module de correction détermine une probabilité d'être suspect pour chaque particule, et seules les particules les plus suspectes sont à valider. Il est également envisageable d'ajouter à cette probabilité des informations extérieures basées sur des critères bio-écologiques. De plus, les erreurs induites par les classificateurs n'étant pas aléatoires mais systématiques, il est possible de modéliser, par la suite, cette erreur afin d'effectuer des corrections statistiques. L'outil envisagé permet donc

de ne valider manuellement que les particules les plus suspectes. Cette approche particulière offre un compromis idéal entre la méthode automatique et la validation manuelle totale, tout en garantissant des performances de reconnaissance similaires ou meilleures dans un laps de temps acceptable.

Un problème important concerne le stockage et la bancarisation des données issues du système FlowCAM/ZooPhytoImage. Il est envisageable, à court terme, de stocker les données sur un disque local ou un espace réseau, puis, à plus long terme, de prévoir une stratégie de bancarisation de ces données sous Quadrigé<sup>2</sup>. En effet, les discussions et réflexions menées tendent à montrer que les données FlowCAM/ZooPhytoImage sont des données sécurisées ayant leur place dans Quadrigé<sup>2</sup>.

Toutefois, la gestion des données numériques associées à une image ainsi que le problème de la flexibilité du système avec plusieurs protocoles pour un jeu de données, peuvent rendre *a priori* complexe la bancarisation des données. Pour cela, il est impératif, dans un premier temps, de pouvoir estimer la volumétrie. L'intégration des données sous Quadrigé<sup>2</sup> passe donc par la rédaction d'un cahier des charges afin de définir les exigences du système (description du format et de la quantité des données, des besoins d'accès ainsi que des objectifs de ces données).

## Perspectives

L'analyse des performances de reconnaissance obtenues grâce à l'utilisation du module de correction statistique de l'erreur pourrait permettre d'optimiser les méthodes de détection des vignettes les plus suspectes. En effet, l'ajout de critères bio-écologiques supplémentaires tels que la probabilité de présence d'une espèce selon la saison, la zone géographique, etc. pourrait être étudié afin d'améliorer la discrimination des espèces ou groupes phytoplanctoniques.

De plus, il serait pertinent de pouvoir améliorer la qualité visuelle des vignettes pour la constitution des banques d'images et des sets d'apprentissage, mais également pour la validation des vignettes suspectes dans le cadre de la correction statistique de l'erreur. La méthode envisagée est basée sur le principe des images couleur avec un format de codage RGB, où R est l'image en niveaux de gris, G le masque binaire et B l'image de densité optique en niveaux de gris (avec une évolution linéaire entre le noir et le blanc). La combinaison de ces images devrait permettre d'obtenir un meilleur rendu visuel des particules (augmentation des contrastes des zones claires et des zones sombres) et de ne pas tenir compte d'une mauvaise calibration des niveaux de gris lors de la numérisation des échantillons au FlowCAM.

Bien que les premiers résultats de comparaison entre les abondances obtenues par lecture au microscope et par le système FlowCAM/ZooPhytoImage soient encourageants pour la mise en oeuvre de l'outil opérationnel, des expérimentations complémentaires doivent être menées. En particulier, il serait intéressant de pouvoir suivre la dynamique d'une espèce particulière au cours d'un cycle annuel ou saisonnier. Pour cela, des outils spécifiques permettant de dénombrer automatiquement (ou semi-automatiquement) le nombre de cellules dans les colonies sont actuellement étudiés pour être intégrés au logiciel. En effet, le FlowCAM, comme le cytomètre de flux, ne dispose pas encore de systèmes qui permettent de mener à bien cette tâche (Olson et Sosik, 2007). Les espèces coloniales étant très fréquentes dans les échantillons, cela représente un travail qui doit être mené afin de pouvoir comparer aisément les résultats obtenus en semi-automatique avec ceux des dénombrements REPHY.

Enfin, les tests de la nouvelle interface graphique utilisateur doivent être poursuivis afin de définir les fonctionnalités requises pour l'optimisation de l'utilisation de l'interface graphique de ZooPhytoImage. Ces fonctionnalités concernent les spécifications d'ergonomie du logiciel, le contenu des différents composants graphiques et la logique d'interaction entre eux :

l'objectif étant de s'orienter vers une interaction avec l'analyste (affichage des vignettes, de graphes, etc.) à chaque étape du traitement des données.

## Conclusion

Le logiciel Zoo/PhytoImage dans sa version 4, et couplé au FlowCAM, devient un outil véritablement opérationnel en 2014. Totalement adapté aux observations du phytoplancton réalisées dans le cadre du réseau d'observation REPHY, il permettra de mieux répondre aux sollicitations présentes et futures concernant l'évaluation de la qualité des eaux littorales et marines dans le cadre des exigences européennes, telles que la DCE et la DCSMM. Un des bénéfices immédiats sera par exemple pour l'acquisition des données nécessaires au calcul de l'indice abondance composant l'indicateur phytoplancton en Manche- Atlantique, indice qui est basé sur la proportion de taxons du micro-phytoplancton présents en quantité importante dans un échantillon.

En effet, un gain important en temps est attendu avec cet outil en comparaison des observations classiques au microscope, ce qui permettra potentiellement d'augmenter le nombre d'échantillons, et donc d'avoir une meilleure représentativité des données acquises sur un site. Le gain en efficacité et en pertinence est également crucial, grâce à la sauvegarde des observations permettant de revenir sur celles-ci *a posteriori*. Il permettra en outre une plus grande homogénéité entre les données acquises sur différents sites, en diminuant l'effet observateur.

Enfin, de nouveaux paramètres pourront être calculés afin d'aller au-delà de l'information couramment accessible via le proxy de biomasse phytoplanctonique qu'est la concentration en chlorophylle. Il s'agira, à partir des nombreuses mesures fournies par le logiciel, d'estimer systématiquement, par exemple, le biovolume des cellules du phytoplancton permettant alors d'estimer des biomasses en terme de contenu en Carbone et donc d'obtenir une information précise quant à la quantité de phytoplancton disponible dans l'écosystème et contribuant à la structure du réseau trophique.

Le présent livrable est composé des rapports et annexes suivants :

Rapport sur la finalisation d'un outil pré-opérationnel FlowCAM/ZooPhytoImage dans le cadre du REPHY, Version 1.3. G. Wacquet, A. Lefebvre, F. Colas & Ph. Grosjean. Février 2014.

Annexe 1 : Mise au point automatique de la cuve du FlowCAM. B. Forest, M. Tardivel, M.-P. Crassous & F. Colas. Juillet 2013

Annexe 2 : Préparation de l'échantillon.

Annexe 3 : Numérisation de l'échantillon au FlowCAM.

Annexe 4 : Spécificités de la pompe seringue.

Annexe 5 : Traitement numérique et importation des données du FlowCAM dans ZooPhytoImage.

Annexe 6 : Modalités d'intégration de l'outil de correction statistique de l'erreur en utilisation en (pré)-routine de PhytoImage. Cahier des charges. G. Wacquet & L. Fiant. Janvier 2013.

Annexe 7 : Interface Graphique Utilisateur et fonctionnalités requises pour l'utilisation en (pré)routine de ZooPhytoImage. Cahier des charges. G. Wacquet & A. Lefebvre. Juillet 2013.

Ce livrable est complémentaire des trois autres livrables fournis pour cette action :

Livrable 1 Guide méthodologique concernant la version actualisée de PhytoImage avec refonte de l'interface graphique.

Livrable 2 Protocole sur les outils de reconnaissance optimisés Manche Atlantique.

Livrable 4 Rapport sur l'utilisation conjointe de FlowCAM / PhytoImage et de la cytométrie en flux. Premiers résultats et perspectives.

# Rapport sur la finalisation d'un outil pré-opérationnel FlowCAM/ZooPhytoImage dans le cadre du REPHY

Version 1.3

*Février 2014*





## SOMMAIRE

<b>1. Stratégie de numérisation</b> .....	<b>5</b>
1.1. <i>Présentation du REPHY</i> .....	5
1.2. <i>Niveaux de stratégie - Types d'observation</i> .....	6
1.3. <i>Numérisation au FlowCAM</i> .....	7
<b>2. Etude et optimisation optique du FlowCAM</b> .....	<b>8</b>
2.1. <i>Introduction</i> .....	8
2.2. <i>Vitesse de numérisation</i> .....	9
2.3. <i>Illumination de l'échantillon</i> .....	11
2.4. <i>Mise au point de l'échantillon</i> .....	12
<b>3. Etude et optimisation du protocole de numérisation du FlowCAM</b> ....	<b>13</b>
3.1. <i>Choix de l'assemblage Objectif/Cellule de flux</i> .....	13
3.2. <i>Homogénéisation et agitation de l'échantillon</i> .....	13
3.3. <i>Vitesse de flux et efficacité</i> .....	14
3.4. <i>Sédimentation des particules</i> .....	19
3.5. <i>Dilution/Concentration des échantillons</i> .....	24
3.6. <i>Comparaison des abondances (Lugolé vs Vivant)</i> .....	26
3.7. <i>Comparaison des abondances (Microscope vs FlowCAM)</i> .....	29
3.8. <i>Protocole dans le cadre du REPHY</i> .....	32
<b>4. Etude du traitement des échantillons par ZooPhytoImage</b> .....	<b>35</b>
4.1. <i>Comparaison des abondances (Microscope vs FlowCAM / ZooPhytoImage)</i> .....	35
4.2. <i>Correction statistique de l'erreur</i> .....	40
<b>5. Exploitation des résultats</b> .....	<b>45</b>
5.1. <i>Abondances et biomasse en milieu naturel</i> .....	45
5.2. <i>Stockage et bancarisation des résultats</i> .....	46
<b>Annexe 1 - Mise au point automatique de la cuve</b> .....	<b>56</b>
<b>Annexe 2 - Préparation de l'échantillon</b> .....	<b>57</b>
<b>Annexe 3 - Numérisation de l'échantillon au FlowCAM</b> .....	<b>59</b>
<b>Annexe 4 - Spécificités de la pompe seringue</b> .....	<b>78</b>
<b>Annexe 5 - Traitement numérique et importation des données du FlowCAM dans ZooPhytoImage.</b> .....	<b>81</b>
<b>Annexe 6 - Modalités d'intégration de l'outil de correction statistique de l'erreur en utilisation en (pré)routine de PhytolImage (cahier des charges - février 2013)</b> .....	<b>87</b>
<b>Annexe 7 - Interface Graphique Utilisateur et fonctionnalités requises pour l'utilisation en (pré)routine de ZooPhytoImage (cahier des charges - juillet 2013).</b> .....	<b>88</b>

## Introduction

Aujourd'hui, plus de 3000 échantillons d'eau sont examinés annuellement pour des lectures de phytoplancton dans le cadre du REPHY et des réseaux régionaux pilotés par l'IFREMER. Les données obtenues (identification et dénombrement par taxon, genre ou espèce) sont utilisées en particulier pour répondre aux exigences de la DCE, puisqu'elle permettent de calculer les indices de l'élément de qualité phytoplancton, qui fait partie des éléments biologiques obligatoires pour l'évaluation de la qualité écologique des masses d'eau.

La méthode la plus utilisée à ce jour est celle de la microscopie. Les identifications des taxons phytoplanctoniques sont effectuées sur 10 mL d'eau de mer fixée au Lugol, au microscope inversé selon la méthode d'Utermöhl (décantation de l'échantillon dans une cuve). L'identification se fait au niveau le plus bas possible (espèce ou genre) sinon au rang supérieur (famille voire ordre ou classe). Cette méthode nécessite une formation approfondie en taxonomie mais le niveau de compétence des opérateurs étant différent, certains taxons peuvent leur être inconnus ou faire l'objet d'identifications erronées. De plus, 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.

De ce fait, des nouveaux systèmes sont mis en place pour l'identification automatique des espèces de phytoplancton contenues dans les échantillons d'eau de mer. Ces nouvelles technologies (FlowCAM, cytomètre en flux, ...) permettent une analyse automatique et rapide des échantillons d'eau de mer (quelques dizaines de minutes pour le FlowCAM contre quelques heures pour l'identification par un taxonomiste) et permettent également de quantifier les erreurs commises. Dans ce contexte, le projet FlowCAM/ZooPhytoImage a été monté il y a quelques années pour améliorer et optimiser ces observations. L'outil est constitué du dispositif FlowCAM destiné à numériser les images de particules phytoplanctoniques et du logiciel ZooPhytoImage qui permet d'identifier automatiquement et de dénombrer le phytoplancton à partir de ces images.

Le présent document traite de l'optimisation du protocole de numérisation des échantillons dans le cadre du REPHY. Il a pour but de présenter les expérimentations menées au cours de l'élaboration du protocole, les résultats associés et leur interprétation. Ce document s'organise de la manière suivante : la première section présente la stratégie générale de numérisation des échantillons dans le cadre du REPHY; l'étude et l'optimisation technologique du FlowCAM sont abordées dans la seconde section; dans la troisième partie seront présentés les expérimentations et résultats ayant conduit à l'élaboration du protocole de numérisation; enfin, la quatrième et dernière section est dédiée à l'exploitation des résultats obtenus par l'analyse conjointe FlowCAM/ZooPhytoImage.



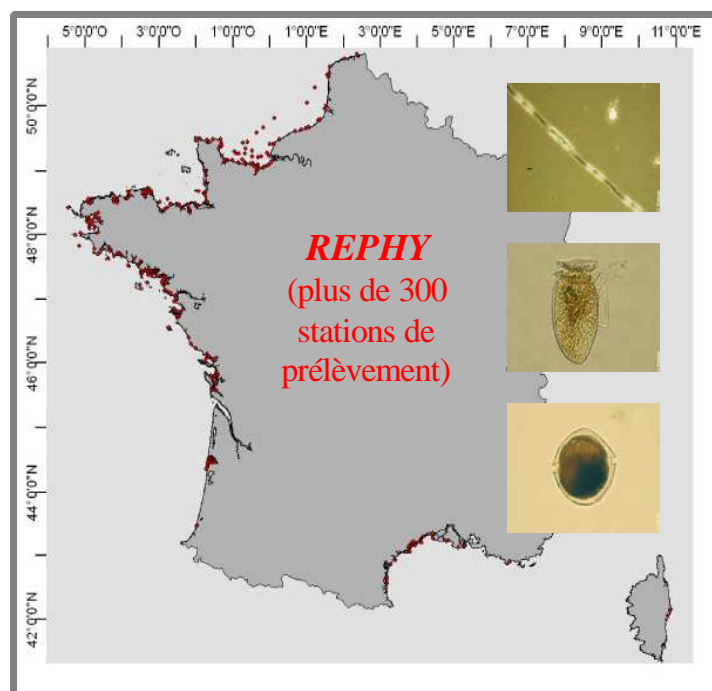
# 1. Stratégie de numérisation

## 1.1. Présentation du REPHY

Pour répondre à ses objectifs, le REPHY doit assurer une surveillance du compartiment "eau" (figure 1). Des mesures *in situ* de paramètres physico-chimiques sont réalisées et des prélèvements d'échantillons sont effectués pour l'observation du phytoplancton et pour les analyses de la chlorophylle-a et des nutriments (Belin, 2012).

Les prélèvements d'eau sont effectués de façon régulière ou événementielle : la périodicité et la fréquence d'échantillonnage sur chacun des points, sont définies dans le contexte d'une mutualisation des différentes stratégies de surveillance, qu'elles soient à visée environnementale ou sanitaire. Ainsi, les paramètres pouvant être utilisés à la fois pour les aspects environnementaux et sanitaires sont également décrits dans ce contexte de mutualisation : les mêmes observations de phytoplancton peuvent ainsi servir les deux aspects sans que l'un d'entre eux soit négligé. Les points peuvent être échantillonnés :

- de façon régulière : toute l'année, ou bien sur une période déterminée de l'année et définie à l'avance ;
- de façon événementielle : en cas de présence d'espèces toxiques ou de toxines sur les points de la zone concernée, ou proches de cette zone, ou bien en cas d'événement particulier (eau colorée, TIAC17, etc) ; l'échantillonnage événementiel est par définition non prévisible : il n'est utilisé que pour renforcer l'effort d'échantillonnage ;
- à une période ou une fréquence différente sur un même point ;
- à une période ou une fréquence différente également selon le paramètre mesuré (par exemple les nutriments mesurés en hiver, alors que le phytoplancton est dénombré toute l'année).



*Figure 1 : Stations de prélèvements REPHY en France métropolitaine.*

## **1.2. Niveaux de stratégie - Types d'observation**

### ***FLORE TOTALE (FLORTOT)***

C'est l'identification et le dénombrement de toutes les espèces phytoplanctoniques pouvant être identifiées dans les conditions d'observation, c'est à dire globalement toutes les espèces dont la taille est supérieure à 20 µm, et celles dont la taille est inférieure mais qui sont en chaîne. Les espèces plus petites sont dénombrées seulement quand elles concernent des espèces potentiellement toxiques.

La liste de référence des espèces potentiellement identifiables se trouve dans le référentiel taxinomique de Quadrige<sup>2</sup>. 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). Dans de nombreux cas, des groupes de genres ou d'espèces ont été créés dans le référentiel, permettant aux observateurs de cibler au mieux l'identification.

Les Flores Totales sont réalisées sur un nombre limité de lieux, suffisant pour fournir à long terme une connaissance satisfaisante de la biodiversité du phytoplancton sur le littoral français, ainsi que de la distribution spatiale et des variations temporelles des divers taxons phytoplanctoniques le long des côtes françaises. Elles servent également à calculer les indicateurs prévus pour l'estimation de la qualité de l'élément phytoplancton dans le cadre de la DCE.

### ***FLORE PARTIELLE (FLORPAR)***

Ce sont des flores simplifiées pour lesquelles aucune contrainte n'est imposée : elles peuvent même être réduites au seul genre toxique qui a déclenché l'alerte sur les lieux voisins.

### ***FLORE INDICATRICE (FLORIND)***

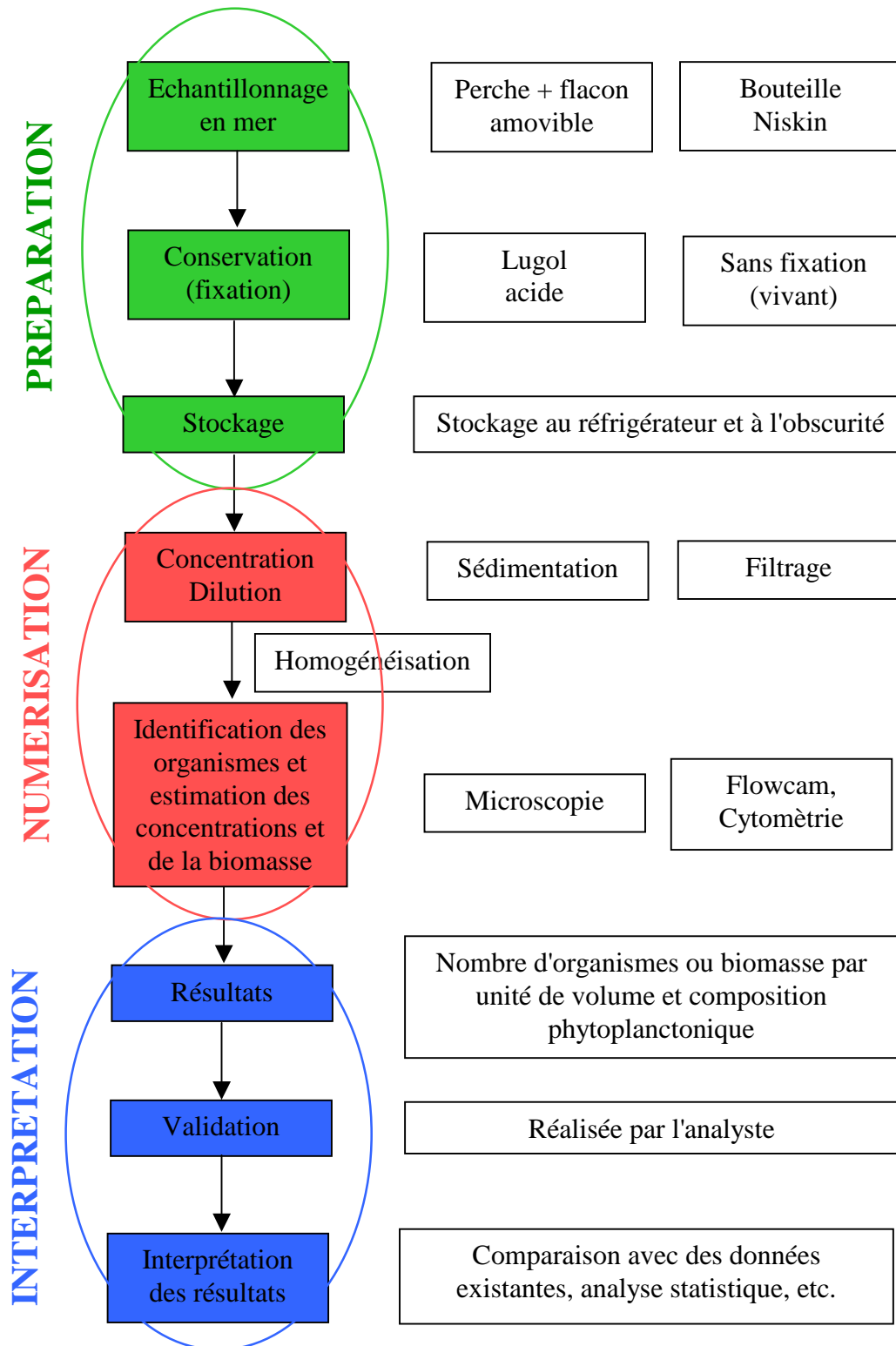
C'est l'identification et le dénombrement :

- de toutes les espèces présentes à une concentration supérieure à 100 000 cellules par litres, toxiques ou non;
- des genres auxquels sont rattachées des espèces avérées toxiques pour le consommateur et présentes sur nos côtes, c'est à dire : *Alexandrium*, *Dinophysis*, *Pseudo-nitzschia* et *Ostreopsis*, quelle que soit leur concentration;
- des espèces suivantes connues pour produire des toxines lipophiles : *Gonyaulax spinifera*, *Lingulodinium polyedra*, *Protoceratium reticulatum*, *Prorocentrum lima*.

Les taxons à dénombrer sont donc moins nombreux que pour les Flores Totales, ce qui conduit à un gain de temps en dénombrement et en saisie des résultats, tout en gardant une information pertinente. En effet, ce type de paramètre permet à la fois :

- de calculer les indicateurs prévus pour l'estimation de la qualité de l'élément phytoplancton dans le cadre de la DCE (Directive Cadre sur l'Eau),
- de répondre à des objectifs sanitaires, puisque les espèces toxiques sont systématiquement dénombrées.

### 1.3. Numérisation au FlowCAM



## 2. Etude et optimisation optique du FlowCAM

### 2.1. Introduction

Le FlowCAM (développé par la société Fluid Imaging Technologies®) est un système d'imagerie en flux de particules (figure 2) dont la taille varie de quelques micromètres à quelques millimètres. Les clichés automatiquement enregistrés permettent une classification des particules suivant des critères morphologiques. Cette opération peut être réalisée directement par le logiciel fourni avec le système (Visual SpreadSheet) ou par d'autres, tel que ZooPhytoImage, logiciel développé par l'Université de Mons (pour l'utilisation des différents modules de ZooPhytoImage, se reporter au livrable n°1 : "Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique"). Dans cette étude, nous utilisons le mode *AutoImage*, qui définit l'acquisition d'un nombre fixe d'images par intervalle de temps. Ce mode 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 (Tunin-Ley et Maurer, 2011).

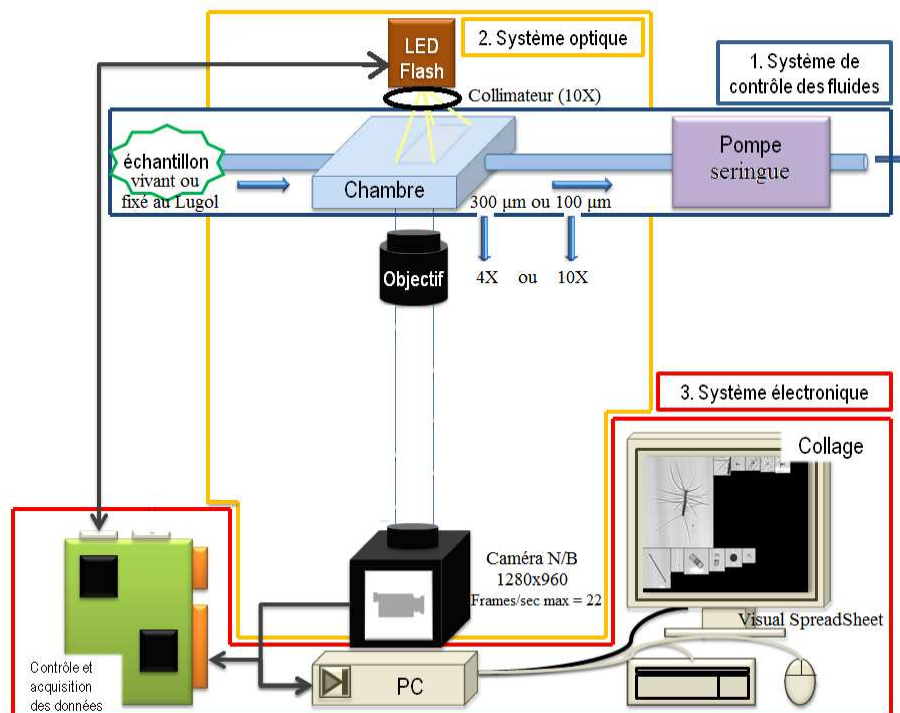


Figure 2 : Schéma fonctionnel du système d'imagerie numérique FlowCAM utilisé au LER de Boulogne-sur-Mer.

L'IFREMER mène des actions pour intégrer cet équipement au REPHY. Suite au travail de post-doctorat de Tunin-Ley (LER Arcachon puis Université de Bordeaux), différents axes d'évolution ont été proposés pour adapter le FlowCAM couplé à ZooPhytoImage aux exigences du REPHY qui sont la justesse et la répétabilité de la mesure (Tunin-Ley et Maurer, 2011; Tunin-Ley *et al.*, 2011). A cela s'ajoute la rapidité ; l'objectif de ces nouveaux outils étant d'offrir un gain de temps tangibles aux observateurs REPHY par rapport aux observations au microscope optique.

Le temps d'analyse d'un échantillon est un point clef de ce système. En effet, pour l'analyse du phytoplancton deux grossissements sont possibles : 10X et 4X. Le premier permet une meilleure description morphologique mais avec un temps d'acquisition, bien plus long que le second. Un gain considérable serait de pouvoir réaliser une acquisition au grossissement 10X avec un temps d'analyse plus court, comparable au 4X. Des pistes d'évolution du système seront étudiées dans ce sens.

Parallèlement à cela, la répétabilité et la justesse de la mesure sont conditionnées par les réglages optiques du système. L'illumination produite par la source lumineuse souffre d'un manque d'homogénéité sur l'ensemble du champ et est très dépendante de son orientation. De plus, la mise au point est réalisée manuellement à l'aide d'une vis micrométrique. Cette opération dure 15-20 min et le réglage final n'est pas répétable.

Il est donc proposé d'étudier :

1. la vitesse de numérisation,
2. l'illumination de l'échantillon,
3. la mise au point du système.

## 2.2. Vitesse de numérisation

Le volume à numériser doit être choisi pour que la mesure soit représentative du milieu. Historiquement, les analyses du REPHY reposent sur l'observation de 10 mL au microscope optique. La transposition de ce protocole au FlowCAM peut se faire en considérant que le volume à numériser doit être de 10 mL. Cependant, l'analyse est alors très longue. Par ailleurs, l'approche FlowCAM/ZooPhytoImage impose une contrainte supplémentaire : l'analyse d'un échantillon est considérée statistiquement représentative si au moins 2000 particules sont comptées et analysées.

	4X	10X
$l_{ch}$ (mm)	1.7	0.7
$h_{ch}$ (mm)	1.3	0.5
$p_c$ (mm)	0.3	0.1
$V_i$ ( $\mu$ L)	0.663	0.035
$\Phi_{num}$ (mL/min)	0.88	0.046
$t_{10mL}$ (min)	11.36	217.39

*Tableau 1 : Temps de numérisation pour une fréquence d'acquisition de 22 i/s.*

Le volume imagé ( $V_i$ ) par trame est donné par la relation 1 :

$$V_i = l_{ch} h_{ch} p_c \quad (1)$$

avec  $l_{ch}$  et  $h_{ch}$  les largeur et hauteur du champ objet respectivement et  $p_c$  la profondeur de la cuve. Le débit numérisé, défini comme le volume de solution numérisée par unité de temps est directement donné par l'équation 2 :

$$\Phi_{num} = N V_i \quad (2)$$

avec  $N$  le nombre de trame acquise par la caméra par unité de temps.

De cette équation, le temps de numérisation minimal de 10 mL peut être calculé pour les deux grossissements à partir des données résumées dans le tableau 1. Il faut environ 217 min avec l'objectif 10X et seulement 11 min avec le 4X à la fréquence d'acquisition maximale du FlowCAM de 22 images/s.

Le nombre de particules alors numérisées dans l'échantillon dépend de la concentration. Le tableau 2 montre le temps nécessaire pour numériser 2000 particules en fonction de la concentration et le temps de numérisation au FlowCAM correspondant pour des objectifs 4X et 10X.

Concentration cellulaire (cellules/L)	Volume à numériser pour compter 2000 particules (mL)	Temps de numérisation au 4X (min)	Temps de numérisation au 10X (min)
$10^4$	200	227	4347
$10^5$	20	22	435
$10^6$	2	2	43

Tableau 2 : Temps de numérisation de 2000 particules en fonction des concentrations cellulaires pour les grossissements 4X et 10X.

Il apparaît des données du tableau 2 que seul dans le cas d'une grande concentration cellulaire, cette approche est difficilement applicable.

#### **Voies d'amélioration :**

Pour pouvoir réaliser des analyses rapides au 10X, il est donc impératif de modifier le système du FlowCAM et éventuellement le protocole d'analyse.

Comme abordé lors de la réunion du comité de pilotage du projet FlowCAM/ZooPhytoImage (à Boulogne-sur-Mer le 19 décembre 2012), actuellement, pour un échantillon non concentré et 10 ml de volume analysé, le temps de numérisation est de 12 minutes à l'objectif 4X, 19 fois plus soit environ 2h30 à l'objectif 10X. En effet, le facteur limitant dans le FlowCAM est la caméra qui ne peut traiter que 22 images par seconde : ce n'est pas suffisant.

Pour accélérer la numérisation des échantillons un moyen efficace est de travailler avec une caméra plus rapide. A titre d'exemple, avec une caméra capable d'acquérir des images à une fréquence de 350 image/s, le temps d'analyse sera réduit par un facteur 16. Cela ne sera pas toutefois pas suffisant pour des concentrations de l'ordre de  $10^4$  cellules/L, le temps d'analyse par le FlowCAM restant de 270 min. De plus, un changement de caméra sur le FlowCAM reviendrait à abandonner la piste achat des FlowCAM chez Fluid Imaging Technologies et construire un matériel entièrement adapté à nos besoins : il y aurait en effet d'autres améliorations à prévoir. Ceci aurait des avantages : personnalisation du matériel, non dépendance vis à vis du fournisseur, SAV moins compliqué (actuellement, toute panne sur un FlowCAM nécessite son renvoi aux Etats Unis). Mais le logiciel implanté sur le FlowCAM pour la numérisation (Visual Spreadsheet) serait entièrement à réécrire, et ceci représente un travail conséquent de programmation. Le plus raisonnable, dans un premier temps, est donc de garder la piste FlowCAM, d'autant qu'il pourrait y avoir une SAV en Europe.

Pour réduire les analyses à 20 min, il faut augmenter la concentration dans l'échantillon. Pour cela, il est nécessaire de pré-concentrer les cellules lorsque la concentration cellulaire est inférieure à  $10^5$  cellule/L (cf. section 3.5, "Dilution/Concentration des échantillons").

### 2.3. Illumination de l'échantillon

Le FlowCAM est un système de microscopie en flux. Le système optique complet comprend :

1. une LED pulsée pour source de lumière,
2. un illuminateur,
3. un objectif,
4. une caméra numérique.

L'ensemble est piloté par un PC industriel intégré.

Pour un grandissement 10X, l'illuminateur du FlowCAM est un collimateur, constitué d'une lentille plan-convexe de distance focale mesurée d'environ 40 mm. Pour le grandissement 4X, il n'y a pas d'optique autre que le moulage plastique de la LED. Les systèmes d'illumination sont donc basiques. Il en découle parfois une mauvaise homogénéité de l'éclairage dans le plan objet.

#### Voies d'amélioration :

Deux configurations ont été testées, essentiellement pour le grandissement 10X. La première consiste à ajouter un dépoli devant la LED et déplacer le collimateur. La seconde consiste à injecter la lumière issue de la LED dans une fibre optique (FO) d'1 mm de diamètre de cœur et d'utiliser l'embout de la FO comme source.

La première configuration induit une perte de signal importante. Nous devons alors travailler avec des impulsions de durées supérieures à 30  $\mu$ s pour garantir une image dont l'histogramme des niveaux de gris est centré sur 180, comme préconisé dans le rapport RST/LER/AR/11/002.

La seconde configuration induit également une perte de signal. Nous devons alors travailler avec des impulsions d'une durée supérieure à 20  $\mu$ s. Ce second montage présente également le désavantage d'être encombrant et de ne pas de se monter dans le FlowCAM tel qu'il est actuellement conditionné. Toutefois, comme nous le verrons, ces deux durées d'impulsion sont suffisantes pour le débit avec lesquels il est nécessaire de travailler.

Le débit de la pompe ne doit pas être trop faible pour ne pas sur-échantillonner, ni trop important pour ne pas induire de flou de bougé. Le flou de bougé est déterminé par la vitesse linéique ( $v$ ) des objets dans la cellule et la durée de l'impulsion lumineuse ( $\Delta t$ ). A 22 image/s, le débit de numérisation est de 70  $\mu$ L/min. Si le débit de la pompe est réglé sur le double, c'est-à-dire 140  $\mu$ L/min, alors, il n'y aura pas de sur-échantillonnage.

La vitesse linéique dans la cuve est donnée par l'équation 3 :

$$v = \frac{\Phi_p}{\pi \cdot d \cdot h} \quad (3)$$

avec  $\Phi_p$  le débit de la pompe, soit  $v=23$  mm/s.

Pour obtenir un flou de bougé inférieur à 1  $\mu\text{m}$ , correspondant à la limite de résolution de cet objectif si le système d'illumination était parfait, il nous faut travailler avec une impulsion lumineuse d'environ 40  $\mu\text{s}$ . Les deux configurations peuvent donc convenir. En raison de l'encombrement de la seconde, la première configuration est préférable et envisageable au regard des LED disponibles sur le marché et à moindre coût.

## 2.4. Mise au point de l'échantillon

La mise au point de l'échantillon est réalisée à l'aide d'une vis micrométrique manuelle réglant le positionnement de la cuve par rapport à l'objectif. La cuve est positionnée de manière à ce qu'une face de la cuve soit nette. La position est repérée. La cuve est translaturée de manière à observer nettement l'autre face de la cuve. La cuve est ensuite placée à mi-chemin de ces deux positions.

Ce procédé montre ces limites. En effet, le microscope a une profondeur de champ de plusieurs dizaines de microns au regard du système d'illumination. Il n'est donc pas possible de positionner la cuve de manière répétable à plusieurs dizaines de micron près. Ainsi, pour s'assurer que la mise au point est bonne, il est alors nécessaire de faire passer un échantillon puis d'ajuster la mise au point en prenant pour critère la netteté des vignettes.

L'ensemble du réglage est donc un processus long dont la répétabilité peut être remise en cause.

### Voies d'amélioration :

L'utilisation d'un actionneur automatique peut permettre d'améliorer la répétabilité du positionnement (figure 3). Actuellement, ce système, qui peut s'apparenter à un système d'autofocus a été installé sur le FlowCAM, ce qui permet un positionnement plus aisé de l'échantillon (Annexe 1).

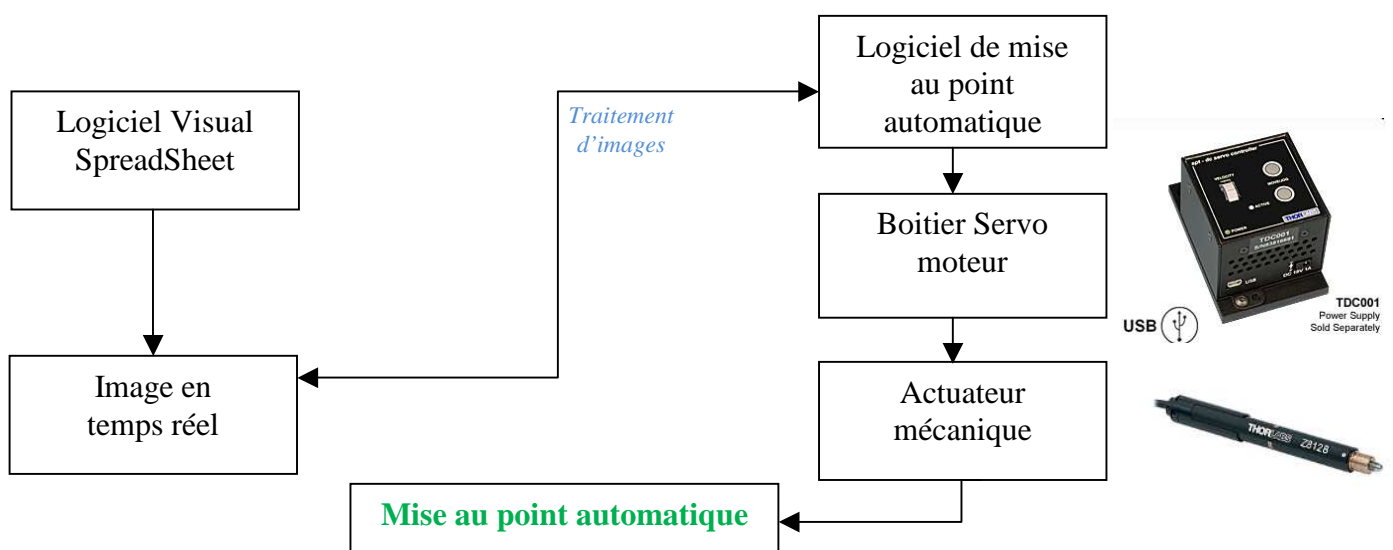


Figure 3 : Schéma de fonctionnement général du principe de mise au point automatique.



## 3. Etude et optimisation du protocole de numérisation du FlowCAM

### 3.1. Choix de l'assemblage Objectif/Cellule de flux

L'utilisation d'un objectif de grossissement faible (4X, ce qui correspond à un grossissement global de 40X) associé à une cellule de grande profondeur (300 ou 600  $\mu\text{m}$ ) s'avère utile dans le cas où certaines espèces de grande taille sont présentes en abondance. Cependant, dans ce cas, la résolution des images n'est pas toujours suffisante pour identifier manuellement certaines espèces présentes dans l'échantillon.

Pour pallier ce problème, l'utilisation de l'assemblage optique 10X/100 $\mu\text{m}$  (grossissement global de 100X) peut être utilisé. 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 (tableau 3). Les plus petits organismes sont également mieux détectés (Poulton, 2011).

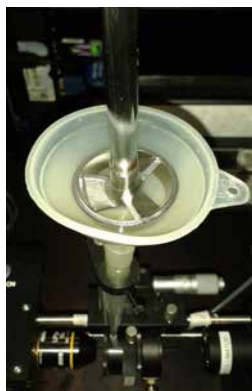
Gamme de taille des particules	Objectif	Cellule de flux	Dimensions de la cellule de flux (Profondeur x Largeur)
5 $\mu\text{m}$ - 1000 $\mu\text{m}$	10 X	FC100	100 $\mu\text{m}$ x 2 mm
15 $\mu\text{m}$ - 2000 $\mu\text{m}$	4 X	FC300	300 $\mu\text{m}$ x 3 mm

*Tableau 3 : Assemblages théoriques recommandés par Fluid Imaging Technologie selon les différentes classes de taille.*

Généralement, l'utilisation d'un assemblage particulier dépend, en grande partie, des objectifs de l'étude et du type d'observation défini par le REPHY. Ces spécifications sont abordées dans la section 3.9 (Protocole dans le cadre du REPHY).

### 3.2. Homogénéisation et agitation de l'échantillon

Nous étudions ici l'importance de l'agitation de l'échantillon durant l'analyse. Pour le FlowCAM, il est nécessaire de choisir un système d'agitation adapté. Dans un premier temps, nous choisissons d'utiliser un agitateur à hélice directement dans l'entonnoir (figure 4).



*Figure 4 : Agitateur à hélices pour l'homogénéisation de l'échantillon durant l'analyse.*

Cet agitateur à hélice électronique convient parfaitement aux applications simples de mélange en laboratoire telles que le traitement de milieux à fortes variations spontanées de viscosité et est idéal pour l'agitation de substances de faible à moyenne viscosité. De plus, la régulation électronique de la vitesse permet de la maintenir constante, même en cas de changement de viscosité.

Afin de démontrer l'importance de l'agitation, un échantillon provenant de Concarneau (Lanvecc Large) fixé au Lugol est utilisé pour les assemblages 4X/300 $\mu$ m et 10X/100 $\mu$ m (numérisation de 3 mL). Les résultats sont présentés en terme de concentrations (figures 5 et 6).

Expérimentation : Assemblages : 4X/300 $\mu$ m et 10X/100 $\mu$ m  
 Echantillon : Concarneau, Lanvecc large (lugolé)  
 Volume numérisé : 3 mL

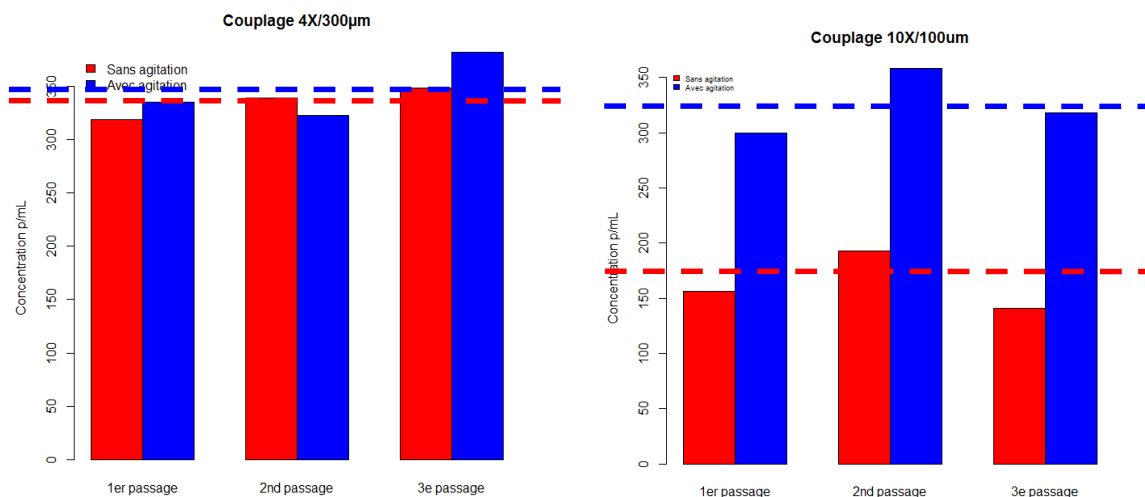


Figure 5 : Concentration pour l'assemblage 4X/300 $\mu$ m (Rouge : sans agitation ; Bleu : avec agitation).

Figure 6 : Concentration pour l'assemblage 10X/100 $\mu$ m (Rouge : sans agitation ; Bleu : avec agitation).

Afin d'obtenir des résultats pertinents et significatifs, trois passages sont réalisés pour chacune des expérimentations. Pour l'assemblage 4X/300 $\mu$ m, nous remarquons que les résultats obtenus sont constants en terme de concentration (particules phytoplanctoniques+débris), avec et sans agitation (+/- 9% d'écart moyen). En effet, il semble que l'agitation n'a pas d'impact sur le nombre de particules numérisées au FlowCAM.

De la même manière, trois passages sont réalisées pour chaque expérimentation avec l'assemblage 10X/100 $\mu$ m. Les résultats obtenus montrent une différence conséquente entre les concentrations relevées (plancton+débris) avec et sans agitation (+/- 50% d'écart moyen). Ici, l'homogénéisation de l'échantillon et l'utilisation d'un agitateur semblent prendre de l'importance pour le 10X. En effet, l'agitation de l'échantillon durant l'analyse permet d'obtenir des concentrations moyennes (en pointillés) similaires à celles obtenues avec l'assemblage 4X/300 $\mu$ m.

### 3.3. Vitesse de flux et efficacité

Pour effectuer des analyses quantitatives des échantillons, il est nécessaire d'optimiser le compromis vitesse de flux/temps de numérisation/efficacité. En effet, dans un système d'imagerie de flux tel que le FlowCAM, le nombre de particules comptées dans un volume donné n'est pas toujours équivalent à la concentration de

l'échantillon. Cela s'explique principalement par le fait que le volume imagé ne correspond pas toujours au volume passé dans la chambre de flux (figure 7).

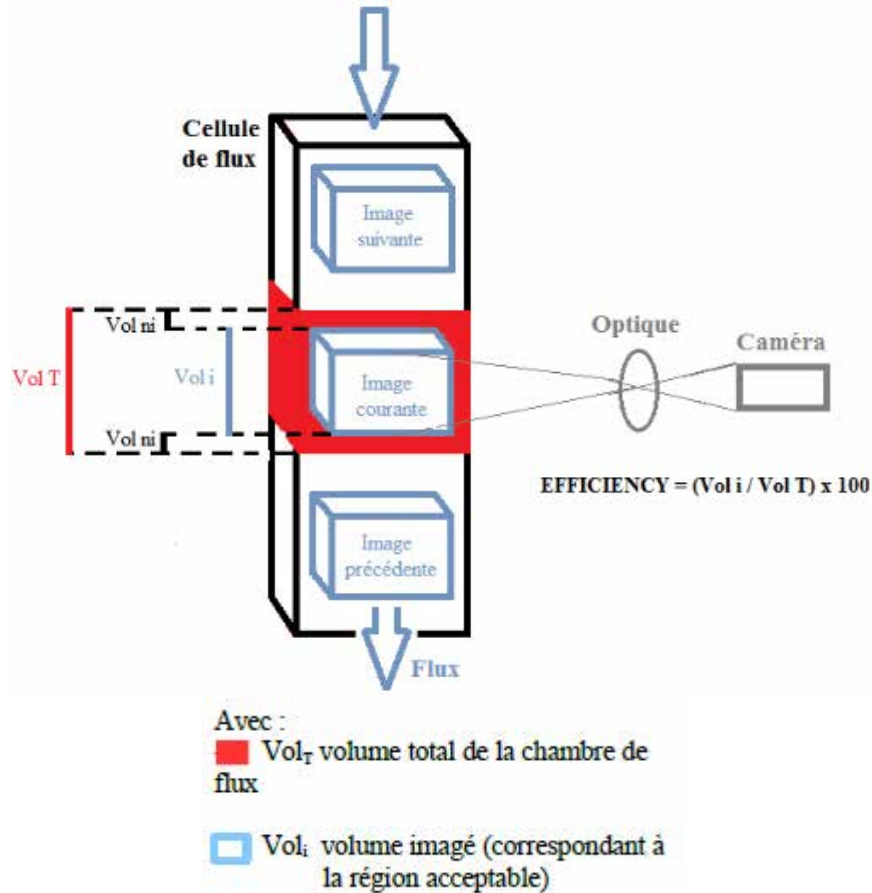


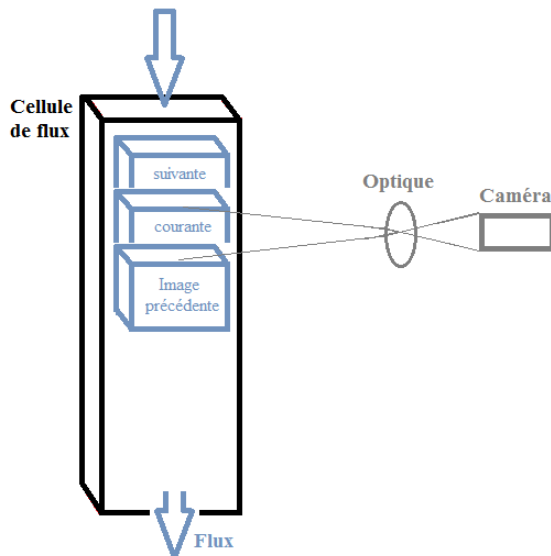
Figure 7 : Volume imagé dans la cellule de flux.

Afin de calculer le pourcentage de volume réellement imagé (nommé "Efficiency"), il est primordial de prendre en considération plusieurs paramètres tels que:

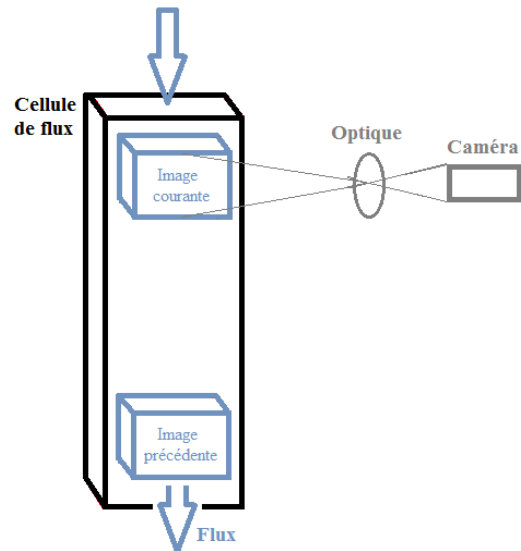
- la région acceptable du champ de la camera (en pixels),
- les dimensions de la chambre de flux (en  $\mu\text{m}$ ),
- le facteur de calibration (taille d'un pixel en  $\mu\text{m}$ ),
- la vitesse de flux (en mL/min),
- le nombre d'images par seconde prises par la caméra (en images/sec).

Ces valeurs permettent alors d'estimer la distance ("de temps") entre deux images :

- Si cette distance est trop faible, le risque de "multi-imaging" augmente (une même particule sera imagée plusieurs fois). Dans ce cas, il est probable que le comptage soit sur-estimé (figure 8).
- Si cette distance est trop importante, le risque de perte de particules augmente. Dans ce cas, il est probable que le comptage soit sous-estimé (figure 9).



*Figure 8 : Distance faible entre deux volumes imagés.*



*Figure 9 : Distance élevée entre deux volumes imagés.*

Par exemple, le FlowCAM du LER de Boulogne-sur-Mer étant équipé d'une caméra noir et blanc haute résolution (1280x960), la région acceptable (ici totale) du champ de la caméra est :

**Left (L) = 0 pixel ; Right (R) = 1279 pixels**  
**Top (T) = 0 pixel ; Bottom (B) = 959 pixels**

#### Assemblage 4X/300µm

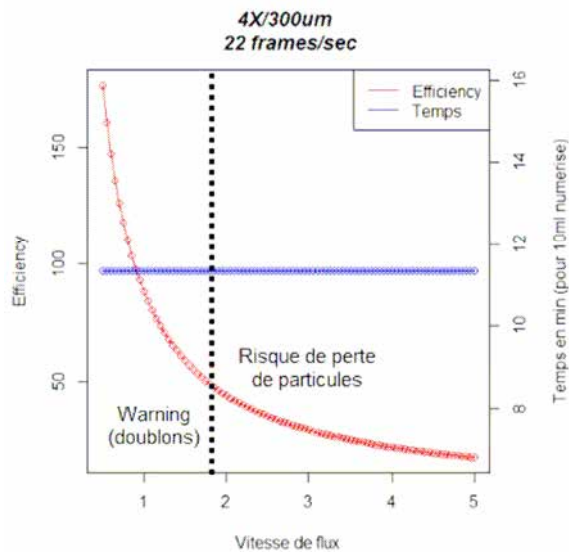
Dans le cas d'un couplage 4X/300µm, le facteur de calibration (défini et fourni par le constructeur) est de **1,3462 µm/pixel** (spécifique au FlowCAM du LER-BL). Grâce à cette valeur, il est possible de déterminer la hauteur et la largeur (en µm) de la région acceptable du champ de vision de la caméra :

**Height (H) = (1+(B-T)) x Cal. Factor = 960 x 1,3462 = 1292,352 µm**  
**Width (W) = (1+(R-L)) x Cal. Factor = 1280 x 1,3462 = 1723,136 µm**  
**Depth (D) = profondeur de la chambre de flux = 300 µm**

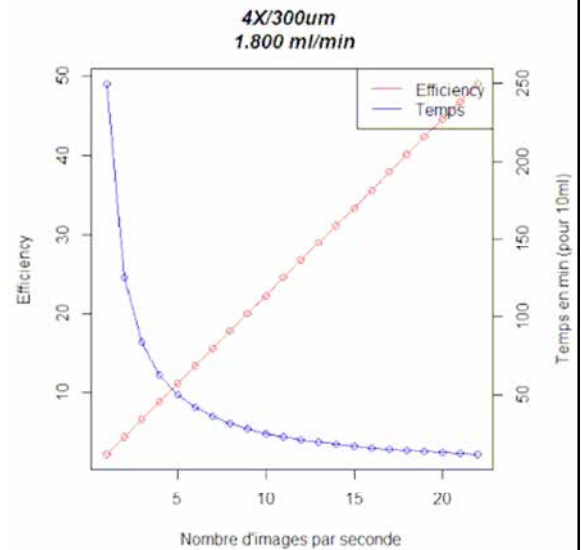
Le volume réellement imagé est alors déterminé par :

**D x H x W = 300 x 1292,352 x 1723,136**  
**= 668069476,7616 µm<sup>3</sup>/image**  
**= 6,680695.10-04 mL/image**

Sachant que l'efficacité maximale conseillée par Fluid Imaging Technologies® est de l'ordre de 50% pour l'assemblage 4X/300µm (déterminée par la combinaison objectif-cellule de flux) et que, dans un souci de gain de temps pour la numérisation, le nombre d'images par seconde acquises par la caméra est maximisé (22 images/sec), il est possible d'étudier l'évolution des valeurs de l'efficacité en fonction de la vitesse de flux. La figure 10 montre que, pour le couplage 4X/300µm, une vitesse de flux de l'ordre de 1,8 mL/min est requise afin de respecter les normes définies par le constructeur. De plus, nous pouvons remarquer que l'analyse de 10 mL peut être effectuée en moins de 12 minutes.



*Figure 10 : Efficacité et temps en fonction de la vitesse de flux pour le couplage 4X/300µm (nombre d'images/sec fixé à 22).*



*Figure 11 : Efficacité et temps en fonction du nombre d'images/sec pour le couplage 4X/300µm (vitesse de flux fixée à 1.8 mL/min).*

En fixant cette vitesse de flux, nous pouvons étudier l'impact du nombre d'images par seconde prises par la caméra sur le temps de numérisation (figure 11). Nous remarquons alors une augmentation de la durée de l'analyse lorsque le nombre d'images par seconde de la caméra diminue.

#### Vérification

Sachant que la vitesse de flux est de 1,800 mL/min (0,03 mL/sec) et que le nombre d'images par seconde est fixé à 22 fr/sec, le volume total passé  $Vol_T = 1,363636 \cdot 10^{-03}$  mL/image et donc le volume non imagé  $Vol_{ni} = 6,955668 \cdot 10^{-04}$  mL/image ( $Vol_T - Vol_i$ ). L'efficacité est donc de :

$$(Vol_i / Vol_T) \times 100 = 48,9\%$$

#### Assemblage 10X/100µm

Dans le cas du couplage 10X/100µm, le facteur de calibration (défini et fourni par le constructeur) est de **0,5515 µm/pixel** (spécifique au FlowCAM du LER-BL). Grâce à cette valeur, il est possible de déterminer la hauteur et la largeur (en µm) de la région acceptable du champ de vision de la caméra :

$$\text{Height (H)} = (1 + (B - T)) \times \text{Cal. Factor} = 960 \times 0,5515 = 529,44 \mu\text{m}$$

$$\text{Width (W)} = (1 + (R - L)) \times \text{Cal. Factor} = 1280 \times 0,5515 = 705,92 \mu\text{m}$$

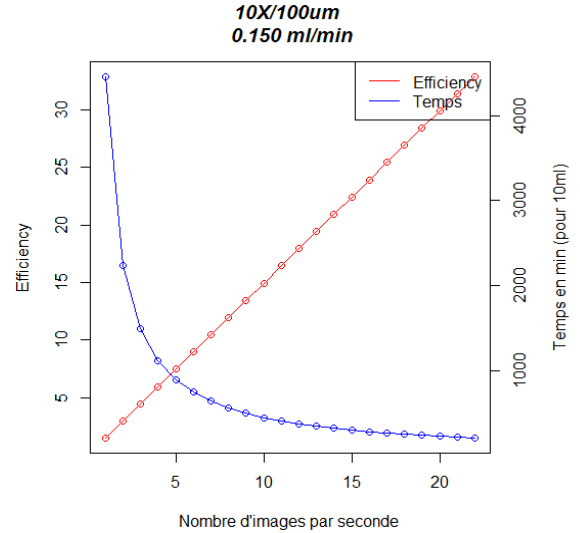
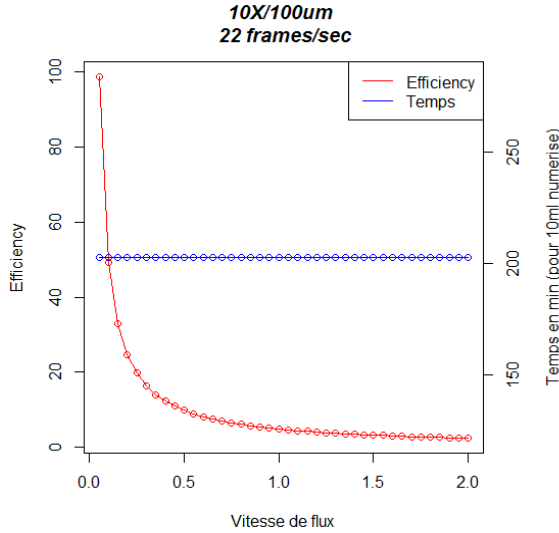
$$\text{Depth (D)} = \text{profondeur de la chambre de flux} = 100 \mu\text{m}$$

Le volume réellement imagé est alors déterminé par :

$$\begin{aligned} D \times H \times W &= 100 \times 529,44 \times 705,92 \\ &= 37374228,48 \mu\text{m}^3/\text{image} \\ &= 3,737423 \cdot 10^{-05} \text{ mL/image} \end{aligned}$$

Sachant que l'efficacité maximale conseillée par Fluid Imaging Technologies® est de l'ordre de 30% pour l'assemblage 10X/100µm et que, dans un souci de gain de temps pour la numérisation, le nombre d'images par seconde acquises par la caméra est maximisé (22 images/sec), il est possible d'étudier l'évolution des valeurs de l'efficacité

en fonction de la vitesse de flux. La figure 12 montre que, pour le couplage 10X/100 $\mu$ m, une vitesse de flux de l'ordre de 0,150 mL/min est requise afin de respecter les normes définies par le constructeur. Cependant, dans ce cas, l'analyse de 10 mL est effectuée en 200 minutes.



*Figure 12 : Efficacité et temps en fonction de la vitesse de flux pour le couplage 10X/100 $\mu$ m (nombre d'images/sec fixé à 22).*

*Figure 13 : Efficacité et temps en fonction du nombre d'images/sec pour le couplage 10X/100 $\mu$ m (vitesse de flux fixée à 0.15 mL/min).*

En fixant cette vitesse de flux, nous pouvons étudier l'impact du nombre d'images par seconde prises par la caméra sur le temps de numérisation (figure 13). Nous remarquons alors une augmentation de la durée de l'analyse lorsque le nombre d'images par seconde de la caméra diminue.

Vérification

Sachant que la vitesse de flux est de 0,150 mL/min (0,0025 mL/sec) et que le nombre d'images par seconde est fixé à 22 fr/sec, le volume total passé  $Vol_T = 1.136363 \cdot 10^{-04}$  mL/image et donc le volume non imagé  $Vol_{ni} = 7,626207 \cdot 10^{-05}$  mL/image ( $Vol_T - Vol_i$ ). L'efficacité est donc de :

$$(Vol_i / Vol_T) \times 100 = 32,8\%$$

Pourquoi ces seuils d'efficacité ?

Les recommandations de seuils d'efficacité sont basées sur le désir de ne pas "multi-imager" une particule ou un volume de fluide. Les quatre paramètres impactant l'efficacité sont la cellule de flux, l'objectif, la vitesse de flux et le nombre d'images par seconde.

La combinaison cellule de flux/objectif détermine le pourcentage de fluide total qui passe dans le champ de vision de l'objectif. La combinaison vitesse de flux/nombre d'images par seconde détermine le pourcentage de fluide passant dans le champ de vision optique et ayant été réellement imagé. Le résultat obtenu est l'efficacité théorique maximale qui peut être atteinte sans possibilité de "multi-imaging".

### Quels risques ?

Une efficacité supérieure au seuil recommandé entraîne un risque de "multi-imaging" des particules (calcul d'abondance et de concentration incorrect) et de "multi-imager" une partie du fluide lui-même. Ceci peut alors causer des erreurs dans les calculs des volumes analysés. Les résultats ainsi obtenus montrent alors une légère sur-estimation du nombre total de particules, et une augmentation importante du volume de fluide analysé.

L'effet d'une efficacité inférieure au seuil recommandé n'est pas aussi extrême. Cependant, dans ce cas, la chance de "rater" une particule est élevée car une partie importante du fluide qui passe dans la cellule de flux n'est pas imagée.

### Pourquoi des seuils d'efficacité différents selon l'assemblage ?

Le seuil d'efficacité recommandé par Fluid Imaging Technologies® (50% pour l'assemblage 4X/300 $\mu$ m et 30% pour l'assemblage 10X/100 $\mu$ m) est dépendant de la combinaison cellule de flux, optique, vitesse de flux et nombre d'images par seconde. En effet, une lentille de grossissement supérieur regarde dans une fraction plus petite de la cellule de flux (champ de vision plus petit). En conséquence, lorsque l'échantillon passe dans la cellule de flux, moins de fluide passe dans le champ de vision de la caméra, conduisant à une efficacité plus faible.

### Comment augmenter ces seuils sans risque de multi-imaging ?

Afin d'obtenir une efficacité plus importante sans risque de "multi-imaging", il est possible d'utiliser des cellules de flux "Field-Of-View" (FOV). Ces cellules de flux ont des canaux beaucoup plus étroits qui sont entièrement dans le champ de vision de la caméra (possibilité d'atteindre 72% d'efficacité avec un assemblage 10X/100 $\mu$ m).

## **3.4. Sédimentation des particules**

L'utilisation de l'assemblage 10X/100 $\mu$ m pose le problème des particules fixées de très grande taille et de certaines espèces coloniales (notamment les diatomées formant de longues chaînes ou présentant des soies très longues). En effet, dans un flux descendant, ces espèces ont tendance à s'accumuler à l'entrée de la cellule de flux (comme montré sur la figure 14), retenant probablement avec elles une partie des particules 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.



*Figure 14 : colmatage de la cellule de flux 100 $\mu$ m.*



En effet, bien que le flux d'entraînement tende à positionner certaines particules de telle sorte que la plus grande longueur soit orientée dans le sens du flux et permette ainsi de faire passer des particules longues et fines (par exemple, *Proboscia alata*) ou des chaînes (par exemple, *Pseudo-Nitzschia*), certaines particules d'autres espèces (par exemple, les colonies de *Phaeocystis*) s'agglomèrent à l'entrée de la cellule de flux de 100 $\mu$ m. Il est important de noter que ce problème n'a pas été rencontré lors des expérimentations menées avec l'assemblage 4X/300 $\mu$ m, même en présence de colonies de *Phaeocystis* mesurant plusieurs millimètres.

**Proposition 1 : Sens du débit d'eau**

Les expérimentations ont donc permis de mettre en évidence des problèmes liés à la fixation des particules au Lugol (en particulier, pour l'assemblage 10X/100 $\mu$ m). En effet, un phénomène de sédimentation des particules engendrant un colmatage partiel ou total de la chambre de flux a pu être observé. C'est pourquoi, des expérimentations ont été menées sur le sens du débit d'eau (figures 15 et 16) avec les couplages 4X/300 $\mu$ m et 10X/100 $\mu$ m. En effet, nous supposons qu'un débit ascendant pourrait éventuellement éviter ce colmatage. Pour cela, deux systèmes d'agitation ont été mis en place selon le sens du débit (agitateur magnétique pour le sens ascendant et agitateur à hélice pour le sens descendant).

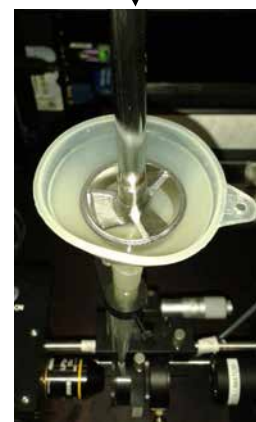


Figure 15 : Débit ascendant et agitateur magnétique.

Figure 16 : Débit descendant et agitateur à hélice.



### Hypothèse n°1 :

Le sens du débit ascendant évite la décantation donc le colmatage (partiel ou total) de la chambre de flux.

### Vérification n°1 :

Assemblage : 4X/300 $\mu$ m

Echantillon : culture de *Ditylum brightwellii* (lugol)

Volume numérisé : 5 mL

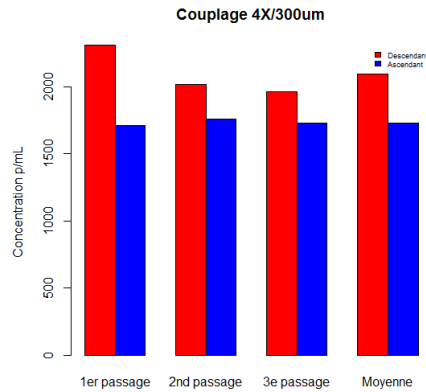


Figure 17 : Concentration pour l'assemblage 4X/300 $\mu$ m (Rouge : sens descendant ; Bleu : sens ascendant).

Assemblage : 10X/100 $\mu$ m

Echantillon : Sète, bloom de Dinophysis (lugolé)

Volume numérisé : 5 mL

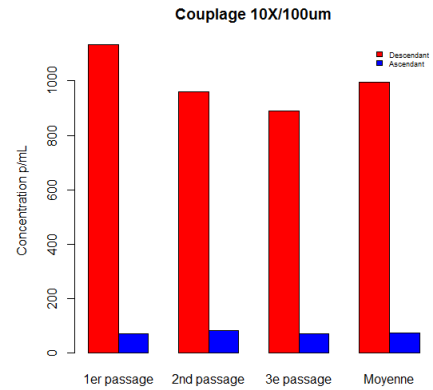
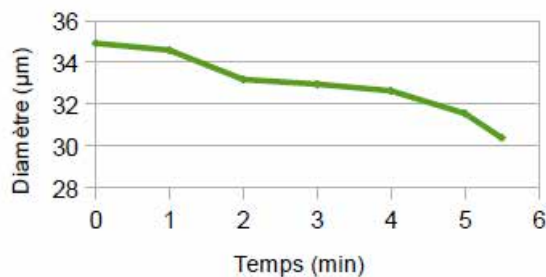


Figure 18 : Concentration pour l'assemblage 10X/100 $\mu$ m (Rouge : sens descendant ; Bleu : sens ascendant).

D'après les résultats présentés sur les figures ci-dessus (figures 17 et 18), nous pouvons constater des pertes moyennes d'environ **17% (pour le couplage 4X/300 $\mu$ m)** et **92% (pour le couplage 10X/100 $\mu$ m)** des particules numérisées pour le sens ascendant du flux par rapport au sens descendant. Deux observations peuvent expliquer ces différences d'abondances (et donc de concentrations). Comme montré sur les figures 19 et 20, le diamètre important (et donc la masse importante) des particules favorise leur décantation. Ceci peut alors engendrer deux phénomènes :

- lorsque le flux est descendant, ces particules sont les premières à être numérisées. Les particules de masse plus faible décantent moins rapidement.
- lorsque le flux est ascendant, ces particules peuvent décanter et sédimenter dans les tuyaux. Seules les particules de masse plus faible sont numérisées.

Diamètre en fonction du temps



Diamètre en fonction du temps

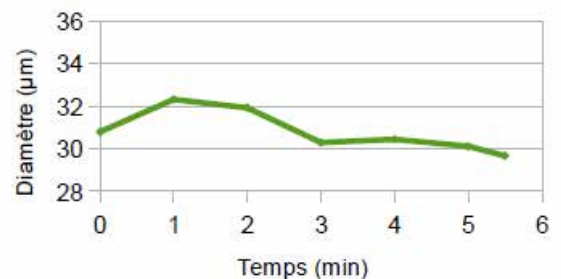


Figure 19 : Diamètre des particules en fonction du temps pour l'assemblage 4X/300 $\mu$ m et pour un flux descendant et ascendant respectivement.

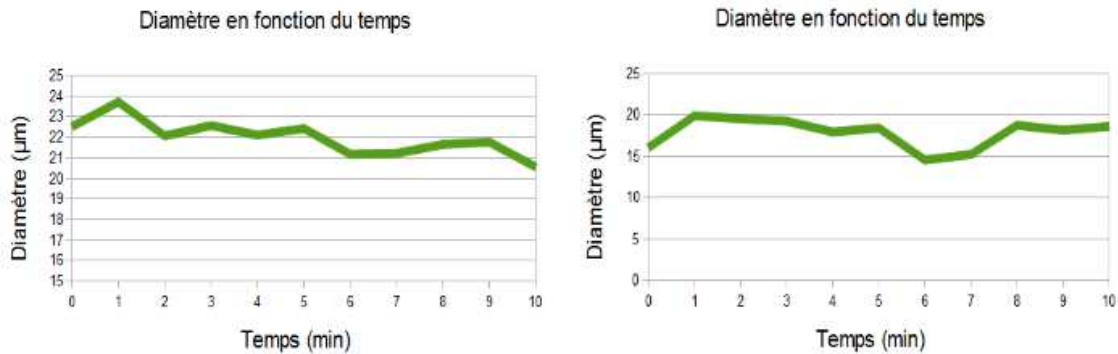


Figure 20 : Diamètre des particules en fonction du temps pour l'assemblage 10X/100µm et pour un flux descendant et ascendant respectivement.

### Hypothèse n°2 :

L'augmentation de la viscosité de l'échantillon limite le facteur de sédimentation dans les tuyaux pour un débit ascendant.

### Vérification n°2 :

Les échantillons utilisés dans la vérification de l'hypothèse 1 sont réutilisés ici. Une solution de Glycérol (en différentes concentrations : 1 à 10%) est ajoutée. Les résultats sont présentés en terme de concentrations (figures 21 et 22).

Assemblage : 4X/300µm  
 Echantillon : culture de *Ditylum brightwellii* (lugol)  
 Volume numérisé : 5 mL

Assemblage : 10X/100µm  
 Echantillon : Sète, bloom de *Dinophysis* (lugolé)  
 Volume numérisé : 5 mL

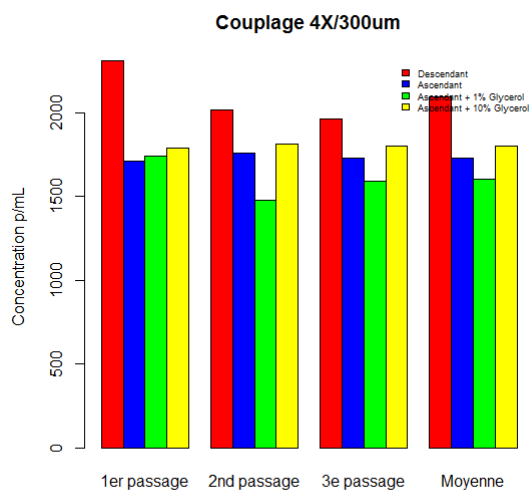


Figure 21 : Concentrations obtenues pour l'assemblage 4X/300µm (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).

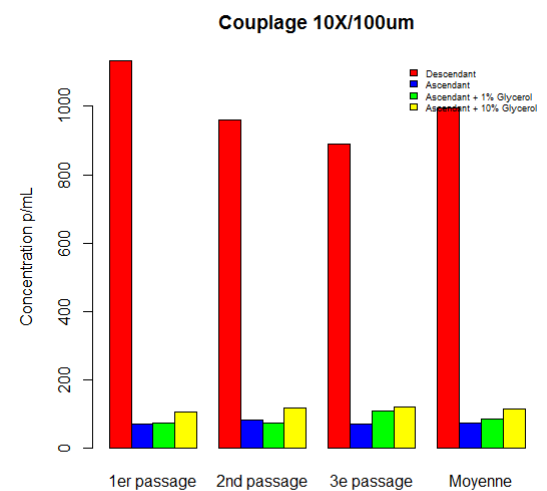


Figure 22 : Concentrations obtenues pour l'assemblage 10X/100µm (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).

Pour l'assemblage 4X/300µm, l'ajout de Glycérol (en plus ou moins forte concentration) permet d'obtenir de meilleurs résultats en terme de concentration. Cependant, cela ne permet pas d'atteindre les performances obtenues en écoulant le fluide du haut vers le bas.

Les mêmes conclusions que pour l'assemblage 4X/300 $\mu$ m peuvent être formulées pour le 10X. En effet, même si l'ajout de 10% de Glycérol permet d'obtenir de meilleures abondances, ces dernières restent largement inférieures à celles obtenues avec un sens descendant.

### En est-il de même pour les particules phytoplanctoniques ?

Les résultats précédents présentent les concentrations observées pour la totalité des particules numérisées (c'est-à-dire phytoplancton+débris). Mais nous pouvons nous demander s'il en est réellement de même pour les seules particules de phytoplancton. La figure 23 présente les concentrations obtenues pour chaque assemblage selon l'ajout ou non de Glycérol et selon le sens du débit.

Assemblage : 4X/300 $\mu$ m

Echantillon : culture de *Ditylum brightwellii* (lugol)

Volume numérisé : 5 mL

Assemblage : 10X/100 $\mu$ m

Echantillon : Sète, bloom de *Dinophysis* (lugolé)

Volume numérisé : 5 mL

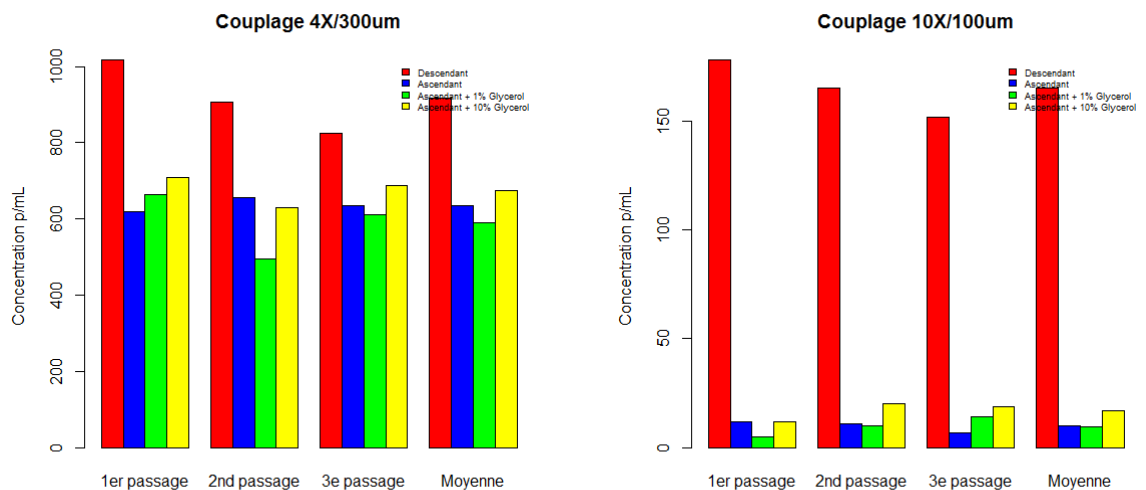


Figure 23 : Concentrations en particules phytoplanctoniques pour les assemblages 4X/300 $\mu$ m et 10X/100 $\mu$ m (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).

Les résultats obtenus sur les particules phytoplanctoniques permettent d'observer les mêmes tendances que pour le traitement de la totalité des particules (plancton+débris). La piste menant à une modification du sens du flux (sens ascendant) pour la numérisation des échantillons est donc abandonnée.

### Proposition 2 : Pré-filtration

Afin de limiter le colmatage éventuel de la cellule de flux de 100 $\mu$ m lors d'une analyse au 10X, une pré-filtration systématique des échantillons s'impose. Une pré-filtration sur filtres de vide de maille spécifique doit alors être appliquée à chaque échantillon (tableau 4).

Gamme de taille des particules	Objectif	Cellule de flux	Dimensions de la cellule de flux (Profondeur x Largeur)
10 $\mu$ m - 1000 $\mu$ m	10 X	FC100	100 $\mu$ m x 2 mm
20 $\mu$ m - 2000 $\mu$ m	4 X	FC300	300 $\mu$ m x 3 mm

Tableau 4 : Assemblages recommandés selon les différentes classes de taille.

Cependant, une pré-filtration de l'échantillon ne permet pas (ou plus) l'analyse de flores totales. Cette solution est donc dépendante des objectifs fixés par l'utilisateur (flores partielles, flores indicatrices, etc.) et doit être appliquée au cas par cas (cf. section 3.10, "Protocole dans le cadre du REPHY").

### 3.5. Dilution/Concentration des échantillons

La limitation de concentration est définie par la capacité du logiciel à discriminer des particules individuelles dans les images caméra (Sieracki et al., 1998). Dans le cas où l'échantillon à analyser présente une très forte concentration en particules ( $\gg 1000$  particules par mL), il est nécessaire de le diluer dans de l'eau de mer filtrée avant le passage au 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.

Les figures 24 et 25 présentent les résultats obtenus pour les expérimentations menées sur la qualité de numérisation en fonction de la dilution (en rouge : sans dilution; en bleu : dilué à 50%; en vert : dilué à 75%), pour les assemblages 4X/300 $\mu$ m et 10X/100 $\mu$ m respectivement. L'échantillon analysé est une culture de *Prorocentrum micans* fixée au Lugol. Trois passages ont été réalisés afin d'obtenir des résultats significatifs.

Expérimentation : Assemblages : 4X/300 $\mu$ m et 10X/100 $\mu$ m  
 Echantillon : culture de *Prorocentrum micans* (lugolée)  
 Volume numérisé : 5 mL

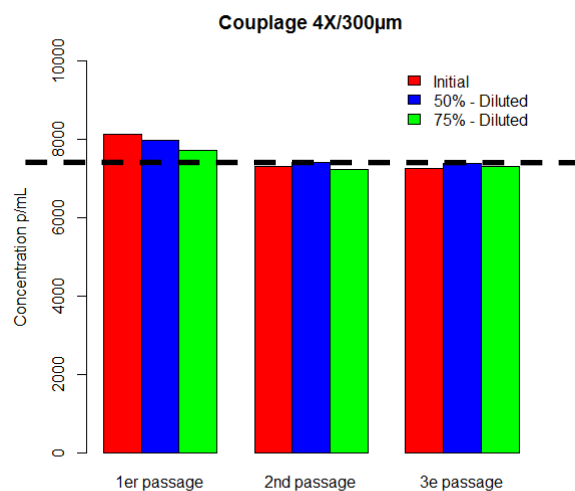


Figure 24 : Concentration pour l'assemblage 4X/300 $\mu$ m selon la dilution appliquée (Rouge : 1er passage ; Bleu : second passage ; Vert : troisième passage).

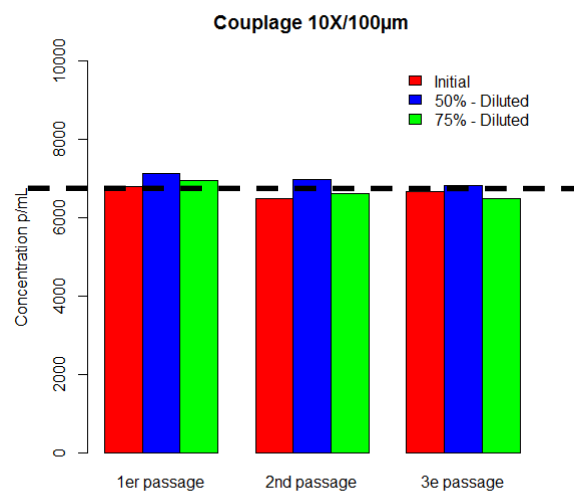


Figure 25 : Concentration pour l'assemblage 10X/100 $\mu$ m selon la dilution appliquée (Rouge : 1er passage ; Bleu : second passage ; Vert : troisième passage).

Les résultats présentés sur les figures 24 et 25 montrent une cohérence dans les concentrations obtenues avec et sans dilution. Les moyennes des 3 passages (en pointillés noirs), pour les assemblages 4X/300 $\mu$ m et 10X/100 $\mu$ m sont relativement proches.



De la même manière, si l'échantillon est très faiblement concentré ( $\ll 1000$  particules par mL), il est envisageable de le concentrer par sédimentation afin de limiter le temps de numérisation. La sédimentation doit avoir lieu à température ambiante ou au réfrigérateur et à l'obscurité. Le temps de sédimentation des particules est dépendant de la hauteur du flacon (Lund *et al.* 1958, Edler 1979). Les temps de sédimentation recommandés sont présentés dans le tableau 5 (McDermott and Raine, 2011). Pour le passage au FlowCAM et en raison des délais de passage, la fixation au Lugol est préférée à la solution de formaldéhyde. En effet, selon Hasle (1978), la fixation au formaldéhyde nécessite un temps de sédimentation supérieure à 40 heures quelque soit la taille du flacon.

*Astuce* : ajouter une goutte de tensioactif (savon) à la surface de l'échantillon, de façon à libérer les particules de phytoplancton coincées dans le film de surface.

Volume du flacon	Hauteur du flacon	Temps de sédimentation
2 mL	1 cm	3 h
10 mL	2 cm	8 h
250 mL	8 cm	24 h

Tableau 5 : Temps de sédimentation des particules en fonction de la hauteur du flacon.

Après sédimentation, et afin de concentrer l'échantillon, le surnageant est récupéré puis éliminé grâce à une pipette selon le tableau 6 (vol initial = 250 mL).

Concentration finale	Vol. surnageant	Vol. final
x2	125 mL	125 mL
x5	200 mL	50 mL
x10	225 mL	25 mL

Tableau 6 : Volume de surnageant et volume final selon la concentration désirée.

Pour connaître la concentration initiale de l'échantillon analysé, il est nécessaire de passer préalablement un faible volume de cet échantillon. Selon les expérimentations menées avec les assemblages 4X/300 $\mu$ m et 10X/100 $\mu$ m, la durée de numérisation permettant d'atteindre une estimation pertinente de la concentration est de l'ordre de 1 à 2 minutes (comme illustré sur la figure 26 pour l'assemblage 4X/300 $\mu$ m). Les mêmes conclusions peuvent être formulées pour l'assemblage 10X/100 $\mu$ m.

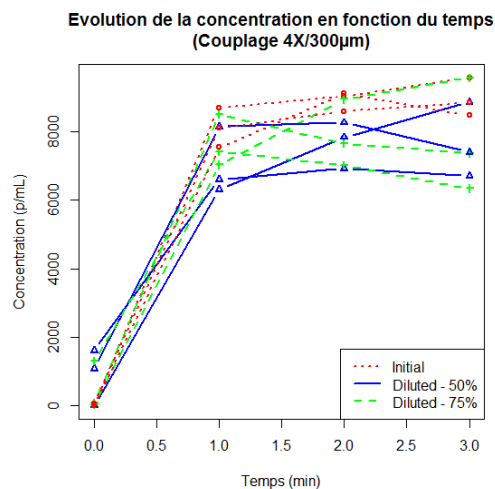


Figure 26 : Evolution de la concentration en fonction du temps pour l'assemblage 4X/300 $\mu$ m.

Dans les deux cas (dilution ou concentration), il est important d'indiquer le ratio de dilution ou de concentration dans les réglages des paramètres de numérisation de Visual SpreadSheet (figure 27). Par exemple, si un échantillon est dilué à 50% de la concentration originale, cette valeur devra être égale à 0.5 (si l'échantillon est concentré 2 fois, cette valeur devra être égale à 2) afin que le logiciel puisse calculer correctement les statistiques par rapport à la concentration initiale.

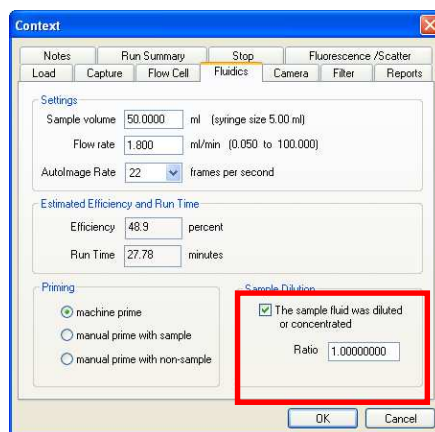


Figure 27 : Onglet "Fluidics" du menu "Context" de Visual SpreadSheet.

### 3.6. Comparaison des abondances (Lugolé vs Vivant)

Dans cette section, nous comparons les abondances et concentrations obtenues pour des échantillons naturels lugolés ou vivants. 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 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.

Parmi les échantillons REPHY 2013, nous avons choisi 9 échantillons (3 dates différentes pour 3 points de prélèvement différents) présentant des abondances intéressantes et une diversité conséquente en cellules non coloniales (ou facilement dénombrables dans les colonies). Un volume de 10 mL a été numérisé pour chacun de ces échantillons, 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 100 cellules/L, ce qui est équivalent à un comptage au microscope inversé. Cependant, pour ce dernier, un facteur multiplicateur peut également être affecté à certains taxons si l'aire qui a été observée est inférieure à la surface totale de la cuve (par exemple, si le comptage a été effectué sur un diamètre, etc.). C'est pourquoi, il est possible d'obtenir des biais pour un comptage manuel.

Pour chaque échantillon, un comptage manuel a été réalisé sur l'ensemble des collages en sortie du FlowCAM. Le tableau 7 présente les abondances et les concentrations obtenues pour les expérimentations menées sur la comparaison de la numérisation d'échantillons lugolés et vivants, pour l'assemblage 4X/300µm.

cells/L	Dunkerque DK4					
	24/04/2013		28/05/2013		19/06/2013	
	Lugol	Vivant	Lugol	Vivant	Lugol	Vivant
<b>B.rhombus</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>C.danicus</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>Coscinodiscus_spp</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>D.brightwellii</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>G.flaccida</b>	0 / 900	0 / 400	0 / 0	0 / 0	5262 / 8300	5262 / 10800
<b>Gyrodinium_spp</b>	877 / 1000	877 / 1200	2631 / 400	2631 / 600	0 / 0	0 / 0
<b>P.micans</b>	877 / 900	877 / 500	0 / 400	0 / 500	200 / 100	200 / 200
<b>Pleuro_Gyro_spp</b>	877 / 700	877 / 700	100 / 400	100 / 500	0 / 0	0 / 100
<b>Protopteridinium_spp</b>	877 / 1200	877 / 1000	877 / 900	877 / 800	200 / 300	200 / 400
<b>Pseudo-Nitzschia_spp</b>	0 / 400	0 / 200	0 / 0	0 / 0	0 / 0	0 / 0
<b>Rhizo_Probo_spp</b>	30695 / 28700	30695 / 26500	0 / 0	0 / 0	7016 / 8400	7016 / 8000
<b>T.rotula</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>TOTAL</b>	<b>34203 / 33800</b>	<b>34203 / 30500</b>	<b>3608 / 2100</b>	<b>3608 / 2400</b>	<b>12678 / 17100</b>	<b>12678 / 19500</b>

cells/L	Boulogne-sur-Mer BL3					
	26/06/2013		23/07/2013		05/08/2013	
	Lugol	Vivant	Lugol	Vivant	Lugol	Vivant
<b>B.rhombus</b>	0 / 100	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>C.danicus</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>Coscinodiscus_spp</b>	0 / 100	0 / 400	0 / 0	0 / 300	0 / 0	0 / 0
<b>D.brightwellii</b>	0 / 100	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>G.flaccida</b>	1754 / 2400	1754 / 3400	800 / 1200	800 / 2000	5262 / 4200	5262 / 4400
<b>Gyrodinium_spp</b>	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
<b>P.micans</b>	200 / 300	200 / 800	700 / 600	700 / 400	877 / 700	877 / 700
<b>Pleuro_Gyro_spp</b>	877 / 1200	877 / 1300	200 / 300	200 / 400	800 / 200	800 / 700
<b>Protopteridinium_spp</b>	1754 / 1000	1754 / 1400	3508 / 1900	3508 / 2400	6139 / 2700	6139 / 3300
<b>Pseudo-Nitzschia_spp</b>	5262 / 4500	5262 / 6200	7016 / 6900	7016 / 5600	7893 / 7300	7893 / 8000
<b>Rhizo_Probo_spp</b>	0 / 0	0 / 0	1754 / 2500	1754 / 1800	149100 / 67600	149100 / 64800
<b>T.rotula</b>	0 / 200	0 / 0	1754 / 300	1754 / 400	0 / 0	0 / 0
<b>TOTAL</b>	<b>9847 / 9900</b>	<b>9847 / 13500</b>	<b>15732 / 13700</b>	<b>15732 / 13300</b>	<b>170071 / 82700</b>	<b>170071 / 81900</b>

cells/L	Baie de Somme ME2					
	08/10/2013		18/11/2013		17/12/2013	
	Lugol	Vivant	Lugol	Vivant	Lugol	Vivant
B.rhombus	1754 / 1200	1754 / 1600	800 / 600	800 / 1200	3500 / 2800	3500 / 4300
C.danicus	300 / 900	300 / 100	0 / 300	0 / 700	5262 / 5100	5262 / 3900
Coscinodiscus_spp	0 / 100	0 / 300	1000 / 700	1000 / 1000	877 / 100	877 / 500
D.brightwellii	0 / 200	0 / 200	877 / 1300	877 / 1300	5262 / 5300	5262 / 8600
G.flaccida	0 / 0	0 / 0	0 / 0	0 / 0	200 / 500	200 / 300
Gyrodinium_spp	877 / 500	877 / 600	0 / 0	0 / 0	0 / 0	0 / 0
P.micans	877 / 600	877 / 800	0 / 100	0 / 0	100 / 400	100 / 200
Pleuro_Gyro_spp	0 / 100	0 / 600	1754 / 1100	1754 / 1300	877 / 600	877 / 800
Protoperdinium_spp	877 / 0	877 / 200	877 / 400	877 / 400	100 / 100	100 / 100
Pseudo-Nitzschia_spp	0 / 0	0 / 0	0 / 0	0 / 0	14900 / 5900	14900 / 3800
Rhizo_Probo_spp	0 / 100	0 / 0	0 / 0	0 / 0	1754 / 1500	1754 / 1400
T.rotula	0 / 0	0 / 0	7016 / 4500	7016 / 5400	1200 / 1000	1200 / 1400
<b>TOTAL</b>	<b>4685 / 3700</b>	<b>4685 / 4400</b>	<b>12324 / 9000</b>	<b>12324 / 11300</b>	<b>34032 / 23300</b>	<b>34032 / 25300</b>

Tableau 7 : Concentrations obtenues pour les expérimentations menées sur la comparaison de la numérisation d'échantillons lugolés et vivants, pour l'assemblage 4X/300µm (comptage microscope / comptage manuel sur les collages du FlowCAM).

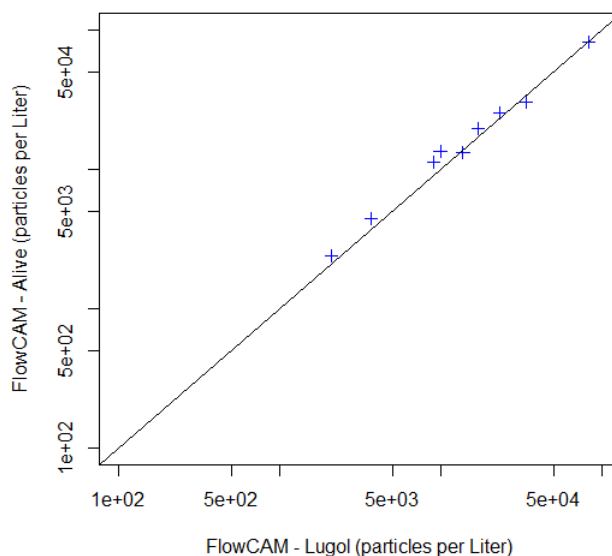


Figure 28 : Comparaison de 9 comptages manuels sur les collages du FlowCAM selon si les échantillons sont fixés ou vivants. La droite noire représente la relation 1:1 ( $r^2=0.994$  et  $p\text{-value}=5.24e-09$ ).

Afin de caractériser la liaison dans la distribution des abondances pour les échantillons lugolés et vivants, il est possible d'étudier la régression linéaire simple, "résumant" graphiquement un nuage de points par une droite, dite de régression. La "qualité" de cette régression est alors mesurée par le coefficient de détermination  $r^2$ . D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre les abondances des échantillons lugolés et vivants.



L'étude de cette relation pour les 9 échantillons (présentée sur la figure 28) montre une adéquation évidente entre les abondances obtenues pour les deux types d'échantillons. De plus, le coefficient de détermination obtenue est très important (supérieur à 0.95) et la p-value (probabilité d'obtenir la même valeur du test si l'hypothèse nulle était vraie) est très faible (inférieur à 0.01). Le résultat obtenu est donc pertinent et statistiquement significatif. Cependant, ces résultats ne concernent qu'une faible quantité d'espèces non coloniales. Il est tout à fait envisageable d'obtenir des différences conséquentes selon le fixateur utilisé et selon la composition des colonies pour certains taxons (en particulier, l'utilisation de solution de Lugol acide peut entraîner une dissolution des membranes muqueuses pour certaines espèces comme la prymnésiophycée *Phaeocystis*).

Dans le rapport de post-doctorat de Tunin-Ley (Tunin-Ley et Maurer, 2011), il est spécifié que : *"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'idée est que le niveau de gris étant un paramètre assez important de ZooPhytoImage, il faut garantir que l'intensité de la coloration est obtenue dans les mêmes conditions"*. Nous avons donc choisi de conserver ces modalités de conservation et de délais de numérisation pour l'établissement d'un protocole de numérisation optimisé.

### **3.7. Comparaison des abondances (Microscope vs FlowCAM)**

Dans cette section, nous comparons les abondances obtenues par lecture au microscope inversé (réalisée dans le cadre du REPHY) et par le FlowCAM sur les 9 échantillons précédents. Comme décrit dans la section 4.7, ces échantillons ont été sélectionnés car ils présentent des abondances intéressantes et une diversité conséquente en cellules non coloniales.

Plusieurs auteurs ont montré que les abondances obtenues par les deux méthodes de dénombrement sont proches sur des échantillons provenant de culture (Buskey and Hyatt, 2006; Ide *et al.*, 2008; Sieracki *et al.*, 1998). Cependant, à notre connaissance, dans la littérature, peu d'expérimentations ont été menées sur des échantillons provenant du milieu naturel. Les résultats de l'étude de la relation entre comptages manuels et lectures au microscope sont présentés sur la figure 29. Celle-ci montre une importante adéquation entre les abondances, renforcée par l'obtention d'une valeur élevée pour le coefficient de détermination  $r^2$  (de nouveau supérieure à 0.95 et avec une p-value inférieure à 0.01).

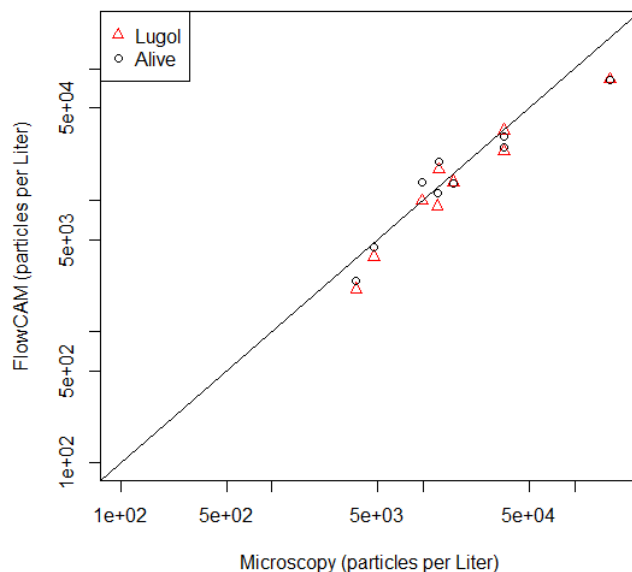
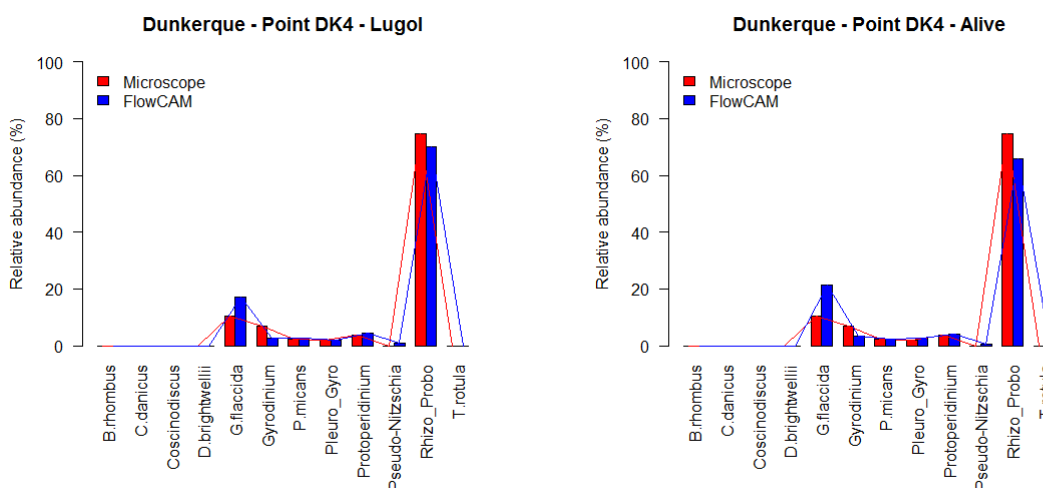


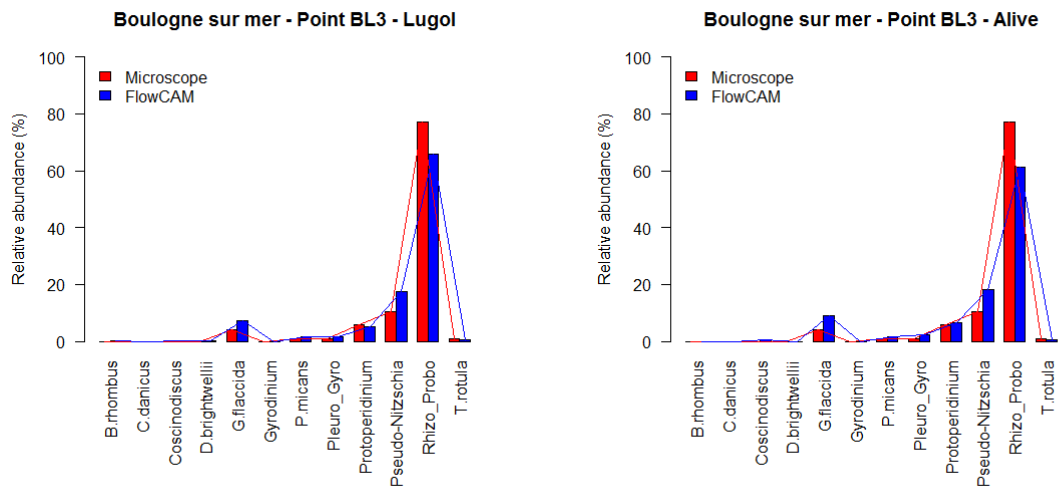
Figure 29 : Comparaison des comptages manuels sur les collages du FlowCAM (4X/300µm) et des lectures au microscope inversé réalisées dans le cadre du REPHY. La droite noire représente la relation 1:1 ( $r^2=0.956$  et  $p\text{-value}=3.06e-12$ ).

Afin de mettre en avant cette corrélation, nous choisissons de comparer les abondances relatives obtenues par comptage manuel et par lecture au microscope pour chacun des taxons et pour chaque point de prélèvement. Les résultats obtenus sont présentés sur la figure 30.



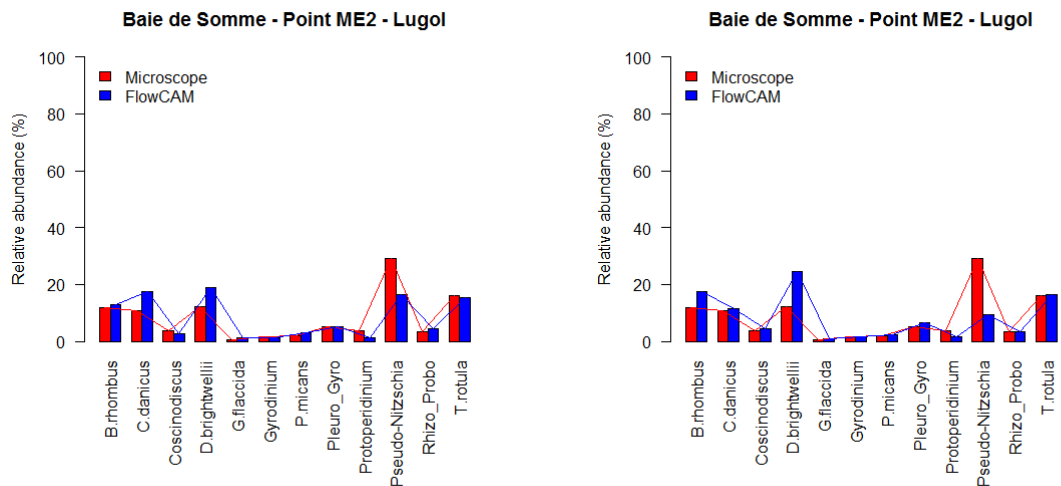
Point DK4 (24/04, 28/05, 19/06) - Echantillons lugolés ( $r^2=0.984$ ,  $p\text{-value} = 2.26e-10$ )

Point DK4 (24/04, 28/05, 19/06) - Echantillons vivants ( $r^2=0.964$ ,  $p\text{-value} = 1.60e-08$ )



Point BL3 (26/06, 23/07, 05/08) - Echantillons  
lugolés ( $r^2=0.981$ ,  $p\text{-value} = 6.97e-10$ )

Point BL3 (26/06, 23/07, 05/08) - Echantillons  
vivants ( $r^2=0.969$ ,  $p\text{-value} = 7.70e-09$ )



Point ME2 (08/10, 18/11, 17/12) - Echantillons  
lugolés ( $r^2=0.650$  (sans PSN : 0.850)  
 $p\text{-value} = 1.53e-03$  (sans PSN : 5.37e-05) )

Point ME2 (08/10, 18/11, 17/12) - Echantillons  
vivants ( $r^2=0.336$  (sans PSN : 0.816)  
 $p\text{-value} = 4.82e-02$  (sans PSN : 1.39e-04) )

Figure 30 : Comparaison des comptages manuels sur les collages du FlowCAM (4X/300 $\mu$ m) et des lectures au microscope inversé réalisées dans le cadre du REPHY pour chacun des taxons et pour chaque point de prélèvement.

Pour les points DK4 et BL3, les coefficients de détermination élevés, associés aux p-values très faibles, montrent une adéquation correcte entre les abondances relatives obtenues par les deux méthodes pour les différents taxons phytoplanctoniques. En revanche, pour le point ME2, nous pouvons remarquer des valeurs de  $r^2$  beaucoup plus faibles ainsi que des p-values plus élevées. Ceci signifie donc qu'une différence significative est observée entre les abondances relatives obtenues par comptage manuel et par lecture au microscope.

En analysant le tableau 7, il est possible de constater qu'il existe une différence considérable dans les abondances obtenues pour *Pseudo-Nitzschia* (point ME2 du 17/12/2013). La suppression de ce groupe pour les différents calculs, entraîne alors une

nette augmentation des coefficients de détermination (0.850 vs 0.650 pour les échantillons ME2 lugolés, et 0.816 vs 0.336 pour les échantillons ME2 vivants) et une baisse significative des p-values ( $5.37e-05$  vs  $1.53e-03$  pour les échantillons ME2 lugolés, et  $1.39e-04$  vs  $4.82e-02$  pour les échantillons ME2 vivants).

Les différences pour ce groupe peuvent s'expliquer principalement par les estimations faites par les observateurs au microscope. En effet, comme décrit dans la section 4.7, un facteur multiplicateur (suivant une table de conversion) peut être affecté à certains taxons selon la surface observée au microscope (par exemple, si le comptage a été effectué sur un diamètre, etc.). C'est pourquoi, il est possible d'obtenir des biais pour un comptage manuel.

### **3.8. Protocole dans le cadre du REPHY**

Il est primordial de garder en tête l'objectif du REPHY lors de l'optimisation du protocole de numérisation. Pour cela, il serait judicieux d'envisager un protocole dédié au 4X pour des flores partielles (avec, par exemple, pour objectif *Dinophysis* et/ou *PseudoNitzschia*) et un protocole dédié au 10X pour les flores totales (avec pour objectif *Alexandrium*). En effet, l'avantage lié au 4X réside dans la possibilité de numériser de grands volumes pour faire une recherche précise de certains taxons. En résumé, le choix de l'un ou de l'autre dépend, en grande partie, de l'objectif fixé (flore totale, flore partielle, recherche, observation, etc.)

Dans ce contexte, nous proposons que les stratégies d'observation du phytoplancton soient :

- pour les flores indicatrices, essentiellement utilisées pour la description d'indicateurs de type DCE, une stratégie simplifiée avec des échantillons numérisés en 4X peut être suffisante,
- pour un certain nombre de flores totales (dont le nombre est à optimiser), l'objectif 10X pourrait être utilisé, afin d'avoir une précision plus grande en biodiversité sur quelques points-clés.

Les sets d'apprentissage construits par Tunin-Ley à l'aide du couplage 10X/100 $\mu$ m (Tunin-Ley et Maurer, 2011) peuvent donc servir pour certains points flore totale ciblés de par leur intérêt d'un point de vue biodiversité, sur lesquels une flore totale détaillée serait requise. Ces points seraient, par exemple, choisis au vu des recommandations émises dans les conclusions de l'étude d'optimisation du REPHY sur la stratégie flore totale, et/ou après la thèse de doctorat de Hernandez Fariñas ("Analyse et modélisation des évolutions à long terme de la biodiversité phytoplanctonique dans les zones côtières sous l'effet des pressions environnementales et anthropiques", IFREMER Nantes). Mais ces sets peuvent également être utilisés pour de l'expertise ou en cas de doute.

Les sets d'apprentissage construits à l'aide du couplage 4X/300 $\mu$ m (livrable n°2) peuvent être utilisés pour les autres flores totales, pour les flores indicatrices (DCE), et pour les flores toxiques. En ce qui concerne ces dernières, un couplage avec les techniques FISH peut être envisagé.

### **FLORe TOTale FLORTOT :**

- Points FLORTOT ciblés (selon les consignes du REPHY ou les résultats de l'étude menée par Hernandez Fariñas à l'IFREMER Nantes)
  - Concentration/dilution de l'échantillon
  - Passage avec l'assemblage 10X/100µm
- Autres points FLORTOT
  - Un passage avec l'assemblage 4X/300µm, puis un passage avec l'assemblage 10X/100µm (si nécessaire).

### **FLORe PARTielle FLORPAR :**

- Pré-filtration sur filtres de vide de maille (si nécessaire)
- Concentration/dilution de l'échantillon (si nécessaire)
- Passage avec assemblage 4X/300µm ou 10X/100µm selon les particules ciblées

### **FLORe INDicatrice FLORIND :**

- *Alexandrium* (figure 31) :
  - Cellules de forme arrondie et de petite taille (entre 10 et 30 µm)
    - Passage avec l'assemblage 10X/100µm
  - Prolifération possible à des concentrations très importantes (plusieurs millions ou même centaines de millions de cellules par litre)
    - Dilution de l'échantillon.



Figure 31 : Concentration maximale d'*Alexandrium* par site (1992-2001).

- *Dinophysis* (figure 32) :
  - Cellules de taille petite ou moyenne (entre 30 et 100  $\mu\text{m}$ )
    - Passage avec l'assemblage 10X/100 $\mu\text{m}$  ou 4X/300 $\mu\text{m}$
  - Faible taux de développement, concentrations maximales généralement comprises entre 1000 et 10 000 cellules par litre (très rarement supérieures à 100 000)
    - Dilution de l'échantillon si nécessaire



Figure 32 : Concentration maximale de *Dinophysis* par site (1992-2001).

- *Pseudo-Nitzschia* (figure 33) :
  - Cellules de forme allongée, souvent assemblées en chaînes, et de taille et largeur très variables d'une espèce à l'autre (généralement, entre 50 et 200  $\mu\text{m}$  et une largeur entre 1.5 et 3.5  $\mu\text{m}$ )
    - Passage avec l'assemblage 10X/100 $\mu\text{m}$  ou 4X/300 $\mu\text{m}$
  - Proliférations observées très régulièrement sur tout le littoral, concentrations le plus souvent importantes (plusieurs centaines de milliers, voire plus d'un million de cellules par litre)
    - Dilution de l'échantillon

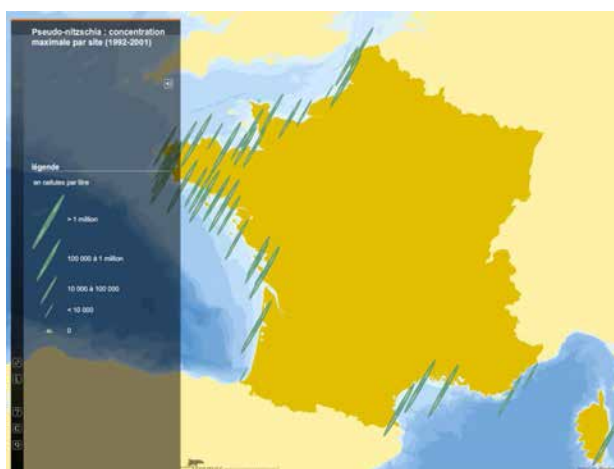


Figure 33 : Concentration maximale de *Pseudo-Nitzschia* par site (1992-2001).

## **4. Etude du traitement des échantillons par ZooPhytoImage**

En suivant les étapes d'importation et de traitement des données décrits dans le livrable n°1 ("Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique") et dans l'Annexe 5, nous comparons, dans cette section, les concentrations obtenues par microscope et par le système couplé FlowCAM/ZooPhytoImage.

### ***4.1. Comparaison des abondances (Microscope vs FlowCAM / ZooPhytoImage)***

Dans le livrable n°2 intitulé "Protocole sur les outils de reconnaissance optimisés Manche-Atlantique", nous avons présenté les performances obtenues sur les groupes de deux sets d'apprentissage ("REPHY\_lugol\_4X" et "REPHY\_alive\_4X"). Toutefois, les scores de performance mesurés l'ont été uniquement sur les données issues des sets d'apprentissage, par une méthode d'estimation non biaisée (validation croisée). Il faut donc considérer que les différents indices de performance sont moins bons sur des échantillons réels pour deux raisons :

1. les échantillons naturels contiennent également des particules n'appartenant pas aux groupes utilisés dans le set d'apprentissage (à commencer par les groupes qui ont été écartés par manque de vignettes). L'effet de ces groupes non inclus dans le set d'apprentissage est, en principe, assez faible si le set d'apprentissage a été élaboré avec soin. En effet, les groupes peu abondants le sont par manque de vignettes obtenues en quantité suffisante à l'issue de l'analyse des échantillons utilisés pour élaborer ces tests d'apprentissage, ce qui traduit la rareté de ces groupes dans les échantillons de départ qui ont été utilisés pour réaliser les sets d'apprentissage.
2. les proportions respectives des différents groupes sont différentes dans les échantillons naturels (et très variables). Or, les performances de reconnaissance par groupe sont influencées par les proportions relatives des différents groupes dans l'échantillon traité.

Il peut en résulter une diminution de 5 à 15% de la valeur des indices, lorsque l'outil est utilisé pour classer des échantillons naturels. Par exemple, il faut s'attendre à ce que le taux de reconnaissance globaux des deux outils de reconnaissance chute entre 65 et 75%. C'est pourquoi, il est nécessaire d'évaluer les performances des outils sur de nouveaux échantillons (non utilisés pour la création des sets d'apprentissage) dont on veut déterminer la composition phytoplanctonique. Cette évaluation est rendue possible grâce aux deux outils de reconnaissance optimisés créés à partir des sets d'apprentissage "REPHY\_lugol\_4X" et "REPHY\_alive\_4X" présentés dans le livrable n°2. Une étude similaire a été réalisée dans le cadre du post-doctorat de Tunin-Ley (Tunin-Ley et Maurer, 2011) pour des échantillons provenant du bassin d'Arcachon et numérisés avec l'assemblage 10X/100µm.

Dans un premier temps, l'étude porte sur la comparaison du nombre de vignettes pour chaque taxon, comptées manuellement sur les collages issus du FlowCAM d'une part, et classées automatiquement par le logiciel ZooPhytoImage d'autre part (à l'aide des sets d'apprentissage "REPHY\_alive\_4X" et "REPHY\_lugol\_4X"). Les étapes

d'importation des données du FlowCAM, ainsi que le traitement des images par ZooPhytoImage sont décrites dans l'Annexe 5. Dans les tableaux 8 et 9, les nombres de Vrais Positifs (notés VP), de Faux Positifs correspondant à des particules cibles (notés FP) et de Faux Positifs appartenant aux débris classés dans chacun des groupes, sont également relevés.

Ici, nous nous intéressons plus particulièrement aux espèces non coloniales (ou facilement dénombrables en colonie). Cette condition est nécessaire car les colonies ne sont pas encore identifiées individuellement par le logiciel ZooPhytoImage. Une colonie est donc considérée comme un objet unique. Afin de comparer les comptages cellulaires avec les comptages manuels des collages du FlowCAM, cette étude se focalise donc sur un nombre limité de taxons (9 espèces).

nb vignettes	Baie de Somme ME1 du 17/12/2013							
	Lugol				Vivant			
	Manuel	Automatique			Manuel	Automatique		
		VP	FP	débris		VP	FP	débris
B.rhombus	31	18	2	29	52	23	2	130
C.danicus	22	22	3	13	27	24	0	3
D.brightwellii	72	59	0	8	47	37	1	4
G.flaccida	0	0	4	18	3	1	0	2
Gyrodinium_spp	0	0	0	2	0	0	0	9
Pleuro_Gyro_spp	5	5	5	16	2	2	0	73
Protoperdinium_spp	3	2	1	14	5	1	1	0
Rhizo_Probo_spp	19	15	1	2	8	2	2	0
T.rotula	15	5	2	57	13	5	11	114
<b>TOTAL</b>	<b>167</b>	<b>126</b>	<b>18</b>	<b>159</b>	<b>157</b>	<b>95</b>	<b>17</b>	<b>335</b>

*Tableau 8 : Nombre de vignettes comptabilisées manuellement et automatiquement pour l'échantillon ME1 du 17/12/2013 (4X/300µm).*

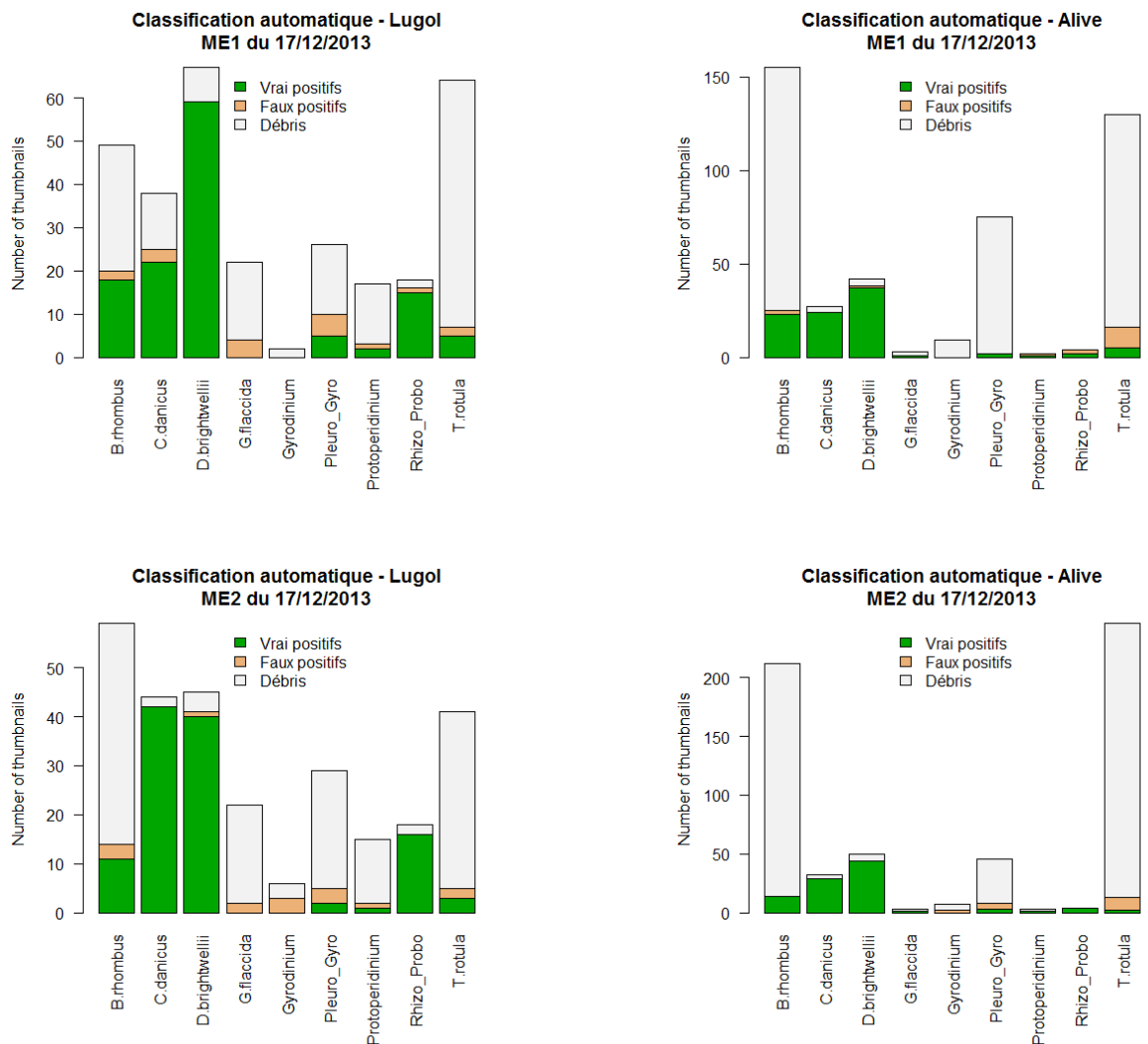
nb vignettes	Baie de Somme ME2 du 17/12/2013							
	Lugol				Vivant			
	Manuel	Automatique			Manuel	Automatique		
		VP	FP	débris		VP	FP	débris
B.rhombus	17	11	3	45	25	14	0	198
C.danicus	51	42	0	2	39	29	0	3
D.brightwellii	49	40	1	4	55	44	0	6
G.flaccida	4	0	2	20	3	1	0	2
Gyrodinium_spp	0	0	3	3	0	0	2	5
Pleuro_Gyro_spp	6	2	3	24	8	3	5	38
Protoperdinium_spp	1	1	1	13	1	1	0	2
Rhizo_Probo_spp	15	16	0	2	14	4	0	0
T.rotula	4	3	2	36	2	2	11	233
<b>TOTAL</b>	<b>147</b>	<b>115</b>	<b>15</b>	<b>149</b>	<b>147</b>	<b>98</b>	<b>18</b>	<b>487</b>

*Tableau 9 : Nombre de vignettes comptabilisées manuellement et automatiquement pour l'échantillon ME2 du 17/12/2013 (4X/300µm).*

Pour le point ME1 (tableau 8), le taux de bonne reconnaissance pour l'échantillon lugolé est de 75.45% et de 60.51% pour l'échantillon vivant. De même pour



le point ME2 (tableau 9), nous constatons un taux de bonne reconnaissance de 78.23% pour l'échantillon lugolé et 66.67% pour l'échantillon vivant.



*Figure 34 : Nombre de vignettes comptabilisées automatiquement pour chaque taxon phytoplanctonique étudié, pour les échantillons ME1 et ME2 (lugolés et vivants) avec l'assemblage 4X/300µm.*

Cependant, en approfondissant l'analyse (figure 34), nous pouvons remarquer que certaines tendances se dégagent:

- pour la plupart des taxons étudiés, la majorité des vignettes est correctement identifiée par les outils de reconnaissance (Vrais Positifs en vert),
- une proportion élevée de faux-positifs, appartenant principalement aux débris, est observée dans la classification automatique pour la majorité des groupes (Débris, en gris),
- 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 (Faux Positifs en orange). Cela signifie qu'il y a peu de confusion entre les groupes ciblés.

Les principales sources de confusion avec les groupe phytoplanctoniques sont donc les particules inertes. Il est donc nécessaire de continuer l'optimisation des sets d'apprentissage afin d'obtenir une discrimination plus pertinente des débris.

L'analyse des vignettes correctement classées (les Vrais Positifs) pour chaque groupe nous permet d'avoir une idée sur l'adéquation entre les lectures au microscope et les comptages automatiques issus du système couplé FlowCAM/ZooPhytoImage. En effet, il est possible qu'une vignette comporte plusieurs cellules (en particulier dans le cas des espèces coloniales). Nous choisissons donc de compter manuellement le nombre de cellules par vignette dans chacune des catégories. Les résultats sont présentés dans les tableaux 10 et 11.

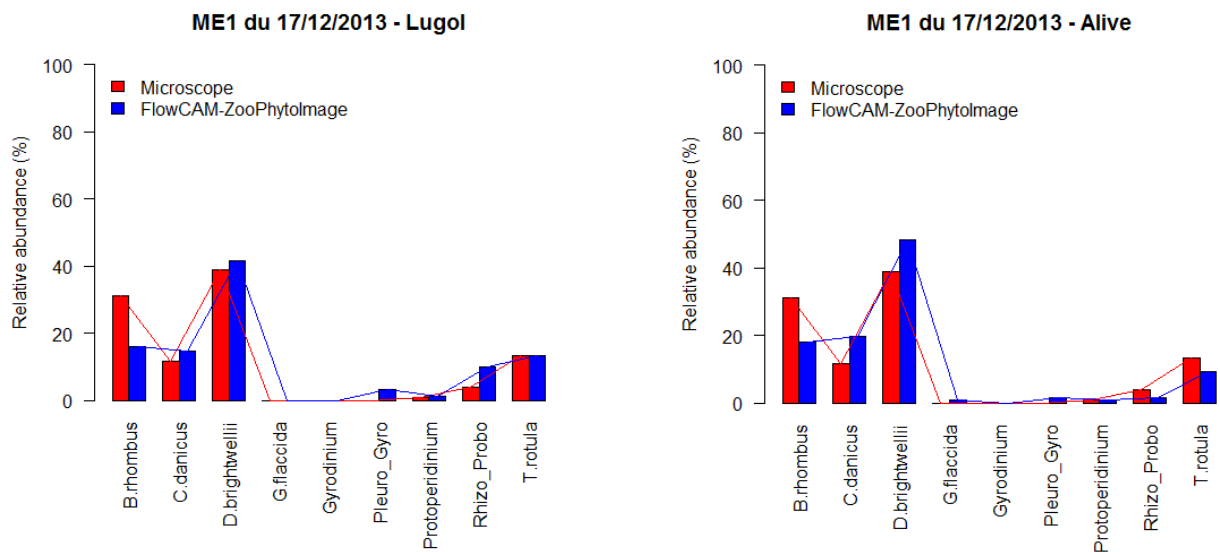
cells/L	Baie de Somme ME1 du 17/12/2013		
	Lugol		Vivant
	Microscope	FlowCAM-ZooPhytoImage	FlowCAM-ZooPhytoImage
B.rhombus	7016	2400	2300
C.danicus	2631	2200	2500
D.brightwellii	8770	6300	6200
G.flaccida	0	0	100
Gyrodinium_spp	0	0	0
Pleuro_Gyro_spp	0	500	200
Protoperidinium_spp	200	200	100
Rhizo_Probo_spp	877	1500	200
T.rotula	3000	2000	1200
<b>TOTAL</b>	<b>22494</b>	<b>15100</b>	<b>12800</b>

*Tableau 10 : Comparaison des concentrations obtenues (en cellules/L) par lecture microscopique et par FlowCAM/ZooPhytoImage (4X/300µm) pour l'échantillon ME1.*

cells/L	Baie de Somme ME2 du 17/12/2013		
	Lugol		Vivant
	Microscope	FlowCAM-ZooPhytoImage	FlowCAM-ZooPhytoImage
B.rhombus	3500	1800	1800
C.danicus	5262	4200	2900
D.brightwellii	5262	4500	7500
G.flaccida	200	0	100
Gyrodinium_spp	0	0	0
Pleuro_Gyro_spp	877	200	300
Protoperidinium_spp	100	100	100
Rhizo_Probo_spp	1754	1600	400
T.rotula	1200	600	1400
<b>TOTAL</b>	<b>18155</b>	<b>13000</b>	<b>14500</b>

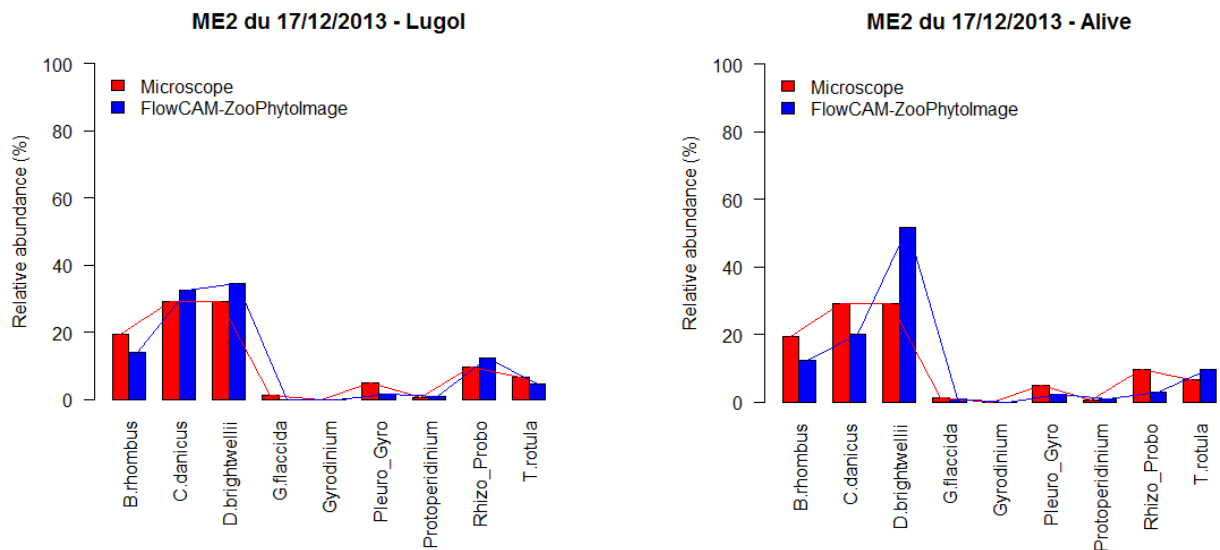
*Tableau 11 : Comparaison des concentrations obtenues (en cellules/L) par lecture microscopique et par FlowCAM/ZooPhytoImage (4X/300µm) pour l'échantillon ME2.*

Pour le point ME1 (tableau 10), le taux de bonne reconnaissance pour l'échantillon lugolé est de 67.13% et de 56.90% pour l'échantillon vivant. De la même façon pour le point ME2 (tableau 11), nous notons un taux de bonne reconnaissance de 71.61% pour l'échantillon lugolé et 79.87% pour l'échantillon vivant.



*Point ME1 du 17/12/2013 - Echantillon lugolé*  
**( $r^2=0.827$ , p-value =  $6.81e-04$ )**

*Point ME1 du 17/12/2013 - Echantillon vivant*  
**( $r^2=0.828$ , p-value =  $6.59e-04$ )**



*Point ME2 du 17/12/2013 - Echantillon lugolé*  
**( $r^2=0.950$ , p-value =  $8.30e-06$ )**

*Point ME2 du 17/12/2013 - Echantillon vivant*  
**( $r^2=0.698$ , p-value =  $5.03e-03$ )**

Figure 35 : Comparaison des comptages obtenus par l'outil FlowCAM/ZooPhytoImage (4X/300 $\mu$ m) et des lectures au microscope inversé réalisées dans le cadre du REPHY pour chacun des taxons étudiés.

La figure 35 présente les comparaisons entre les abondances relatives obtenues par lecture microscopique et par l'outil FlowCAM/ZooPhytoImage. Pour les échantillons lugolés, comme pour les échantillons vivants, les proportions relatives de chaque groupe semblent suivre une même distribution pour les deux méthodes de comptage utilisées. De plus, les coefficients de détermination sont supérieurs à 0.8 et sont associés à des p-values faibles.

Cependant, de tels scores sont trop faibles pour donner des classifications fiables en routine. Il faut donc valider les données, grâce aux options nouvellement ajoutées au logiciel (étape de vérification manuelle des classements effectués par l'ordinateur). Cette opération est accélérée par le fait que l'ordinateur a effectué un classement correct à 65-75%, par rapport à une classification purement manuelle des vignettes. Les études menées par l'Université de Mons (Grosjean et Denis, 2010; Dereume-Hancart, 2013) ouvrent la voie vers des techniques plus pointues de reconnaissance automatisée permettant une estimation moins biaisée de l'abondance sans avoir à valider manuellement la totalité des vignettes. En effet, il n'est pas question pour l'instant, et il n'en sera probablement pas question pour un certain temps encore, de se fier uniquement aux classements effectués par l'ordinateur. Mais différentes techniques sont envisageables pour permettre une validation adéquate sans nécessairement devoir vérifier manuellement toutes les vignettes.

## 4.2. Correction statistique de l'erreur

Deux choix étaient possibles au départ : une validation manuelle totale de l'échantillon (comme pour le zooplancton traité avec le ZooSCAN - Ledée, 2013), ou une validation partielle. Pour le phytoplancton, le choix s'est porté sur la deuxième solution. Le module de correction détermine une probabilité d'être suspect pour chaque particule, et seules les particules les plus suspectes sont à valider. Il est également envisageable d'ajouter à cette probabilité des informations extérieures basées sur des critères bio-écologiques (par exemple la pertinence d'une identification dans une zone géographique donnée).

Les erreurs induites par les classifieurs n'étant pas aléatoires mais systématiques, il est ensuite possible de modéliser cette erreur afin d'effectuer des corrections statistiques. L'outil envisagé permet donc de ne valider manuellement que les particules les plus suspectes (figure 36). Cette approche particulière offre un compromis idéal entre la méthode automatique et la validation manuelle totale, tout en garantissant des performances de reconnaissance similaires ou meilleures dans un laps de temps acceptable.

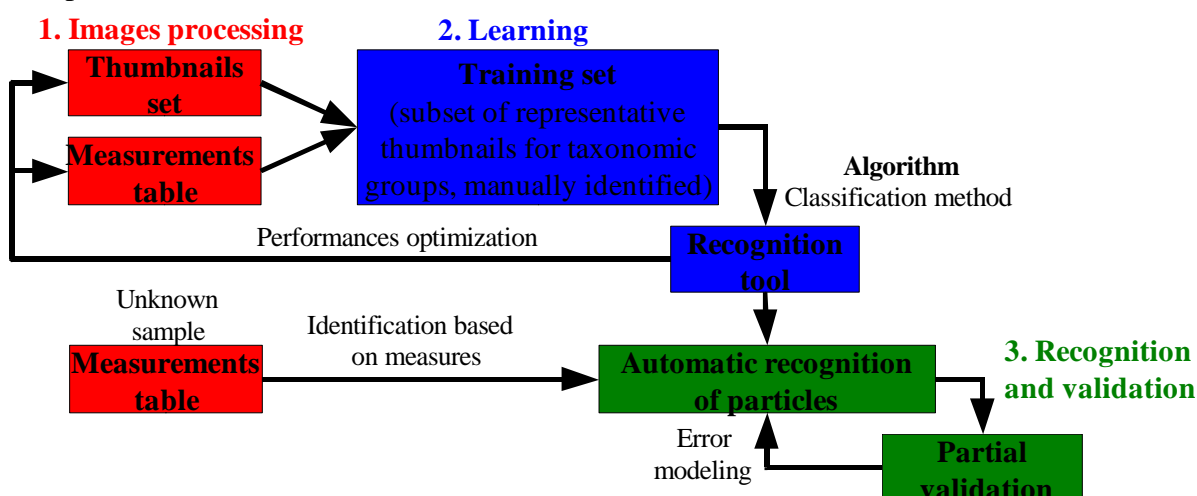
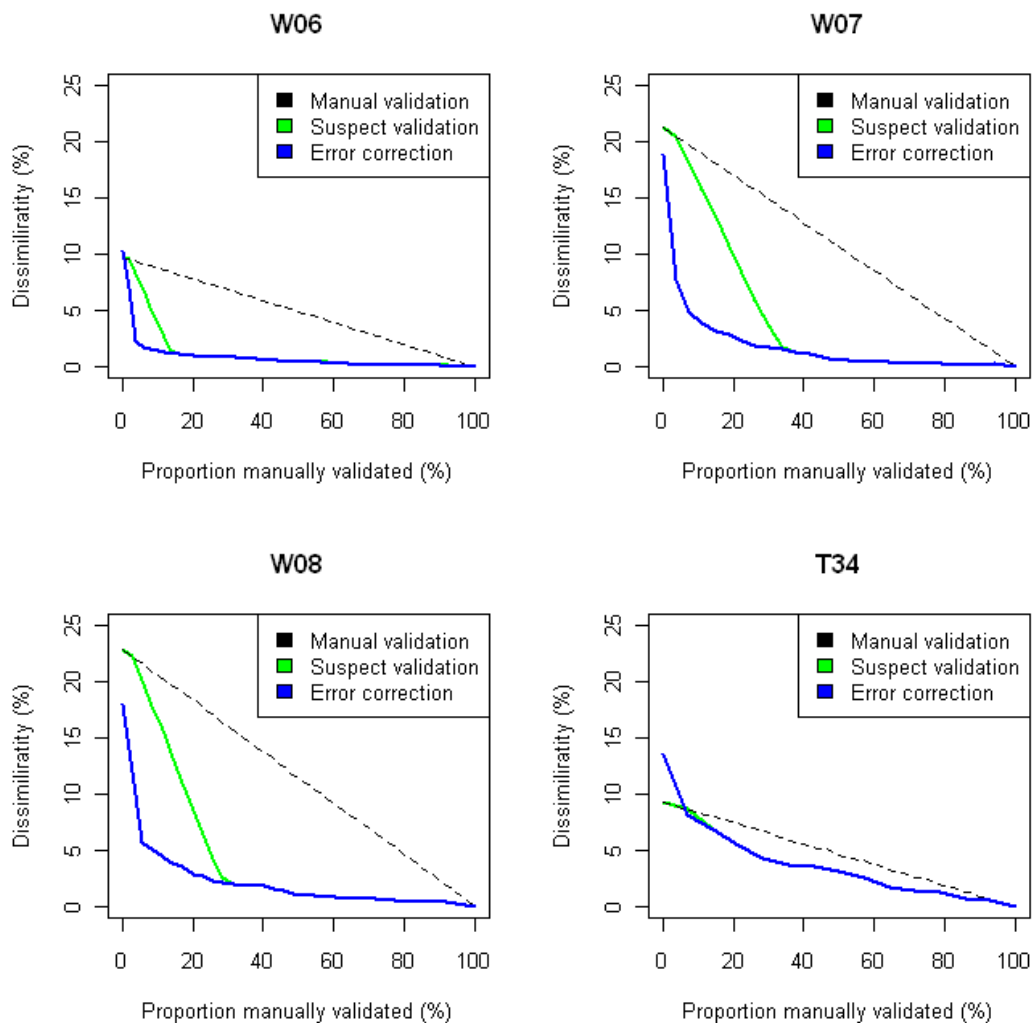


Figure 36 : Principe de fonctionnement du module de correction statistique de l'erreur, basé sur la validation manuelle des vignettes les plus "suspectes".

Les graphes présentés sur la figure 37 (fournis par l'Université de Mons), montrent l'évolution du score de dissimilarité entre l'identification automatique obtenue et l'identification manuelle considérée comme "vérité-terrain", en fonction de la proportion de vignettes validées pour des échantillons de culture. En noire est représentée la validation manuelle de vignettes aléatoires; en vert, nous avons la validation manuelle des vignettes les plus "suspectes" et en bleu, la validation des vignettes suspectes suivie de l'étape de modélisation et de correction de l'erreur. Ces résultats préliminaires montrent la pertinence de la méthode utilisée. En effet, les premières expérimentations ont montré que, pour des échantillons provenant de cultures, la validation de 5 à 20% de vignettes permettait d'obtenir moins de 5% d'erreur finale. La correction statistique de l'erreur apporte donc un gain substantiel. La validation partielle permet de mener alors à une estimation satisfaisante des abondances par groupe. En d'autres termes, la diminution de la dissimilarité étant linéaire avec la validation manuelle, et exponentielle avec la correction d'erreur, une validation manuelle des particules suspectes, combinée à la technique de correction d'erreur originale qui y est associée, permet une correction plus rapide des erreurs que la validation purement manuelle.



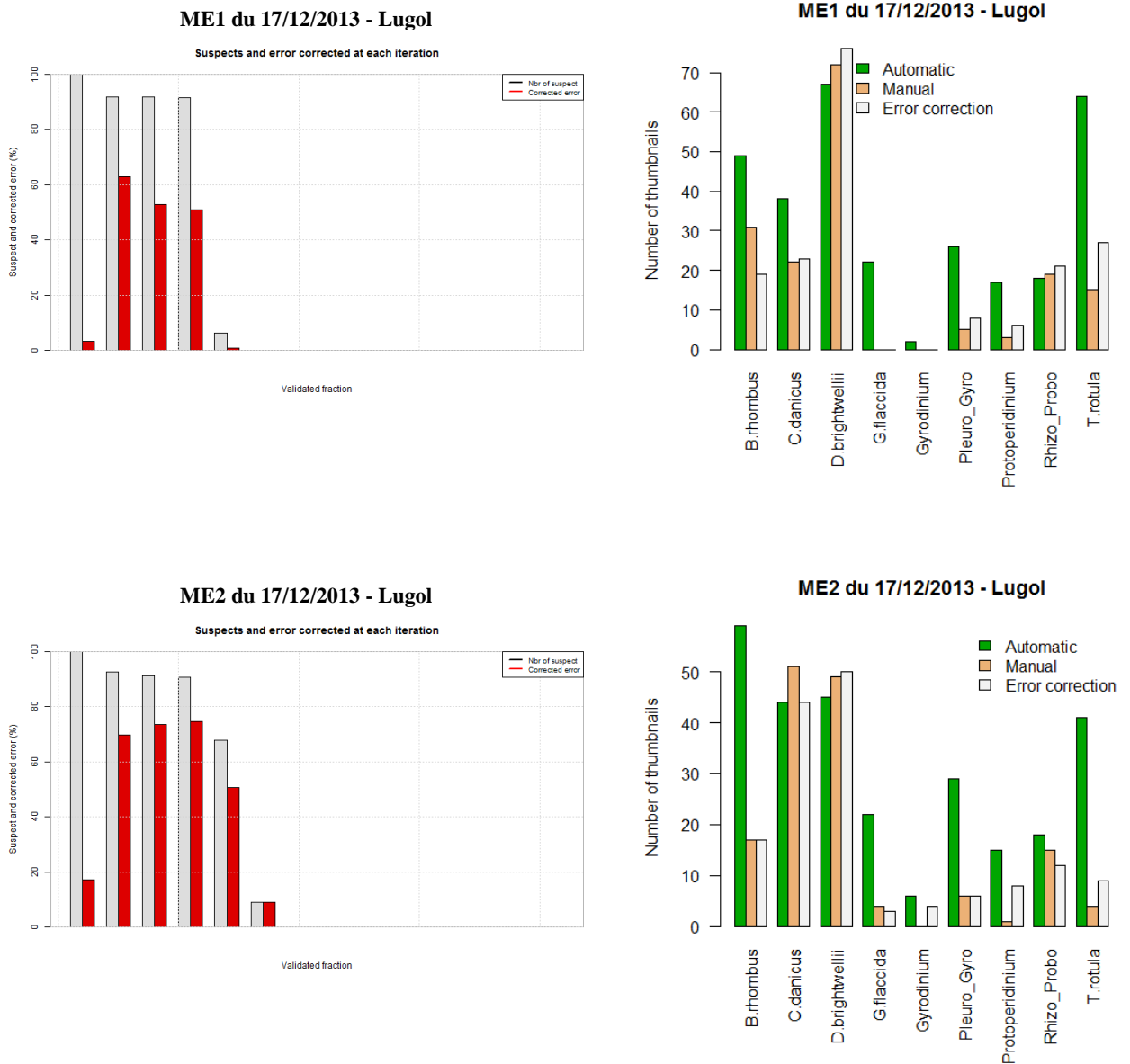
*Figure 37 : Evolution des scores de dissimilarité entre l'identification automatique et la "vérité-terrain", en fonction de la proportion de vignettes validées manuellement.*

Un cahier des charges a été rédigé en février 2013 (Annexe 6). L'objet de celui-ci est de définir la liste des fonctionnalités à développer afin de disposer d'un logiciel permettant de gérer complètement le module de correction statistique de l'erreur par validation partielle et manuelle des vignettes suspectes. En effet, plusieurs fonctions nouvelles doivent être ajoutées au logiciel ZooPhytoImage existant afin de faciliter la manipulation du nouvel outil et en particulier, la validation manuelle des classifications automatiques générées. Ce document spécifie également les différentes phases de ce nouveau module en utilisation en (pré)routine, qu'il serait nécessaire de comprendre et de suivre afin d'assurer une transition douce entre les versions antérieures du logiciel et le nouveau système. Pour comprendre les mécanismes du module de correction de l'erreur et son utilisation en routine, se reporter au livrable n°1 ("Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique").

Nous souhaitons maintenant, expérimenter ce module de validation en routine sur les échantillons naturels utilisés précédemment (section 4.1). Le tableau 12 présente les résultats obtenus après validation des prédictions automatiques par la méthode de correction statistique de l'erreur, sur les échantillons lugolés ME1 et ME2. Alors que la classification purement automatique montre une sur-estimation du nombre de vignettes pour une grande majorité des groupes phytoplanctoniques (due principalement à la confusion entre les vignettes de chaque groupe et les particules détritiques), la validation partielle manuelle des vignettes permet d'atteindre des scores de performance intéressants puisque les nombres de vignettes après correction de l'erreur sont relativement proches de ceux comptés manuellement sur les collages issus du FlowCAM.

nb vignettes	Baie de Somme ME1 du 17/12/2013			Baie de Somme ME2 du 17/12/2013		
	Lugol			Lugol		
	Manuel	Automatic	Error correction	Manuel	Automatic	Error correction
B.rhombus	31	49	19	17	59	17
C.danicus	22	38	23	51	44	44
D.brightwellii	72	67	76	49	45	50
G.flaccida	0	22	0	4	22	3
Gyrodinium_spp	0	2	0	0	6	4
Pleuro_Gyro_spp	5	26	8	6	29	6
Protoperidinium_spp	3	17	6	1	15	8
Rhizo_Probo_spp	19	18	21	15	18	12
T.rotula	15	64	27	4	41	9
<b>TOTAL</b>	<b>167</b>	<b>303</b>	<b>180</b>	<b>147</b>	<b>279</b>	<b>153</b>

*Tableau 12 : Comparaison du nombre de vignettes comptées manuellement, automatiquement, et après correction de l'erreur pour les échantillons lugolés ME1 et ME2, pour l'assemblage 4X/300µm.*



*Figure 38 : Nombre de vignettes comptabilisées manuellement, automatiquement et après correction de l'erreur pour chaque taxon phytoplanctonique étudié, pour les échantillons ME1 et ME2 lugolés avec l'assemblage 4X/300µm.*

La figure 38 présente les résultats de comparaisons des nombres de vignettes pour chaque méthode utilisée ainsi que les pourcentages de suspects et d'erreur corrigée à chaque itération du processus de validation des vignettes suspectes. Sur les graphes de gauche, les barres grises représentent la proportion d'objets suspects dans la fraction validée. Au cours de la première étape, aucun modèle n'est encore calculé, donc toutes les vignettes sont considérées comme suspectes. Les barres rouges indiquent, quant à elles, la fraction d'objets qui ont été mal classés et qui ont été corrigés.

Pour l'échantillon ME2 lugolé, après la sixième itération, la détection de suspects correspond étroitement à l'erreur réelle. La détection est alors améliorée grâce à la fraction de l'échantillon déjà validé qui peut être utilisé pour l'apprentissage de l'algorithme de détection. D'autre part, l'erreur résiduelle est ici inférieure à 10%. C'est pourquoi, nous avons choisi raisonnablement de considérer que la validation peut se terminer à la sixième itération et que la correction apportée par cette validation partielle grâce à la correction statistique de l'erreur, peut être considérée comme pertinente.

En ce qui concerne l'échantillon ME1 lugolé, la validation partielle est considérée comme suffisante à la cinquième itération. Dans ce cas, nous avons une erreur résiduelle de moins de 5%.

En s'appuyant sur les travaux réalisés par les différents partenaires impliqués dans le projet FlowCAM/ZooPhytoImage, un protocole de numérisation a été mis en place (Annexes 2 et 3). Ce dernier permet alors une analyse rapide et répétable des échantillons étudiés dans le cadre d'un réseau de surveillance et d'observation tel que le REPHY. Les études quantitatives et qualitatives menées dans ce rapport, montrent également la pertinence de l'utilisation d'un système d'analyse d'images et de reconnaissance semi-automatisée des données tel que ZooPhytoImage. En ce sens, les étapes de numérisation au FlowCAM, d'importation des données dans ZooPhytoImage, de traitement d'images, de création et d'optimisation d'outils de reconnaissance pour la reconnaissance semi-automatique devront donc suivre les modalités décrites dans les Annexes 2, 3 et 5, ainsi que dans les livrables n°1 ("Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique") et n°2 ("Protocole sur les outils de reconnaissance optimisés Manche-Atlantique").



## 5. Exploitation des résultats

### 5.1. *Abondances et biomasse en milieu naturel*

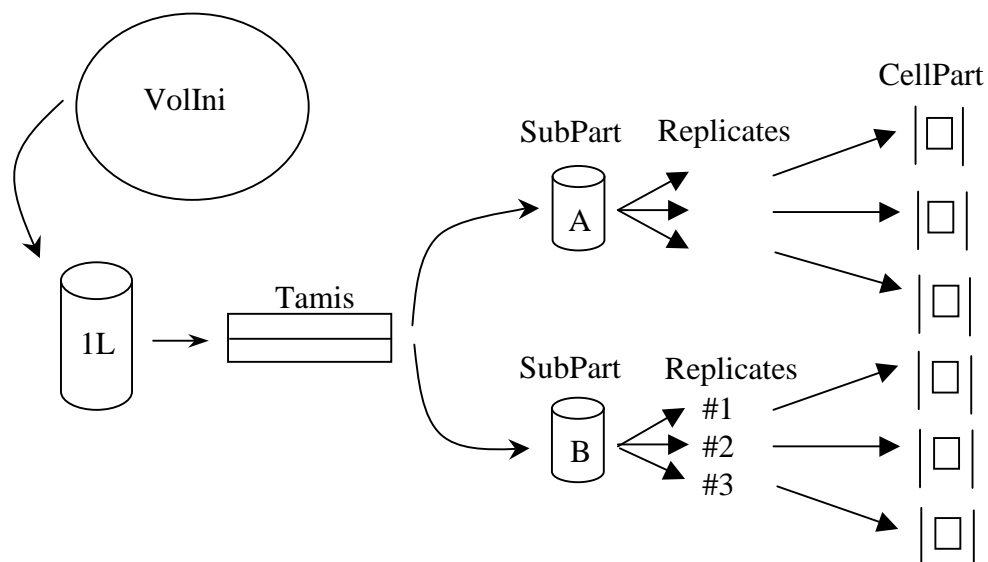
Le logiciel ZooPhytoImage est destiné à l'élaboration de séries spatio-temporelles de plancton en automatisant les processus de traitement des échantillons. Une fois les différents organismes phytoplanctoniques dénombrés et classés, le logiciel calcule des variables d'importance écologique comme les abondances, les biomasses et les spectres de taille par groupe taxonomique ou pour l'ensemble de l'échantillon.

Cependant, pour le calcul de ces variables, certaines métadonnées doivent être fournies par l'utilisateur. Ces informations permettent de calculer le coefficient de dilution (équation 4) nécessaire à la conversion des fréquences observées en nombre d'individus par unité de volume d'eau de mer.

$$Dil = \frac{1}{SubPart.CellPart.replicates.VolIni} \quad (4)$$

Les données à fournir concernent donc (figure 39) :

- **VolIni** = volume échantillonné en mer (bouteille Niskin, flacon sur perche, etc.) → unité choisie par l'utilisateur;
- **SubPart** = fraction introduit dans l'appareil (si *volIni* = 1L et 10 mL introduit dans l'appareil, alors *SubPart* = 10mL/1000mL = 0.01);
- **Replicates** = nombre de réplicats (si l'expérimentation est répétée 3 fois, alors *Replicates* = 3);
- **CellPart** = fraction réellement numérisé (si *SubPart* = 10 mL mais que le volume réellement numérisé est de 5 mL, alors *CellPart* = 10mL/5mL = 2).



*Figure 39 : Mesures associées aux différentes étapes de l'échantillonnage.*

## 5.2. Stockage et bancarisation des résultats

Un problème important concerne le stockage et la bancarisation des données issues du système couplé FlowCAM/ZooPhytoImage. Il est envisageable, à court terme, de stocker les données sous un format "Collages/Zidb/Resultats" (sur disque local ou réseau), puis, à plus long terme, de prévoir une stratégie de bancarisation de ces données sous Quadrigé<sup>2</sup>. En effet, les discussions et réflexions menées avec Antoine Huguet du Service "Valorisation de l'Information pour la Gestion Intégrée et la Surveillance" de l'IFREMER Nantes, tendent à montrer que les données FlowCAM/ZooPhytoImage sont des données sécurisées ayant leur place dans Quadrigé<sup>2</sup>.

Toutefois, la gestion des données numériques associées à une image ainsi que le problème de la flexibilité du système avec plusieurs protocoles pour un jeu de données, peuvent rendre *a priori* complexe la bancarisation des données sous Quadrigé<sup>2</sup>. Pour cela, il est impératif, dans un premier temps, de pouvoir estimer la volumétrie. L'intégration des données sous Quadrigé<sup>2</sup> passera donc par la rédaction d'un cahier des charges (avec une collaboration entre les administrateurs Quadrigé<sup>2</sup> et les utilisateurs directs du système couplé FlowCAM/ZooPhytoImage) pour le début de l'année 2014 afin de définir les exigences du système (description du format et de la quantité des données, des besoins d'accès ainsi que des objectifs de ces données). Ce document fera alors partie intégrante du passage en routine.

Des premières réflexions ont été menées depuis début 2014. Voici les conclusions et perspectives en résultant :

- il est nécessaire que les métadonnées respectent le même format quelque soit l'appareil, l'analyste, le mode de prélèvement, le mode d'acquisition, etc.
- il serait intéressant de connaître les stratégies de bancarisation des méthodes similaires telles que la cytométrie en flux, le ZooSCAN, etc.  
→ la bancarisation des données cytométriques dans Quadrigé<sup>2</sup> a déjà été réalisée dans les DOMs. Le référentiel de stockage est donc existant.
- plusieurs points concernant la flexibilité du système peuvent rendre difficile et complexe la phase de bancarisation des données : problème de « machine-dépendance », stockage des résultats obtenus avec plusieurs protocoles pour un même jeu de données, etc.

## Conclusions

Le système couplé FlowCAM/ZooPhytoImage utilisé lors de cette étude, offre plusieurs avantages par rapport au dénombrement des particules au microscope par un opérateur. Le stockage peu encombrant et quasi-illimité des données issues du FlowCAM peut permettre une vérification ou une nouvelle analyse des échantillons. De plus, ce système permet également l'obtention d'un nombre considérable de nouvelles informations (telles que les paramètres de taille, de forme, etc.) qui pourront servir pour des études écologiques plus poussées. Enfin, le temps d'analyse associé à la possibilité de quantification des erreurs liées au dénombrement apparaissent comme des avantages non négligeables des méthodes d'analyse automatisées. Ce facteur de temps offre également l'opportunité de multiplier les analyses et donc d'obtenir un suivi plus régulier et plus précis de la composition phytoplanctonique dans les zones étudiées par le REPHY.

Pour que cet outil opérationnel puisse être finalisé, de nombreuses tâches ont été réalisées : apporter des améliorations technologiques au matériel FlowCAM d'un point de vue mécanique et optique, améliorer et optimiser le logiciel ZooPhytoImage, continuer ou initier l'acquisition d'images, constituer des sets d'apprentissage et des outils de reconnaissance optimisés pour chacune des banques d'images, comparer les résultats obtenus par l'analyse FlowCAM/ZooPhytoImage avec ceux des lectures réalisées au microscope, etc.

Les travaux effectués au sein du Laboratoire Détection, Capteurs et Mesures de l'IFREMER Brest, concernent plusieurs points technologiques importants à prendre en compte pour l'utilisation en (pré)routine du système FlowCAM. Premièrement, l'étude de la vitesse de numérisation en fonction de la concentration cellulaire des échantillons a permis d'accélérer les temps d'analyses au FlowCAM, et particulièrement pour l'assemblage 10X/100 $\mu$ m. Dans un second temps, l'étude de l'homogénéité de l'éclairage dans le plan objet nous a conduit à expérimenter et adopter une nouvelle configuration afin d'améliorer la qualité des images obtenues. La méthode retenue consiste alors à ajouter un dépoli devant la LED et déplacer le collimateur. Enfin, il est important de prendre en considération la répétabilité des réglages. Pour cela, un système de mise au point automatique a été développé et installé sur le FlowCAM. La vis micrométrique originale permettant une mise au point fine manuelle, est remplacée par un actionneur automatique et un coefficient de netteté des particules est calculé afin d'ajuster la mise au point.

Comme présenté dans les différentes sections de ce rapport, différents problèmes liés au système fluide du FlowCAM ont été mis en évidence au Laboratoire Environnement Ressources de Boulogne-sur-Mer. Notamment, un impact important sur la circulation des particules phytoplanctoniques fixées dans le système de flux a été observé. Cependant, compte tenu des contraintes qu'imposeraient des numérisations uniquement à partir d'échantillons vivants (en termes de conservation et de délais de passage au FlowCAM), il semble difficile d'envisager de ne plus fixer les échantillons du REPHY. De plus, différents facteurs en rapport avec la forme, la flottabilité, la capacité d'adhérence et le taux de sédimentation naturel de chaque espèce, peuvent probablement expliquer ce phénomène. Dans ce sens, les expérimentations menées dans les parties 3.2, 3.3 et 3.4 montrent l'importance du sens du débit ainsi que de la cadence

de numérisation pour éviter les problèmes liés à la décantation des particules fixées. En effet, il est probable que, dans un flux descendant, un débit d'eau élevé dans les tuyaux et la chambre de flux limite la décantation et donc la sédimentation des particules fixées. Ces différentes expérimentations ont donc permis de mettre au point un protocole de numérisation efficace et pertinent pour l'analyse d'échantillons lugolés ou vivants.

Pour le moment, la comparaison des comptages obtenues par lecture microscopique et par le système FlowCAM/ZooPhytoImage ne peut être réalisée avec précision que pour les espèces non coloniales ou facilement dénombrables. Pour nos expérimentations, les moyennes des taux de reconnaissance des espèces sont d'environ 69% pour les échantillons lugolés et 67.50% pour les échantillons vivants. Ces indices de performances représentent donc des scores de reconnaissance encourageants. De plus, comme montré dans la section 4.2, l'approche semi-automatisée basée sur la validation partielle des identifications obtenues (grâce au module de correction statistique de l'erreur) devrait permettre une amélioration considérable dans la discrimination des groupes phytoplanctoniques.

## **Perspectives**

L'analyse des performances de reconnaissance obtenues grâce à l'utilisation du module de correction statistique de l'erreur pourrait permettre d'optimiser les méthodes de détection des vignettes les plus suspectes. En effet, l'ajout de critères bio-écologiques tels que la probabilité de présence d'une espèce selon la saison, la zone géographique, etc. pourrait être étudié afin d'améliorer la discrimination des espèces ou groupes phytoplanctoniques.

De plus, il serait pertinent de pouvoir améliorer la qualité visuelle des vignettes pour la constitution des banques d'images et des sets d'apprentissage. La méthode envisagée est basée sur le principe des images RGB, avec R l'image en niveaux de gris, G le masque binaire et B l'image de densité optique en niveaux de gris (avec une évolution linéaire entre le noir et le blanc). La combinaison de ces images devraient permettre d'obtenir un meilleur rendu visuel des particules (augmentation des contrastes des zones claires et des zones sombres) et de ne pas tenir compte d'une mauvaise calibration des niveaux de gris lors de la numérisation des échantillons au FlowCAM.

Bien que les premiers résultats de comparaison entre les abondances obtenues par lecture au microscope et par le système FlowCAM/ZooPhytoImage soient encourageants pour la mise en oeuvre de l'outil opérationnel, des expérimentations complémentaires doivent être menées. En particulier, il serait intéressant de pouvoir suivre la dynamique d'une espèce particulière au cours d'un cycle annuel ou saisonnier. Pour cela, des outils spécifiques permettant de dénombrer automatiquement (ou semi-automatiquement) le nombre de cellules dans les colonies sont actuellement étudiés pour être intégrés au logiciel et ainsi obtenir des résultats comparables à ceux obtenus grâce aux lectures microscope réalisées dans le cadre du REPHY (livrable n°2). En effet, le FlowCAM, comme le cytomètre de flux, ne dispose pas encore de systèmes qui permettent de mener à bien cette tâche (Olson et Sosik, 2007). Les espèces coloniales étant très fréquentes dans les échantillons, cela représente un travail qui doit être mené afin de pouvoir comparer aisément les résultats obtenus en semi-automatique avec ceux des dénombrements REPHY.

Enfin, les tests de la nouvelle interface graphique utilisateur doivent être poursuivies afin de définir les fonctionnalités requises pour l'optimisation de l'utilisation de l'interface graphique de ZooPhytoImage. Ces fonctionnalités concernent les spécifications d'ergonomie du logiciel, le contenu des différents composants graphiques et la logique d'interaction entre eux. L'objectif est donc de s'orienter vers une interaction avec l'analyste (affichage des vignettes, de graphes, etc.).

## Bibliographie

- Belin C., 2012. Cahier de Procédures et de Programmation REPHY 2012-2013. *Document de prescription*, 82p.
- Buskey E., Hyatt C., 2006. Use of the FlowCAM for semi-automated recognition and enumeration of red tide cells (*Karenia brevis*) in natural plankton samples. *Harmful Algae*, vol. 5, 685-692.
- Choi J.W. Stoecker D.K., 1989. Effects of Fixation on Cell Volume of Marine Planktonic Protozoa. *Applied and Environmental Microbiology*, vol. 55, 1761-1765.
- Dereume-Hancart F, 2013. Correction statistique de l'erreur dans le cadre de la classification automatique du plancton. Rapport de projet de fin d'année, Université de Mons, 71 pp.
- Directive 2000/60/ec of the European parliament and of the council of 23 October 2000 establishing a framework for community action in the field of water policy. In *Official Journal of the European Communities EN*, vol. 2000/60/EC.
- Edler L., 1979. Recommendations on methods for Marine Biological Studies in the Baltic Sea. *Phytoplankton and Chlorophyll. Baltic Marine Biologists Publication*, vol. 5, 38p.
- Grosjean P., Denis K, 2010. Rapport Université de Mons - ZooPhytoImage - Optimisation du set d'apprentissage REPHY en 2010.
- Hasle G.R., 1978. The inverted-microscope method. *Chapter 5.2.1 in Phytoplankton Manual.*, ed. United Nations Educational, Scientific and Cultural Organization. 337 pp.
- Ide K., Takahashi K., Kuwata A., Nakamachi M., Saito H., 2008. A rapid analysis of copepod feeding using FlowCAM. *Journal of plankton research*, vol. 30, 275-281.
- Karlson B., Godhe A., Cusack C., Bresnan E., 2011. Introduction to methods for quantitative phytoplankton analysis. *Microscopic and molecular methods for quantitative phytoplankton analysis. IOC Manuals and Guides, no. 55. UNESCO*, 5-12.
- Ledée M, 2013. Modification et amélioration du logiciel Plankton Identifier. Rapport de projet de fin d'étude, IUT Calais-Boulogne, 88 pp.
- Lund J.W.G., Kipling C. and Le Cren E.D., 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiol.*, vol. 11, 143-170.
- McDermott G., Raine R., 2011. Settlement bottle method for quantitative phytoplankton analysis. *Microscopic and molecular methods for quantitative phytoplankton analysis. IOC Manuals and Guides, no. 55. UNESCO*, 21-24.
- 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*, vol. 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*, vol. 5, 195-203.
- Poulton N., Martin J., 2011. Imaging flow cytometry for quantitative phytoplankton analysis - FlowCAM. *Microscopic and molecular methods for quantitative phytoplankton analysis. IOC Manuals and Guides, no. 55. UNESCO*, 47-54.
- Sieracki C.K., Sieracki M.E., Yentsch C.S., 1998. An imaging-in-flow system for automated analysis of marine microplankton. *Mar. Ecol. Prog. Ser.*, 168, 285-296.

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

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

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

## Table des figures

Figure 1 : Stations de prélèvements REPHY en France métropolitaine.....	5
Figure 2 : Schéma fonctionnel du système d'imagerie numérique FlowCAM utilisé au LER de Boulogne-sur-Mer.....	8
Figure 3 : Schéma de fonctionnement général du principe de mise au point automatique.....	12
Figure 4 : Agitateur à hélices pour l'homogénéisation de l'échantillon durant l'analyse.....	13
Figure 5 : Concentration pour l'assemblage 4X/300µm.....	14
Figure 6 : Concentration pour l'assemblage 10X/100µm.....	14
Figure 7 : Volume imagé dans la cellule de flux.....	15
Figure 8 : Distance faible entre deux volumes imagés.....	16
Figure 9 : Distance élevée entre deux volumes imagés.....	16
Figure 10 : Efficacité et temps en fonction de la vitesse de flux pour le couplage 4X/300µm (nombre d'images/sec fixé à 22).....	17
Figure 11 : Efficacité et temps en fonction du nombre d'images/sec pour le couplage 4X/300µm (vitesse de flux fixée à 1.8 mL/min).....	17
Figure 12 : Efficacité et temps en fonction de la vitesse de flux pour le couplage 10X/100µm (nombre d'images/sec fixé à 22).....	18
Figure 13 : Efficacité et temps en fonction du nombre d'images/sec pour le couplage 10X/100µm (vitesse de flux fixée à 0.15 mL/min).....	18
Figure 14 : colmatage de la cellule de flux 100µm.....	19
Figure 15 : Débit ascendant et agitateur magnétique.....	20
Figure 16 : Débit descendant et agitateur à hélice.....	20
Figure 17 : Concentration pour l'assemblage 4X/300µm.....	21
Figure 18 : Concentration pour l'assemblage 10X/100µm.....	21
Figure 19 : Diamètre des particules en fonction du temps pour l'assemblage 4X/300µm et pour un flux descendant et ascendant respectivement.....	21
Figure 20 : Diamètre des particules en fonction du temps pour l'assemblage 10X/100µm et pour un flux descendant et ascendant respectivement.....	22
Figure 21 : Concentrations obtenues pour l'assemblage 4X/300µm (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).....	22
Figure 22: Concentrations obtenues pour l'assemblage 10X/100µm (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).....	22
Figure 23 : Concentrations en particules phytoplanctoniques pour les assemblages 4X/300µm et 10X/100µm (Rouge : flux descendant ; Bleu : flux ascendant ; Vert : flux ascendant + 1% Glycérol ; Jaune : flux ascendant + 10% Glycérol).....	23
Figure 24 : Concentration pour l'assemblage 4X/300µm selon la dilution appliquée (Rouge : 1er passage ; Bleu : second passage ; Vert : troisième passage).....	24



Figure 25 : Concentration pour l'assemblage 10X/100µm selon la dilution appliquée (Rouge : 1er passage ; Bleu : second passage ; Vert : troisième passage).....	24
Figure 26 : Evolution de la concentration en fonction du temps pour l'assemblage 4X/300µm.....	25
Figure 27 : Onglet "Fluidics" du menu "Context" de Visual SpreadSheet.....	26
Figure 28 : Comparaison de 9 comptages manuels sur les collages du FlowCAM selon si les échantillons sont fixés ou vivants.....	28
Figure 29 : Comparaison des comptages manuels sur les collages du FlowCAM (4X/300µm) et des lectures au microscope inversé réalisées dans le cadre du REPHY.....	30
Figure 30 : Comparaison des comptages manuels sur les collages du FlowCAM (4X/300µm) et des lectures au microscope inversé réalisées dans le cadre du REPHY pour chacun des taxons et pour chaque point de prélèvement.....	31
Figure 31 : Concentration maximale d' <i>Alexandrium</i> par site (1992-2001).....	33
Figure 32 : Concentration maximale de <i>Dinophysis</i> par site (1992-2001).....	34
Figure 33 : Concentration maximale de <i>Pseudo-Nitzschia</i> par site (1992-2001).....	34
Figure 34 : Nombre de vignettes comptabilisées automatiquement pour chaque taxon phytoplanctonique étudié, pour les échantillons ME1 et ME2 (lugolés et vivants) avec l'assemblage 4X/300µm.....	37
Figure 35 : Comparaison des comptages obtenus par l'outil FlowCAM/ZooPhytoImage (4X/300µm) et des lectures au microscope inversé réalisées dans le cadre du REPHY pour chacun des taxons étudiés.....	39
Figure 36 : Principe de fonctionnement du module de correction statistique de l'erreur, basé sur la validation manuelle des vignettes les plus "suspectes".....	40
Figure 37 : Evolution des scores de dissimilarité entre l'identification automatique et la "vérité-terrain", en fonction de la proportion de vignettes validées manuellement.....	41
Figure 38 : Nombre de vignettes comptabilisées manuellement, automatiquement et après correction de l'erreur pour chaque taxon phytoplanctonique étudié, pour les échantillons ME1 et ME2 lugolés avec l'assemblage 4X/300µm.....	43
Figure 39 : Mesures associées aux différentes étapes de l'échantillonnage.....	45

## Liste des tableaux

Tableau 1 : Temps de numérisation pour une fréquence d'acquisition de 22 i/s.....	9
Tableau 2 : Temps de numérisation de 2000 particules en fonction des concentrations cellulaires pour les grossissements 4X et 10X.....	10
Tableau 3 : Assemblages théoriques recommandés par Fluid Imaging Technologie selon les différentes classes de taille.....	13
Tableau 4 : Assemblages recommandés selon les différentes classes de taille.....	23
Tableau 5 : Temps de sédimentation des particules en fonction de la hauteur du flacon.....	25
Tableau 6 : Volume de surnageant et volume final selon la concentration désirée.....	25
Tableau 7 : Concentrations obtenues pour les expérimentations menées sur la comparaison de la numérisation d'échantillons lugolés et vivants, pour l'assemblage 4X/300µm (comptage microscope / comptage manuel sur les collages du FlowCAM).....	28
Tableau 8 : Nombre de vignettes comptabilisées manuellement et automatiquement pour l'échantillon ME1 du 17/12/2013 (4X/300µm).....	36
Tableau 9 : Nombre de vignettes comptabilisées manuellement et automatiquement pour l'échantillon ME2 du 17/12/2013 (4X/300µm).....	36
Tableau 10 : Comparaison des concentrations obtenues (en cellules/L) par lecture microscopique et par le système FlowCAM/ZooPhytoImage (4X/300µm) pour l'échantillon ME1.....	38
Tableau 11 : Comparaison des concentrations obtenues (en cellules/L) par lecture microscopique et par le système FlowCAM/ZooPhytoImage (4X/300µm) pour l'échantillon ME2.....	38
Tableau 12 : Comparaison du nombre de vignettes comptées manuellement, automatiquement, et après correction de l'erreur pour les échantillons lugolés ME1 et ME2, pour l'assemblage 4X/300µm.....	42

# ANNEXES

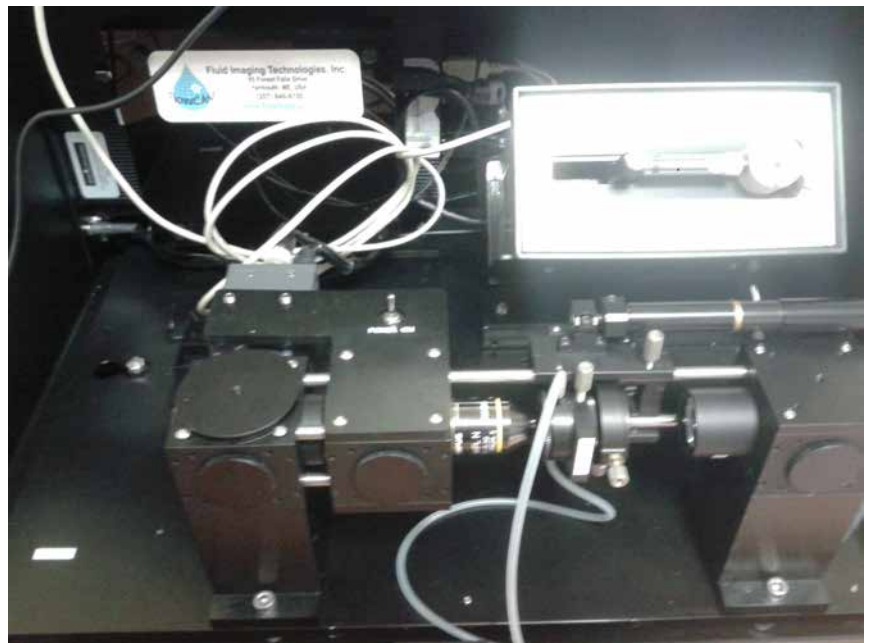
## **Annexe 1 - Mise au point automatique de la cuve**

Auteur : Bertrand FOREST  
Morgan TARDIVEL  
Marie-pierre CRASSOUS  
Florent COLAS

....  
Juillet 2013 – RDT/MS – RDT/LDCM

## Mise au point automatique de la cuve du FlowCam

Mode opératoire du logiciel de mise au point  
automatique



# Sommaire

<b>1. Composition du système .....</b>	<b>3</b>
<b>2. Le logiciel Dpl_Actuateur .....</b>	<b>3</b>
2.1. Lancement du programme .....	4
2.2. Déplacement manuel du vérin .....	5
2.3. Sélection de la zone de travail.....	5
2.4. Premier scan .....	5
2.5. Deuxième scan .....	5
<b>3. Protocole de l'autofocus.....</b>	<b>6</b>



# 1. Composition du Flowcam

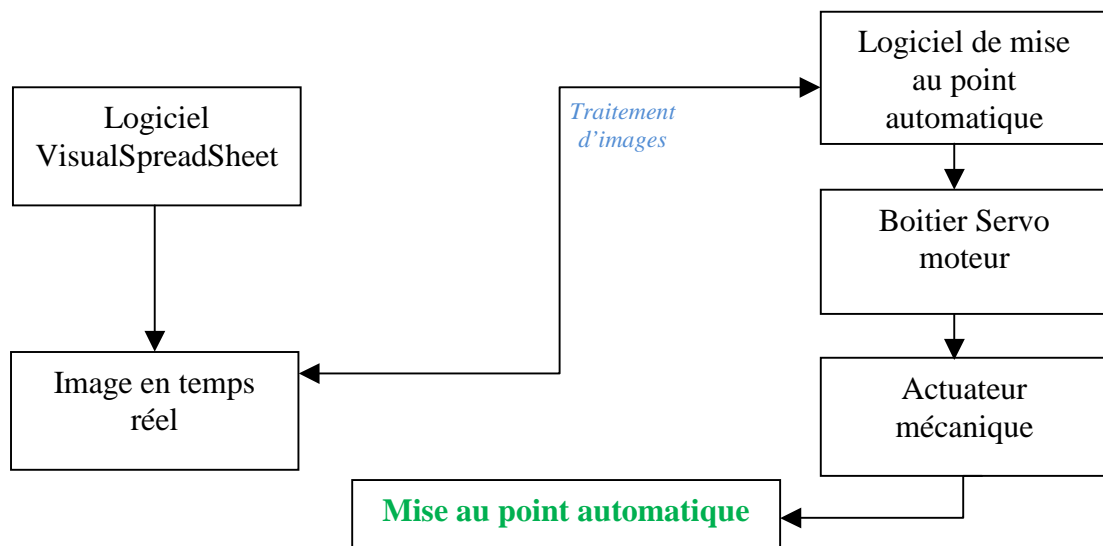
Le FlowCam se compose désormais :

- D'un boîtier de commande « Apt - Dc servo controler » composé de deux boutons qui permet de faire avancer ou reculer la vis manuellement.
- Un vérin à vis THORlabs modèle Z812B qui déplace la platine sur laquelle les objets sont à étudier. Sa plage de travail est de 12 millimètres, d'une résolution inférieure au dixième de micron.



# 2 Le logiciel de mise au point

Le logiciel permet de déplacer la platine, d'effectuer la capture de la fenêtre vidéo de VisualSpreadSheet, d'analyser l'image et d'en calculer un coefficient de netteté.



## 2.1. Lancement du programme

Avant de lancer Dpl\_Actuateur, il est nécessaire de lancer VisualSpreadSheet, et d'avoir une image dans la fenêtre « Setup And Focus Mode » (menu Setup/Setup and Focus).

Il est nécessaire que cette fenêtre reste ouverte. Ne pas la mettre sous forme d'icône.

On peut utiliser les deux tailles d'images (Z+, Z-)

Lancer le logiciel « Dpl\_Actuateur », il faut patienter quelques secondes.

Si la fenêtre vidéo de VisualSpreadSheet n'a pas été lancée correctement, un message d'avertissement s'affichera en bas de l'écran.

De même, si le contrôleur THORlabs n'est pas initialisé correctement, une fenêtre popup d'ouvrira, et le programme se terminera. Remarque : il peut arriver qu'il y ait un problème après l'allumage du PC. Il faut débrancher et rebrancher l'alimentation du contrôleur et relancer le logiciel.

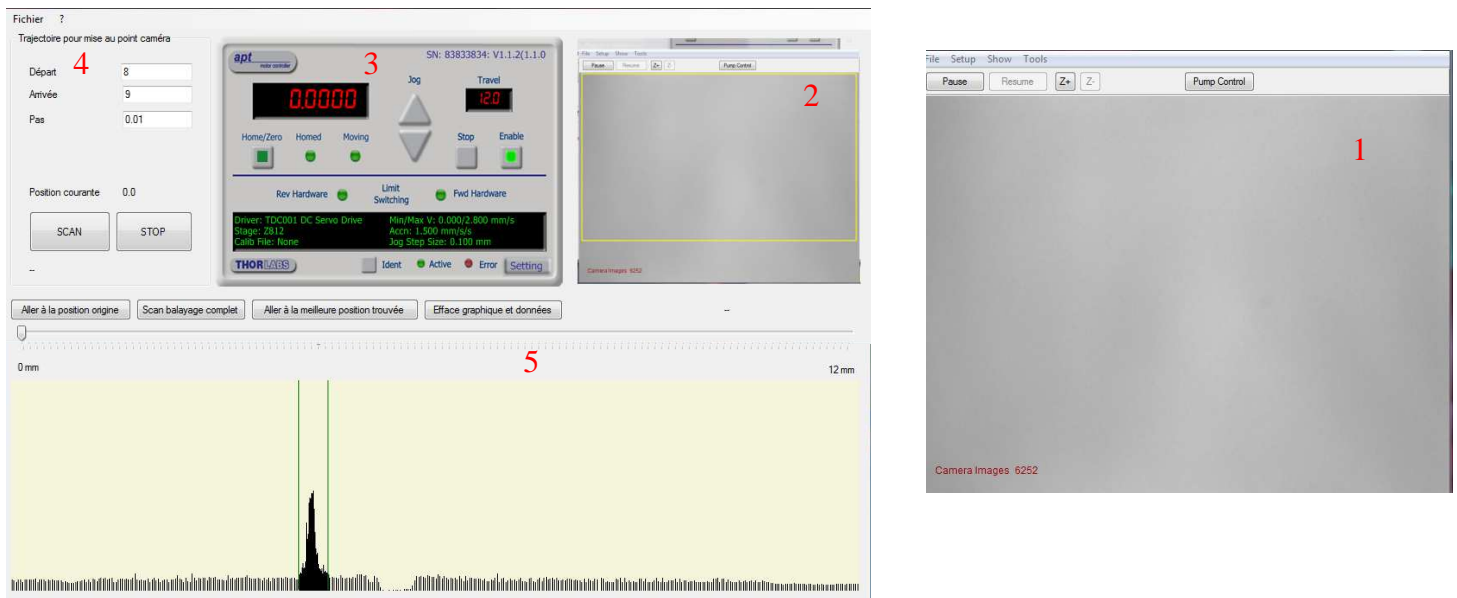


Figure 1 : interface utilisateur

1. Fenêtre “Set and Focus Mode “de VisualSpreadsheet.
2. En haut à droite est recopié le contenu de la fenêtre « Set and Focus Mode » de VisualSpreadsheet.
3. On reconnaît le panneau de contrôle de thorlabs utilisé dans le logiciel APT User. Il est possible de manœuvrer le vérin par ce contrôle.

Remarque : le numéro de série du contrôleur est nécessaire pour y avoir accès. Celui-ci est codé « en dur » dans le logiciel. Cela signifie que ce logiciel ne fonctionnera pas avec un vérin du même modèle avec un numéro de série différent.



4. Fenêtre permettant de gérer l'intervalle d'analyse. L'utilisateur choisit la position de départ, de fin ainsi que le pas.
5. Curseur de sélection + zone de résultat du balayage

## 2.2. Déplacement manuel de l'actuateur

On peut déplacer l'actuateur soit avec le contrôle « Jog » ou en cliquant dans l'afficheur de position (en rouge) puis saisir une valeur dans le contrôle APT ou utiliser le grand curseur situé entre la rangée des quatre boutons et la zone graphique du bas (cf.figure 1)

## 2.3. Sélection de la zone de travail

Un grand cadre jaune représentant la zone d'image de travail est tracé dans la fenêtre « vidéo » (copie de VisualSpreadsheet)

Par défaut, il est aussi grand que possible en évitant certaines zones (menus, zone d'affichage de texte en rouge...)

En déplaçant le curseur de la souris dans cette zone, un zoom est affiché en bas de la fenêtre. On peut sélectionner une zone particulière en traçant un « élastique ».

## 2.4. Premier scan

Ce premier SCAN n'est nécessaire que lorsque l'intervalle de mise au point n'est pas connu.

Appuyer sur le bouton « SCAN balayage complet ». L'actuateur va se déplacer de 0 à 12 mm pendant 100 pas et effectuera ses mesures de netteté. Si la hauteur de la zone de travail le permet, 16 lignes de pixels horizontales sont examinées. La moyenne des 16 pentes des intensités les plus raides est conservée pour chaque arrêt de l'actuateur.

A partir d'une certaine position, un pic significatif apparaît sur le graphique, et l'image est a priori relativement nette.

On peut stopper le scan (bouton STOP) une fois le pic dépassé.

## 2.5. Deuxième scan

1 - Restreindre la zone de mise au point :

On peut cliquer sur le bouton gauche de la souris au commencement du pic et sur le bouton droit de la souris à la fin du pic (deux barres vertes verticales sont affichées). Les deux positions correspondantes sont saisies dans les champs de saisie « Départ » et « Arrivée »

2 - Appuyer sur le bouton « Efface graphique et données »

3 - Appuyer sur le bouton SCAN.

Le deuxième scan, lui aussi de 100 pas (si on n'a pas modifié le champ « pas »), s'effectue. Le pic est plus précis.

Une fois la hauteur maximale du pic franchie, on peut stopper le scan et appuyer sur le bouton « **Aller à la meilleure position** » pour que le vérin se déplace à la position correspondant à l'image la plus nette.

### 3. Réglage de la mise au point pour le phytoplancton.

Lancer VisualSpreadsheet situé sur le bureau



Lancer l'acquisition en Temps réel : Menu **Setup** → **Setup and Focus**



Figure 2 : fenêtre setup and focus

Vérifier la valeur moyenne des niveaux de gris (**Setup and Focus** → **tools** → **raw pixel intensity**) ; elle doit avoisiner les 180 en intensité, sinon modifier les paramètres liés à l'illumination (flash duration et flash delay).

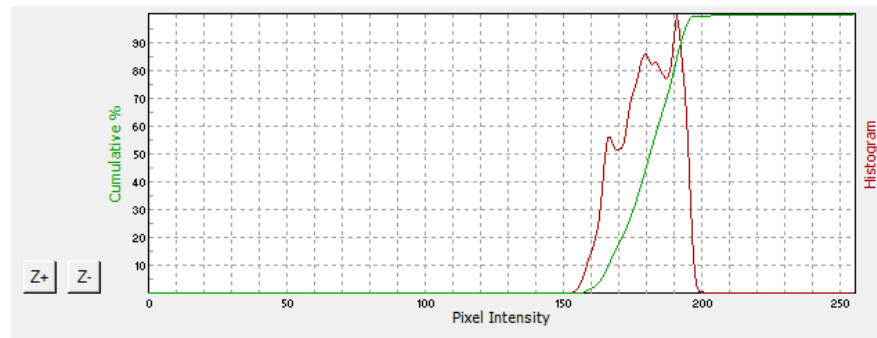
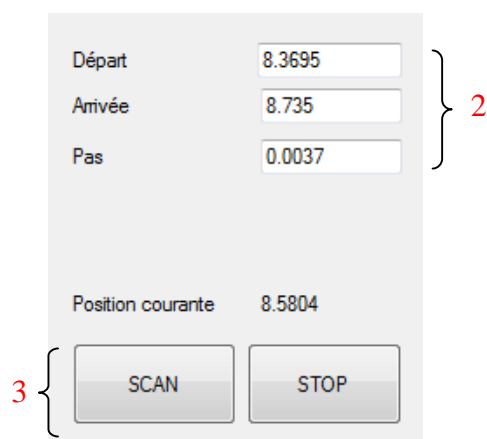
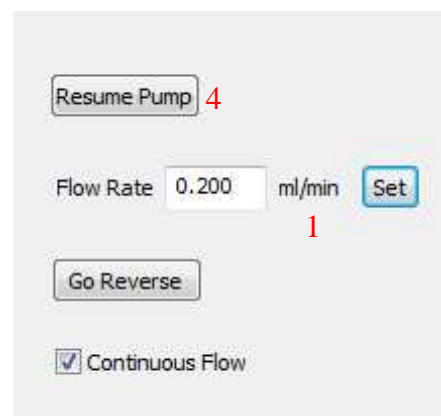


Figure 3 : Histogramme des niveaux de gris

- Placer la cuve du Flowcam (100 $\mu$ m  $\rightarrow$  X10 et 300  $\mu$ m  $\rightarrow$  X4).
- Vérifiez que la cuve soit totalement **propre** (nettoyage à l'éthanol)
- La fenêtre « setup and focus » doit rester ouverte.
- Lancer le logiciel qui permet de déplacer l'actuateur mécanique (voir chapitre 2 pour l'utilisation du programme)
- Vérifier que la zone de travail soit la plus grande possible. Pour que la mise au point soit la meilleure possible. Si des salissures persistent, définissez la zone de travail (voir section 2.3 §3)
- Préparer un échantillon concentré de phytoplancton
- Ouvrir la fenêtre permettant de gérer le débit de la pompe
- Fixer le débit à 0.2ml/min pour X10 et 1.8ml/min au X4 (Menu **Setup  $\rightarrow$  Setup and Focus  $\rightarrow$  pump control**)
- Passer l'échantillon de phytoplancton
- Choisir l'intervalle sur lequel doit être effectué le déplacement de l'actuateur
- Appuyer sur le bouton **SCAN**
- Mettre en marche la pompe en cliquant sur « **resume pompe** ».



Interface logiciel de déplacement de l'actuateur



Interface de contrôle de la pompe dans VisualSpreadsheet

Une fois le **SCAN** lancé, une série de pics apparaît dans la partie inférieure de la fenêtre. A partir d'une certaine position, un pic significatif apparaît sur le graphique



*Figure 4 : Pic représentatif de l'image la plus nette*

Appuyer sur le bouton « **Aller à la meilleure position** » pour que le vérin se déplace à la position correspondant à l'image la plus nette. La mise au point automatique est ainsi faite.



*Figure 5 : Mise au point effectuée sur des cellules de phytoplancton.*

## Annexe 2 - Préparation de l'échantillon

### **Prélèvement d'eau**

Les prélèvements d'eau destinés aux observations du phytoplancton sont effectués en sub-surface (-0,5 a -1 m) à partir d'un bateau et à l'aide d'une perche équipée d'un flacon amovible de 1L ou d'une bouteille Niskin. L'eau est ensuite transvasée directement dans un flacon plastique opaque brun de 1L.

### **Echantillons fixés**

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

- a. Enregistrer l'échantillon dans le registre (cf. modèle page 56),
- b. Homogénéiser l'échantillon de 1L par 30 retournements doux (des retournements violents peuvent produire des bulles difficiles à éliminer),
- c. Remplir rapidement un flacon en verre brun de 250 mL,
- d. Ajouter 2,5 mL de Lugol (concentration finale 1 %).
- e. Identifier le flacon et le stocker au réfrigérateur.

La fixation des échantillons suit le protocole établi pour le REPHY, soit l'utilisation de Lugol acide pour la conservation des échantillons.

*Solution de Lugol utilisée :*

*100g KI + 50 g I<sub>2</sub> (iode bisublimé) + 100 mL acide acétique glaciale qsp 1000 mL à l'eau distillée.*

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** (Belin, 2012).

### **Echantillons vivants**

Dès l'arrivée des échantillons destinés à l'analyse d'une flore totale, identifier le flacon et le stocker à l'obscurité au réfrigérateur.

Les échantillons vivants ne peuvent pas être conservés sur le long terme. Pour l'usage du FlowCAM, il est nécessaire de passer l'échantillon **dans les 48 heures suivant le prélèvement**, pour éviter le broutage des particules de phytoplancton par le zooplancton ou la détérioration naturelle des cellules (Belin, 2012).

Numéro Enregistrement	Point ou Réf. du Labo fournisseur	Date	Echantillon accompagné de la Fiche Résultat	Observations	Visa Agent
2013 /			Oui / Non		
2013 /			Oui / Non		
2013 /			Oui / Non		
2013 /			Oui / Non		
2013 /			Oui / Non		

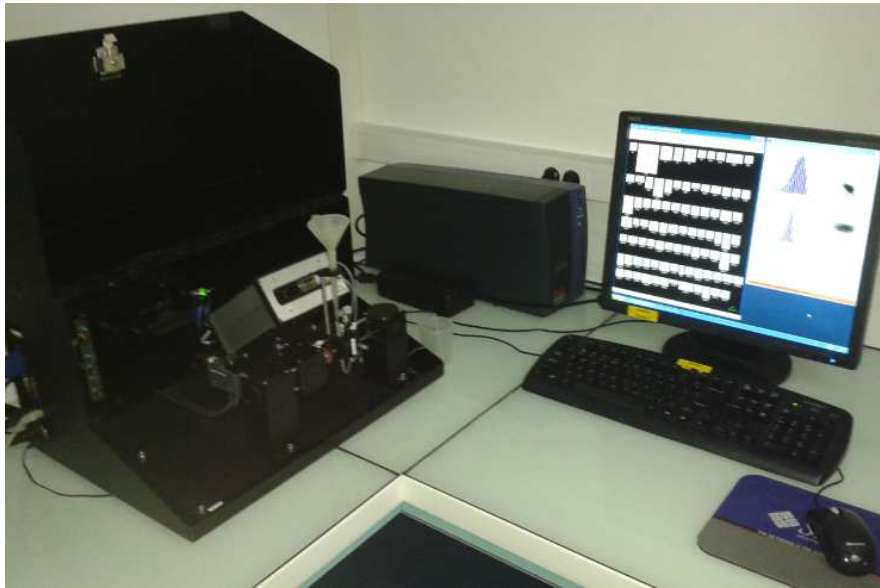
## Annexe 3 - Numérisation de l'échantillon au FlowCAM

### *Enregistrement de l'analyse*

Pour chaque expérimentation menée, il est important de laisser une trace dans le registre dédié. Un modèle pour le cahier d'enregistrement est fourni en page 58.

### *Enregistrement de l'analyse*

Cf. Préparation de l'échantillon - Annexe 2.



### *Assemblage Objectif / cellule*

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

### *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.

Le système optique du FlowCAM doit être le plus propre possible. Ne jamais toucher directement la cellule de flux, l'objectif et le collimateur afin d'éviter toutes traces de doigt.

Numéro Enregistrement	Date	Dilution Concentration	Objectif Nb de passage	Observations, nom du fichier conservé	Visa Agent
2013 /			10X  4X		
2013 /			10X  4X		
2013 /			10X  4X		

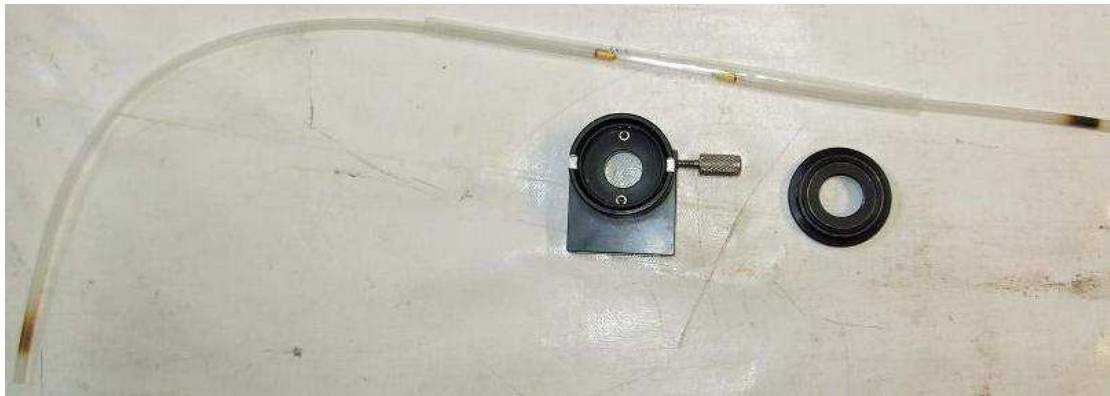


## Montage de la cellule de flux

Pour les cellules de flux calibrées de 300 $\mu$ m et 100 $\mu$ m, 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.



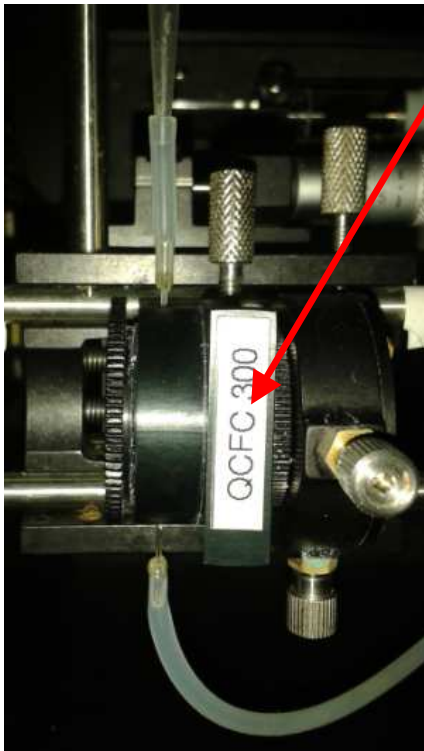
Pour la cellule de 300  $\mu$ m, l'anneau noir situé à l'intérieur du porte-cellule (vissé dans la bague inférieure) doit être vissé de manière à ne pas dépasser de l'encoche. Au contraire, pour la cellule de 100  $\mu$ m, il faut dévisser cet anneau pour qu'il dépasse de l'encoche d'environ 0,5 mm. 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.



Remarque : Le FlowCAM du LER/BL possédant une pompe seringue, les longueurs des tubes en silicone doivent être indiquées précisément pour connaître le volume de liquide qui est aspiré. Ces longueurs sont de 28cm pour le tube inférieur et 6cm pour le tube supérieur.

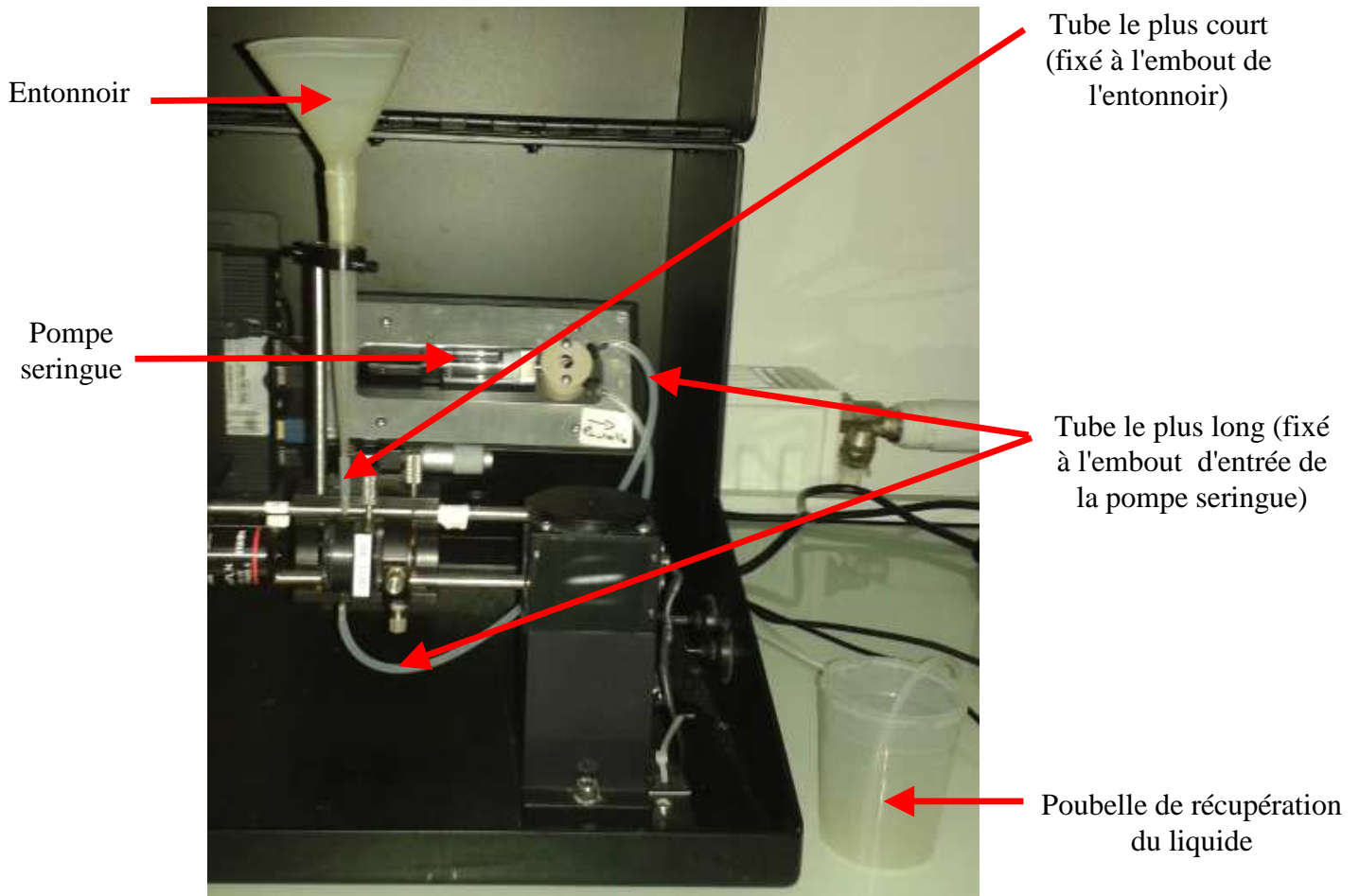


L'ensemble porte-cellule + cellule de flux est ensuite inséré sur le bloc central, la vis et le tube court dirigés vers le haut. Pour cette étape, procéder comme suit :

- insérer horizontalement le porte-cellule sur son pivot et serrer la vis pour immobiliser l'ensemble.
- fixer l'extrémité du tube le plus long sur l'embout d'entrée de la pompe seringue.
- fixer l'extrémité du tube le plus court à l'embout de l'entonnoir.
- mettre le pot de récupération du liquide en sortie

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.



## Réglages du FlowCAM

### - Démarrage du FlowCAM

Allumer le FlowCAM à l'aide du commutateur situé à gauche du bloc central.

Allumer l'écran.



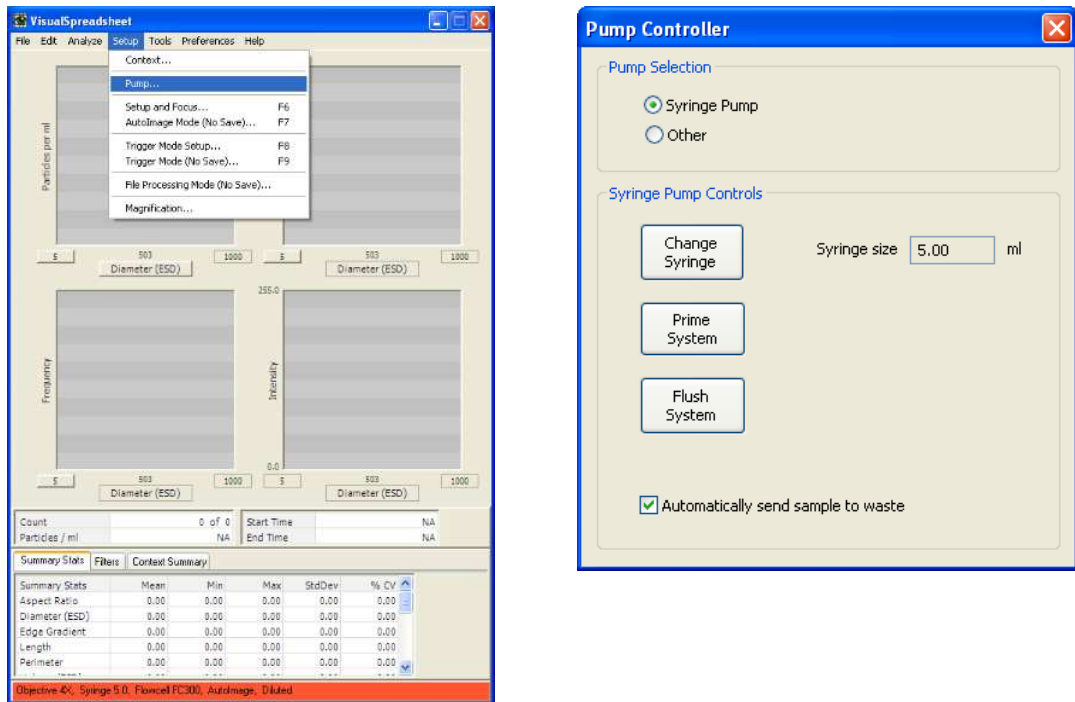
Lancer le programme VisualSpreadsheet.



### - Introduction de l'échantillon

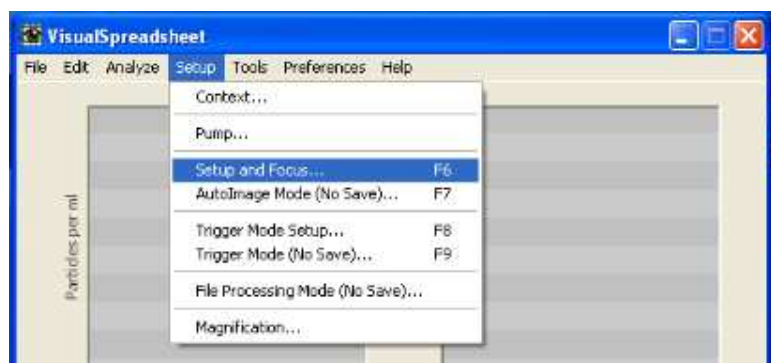
Le FlowCAM du LER/BL est équipé d'une pompe seringue. Les spécificités de cette pompe sont détaillées en Annexe 4.

Remplir l'entonnoir avec l'échantillon (vivant ou 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 placer la cellule de flux dans le champ de vision.

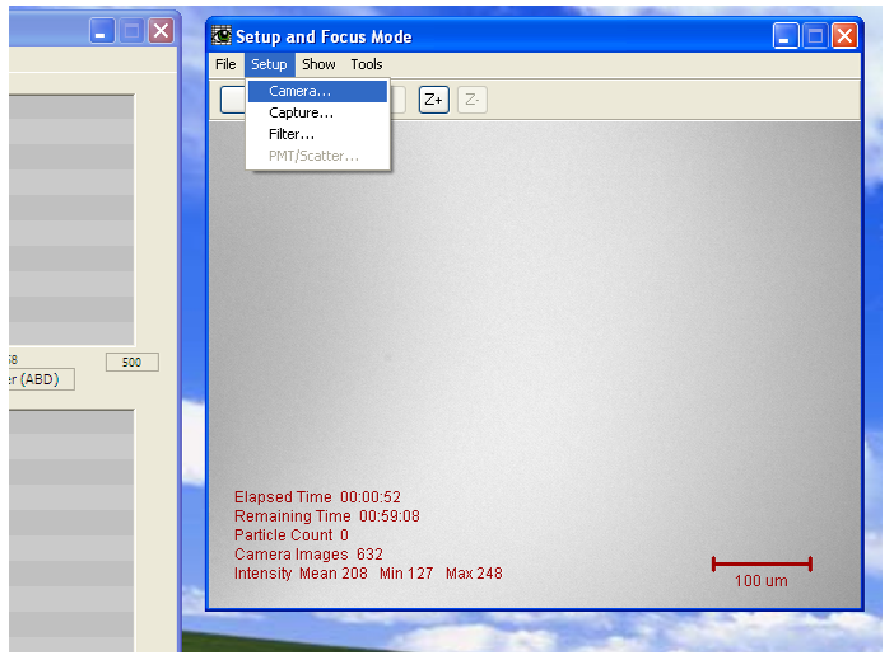
A l'aide de la vis de déplacement latéral, cibler les bords de la cellule de flux et placer la distance focale entre eux (centrer la vue).

Si des imperfections sont visibles, ne pas hésiter à changer la position de la cellule de flux.

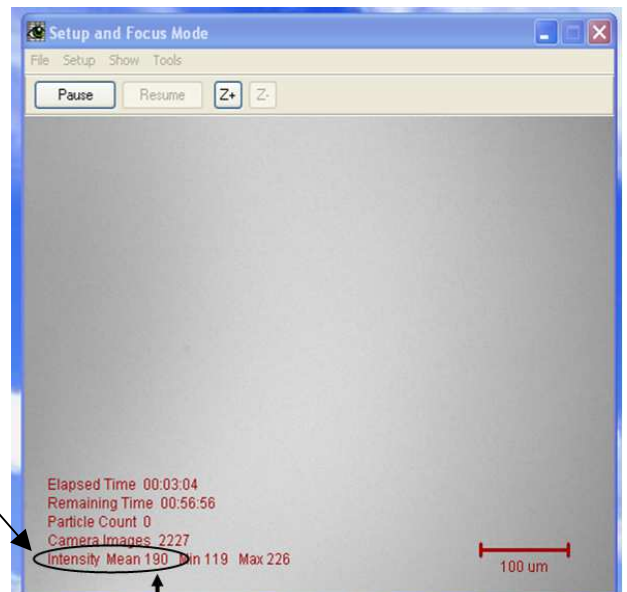
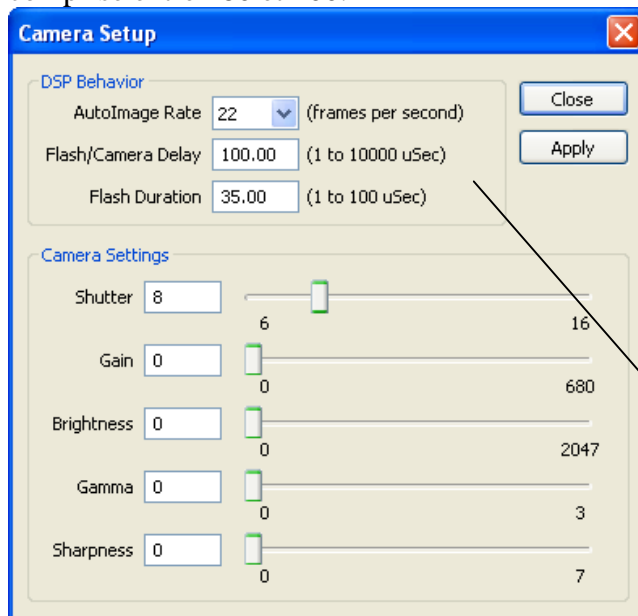


- Réglage de l'illumination

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

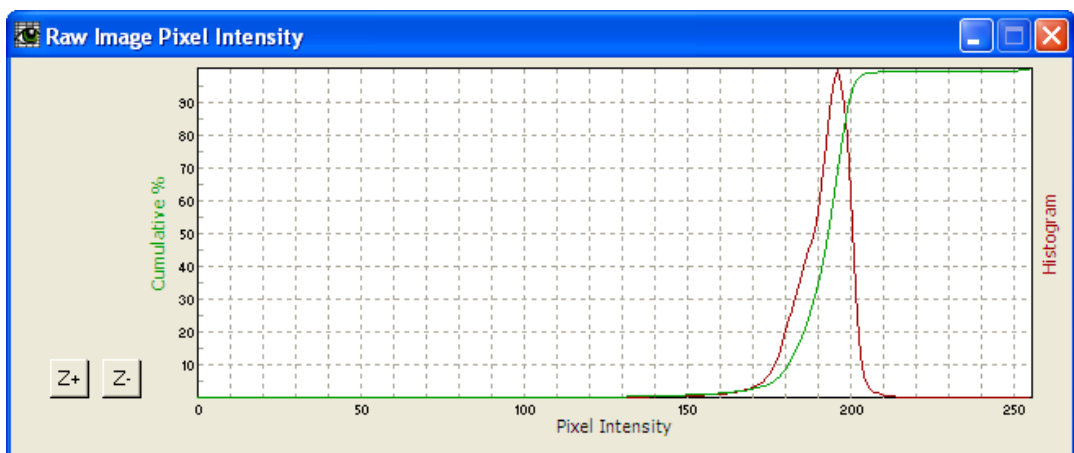
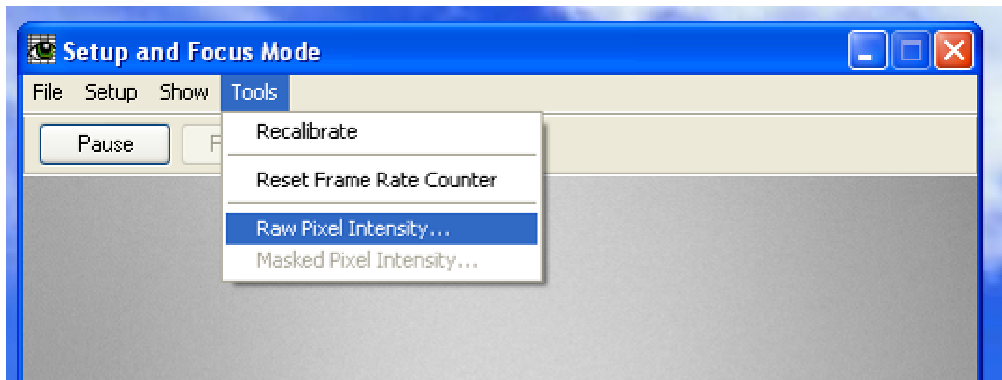


Modifier les paramètres « Flash duration » et « Gain » pour obtenir une intensité moyenne comprise entre 180 et 200.



L'objectif est de minimiser la durée du flash ainsi que le gain, tout en conservant une intensité moyenne comprise entre 180 et 200 (cf. onglet "Camera" du menu "Context").

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



La courbe rouge représente l'histogramme des niveaux de gris dans le champ de vision. Tout ce qui se trouve à gauche de cette courbe correspond aux niveaux de gris disponible pour les particules. Tout ce qui se trouve à droite représente les niveaux de gris de l'image de fond.

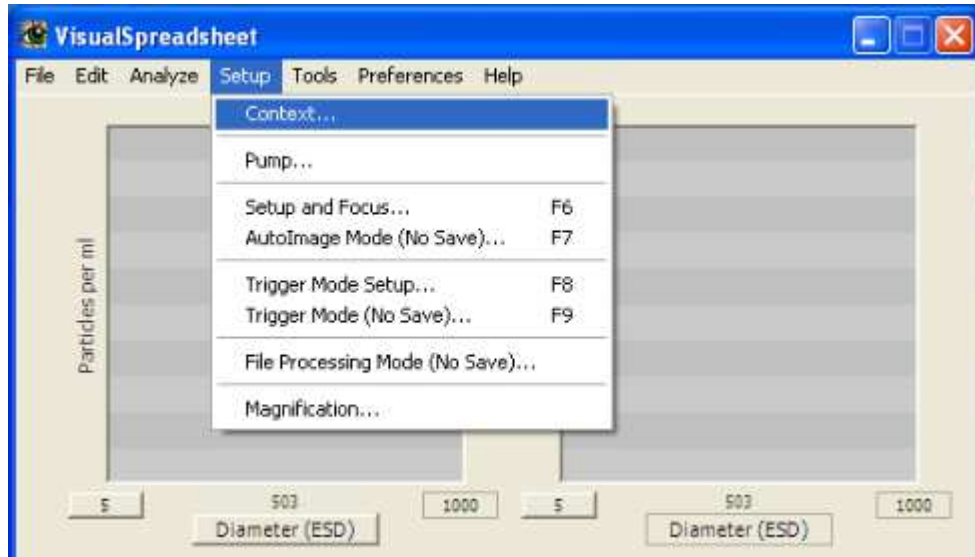
Vérifier régulièrement grâce à cette option que la courbe des niveaux de gris présente une forme correcte (pic le plus étroit possible, situé entre 180 et 200).

Avec le 4x et la cellule de 300  $\mu\text{m}$ , 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.

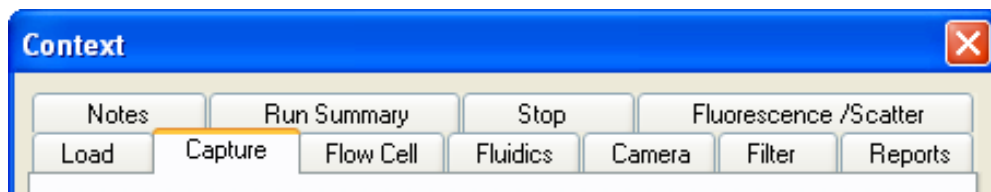
## Réglages des paramètres d'acquisition

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

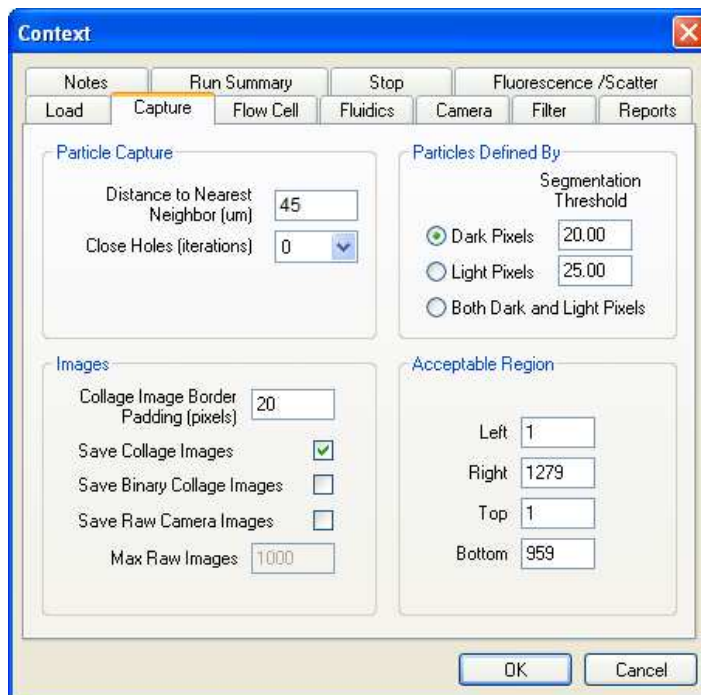
Les paramètres d'acquisition sont exportés à la fin de chaque exécution dans un fichier contexte (.ctx).



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



## - Onglet « Capture »

**Max Raw Images = No**

Nombre maximum d'images brutes à sauvegarder durant l'exécution.

**Segmentation Threshold – Dark Pixel = 20**

Nombre minimum de niveaux de gris plus foncés que le fond pour détecter une particule. Utiliser uniquement les pixels sombres car les particules d'intérêt sont plus sombres que le fond d'image.

**Acceptable Region - L=1, R=1279, T=1, B=959**

Ces valeurs déterminent la région d'intérêt dans le champ de vue total. Toutes les particules touchant les contours rouges ne sont pas comptées par le logiciel. Cette option permet donc de ne numériser que les particules complètes.

**Distance to Nearest Neighbor = 45**

Distance au plus proche voisin ; seuil minimal en  $\mu\text{m}$  entre deux particules distinctes. En dessous de cette distance, deux particules sont détectées comme une. Cette valeur semble être un bon compromis pour détecter à la fois les cellules seules et les cellules en colonies.

**Close Holes = 0**

Ajout de pixels autour de la particule afin de compléter le périmètre. Cette option peut engendrer la fusion de particules proches.

**Collage Image Border Padding = 20**

Nombre de pixels détournés autour de la particule. Cette valeur est importante pour faciliter la reconnaissance visuelle et pour incruster une barre d'échelle.

**Save Collage Images = Yes**

Sauvegarde des collages (consommant moins d'espace mémoire que les images brutes).

**Save Binary Collage Images = No**

Sauvegarde des images utilisées pour la segmentation (pas nécessaires).

**Save Raw Camera Images = No**

Sauvegarde des images brutes (consommant beaucoup d'espace mémoire et non utilisées).



- Onglet « FlowCell »

- Onglet « Fluidics »

**Sample Dilution = Yes - Ratio = 1.00000**

L'échantillon est dilué ou concentré. Par exemple, si l'échantillon est dilué à 50% de la concentration originale, cette valeur doit être réglée à 0.5 pour que le logiciel puisse calculer des statistiques sur la concentration originale. Généralement, pour le mode AutoImage, aucune dilution n'est appliquée.

**Flow cell**

Choisir le type de cellule de flux selon l'objectif utilisé. La profondeur et la largeur sont alors automatiquement paramétrées.

Flow Cell Type	Depth (µm)	Width (µm)	FOV	Tubing Inner Diameter (cm)	Matching Objective(s)
FC 50	50	1000		0.16	20X or 10X
FC80FV	80	570	YES	0.16	10X
FC80-7FV	80	700	YES	0.16	10X
FC90FV	90	570	YES	0.16	10X
FC100	100	2000		0.16	10X or 20X
FC200	200	2000		0.16	10X or 4X
FC300	300	3000		0.16	4X
FC300FV	300	1500	YES	0.16	4X
FC600	600	6000		0.40	4X or 2X
FC700	700	7000		0.40	4X or 2X
FC800	800	8000		0.50	4X or 2X

**Tubing**

Longueurs des tubes au-dessus et en-dessous de la cellule de flux. Ces longueurs sont primordiales pour obtenir des mesures précises de fluide (mode "Machine Prime" dans l'onglet "Fluidics").

**Settings**

Volume à analyser, vitesse de flux (**1.8 ml/min pour le 4X; 0.150 ml/min pour le 10X**), et nombre d'images par seconde de la caméra (**22 frames/sec**). Ces paramètres doivent être optimisés de façon à obtenir une efficacité pertinente et un temps d'analyse minimal.

**Estimated Efficiency and Run Time**

L'efficacité et le temps d'exécution sont automatiquement calculés, de manière à aiguiller l'utilisateur dans son choix.

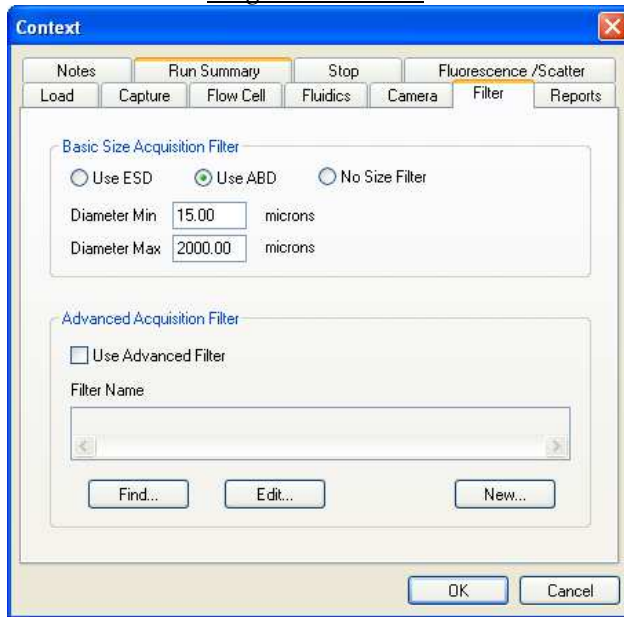
Si les réglages sont en dehors des spécifications opérationnelles recommandées pour le FlowCAM :

- **Recommended Efficiency Exceeded:** possibilité de 'double-image' des particules et biais sur les calculs de concentrations
- **Calibration to sample ratio too high:** la combinaison volume, vitesse de flux et nombre d'images/sec engendre l'utilisation de plus de 50% du volume initial pour le calcul du fond d'image (calibration).

**Priming = Manual prime with sample**

L'échantillon est conduit manuellement de l'entonnoir à la cellule de flux (avec la pompe seringue). Cette option est préférée à l'option automatique (Machine prime) car de l'air peut être piégé entre le liquide et l'entrée de la cellule de flux.

## - Onglet « Filter »

**Basic Size Acquisition Filter = Use ABD**

Le diamètre basé sur l'aire est le diamètre d'un cercle qui aurait la même surface que la particule d'intérêt. Le diamètre sphérique équivalent n'est pas utilisé car les images acquises sont bi-dimensionnelles.

**Diameter Min = 15.00**

Les particules ayant une ABD plus faible que cette valeur (**entre 15 et 20µm pour le 4X; entre 5 et 15µm pour le 10X**) ne sont pas conservées dans le collage.

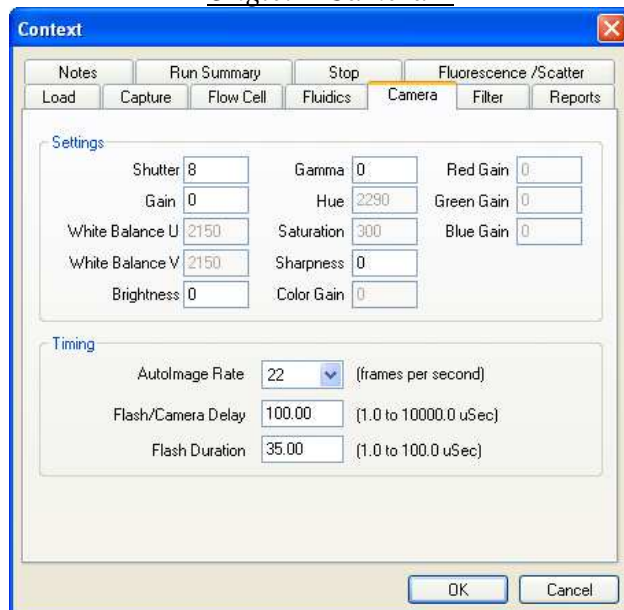
**Diameter Max = 2000.00**

Les particules ayant une ABD supérieure à cette valeur ne sont pas conservées dans le collage.

**Advanced Acquisition Filter = No**

Appliquer uniquement un filtre de taille.

## - Onglet « Camera »

**Settings**

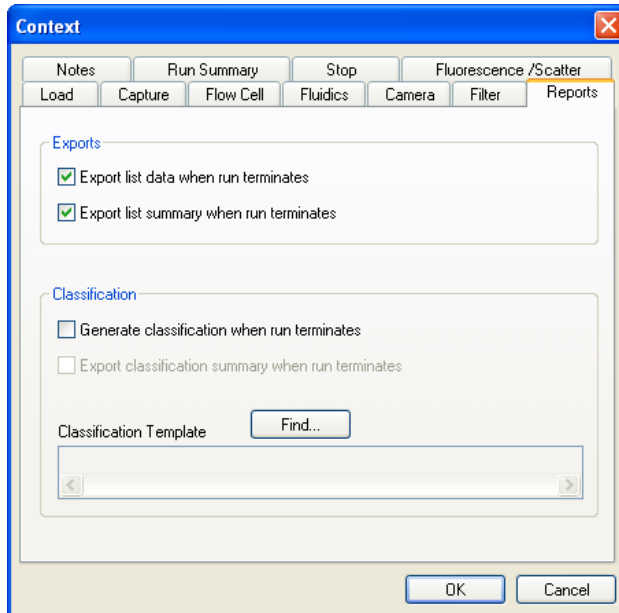
- **Shutter = 8.** Durée d'ouverture de l'obturateur
- **Gain = 0.** Valeur entrée lors de l'ajustement de l'illumination. Cette valeur doit être adaptée à l'objectif et à la cellule de flux utilisés.
- **WhiteBalance U-V = Nothing.** Option utilisée pour le FlowCAM couleur.
- **Brightness = 0.** Luminosité appliquée à l'image.
- **Gamma = 0.** Option utilisée pour le FlowCAM couleur.
- **Hue = Nothing.** Option utilisée pour le FlowCAM couleur.
- **Saturation = Nothing.** Option utilisée pour le FlowCAM couleur.
- **Sharpness = 0.** Ajustement de la netteté de l'image. Cette option augmente le contraste des particules par rapport au fond d'image.
- **Color (Red-Green-Blue) Gain = Nothing.** Option utilisée pour le FlowCAM couleur.

**Timing**

- **AutoImage Rate = 22.** Nombre d'images brutes acquises par seconde. Cette valeur peut-être maximisée selon les performances de l'ordinateur.
- **Flash/Camera Delay = 100.00.** Durée entre le flash et la segmentation de l'image.
- **Flash Duration = 35.00.** Durée du flash en msec.

L'objectif est de minimiser la durée du flash ainsi que le gain, tout en conservant une intensité moyenne comprise entre 180 et 200. Ceci peut être vérifié en utilisant le menu "Setup and Focus".

- Onglet « Reports »



Exports

- **Export list data when run terminates = Yes.** Exporter automatiquement toutes les données mesurées à la fin de l'analyse.
- **Export list summary when run terminates = Yes.** Exporter automatiquement un résumé de la numérisation à la fin de l'analyse.

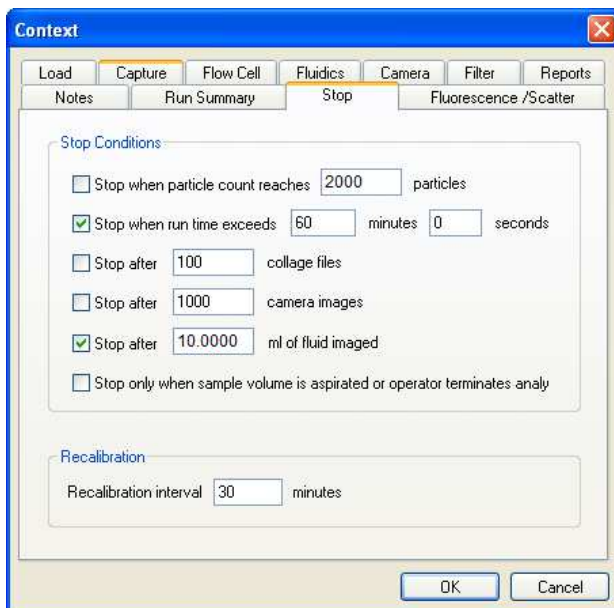
Classification Template = Nothing

Classification des particules par Visual SpreadSheet. Dans le cadre du projet, le logiciel de classification utilisé est Zoo/ZooPhytoImage.

- Onglet « Stop »

Définir le mode d'arrêt de la numérisation :

- pour le couplage **4X/300um**, il est possible de numériser **10 mL** d'échantillon.
- pour le couplage **10X/100um**, l'objectif à atteindre est la numérisation de **2000 particules en moins de 30 minutes**. (cf. section 3.5 - "Dilution/Concentration de l'échantillon").



Stop Conditions

Définir un critère d'arrêt selon l'objectif de l'étude.

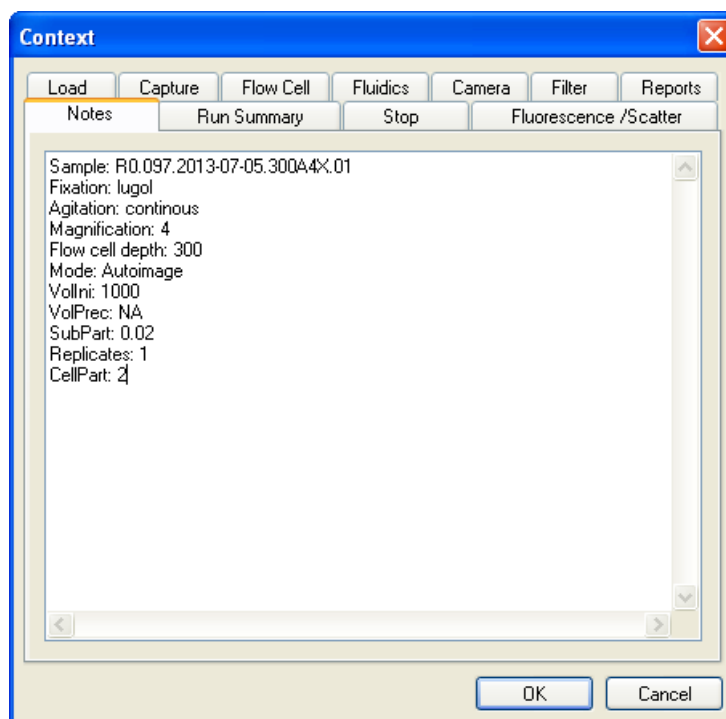
Recalibration interval = 30 min

Définir le temps au bout duquel une recalibration du fond de l'image sera réalisée. Ceci évite d'utiliser une image de fond avec des imperfections qui pourraient apparaître après une longue exécution.

- Onglet « Notes »

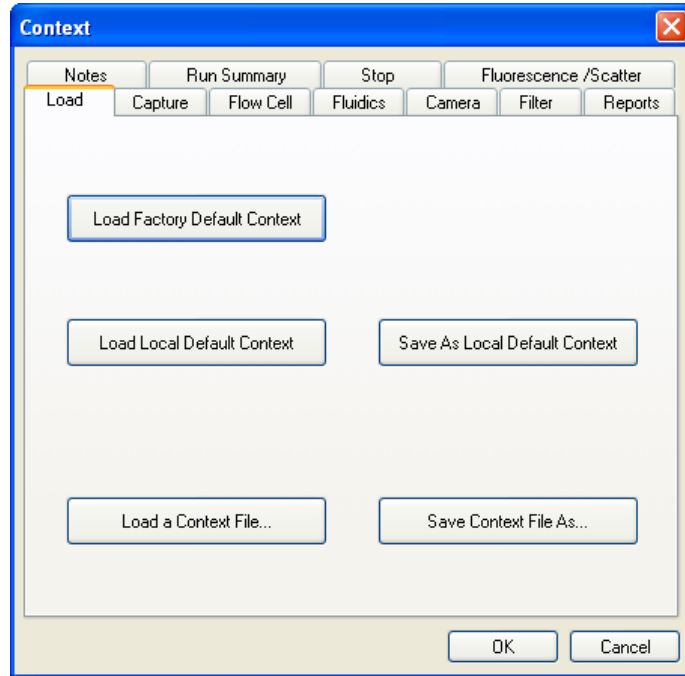
Pour le traitement des données dans ZooPhytoImage, des informations, relatives à l'échantillon et à la numérisation doivent être fournies par l'utilisateur (notamment pour le calcul des abondances, de la biomasse et des spectres de taille). L'importation des données dans ZooPhytoImage étant transparente, l'information doit être contenue dans les fichiers bruts extraits du FlowCAM. C'est pourquoi, il est nécessaire d'entrer ces renseignements dans les notes de l'expérimentation (lors de la configuration du contexte de numérisation dans Visual SpreadSheet). Ces informations concernent :

- **Sample** = nom de l'échantillon analysé;
- **Fixation** = échantillon fixé au Lugol ("lugol") ou non ("none");
- **Agitation** = type d'agitation de l'échantillon ("continuous");
- **Magnification** = objectif utilisé ("4" ou "10");
- **Flow cell depth** = profondeur de la cellule de flux utilisé en  $\mu\text{m}$  ("300" pour le 4X ou "100" pour le 10X);
- **Mode** = mode d'acquisition des images ("Autoimage" ou "fluorescence");
- **Vollni** = volume échantillonné en mer (bouteille Niskin, flacon sur perche, etc.)  
→ unité choisie par l'utilisateur;
- **SubPart** = fraction introduit dans l'appareil (si *vollni* = 1L et 10 mL introduit dans l'appareil, alors *SubPart* = 10mL/1000mL = 0.01);
- **Replicates** = nombre de réplicats (si l'expérimentation est répétée 3 fois, alors *Replicates* = 3);
- **CellPart** = fraction réellement numérisé (si *SubPart* = 10 mL mais que le volume réellement numérisé est de 5 mL, alors *CellPart* = 10mL/5mL = 2).



- Onglet « Load »

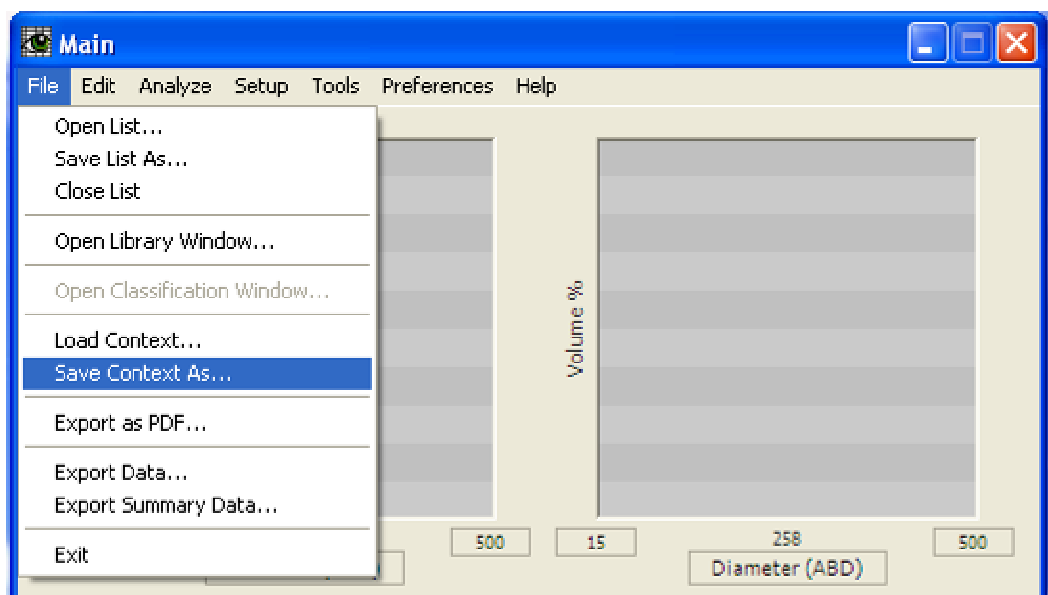
Il permet de sauvegarder les valeurs des paramètres en un clic, et de les ouvrir aussi rapidement



Cliquer sur « Save as Local Defaults » pour enregistrer les valeurs de paramètres.

Cliquer sur « Load Local Defaults » pour ouvrir les valeurs de paramètres précédemment enregistrés.

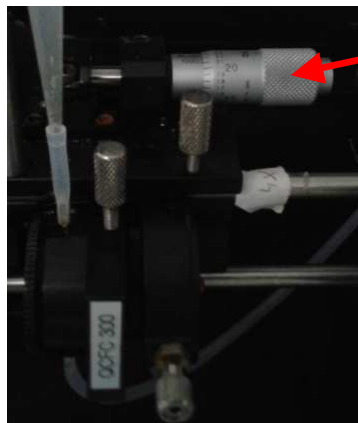
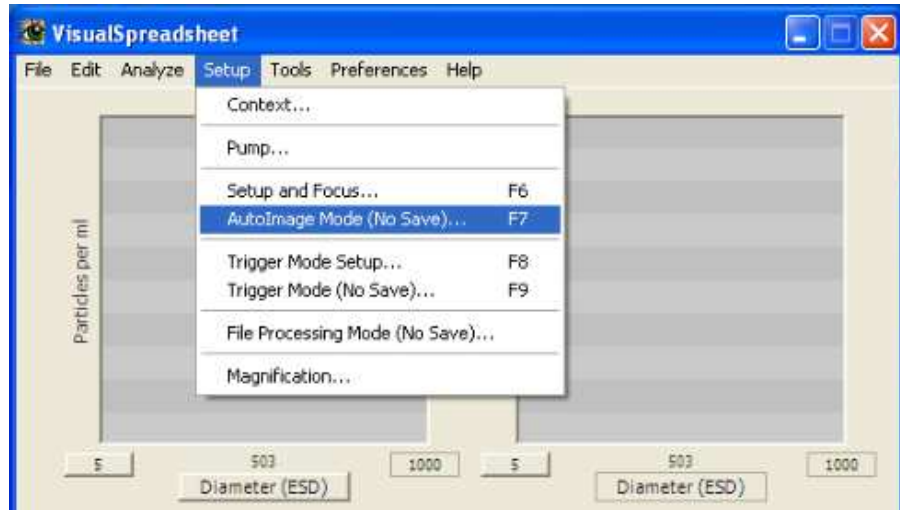
Il est également possible d'utiliser les deux fonctions situées dans le menu déroulant de la fenêtre principale, notamment pour sauvegarder plusieurs modèles de réglages des paramètres, selon l'étude concernée.



## Numérisation de l'échantillon

### - Mise au point fine

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



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

Ouvrir le répertoire "C:\Numerisation\REPHY\

- **Créer un répertoire** portant la **date du jour** suivie des **initiales de l'analyste** (Ex : 2011-03-15-GW), puis dans ce répertoire :
  - **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.yyyy-mm-dd.100A10X.01
- Pour une numérisation au 4X avec une cellule de 300µm  
XX.Ynn.yyyy-mm-dd.300A4X.01

Avec :  
XX initial de la provenance  
Ynn le numéro de l'échantillon

Ex : BL1.032.2013-01-29.300A4X.01

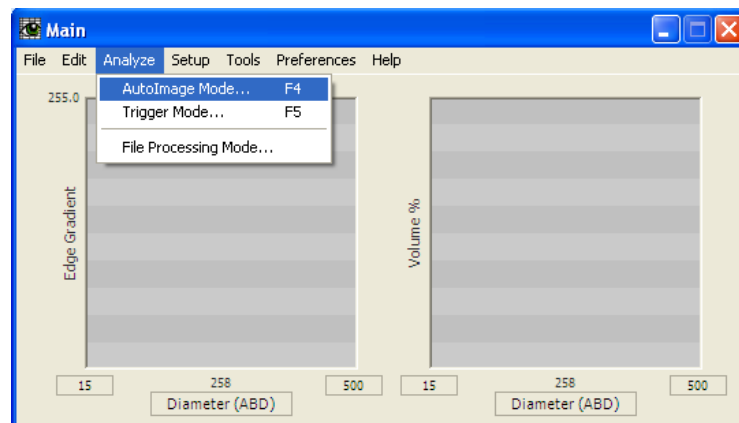


### - Calibration des niveaux de gris

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**.

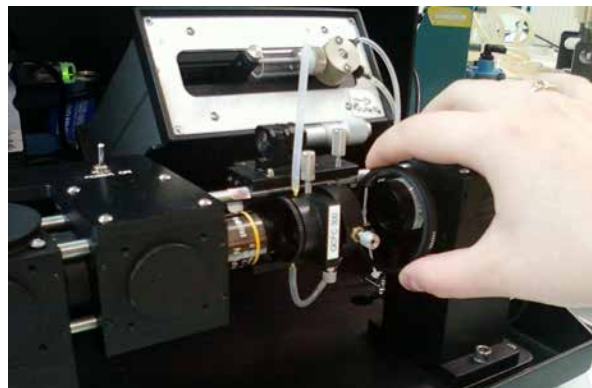


Ces images seront utiles pour calibrer les niveaux de gris des images en densité optique. Lancer une numérisation en mode AutoImage : Menu *Analyse* → *AutoImage Mode*. Cliquer sur « Ok » et sauvegarder le fichier de calibration des gris dans le sous-répertoire <grey> du répertoire de la date de numérisation.



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 : BL1.032.2013-01-29.300A4X.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.

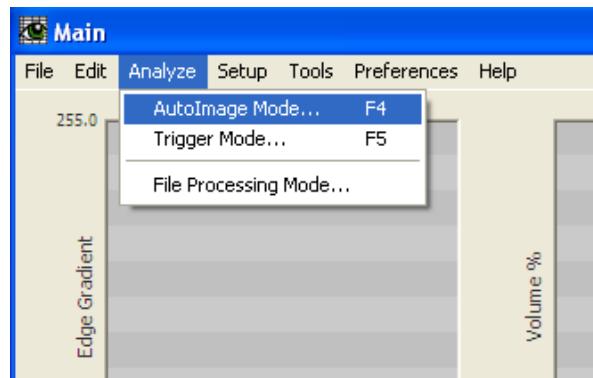


Fermer la fenêtre de la caméra une fois l'opération terminée.

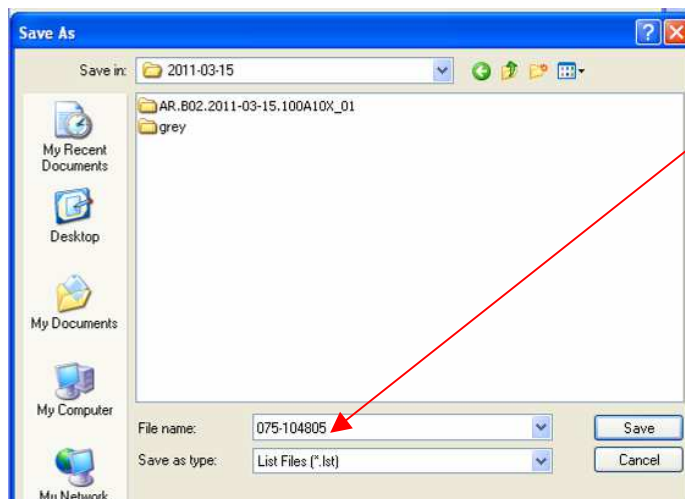
- Acquisition des images de l'échantillon

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 : BL1.032.2013-01-29.300A4X.01)

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

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

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.



## 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 :

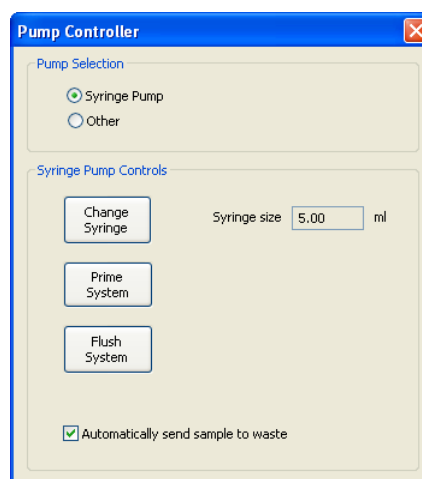
- Pipeter le reste d'échantillon dans l'entonnoir
- Démarrer la pompe seringue pour vider les tubes
- Démonter le tube connecteur de la pompe
- Baisser le porte entonnoir, déconnecter l'entonnoir
- Sortir l'entonnoir, le rincer à l'eau distillée puis à l'éthanol à 95%
- Mettre l'entonnoir à sécher
- Déconnecter le tube connecteur de la pompe du tube inférieur de la cellule, et le re-connecter à la place au tube supérieur
- A l'aide des pissettes d'eau distillée dont l'embout est introduit directement dans le tube inférieur, de la cellule de flux, procéder à un 1<sup>er</sup> rinçage à l'eau distillée, évacuer le liquide avec la poire, faire un 2<sup>e</sup> rinçage avec la pissette d'alcool, évacuer le liquide avec la poire, et terminer par une dernier rinçage à l'eau distillée.

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

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

Pour une durée plus longue, ôter la cellule après un dernier rinçage à l'alcool, la placer dans son étui de protection et la stocker à l'abri de la poussière. Laisser un objectif en place afin d'éviter une accumulation de poussière au niveau de la caméra, et fermer le portillon.

Une fois par semaine, la pompe seringue doit être démontée puis nettoyée à l'eau et à l'alcool. Pour cela, suivre le protocole de changement de la seringue.

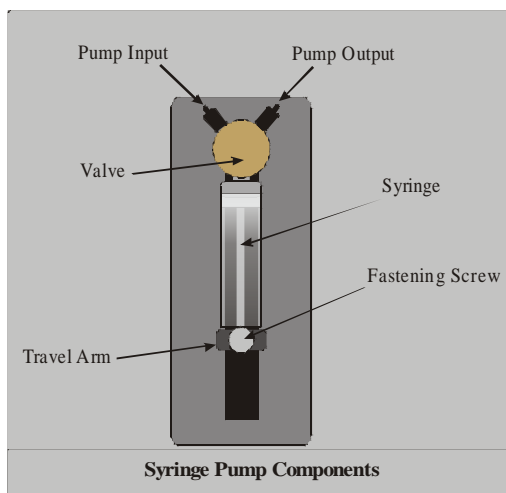


## Sauvegarde des fichiers

Une fois la séquence d'analyse terminée, ouvrir le disque réseau (N :) et sauvegarder les fichiers de la journée sous N:\FlowCAM\Numerisations\.

## Annexe 4 - Spécificités de la pompe seringue

La pompe seringue installé sur l'appareil du LER-BL (pompe interne C70) donne à VisualSpreadSheet un contrôle complet sur les débits, offrant une plus grande précision de contrôle des fluides et de meilleures performances globales du FlowCAM. Grâce à cela, VisualSpreadSheet peut estimer le taux de rendement (efficacité), le temps de passage de l'échantillon et ainsi aider l'utilisateur à optimiser les paramètres pour les expérimentations.



Deux tailles de seringue sont disponibles : 5,0 ml et 12,5 ml. Ces dernières permettent de sélectionner des volumes d'échantillon aussi faible que 35  $\mu$ l et des vitesses de flux dans les gammes suivantes : 0,05-100 ml/min (pour la seringue de 5,0 ml) et 0,125-250 ml/min (pour la seringue de 12,5 ml).

Syringe Size	Flow Rate Range
0.5 ml	0.005 – 10 ml/min
1.0 ml	0.01 – 20 ml/min
5.0 ml	0.05 – 100 ml/min
12.5 ml	0.125 – 250 ml/min

De nouvelles fonctionnalités sont donc disponibles dans VisualSpreadSheet et concernent le contrôle manuel de la pompe, les réglages liés à la cellule de flux et au débit d'eau, mais également la mise au point.

### Contexte – Onglet “Flow Cell”

Permet à l'utilisateur de sélectionner le type de cellule de flux et d'entrer les longueurs de tube au-dessus et en-dessous de la cellule de flux.

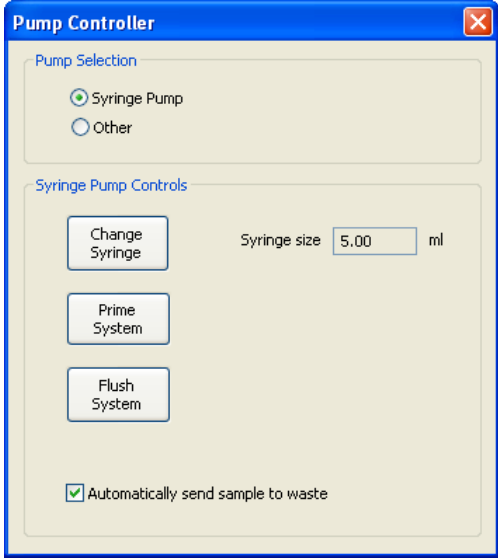
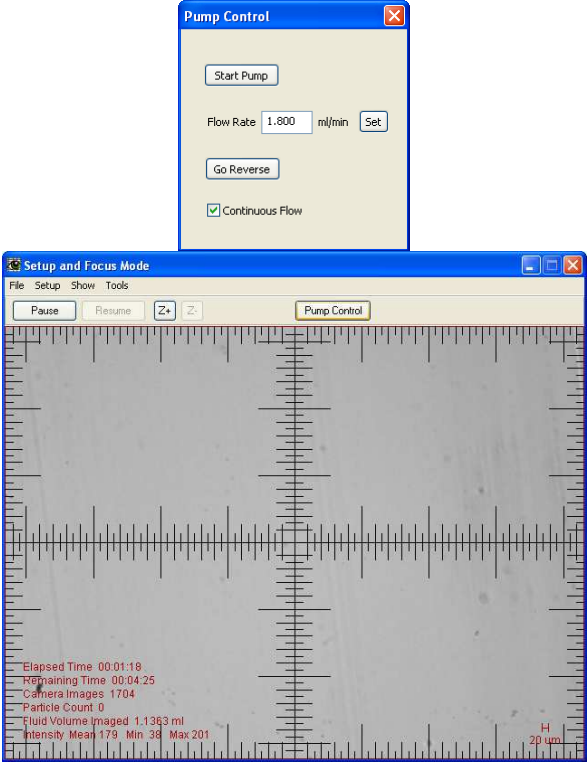
*Remarque :* avec la pompe seringue, ces longueurs de tube (en cm) sont primordiales pour obtenir des mesures précises de fluide. Il est recommandé une résolution minimale de 0,25 cm.

### Contexte – Onglet “Fluidics”

*Settings:* volume à analyser, vitesse de flux, et nombre d'images par seconde (pour la caméra).

*Remarque :* avec la pompe seringue, le FlowCAM arrête automatiquement l'analyse quand le volume spécifié a été aspiré.

*Estimated Efficiency and Run Time:* L'efficacité et le temps d'exécution sont automatiquement calculés, de manière à ce que l'utilisateur puisse utiliser les meilleurs réglages.

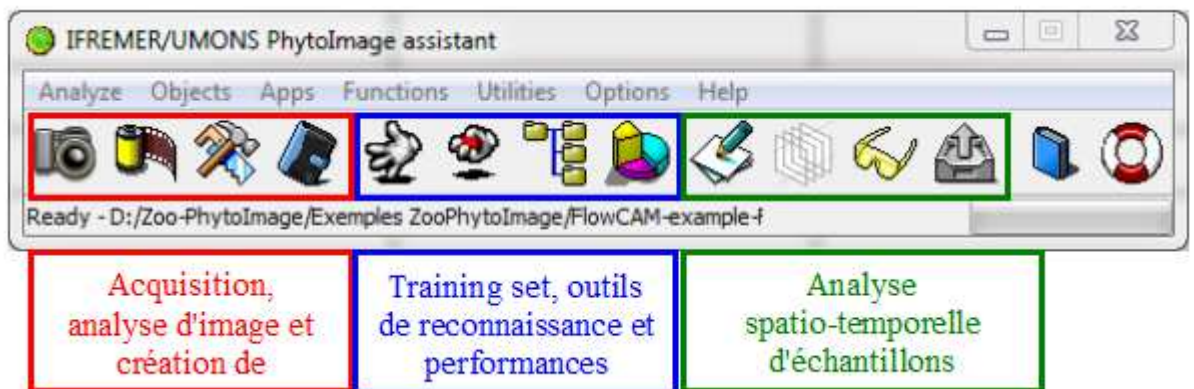
<p><b>Pump Controller</b></p> <p>Installation et contrôle manuel de la pompe seringue pour :</p> <ul style="list-style-type: none"> <li>• changer la seringue,</li> <li>• amorcer le système manuellement ou automatiquement,</li> <li>• vider/nettoyer automatiquement la cellule de flux.</li> </ul>	
<p><b>Setup and Focus Pump Control</b></p> <p>Contrôle manuel de la pompe seringue utilisé pour aider la mise au point du système optique.</p> <p>La fenêtre "setup and focus" ouvre automatiquement une fenêtre de contrôle de la pompe et permet à l'utilisateur de choisir la vitesse de flux, la direction du flux et le démarrage de la pompe.</p>	

Cependant, un inconvénient majeur de la pompe seringue réside dans le fait que, contrairement à la pompe péristaltique, un échantillon analysé ne peut pas être récupéré pour une seconde analyse (échantillon « abîmé », particules détruites ou coincées dans la pompe).

## Annexe 5 - Traitement numérique et importation des données du FlowCAM dans ZooPhytoImage.

ZooPhytoImage est décomposé en 3 modules de base :

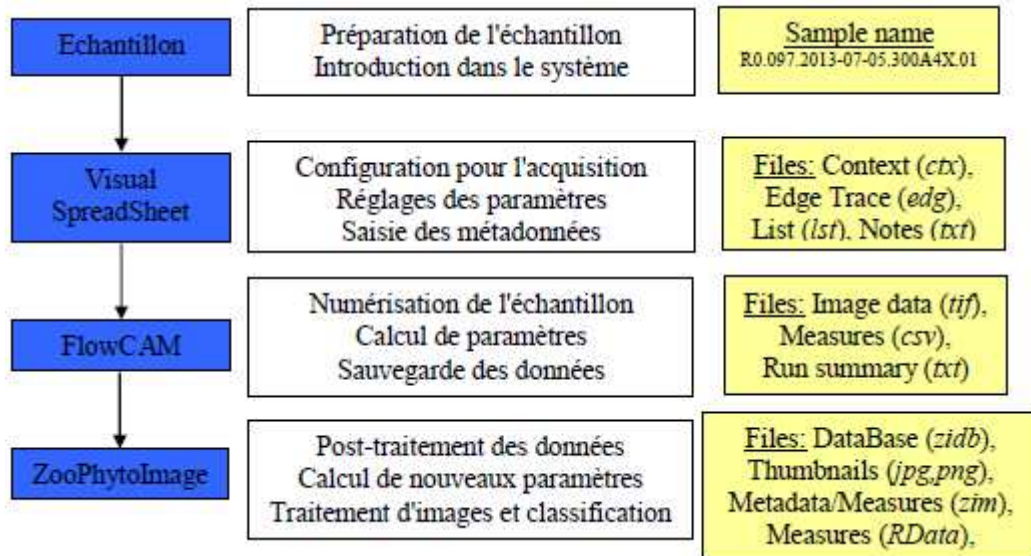
- Acquisition, analyse d'image et création de vignettes. C'est dans ce module que le logiciel sollicite ImageJ pour le traitement automatique des images,
- Création du training-set, des outils de reconnaissance automatique par classification supervisée et test de leurs performances par validation croisée,
- Analyse spatio-temporelle d'échantillons inconnus. L'identification automatique est réalisée grâce aux outils de reconnaissance construits. Actuellement, il est possible d'étudier l'abondance relative totale et/ou par classe, la biomasse totale et/ou par classe et la fréquence des individus d'un groupe ou d'une station classés en catégorie de tailles.



Pour chaque numérisation, les données brutes (issues du FlowCAM et de Visual SpreadSheet) sont stockées dans un répertoire unique qui contient : l'(les) image(s) de calibration, les mesures sur chaque images (data\_export), le fichier de contexte (les paramètres utilisés pour la numérisation), le fichier list (les données brutes de sortie), les collages et le fichier de notes. Tous ces fichiers doivent être présents dans le répertoire afin d'analyser les images avec ZooPhytoImage.

### ***Fichiers disponibles en sortie du FlowCAM (Raw Data)***

A la fin de chaque numérisation, 7 types de fichier sont générés : \*.lst, \*.ctx, \*.edg, \*\_notes.txt, \*\_run summary.txt, \*.csv et des fichiers \*.tif où le caractère '\*' représente le préfixe commun que l'utilisateur a choisi pour nommer l'expérimentation (généralement, le nom de l'échantillon). Ces fichiers sont sauvegardés automatiquement dans le répertoire sélectionné par l'utilisateur.



En effet, en début d'analyse, Visual SpreadSheet invite l'utilisateur à entrer un nom pour l'expérimentation et un répertoire qui servira à la sauvegarde des fichiers liés à cette analyse. Visual SpreadSheet utilise alors ce nom comme préfixe de tous les fichiers de l'analyse. Enfin, le logiciel stocke les images (image de calibration et collages) dans un format non-compressé sans perte (format TIFF). Ces fichiers sont nommés grâce au préfixe commun et à un numéro. Par exemple, si le préfixe commun est "my\_experiment", le répertoire dédié à l'analyse contiendra les fichiers :

- Fichier Context : "my\_experiment.ctx",
- Fichier Particle Edge Trace : "my\_experiment.edg",
- Fichier List : "my\_experiment.lst",
- Fichier Notes : "my\_experiment\_notes.txt",
- Fichier de mesures : "my\_experiment.csv",
- Fichier Run Summary : "my\_experiment\_run\_summary.txt",
- Fichiers Image (Calibration et Collages) : "cal\_image\_000001",  
my\_experiment\_000000001.tif", "my\_experiment\_000000002.tif", etc.

Un fichier List avec l'extension \*.lst est généré pour chaque analyse. Les fichiers List sont des fichiers de données primaires des expérimentations pour le FlowCAM. Ce sont des fichiers texte ASCII qui enregistrent les champs des données collectées pour chacune des particules.

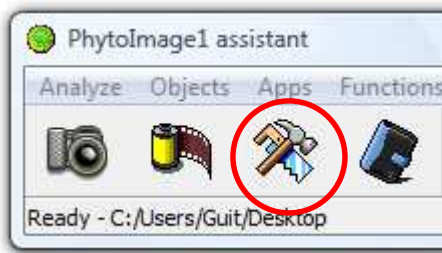
Un fichier Context avec l'extension \*.ctx est généré avec chaque fichier List. Il enregistre les informations de contexte (réglages) qui était en vigueur lorsque le fichier List a été créé ou lorsque l'expérimentation a débuté. Les fichiers texte "Notes" et "Run summary" accompagnent le fichier Context. Ces fichiers contiennent un bref descriptif des paramètres de numérisation et du flux, et les notes entrées dans la boîte de dialogue de Visual SpreadSheet (le plus souvent, les métadonnées).

Le fichier Particle Edge Trace accompagne le fichier .lst et enregistre la séquence de pixels qui forment le bord de chaque particule. Visual Spreadsheet utilise le suivi des bords afin de calculer plusieurs propriétés nouvelles des particules et tracer le contour de la celle-ci dans l'image affichée (si sélectionnée).

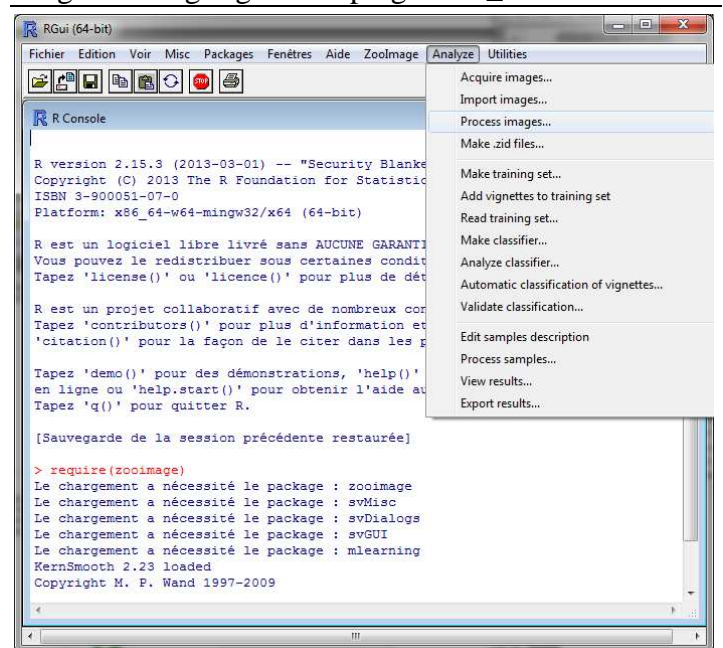


## Traitement des images

L'analyse d'image est réalisée par le logiciel ImageJ grâce au plugin FIT\_VIS.



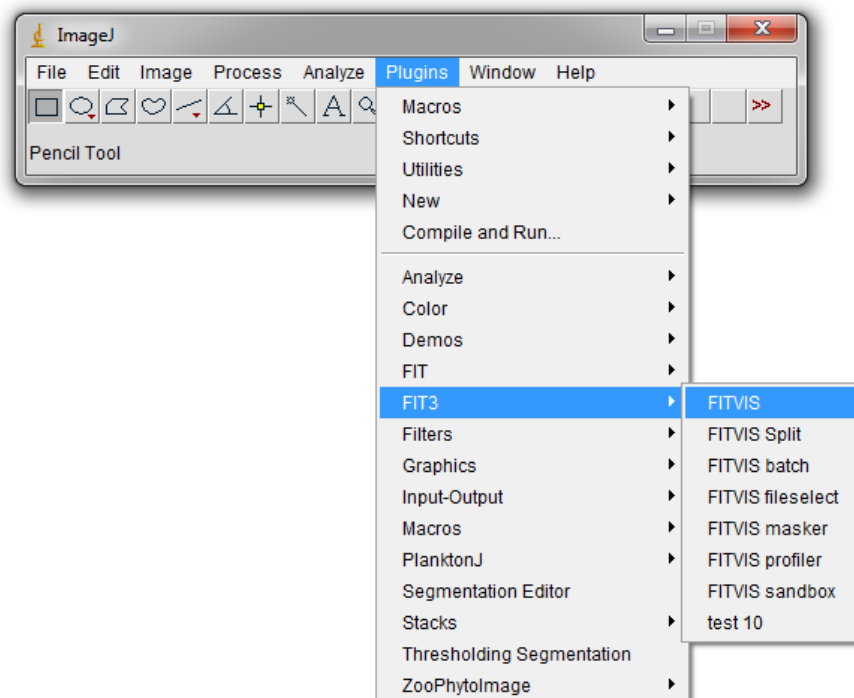
Version 1.2-3



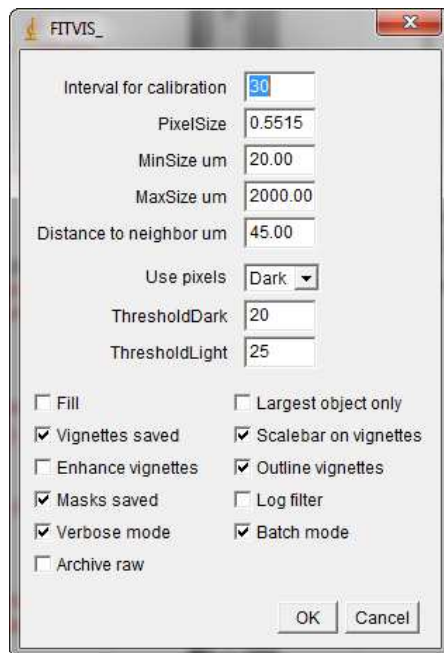
Version 4.0-0

Démarrer ImageJ en cliquant sur l'icône représenté ci-dessus (version 1.2-3) ou en utilisant le menu « Analyze » de la version 4.0-0. L'analyse d'image étant un processus coûteux en terme de ressource mémoire, il est possible de fermer ZooPhytoImage et de n'ouvrir que ImageJ si la quantité de mémoire RAM de l'ordinateur est limitée pour la tâche demandée.

Lorsque ImageJ est lancé, sélectionner le plugin FIT\_VIS dans le menu Plugins (Plugins/FIT3/FITVIS).



Puis, sélectionner le fichier d'export de résultats (fichier au format csv) dans le répertoire contenant les fichiers issus du FlowCAM, et entrer les paramètres utilisés pour la numérisation dans la boîte de dialogue qui apparaît.



#### **Archive raw**

Compression des fichiers bruts dans un ZIP.

#### **Largest object only**

Utilisation des particules les plus grandes.

#### **Scalebar on vignettes**

Ajout d'une barre d'échelle sur chaque vignette.

#### **Outline vignettes**

Sauvergarde des contours des particules pour chacune des vignettes (format PNG).

#### **Log filter**

Transformation logarithmique avant soustraction du fond d'image.

#### **Batch mode**

Exécution du traitement d'images sans afficher chacune des étapes (environ 20 fois plus rapide que le mode normal).

#### **Masks saved**

Sauvergarde des masques binaires utilisés pour segmenter chaque sous-image (format PNG).

#### **Verbose mode**

Affichage de messages durant le processus de traitement d'images.

#### **Interval for calibration = 30**

Temps au bout duquel une recalibration du fond de l'image sera réalisée.

#### **PixelSize**

Taille du pixel en  $\mu\text{m}$ , définie par le constructeur.

#### **MinSize um**

Taille minimale des particules numérisées en  $\mu\text{m}$ .

#### **MaxSize um**

Taille maximale des particules numérisées en  $\mu\text{m}$ .

#### **Distance to neighbor um**

Distance au plus proche voisin ; seuil minimal en  $\mu\text{m}$  entre deux particules distinctes.

#### **Use pixels**

Utilisation des pixels sombres et/ou claires pour la détection des particules.

#### **ThresholdDark**

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

#### **ThresholdLight**

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

#### **Fill**

Ajout de pixels autour de la particule afin de compléter le périmètre.

#### **Vignettes saved**

Création des vignettes nécessaires à la classification manuelle et automatique des images (format JPEG et fichier ZIM de métadonnées).

#### **Enhance vignettes**

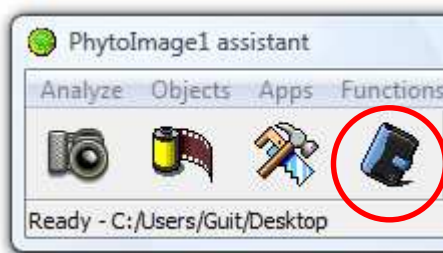
Amélioration de la qualité visuelle des vignettes.



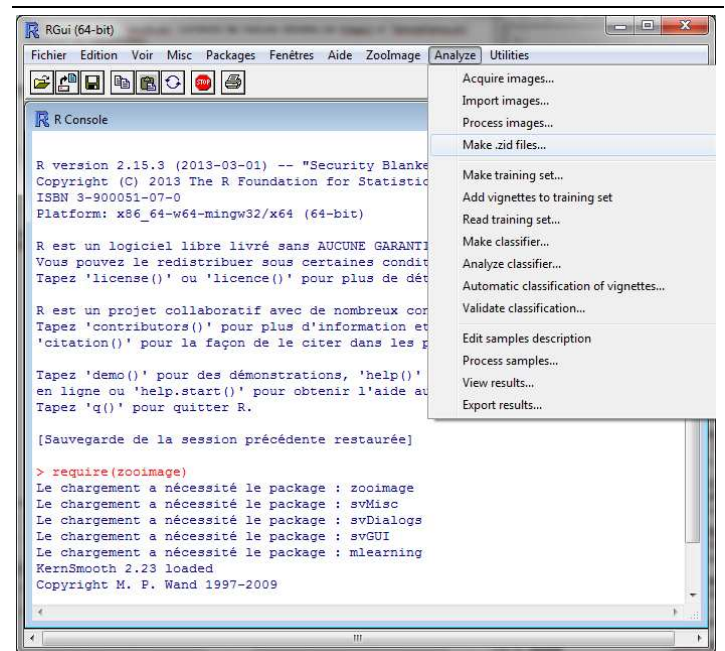
A la suite de ce traitement d'images, et selon les options sélectionnées, un ou plusieurs répertoires sont créés : "\_masks" (contenant les masques binaires), "\_outlines" (contenant les contours des particules) et "SampleName" (contenant les vignettes). Deux fichiers sont également générés : "results.csv" qui contient les mesures réalisées par ImageJ et "SampleName.zim" qui contenant les métadonnées (pour renseigner les champs du fichier ZIm, se reporter au livrable n°1 : "Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique").

## Création du fichier ZID

Les fichiers ZID contiennent toutes les informations nécessaires au reste de l'analyse. Un fichier ZID correspond donc à un échantillon.



Version 2.1-3



Version 4.0-0

Démarrer la création du fichier ZID en cliquant sur l'icône représenté ci-dessus ou en utilisant le menu « Analyze » de l'interface graphique. Il est possible de vérifier si toutes les vignettes sont associées aux entrées de la table de mesures avant la création du fichier ZID.

Sélectionner le fichier de contexte (fichier au format ctx) dans le répertoire contenant les fichiers issus du FlowCAM, et renseigner les deux nouvelles boîtes de dialogue qui s'affichent successivement :

- nom de l'opérateur ayant effectué la numérisation (par défaut : IFREMER),
- volume d'échantillon passé dans le FlowCAM (en mL).

Question

Author of the FlowCAM run:

OK Annuler

Question

Volume of sample passed through the FlowCAM (mL):

OK Annuler

Un fichier ZID est alors créé dans le répertoire. Ce fichier compressé contient :

- toutes les vignettes (nommées "SampleName\_id.jpg"),
- un fichier ZIM contenant les métadonnées ainsi que les mesures sur chacune des vignettes (nommé "SampleName\_dat1.zim"),
- un fichier RData contenant les mesures sur chacune des vignettes (nommé "SampleName\_dat1.RData").

Ce fichier contient l'ensemble des informations nécessaires au traitement des données pour la reconnaissance automatique des particules par ZooPhytoImage. Pour la création des sets d'apprentissage, la génération et l'optimisation des outils de reconnaissance, l'identification automatique et la validation manuelle partielle des prédictions, se reporter au livrable n°1 ("Guide méthodologique concernant la version actualisée de ZooPhytoImage avec refonte de l'interface graphique") et au livrable n°2 ("Protocole sur les outils de reconnaissance optimisés Manche-Atlantique").

Un cahier des charges définissant les fonctionnalités requises pour la refonte de l'interface graphique utilisateur de ZooPhytoImage, a été rédigé en juillet 2013 (Annexe 7). Il a pour objet d'exposer tous les aspects graphiques pour le logiciel. Ce document contient le détail de tous les éléments concernant les spécifications d'ergonomie, le détail des différents composants graphiques et la dynamique entre eux. Il comprend également le découpage de l'interface graphique en zones fonctionnelles. La description des zones fonctionnelles renseigne sur leur aspect et leur rôle. Ce cahier des charges présente également tous les éléments graphiques composant les rubriques et sous-rubriques de l'application ainsi que la logique d'interaction entre eux.

**Annexe 6 - Modalités d'intégration de l'outil de correction statistique de l'erreur en utilisation en (pré)routine de Phytolmage (cahier des charges - février 2013)**

Laboratoire Environnement Ressources (LER)

Auteurs : Guillaume WACQUET, Liliane Fiant

Janvier 2013 - ODE/RST/LERBL/13.03



**Ifremer**

---

# Projet FlowCAM/Phytolmage

## Cahier des charges

### 15/01/2013

Modalités d'intégration de l'outil de correction statistique de l'erreur en utilisation en (pré)routine de Phytolmage.



# Projet FlowCAM/PhytoImage - Cahier des charges - 15/01/2013

Modalités d'intégration de l'outil de correction statistique de l'erreur en utilisation en (pré)routine de PhytoImage.



# sommaire

## Sommaire

Introduction.....	5
Contexte.....	7
• Projet.....	7
• Sous-action I.....	7
• Lot n°3.....	7
Organigrammes pour le module de correction statistique de l'erreur.....	9
• Organigramme général.....	9
• Organigramme détaillé.....	10
Interface graphique utilisateur .....	12
Bibliographie indicative.....	15

## Introduction

L'objectif du module de correction statistique de l'erreur réside dans l'amélioration significative des performances de reconnaissance automatique en ne reclassant qu'une faible proportion des vignettes considérées comme les plus suspectes par l'algorithme. Cette approche particulière offre un compromis idéal entre la méthode automatique et la validation manuelle totale, tout en garantissant des performances de reconnaissance identiques ou supérieures dans un laps de temps acceptable.

L'objet de ce cahier des charges est de définir **la liste des fonctionnalités à développer** afin de disposer d'un logiciel permettant de gérer complètement le module de correction statistique de l'erreur par validation partielle et manuelle des vignettes suspectes. En effet, plusieurs fonctions nouvelles doivent être ajoutées au logiciel PhytoImage existant afin de faciliter la manipulation du nouvel outil et en particulier, la validation manuelle des classifications automatiques générées.

Ce cahier des charges spécifie également **les différentes phases de ce nouveau module en utilisation en (pré)routine**, qu'il sera nécessaire de comprendre et de suivre afin d'assurer une transition douce entre les versions antérieures du logiciel PhytoImage et le nouveau système.





# 1. Contexte

## 1.1. Projet - FlowCAM/PhytoImage

Optimisation de l'identification et du dénombrement du micro-phytoplancton avec le système couplé de numérisation et d'analyse d'images FlowCAM/PhytoImage (système innovant).

## 1.2. Sous-action I - Finalisation de l'outil opérationnel (2013)

Pour qu'un outil opérationnel puisse être finalisé, de nombreuses tâches restent à réaliser, par exemple : améliorations technologiques au matériel FlowCAM d'un point de vue mécanique et optique, amélioration et optimisation du logiciel PhytoImage, continuer ou initier l'acquisition d'images, finaliser un set d'apprentissage et un outil de reconnaissance optimisés pour chacune de ces banques d'images, comparer les résultats obtenus par l'analyse FlowCAM/PhytoImage avec ceux des lectures réalisées au microscope, etc. Cette action se fera dans le cadre de la collaboration Ifremer / Université de Mons.

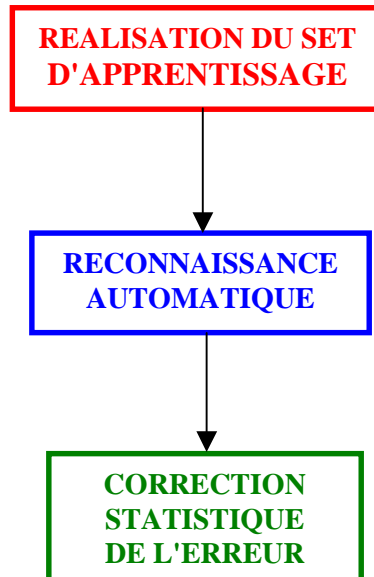
## 1.3. Lot n°3 - Description de la recherche

Les nouvelles techniques de correction d'erreur introduites récemment permettent une diminution drastique du taux d'erreur, à condition de réaliser un outil de classification flexible capable de s'adapter localement en tenant compte des validations partielles de chaque échantillon, et en introduisant la notion de correction statistique de l'erreur par validation des suspects. Jusqu'à présent, les outils de classification du phytoplancton utilisés par PhytoImage étaient créés et optimisés une fois pour toutes. C'est pourquoi, il est nécessaire de redéfinir le flux de traitement de données issues du FlowCAM et d'intégrer la validation partielle nécessaire pour les nouvelles techniques de correction d'erreur dans ce flux de traitement. Sur base de spécifications fournies par l'Ifremer, et en relation avec ce nouveau flux de données, les parties du logiciel concernant la réalisation du set d'apprentissage, la création d'outils de classification et la validation manuelle doivent être intégrées, non plus sous forme linéaire, mais sous forme itérative. Ce lot vise donc à remanier tout le flux de données concerné par ces modifications, plus précisément, depuis l'élaboration du set d'apprentissage jusqu'à la classification des particules dans les différents groupes taxonomiques.



## 2. Organigrammes pour le module de correction statistique de l'erreur

### 2.1. Organigramme général

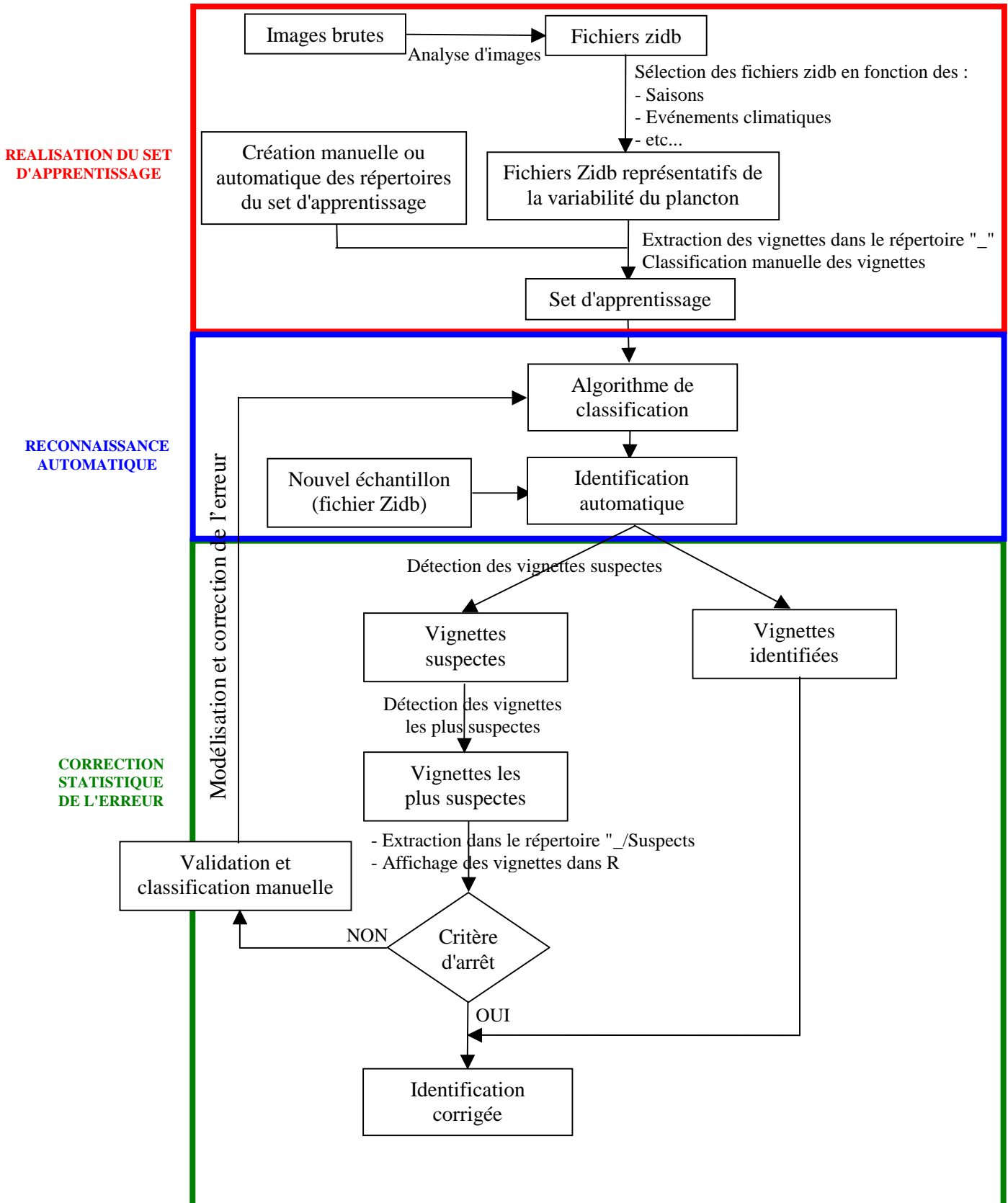


*Organigramme global de reconnaissance supervisée automatique et correction de l'erreur*

Pour le phytoplancton, le module de correction détermine une probabilité d'être suspect pour chaque particule, et seules les particules les plus suspectes sont validées. Ensuite, l'erreur est modélisée et corrigée en fonction des résultats de cette validation partielle. Il est possible d'ajouter à cette probabilité des informations extérieures (par exemple, la pertinence d'une identification dans une zone géographique donnée).

L'outil envisagé permettra de ne valider manuellement que les particules les plus suspectes. Un aspect important est que la correction statistique de l'erreur diminue très rapidement le nombre de particules mal classées : suite aux premières expérimentations, avec 10 à 15% de particules validées, il est possible de descendre jusqu'à 5% d'erreurs.

## 2.2. Organigramme détaillé



*Organigramme détaillé de reconnaissance supervisée automatique et correction de l'erreur*

Dans le schéma ci-dessus, la partie "Réalisation du set d'apprentissage" est tiré du manuel d'utilisation du logiciel PhytoImage édité par Philippe Grosjean et Kévin Denis. Lors de cette phase, l'analyse d'image permet d'obtenir des sous-images des particules (appelées "vignettes") destinées à une future identification visuelle des organismes. L'information numérique est collectée dans des objets "ZIDat" (ZooImage Data) contenant les mesures sur chacune des particules. Les vignettes et l'objet "ZIDat" sont ensuite stockés sur le disque dans un fichier de type "Zidb" (ZooImage DataBase). Un unique fichier "Zidb" est créé pour chaque échantillon et contient l'information sur plusieurs centaines ou milliers de particules. Le set d'apprentissage est ensuite réalisé manuellement en utilisant le logiciel libre "XnView".

Suite à la phase de "Reconnaissance automatique", le module de "Correction statistique de l'erreur" (basé sur la modélisation et l'intégration des biais systématiques mesurables) utilisé pour l'optimisation de la précision de l'outil de reconnaissance, doit permettre d'optimiser le temps alloué à la vérification des classifications automatiques générées. Ce module comporte :

- une étape d'application d'un algorithme de classification (A) : dans un premier temps, des règles sont déterminées afin de discriminer au mieux chaque groupe du set d'apprentissage, et sont ensuite appliquées au nouvel échantillon afin d'obtenir une identification automatique des particules.
- une étape de détection des vignettes suspectes et de test (B) pour laquelle le système propose un critère d'arrêt. Ce critère d'arrêt subjectif est basé sur l'avis expert de l'analyste grâce à l'affichage des vignettes sélectionnées comme étant les plus suspectes (sous forme de collage).

A la suite de l'étape (A), le système considère des vignettes comme suspectes en se basant sur leur probabilité de reconnaissance incorrecte (probabilité d'appartenance à un groupe) retournée *a posteriori* par l'algorithme lui-même. De plus, il est possible d'utiliser également des critères bio-écologiques (par exemple, possibilité de rencontrer une espèce à un endroit précis, à un moment précis, etc.) afin de détecter de manière plus précise les vignettes suspectes.

Ces images sont ensuite extraites dans un répertoire dédié (par exemple, "\_/Suspects"), affichées dans R et l'utilisateur peut alors juger de la pertinence de la classification et décider d'arrêter ou de continuer la correction manuelle (grâce au logiciel "XnView" ou "Plankton Identifier"). En bref, au fur et à mesure des corrections, l'utilisateur peut visualiser l'impact de la modélisation et de la correction de l'erreur.

Ce traitement permet ainsi d'estimer plus précisément la composition finale du nouvel échantillon en modélisant l'erreur et en la corrigeant.

Remarques (à développer) :

1. Traitement des vignettes les plus suspectes → par tranche de 2-4% ?
2. Nécessité de connaître le groupe d'appartenance (résultant de l'identification automatique) de chacune des vignettes suspectes ?

### 3. Interface Graphique Utilisateur

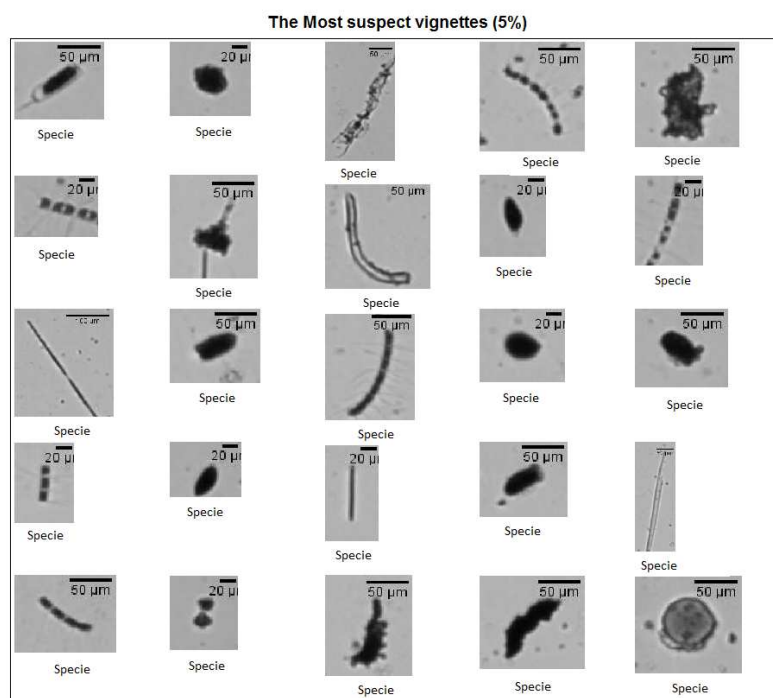
Réalisation de l'outil de classification supervisée : l'utilisateur a la possibilité de choisir entre six algorithmes de classification

- Analyse Discriminante Linéaire : estimation rapide de l'efficacité de l'outil de reconnaissance, recherche de la meilleure discrimination entre groupes par calcul de combinaisons linéaires de mesures.
- Arbre de décision et Random Forest : création d'un ou plusieurs arbres décisionnels.
- Réseaux de neurones artificiels contenant une ou plusieurs couches cachées.
- K-NN et Quantification vectorielle d'apprentissage : minimisation des distances entre des objets devant être classés et ceux appartenant au set d'apprentissage.

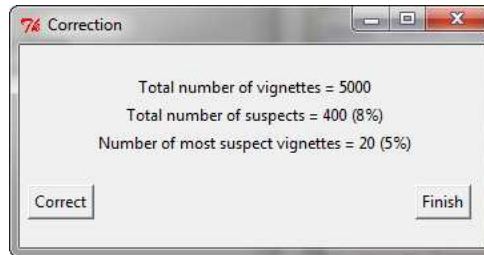
Les erreurs induites par les classifieurs ne sont pas aléatoires mais systématiques. Il est donc possible de modéliser cette erreur afin d'effectuer des corrections statistiques. En effet, une fraction des particules sont difficiles à classer mais peuvent être étiquetées comme suspectes pour une validation manuelle partielle de l'identification. De cette manière, le dispositif FlowCAM/PhytoImage pourrait être en mesure de déterminer avec précision l'abondance des taxa phytoplanctoniques majeurs présents dans un échantillon et pourrait être utilisé pour valider les modèles écologiques plus rapidement que le comptage manuel.

Application de l'outil de classification sur un nouvel échantillon :

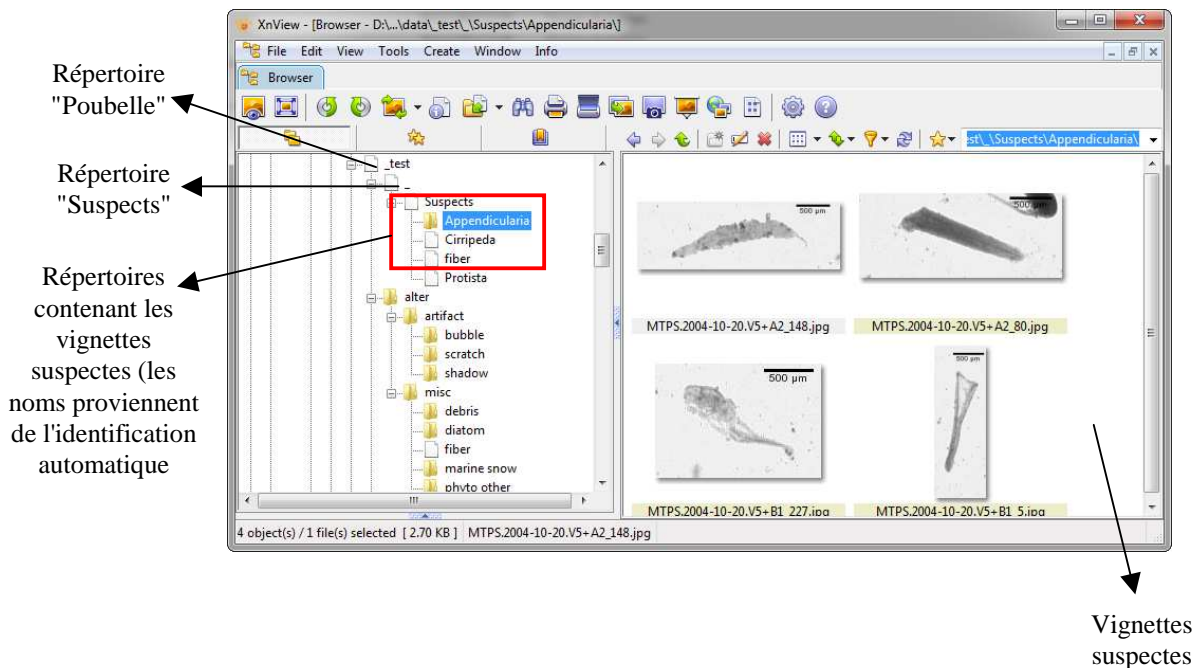
- Identification automatique et détection de la totalité des vignettes suspectes.
- Proposition automatique (en fin d'identification automatique) à l'utilisateur des vignettes les plus suspectes.



- Proposition automatique à l'utilisateur de correction de l'erreur.



- Si non (bouton "Finish"), présentation habituelle des résultats (abondances, biomasses, etc.).
- Si oui (bouton "Correct"), le processus de correction statistique de l'erreur est exécuté. La première phase consiste alors à extraire les vignettes dans un répertoire dédié (par exemple, "\_/Suspects"). La seconde phase consiste en la correction manuelle des erreurs possibles (grâce à "XnView" ou "Plankton Identifier").



- Lorsque l'expert estime que l'erreur est "négligeable" (après une ou plusieurs itérations du processus de correction), l'utilisateur clique sur 'Finish'. Cette action entraîne l'obtention d'une identification finale (corrigée partiellement) et l'affichage des résultats en termes de biomasse, d'abondances, etc.



### **Bibliographie indicative**

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.

Grosjean P. & Denis K., 2010. Rapport Université de Mons – Zoo / PhytoImage – Optimisation du set d'apprentissage REPHY en 2010.

Tunin-Ley A. & Maurer D., 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. Rapport RST/LER/AR/11/002

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.

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.

## **Annexe 7 - Interface Graphique Utilisateur et fonctionnalités requises pour l'utilisation en (pré)routine de ZooPhytoImage (cahier des charges - juillet 2013).**

Laboratoire Environnement Ressources (LER)

Auteurs : Guillaume WACQUET, Alain LEFEBVRE

Juillet 2013 - ODE/RST/LERBL/13.10



**Ifremer**

---

# Projet FlowCAM/PhytoImage Cahier des charges 15/07/2013

Interface Graphique Utilisateur et  
fonctionnalités requises pour l'utilisation en  
(pré)routine de ZooPhytoImage.



# Projet FlowCAM/PhytoImage

## Cahier des charges

15/07/2013

Interface Graphique Utilisateur et  
fonctionnalités requises pour l'utilisation  
en (pré)routine de ZooPhytoImage.



# sommaire

## Sommaire

<b>1. Contexte .....</b>	<b>7</b>
1.1. Projet - FlowCAM/PhytoImage .....	7
1.2. Description de l'action .....	7
1.3. Description de la recherche - Livrables .....	8
<b>2. Pré-requis .....</b>	<b>11</b>
2.1. Etapes de travail.....	11
2.2. Saisie des métadonnées .....	13
<b>3. Interface Graphique Utilisateur .....</b>	<b>15</b>
3.1. Présentation générale .....	15
3.2. Importation des données.....	16
3.3. Création d'un training/test set.....	18
3.4. Création d'un outil de reconnaissance .....	20
3.5. Reconnaissance automatique .....	22
3.6. Validation (correction statistique de l'erreur) .....	25
<b>4. Outils complémentaires .....</b>	<b>26</b>
4.1. Fichier de configuration .....	26
4.2. Historique des actions .....	26
4.3. Dénombrement des cellules en colonie .....	27

## Introduction

Le présent document est le cahier des charges définissant les fonctionnalités requises pour la refonte de l'interface graphique utilisateur de ZooPhytoImage, comme définie dans le lot n°4 de la convention IFREMER/UMONS (Convention de Collaboration pour l'évolution d'outils logiciels couplés au FlowCAM : développement d'une pré-version pour l'utilisation en routine; suite du contrat n°10/5210681). Il a pour objet d'exposer tous les aspects graphiques pour le logiciel.

Ce document contient le détail de tous les éléments concernant les spécifications d'ergonomie, le détail des différents composants graphiques et la dynamique entre eux. Il comprend également le découpage de l'interface graphique en zones fonctionnelles.

La description des zones fonctionnelles renseigne sur leur aspect et leur rôle. Cette partie présentera tous les éléments graphiques composant les rubriques et sous-rubriques de l'application ainsi que la logique d'interaction entre eux.

# sommaire



# 1. Contexte

## 1.1. Projet - FlowCAM/PhytoImage

### Titre du Projet

Optimisation de l'identification et du dénombrement du micro-phytoplancton avec le système couplé de numérisation et d'analyse d'images FlowCAM/PhytoImage (système innovant).

### Contexte de l'action

Plus de 3000 échantillons d'eau sont examinés annuellement pour des lectures de phytoplancton dans le cadre du REPHY et des réseaux régionaux pilotés par l'Ifremer. Les données obtenues (identification et dénombrement par taxon, genre ou espèce) sont utilisées en particulier pour répondre aux exigences de la DCE, puisqu'elles permettent de calculer les indices de l'élément de qualité phytoplancton, qui fait partie des éléments biologiques obligatoires pour l'évaluation de la qualité écologique des masses d'eau.

Le projet FlowCAM/PhytoImage a été monté il y a quelques années pour améliorer et optimiser ces observations. L'outil est constitué (i) du matériel FlowCAM destiné à numériser les images du phytoplancton, (ii) du logiciel ZooPhytoImage, qui permet d'identifier automatiquement et de dénombrer le phytoplancton à partir de ces images numérisées. L'ensemble du projet résulte de la collaboration de l'Ifremer avec l'Université de Mons (Belgique), qui a développé le logiciel ZooPhytoImage, et qui l'a considérablement amélioré et adapté aux contraintes de l'Ifremer.

## 1.2. Description de l'action

La numérisation d'un échantillon au FlowCAM conduit à des images brutes, ensuite transformées en vignettes. Ces vignettes sont ensuite classées manuellement après identification visuelle, dans des groupes ou catégories taxinomiques hiérarchisés, constituant ainsi une banque d'images. Puis un set d'apprentissage est constitué par une sélection dans la banque d'images. A partir du set d'apprentissage, un outil de reconnaissance optimisé est construit avec le logiciel ZooPhytoImage, en liaison avec un set lui-même optimisé. Deux sous-actions sont proposées ici :

### **Sous-action I. Finalisation de l'outil opérationnel**

Pour qu'un outil opérationnel puisse être finalisé, de nombreuses tâches restent à réaliser, dont certaines ont débuté fin 2012, par exemple : améliorer le matériel FlowCAM d'un point de vue mécanique et optique, améliorer et optimiser le logiciel ZooPhytoImage, continuer ou initier l'acquisition d'images, constituer un set d'apprentissage et un outil de reconnaissance optimisés pour chacune de ces banques d'images, comparer les résultats obtenus par l'analyse FlowCAM/PhytoImage avec ceux des lectures réalisées au microscope, etc.

## sommaire

### **Sous-action II. Optimisation et évolutions d'un dispositif intégré FlowCAM/PhytoImage d'analyse automatisée du phytoplancton (2013-2015)**

L'utilisation d'un système tel que FlowCAM/PhytoImage dans le cadre d'un réseau national n'a actuellement pas d'équivalent. Il s'agit donc d'un système totalement innovant. De nombreuses perspectives d'évolution et d'amélioration sont d'ores et déjà envisageables : couplage avec des techniques de type FISH (biologie moléculaire), et avec la cytométrie en flux, qui permet de définir des grands groupes de cellules phytoplanctoniques et de les dénombrer (cf. CytoSense par exemple, utilisé par le CNRS Wimereux, qui permet d'acquérir, en plus de la fluorescence globale de la cellule, son profil/spectre de fluorescence). L'évaluation des apports respectifs des méthodes en utilisation conjointe (mise en oeuvre du CytoSense et du FlowCAM/PhytoImage en parallèle sur les mêmes échantillons et analyse de la complémentarité de l'information) sera donc réalisée.

#### **1.3. Description de la recherche - Livrables**

Les travaux qui visent la finalisation d'un outil utilisable en routine se subdivisent en quatre lots.

- **Lot N° 1 (2011)** : La version 2.1 de ZooPhytoImage est constituée de la version GPL publique 1.2-1 de ce logiciel amendée de nombreux patches développés spécifiquement dans le cadre de la collaboration IFREMER-UMONS. Ce lot vise à intégrer complètement tous ces patches dans une version 3.0 complètement refactorisée du logiciel, et ce, à la fois au niveau de son code et de sa documentation. Contrairement à la version précédente qui ne fonctionne que sous Windows XP, cette nouvelle version devra être plus largement utilisable, soit sur des systèmes d'exploitation incluant Windows XP, Vista et Seven, Linux et Mac OS X Leopard, Snow Leopard ou Lion.

- **Lot N° 2 (2011)** : Ce lot vise à redéfinir le format de stockage des données afin de réduire d'au moins un facteur dix le temps d'accès aux données de type vignettes et le développement de code permettant l'interactivité avec les graphes de type matrices de confusion par visualisation des vignettes correspondant aux zones cliquées à la souris sur les graphes. Des routines de conversion de, et vers l'ancien format de stockage devront également être fournies.

- **Lot N° 3 (15 janvier 2013)** : Redéfinition du flux de traitement des données issues du FlowCAM et intégration de la validation partielle nécessaire pour les nouvelles techniques de correction d'erreur (travail réalisé fin 2011 et début 2012) dans ce flux de traitement. Jusqu'à présent, les outils de classification du phytoplancton utilisés par ZooPhytoImage étaient créés et optimisés une fois pour toutes. Les nouvelles techniques de correction d'erreur introduites récemment permettent une diminution drastique du taux d'erreur, à condition de réaliser un outil de classification flexible capable de s'adapter localement en tenant compte des validations partielles de chaque échantillon, et en introduisant la notion de correction statistique de l'erreur par validation des suspects. Sur base de spécifications fournies par l'IFREMER, et en relation avec ce nouveau flux de données, les parties du logiciel concernant la réalisation du set d'apprentissage, la création d'outils de classification et la validation manuelle doivent être intégrées, non plus sous forme linéaire, mais sous forme itérative. Ce lot vise donc à remanier tout le flux de données concerné par ces modifications, plus précisément, depuis l'élaboration du set d'apprentissage jusqu'à la classification des particules dans les différents groupes taxonomiques.

- **Lot N° 4 (15 juillet 2013)** : Dans le cadre du Contrat, l'interface utilisateur du logiciel a été étudiée du point de vue de son adéquation pour les traitements spécifiques IFREMER, et des modifications ont été apportées (boite de dialogue de sélection des variables, réalisation de « test sets » où les vignettes sont classées automatiquement par l'ordinateur en vue de validation, spécification de seuils maximum et minimum pour au moins une variable mesurée). Il apparaît cependant désirable d'augmenter considérablement l'interactivité visuelle avec le logiciel, notamment au niveau de la visualisation de collages de vignettes directement dans R, et en liaison avec les graphiques (clic souris sur une zone, visualisation des vignettes correspondantes). Pour cela, une refonte complète de l'interface graphique utilisateur de ZooPhytoImage sur base de définition des "use cases" dans le cadre d'une perspective d'exploitation REPHY, est nécessaire. La version 3 du logiciel étant dépourvue d'interface graphique (choix délibéré afin de ne pas influencer les utilisateurs chargés de définir l'aspect général et les fonctionnalités de la nouvelles interface), l'ensemble des opérations s'effectue alors via des lignes de commande en langage R, ainsi une vision neuve sur la construction et l'utilisation de l'interface de ZooPhytoImage peut être fournie. Cette réflexion doit être menée en parallèle de la définition du protocole et de l'écriture des documents de spécifications de la ou des méthodes d'acquisition des données.

***Le présent document est le cahier des charges détaillant les fonctionnalités requises pour l'aboutissement de ce lot.***



## 2. Pré-requis

Un des objectifs principal de la refonte de l'interface graphique utilisateur de ZooPhytoImage réside dans la réduction du nombre de tâches de l'utilisateur pour l'importation et la mise en forme des données issues du FlowCAM. A cette fin, les outils de traitement d'images et de transformation des données brutes devront et pourront être appliqués en tâches de fond (ou en arrière-plan).

### 2.1. Etapes de travail

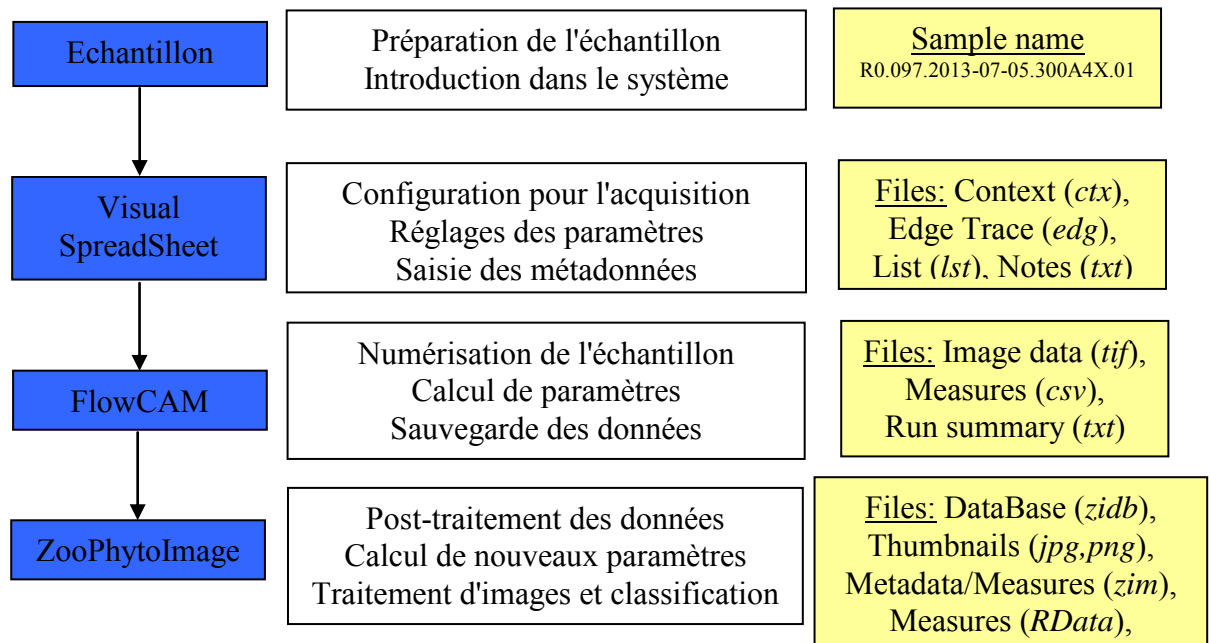


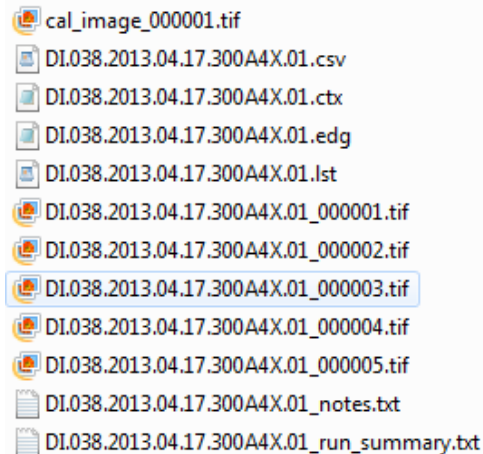
Figure 2.1-1 : Schéma fonctionnel des différentes étapes de travail

#### Fichiers disponibles en sortie du FlowCAM (Raw Data)

Comme montré sur la figure 2.1-1, à la fin de chaque numérisation, 7 types de fichier sont générés : \*.lst, \*.ctx, \*.edg, \*\_notes.txt, \*\_run summary.txt, \*.csv et des fichiers \*.tif où le caractère '\*' représente le préfixe commun que l'utilisateur a choisi pour nommer l'expérimentation (généralement, le nom de l'échantillon). Ces fichiers sont sauvegardés automatiquement dans le répertoire sélectionné par l'utilisateur.

En effet, en début d'analyse, Visual Spreadsheet invite l'utilisateur à entrer un nom pour l'expérimentation et un répertoire qui servira à la sauvegarde des fichiers liés à cette analyse. Visual Spreadsheet utilise alors ce nom comme préfixe de tous les fichiers de l'analyse. Enfin, le logiciel stocke les images (image de calibration et collages) dans un format non-compressé sans perte (format TIFF). Ces fichiers sont nommés grâce au préfixe commun et à un numéro. Par exemple, si le préfixe commun est "my\_experiment", le répertoire dédié à l'analyse contiendra les fichiers (cf. figure 2.1-2) :

- Fichier Context : "my\_experiment.ctx",
- Fichier Particle Edge Trace : "my\_experiment.edg",
- Fichier List : "my\_experiment.lst",
- Fichier Notes : "my\_experiment\_notes.txt",
- Fichier de mesures : "my\_experiment.csv",
- Fichier Run Summary : "my\_experiment\_run\_summary.txt",
- Fichiers Image (Calibration et Collages) : "cal\_image\_000001", "my\_experiment\_00000001.tif", "my\_experiment\_00000002.tif", etc.



*Figure 2.1-2 : Contenu du répertoire dédié aux fichiers bruts issus du FlowCAM et de Visual SpreadSheet*

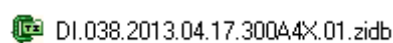
Un fichier List avec l'extension \*.lst est généré pour chaque analyse. Les fichiers List sont des fichiers de données primaires des expérimentations pour le FlowCAM. Ce sont des fichiers texte ASCII qui enregistrent les champs des données collectées pour chacune des particules.

Un fichier Context avec l'extension \*.ctx est généré avec chaque fichier List. Il enregistre les informations de contexte (réglages) qui était en vigueur lorsque le fichier List a été créé ou lorsque l'expérimentation a débuté. Les fichiers texte "Notes" et "Run summary" accompagnent le fichier Context. Ces fichiers contiennent un bref descriptif des paramètres de numérisation et du flux, et les notes entrées dans la boîte de dialogue de Visual SpreadSheet (le plus souvent, les métadonnées).

Le fichier Particle Edge Trace accompagne le fichier .lst et enregistre la séquence de pixels qui forment le bord de chaque particule. Visual Spreadsheet utilise le suivi des bords afin de calculer plusieurs propriétés nouvelles des particules et tracer le contour de la celle-ci dans l'image affichée (si sélectionnée).

### **Fichiers disponibles en sortie de ZooPhytoImage (Compressed Data)**

Suite au traitement d'images effectué lors de l'étape d'importation des données dans ZooPhytoImage, trois types de fichier sont disponibles : \*\_dat1.zim, \*\_dat1.RData et des fichiers vignette \*.jpg (ou \*.png). Ces derniers sont alors réunis dans un fichier compressé \*.zidb (le caractère '\*' représente le préfixe commun que l'utilisateur a choisi pour nommer l'expérimentation).



Par exemple, si le préfixe commun est "my\_experiment", le fichier compressé nommé "my\_experiment.zidb" contiendra les fichiers (cf. figure 2.1-3) :

- Fichier ZIM (**ZooImage Metadata/Measurements**) : "my\_experiment\_dat1.zim",
- Fichier RData : "my\_experiment\_dat1.RData",
- Fichiers Image (vignettes) :  
"my\_experiment\_1.jpg", "my\_experiment\_2.jpg", etc  
**ET/OU**  
"my\_experiment\_1.png", "my\_experiment\_2.png", etc.

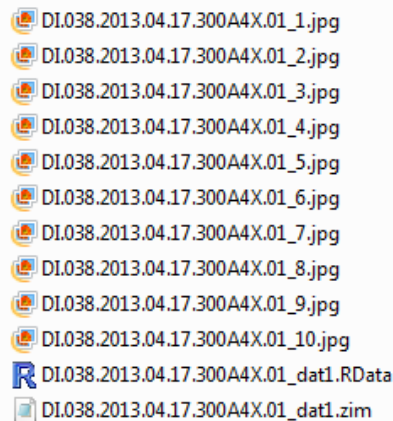


Figure 2.1-2 : Contenu du fichier compressé ZIDB en sortie de ZooPhytoImage

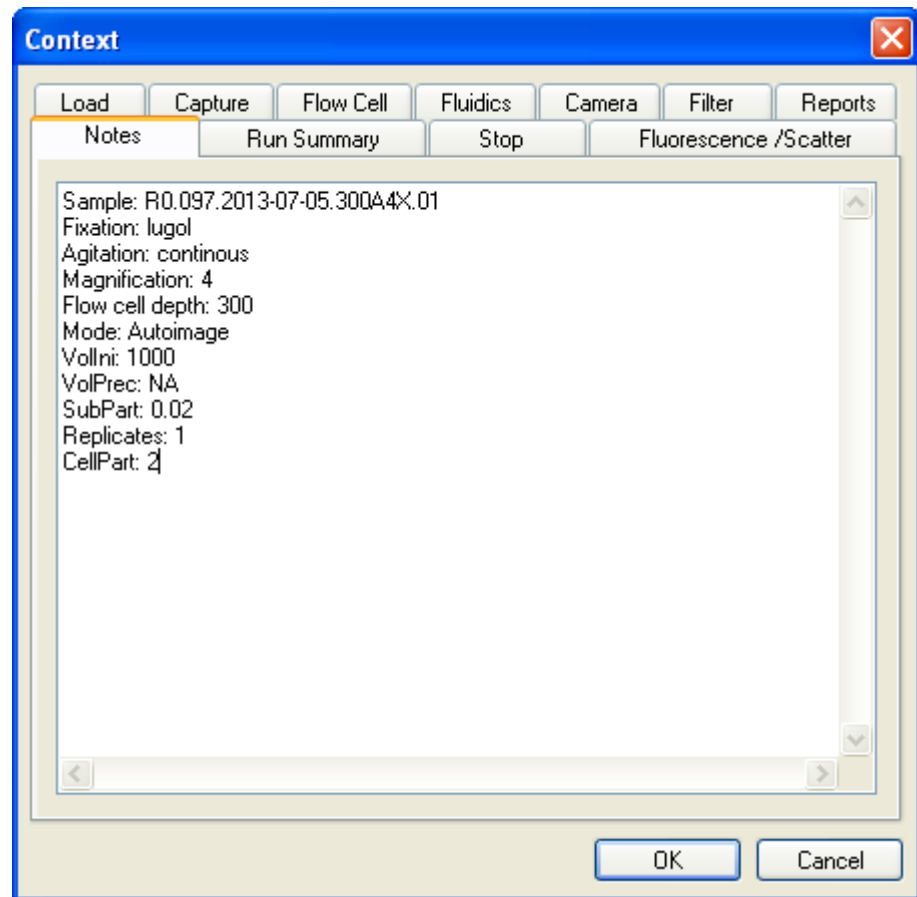
Ce format de compression (ZIDB) permet d'une part, de réduire la taille mémoire des données pour le stockage mais également de réduire le temps d'accès aux données de type vignettes pour un affichage rapide et une interactivité aisée sous R.

## 2.2. Saisie des métadonnées

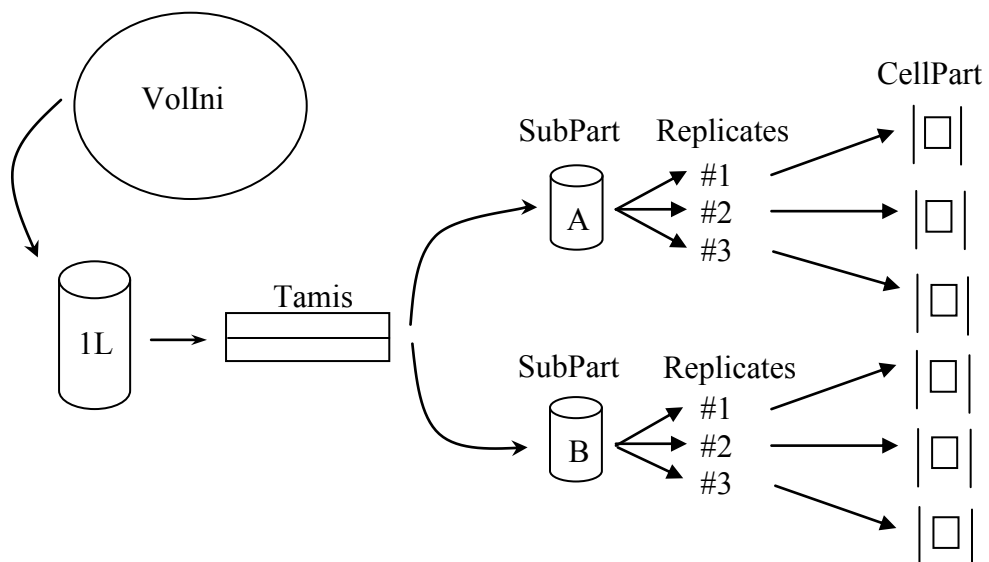
La totalité des informations essentielles pour le traitement des données brutes (transformation des collages en vignettes, calcul de nouvelles mesures, etc.) sont stockés dans le fichier "Contexte" (.ctx) fourni en fin de numérisation par Visual SpreadSheet (logiciel de configuration et d'acquisition associé au FlowCAM).

Cependant, pour la création du fichier ZIM, d'autres informations, relatives à l'échantillon et à la numérisation, doivent être fournies par l'utilisateur (notamment pour le calcul des abondances, de la biomasse et des spectres de taille). La solution envisagée consiste donc à entrer ces renseignements dans les notes de l'expérimentation (lors de la configuration du contexte de numérisation dans Visual SpreadSheet, cf. figure 2.2-1). Ces informations concernent (cf. figure 2.2-2) :

- ***Vollni*** = volume échantillonné en mer (bouteille Niskin, flacon sur perche, etc.)  
→ unité choisie par l'utilisateur (**Litre par défaut ???**);
- ***volPrec*** = précision sur l'échantillonnage réalisé en mer (non utilisée);
- ***SubPart*** = fraction introduit dans l'appareil (si *vollni* = 1L et 10 mL introduit dans l'appareil, alors  $SubPart = 10\text{mL}/1000\text{mL} = 0.01$ );
- ***Replicates*** = nombre de réplicats (si l'expérimentation est répétée 3 fois, alors  $Replicates = 3$ );
- ***CellPart*** = fraction réellement numérisé (si *SubPart* = 10 mL mais que le volume réellement numérisé est de 5 mL, alors  $CellPart = 10\text{mL}/5\text{mL} = 2$ ).



*Figure 2.2-1 : Onglet "Notes" du menu "Context" de Visual Spreadsheet.*



*Figure 2.2-2 : Mesures associées aux différentes étapes de l'échantillonnage.*



### 3. Interface Graphique Utilisateur

Ce cahier des charges a été élaboré dans le but de définir les besoins nouveaux des utilisateurs en ce qui concerne le traitement des données brutes issues du FlowCAM, mais également de cibler le type d'utilisateur susceptible d'interagir avec le système. Pour cela, deux modes d'utilisation sont disponibles : "*Standard*" (utilisation basique et restreinte du logiciel, configuration des algorithmes par défaut) ou "*Avancé*" (utilisation avancée des différents modules, paramétrage des algorithmes).

#### 3.1. Présentation générale

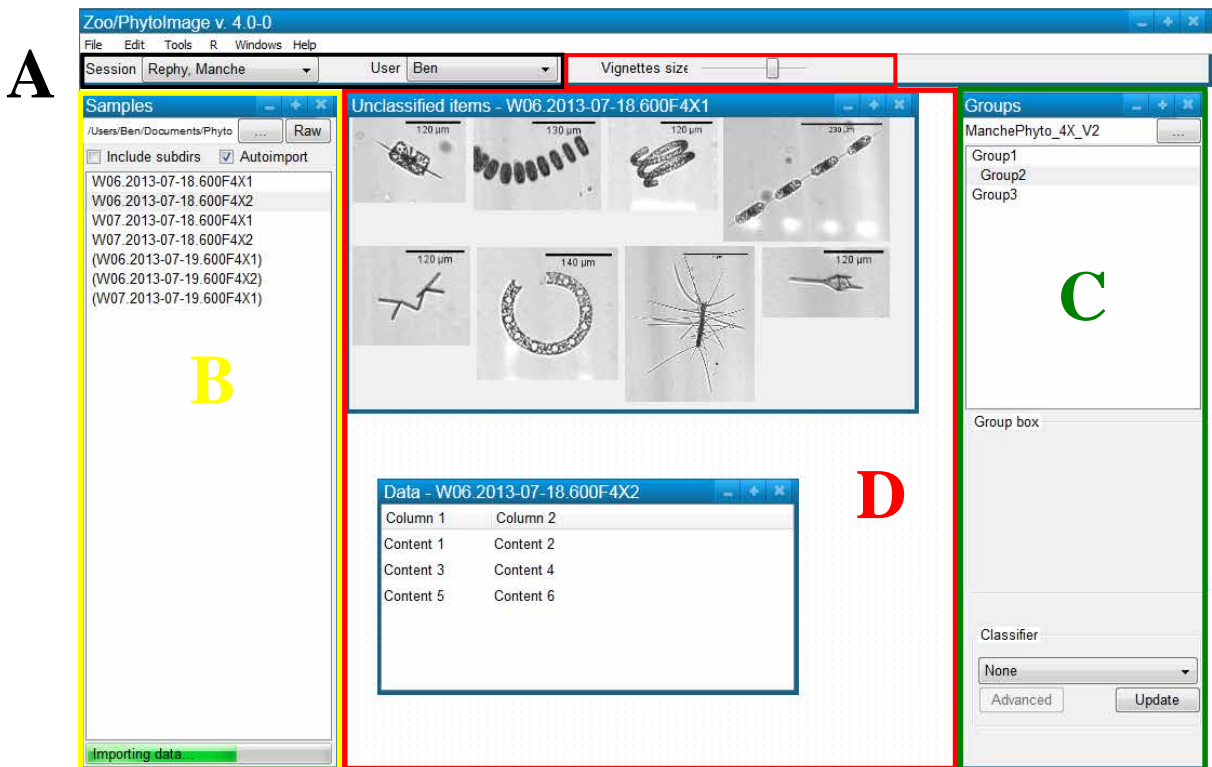


Figure 3.1-1 : Décomposition de l'application en modules

**A** Comme montré sur la figure 3.1-1, la barre d'outils générale est composée de 2 éléments :

- **Session** : Sélection de la session de travail (Rephy Manche, Rephy Méditerranée, Rephy Atlantique, etc.) → disponibilité des outils créés dans le contexte souhaité.
- **User** : Choix de l'utilisateur → chargement du fichier de configuration contenant les préférences et les options de l'utilisateur sélectionné.

L'interface graphique est décomposée en trois zones fonctionnelles :

**B** Zone de départ (échantillons, données à traiter),

**C** Zone d'arrivée (objectifs : training set, test set, classifier),

**D** Zone de travail (fenêtres interactives).

La solution qui a été proposée pour le logiciel ZooPhytoImage version 4.0-0 fait ressortir des concepts importants qui justifient le découpage de l'application en modules. Un module regroupe un ensemble de fonctionnalités en rapport avec le même concept. Voici comment va être découpé de manière générale l'interface graphique utilisateur du logiciel ZooPhytoImage v.4.0-0 :

- **Importation des données** : dans ce module sera présenté tout ce qui est relatif à l'échantillon sélectionné (vignettes, mesures, graphes, etc.).
- **Création d'un training/test set** : ce module concerne la création d'un training set ou d'un test set, ou d'un classifieur.
- **Création d'un outil de reconnaissance** : ce module concerne la création d'un classifieur.
- **Reconnaissance automatique** : lorsqu'un classifieur a été construit (ou sélectionné), l'utilisateur peut entrer dans la phase de reconnaissance automatique des vignettes d'un (ou plusieurs) échantillon(s).
- **Validation (correction statistique de l'erreur)** : lorsque qu'un (ou plusieurs) échantillon(s) a (ont) été traité(s), l'utilisateur peut valider la reconnaissance automatique et visualiser l'effet des corrections apportées.

*Dans les sections suivantes, les détails et concepts de chacun de ces modules sont présentés.*

### 3.2. Importation des données

Ce premier module concerne l'importation des données. Comme montré sur la figure 3.2-1, la fenêtre "**Samples**" contient tous les échantillons disponibles (traités ou à traiter).

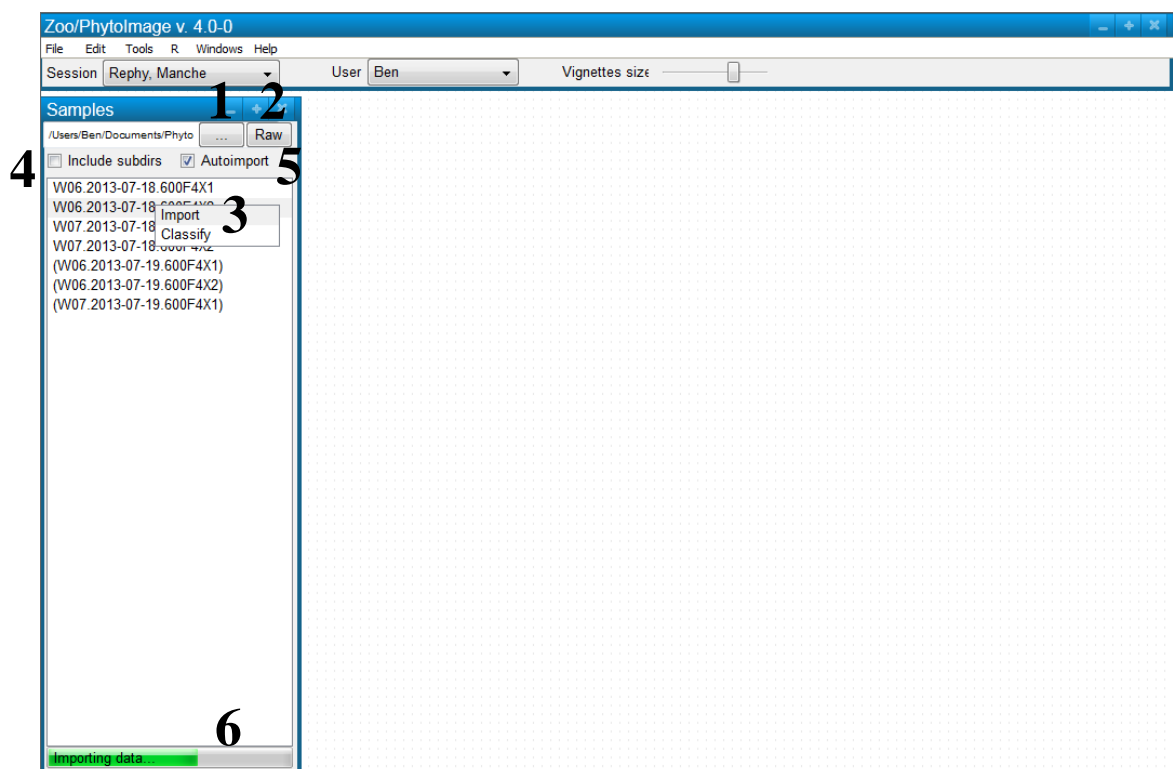


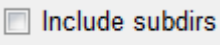
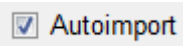



Figure 3.2-1 : Fenêtre d'importation des données

- 1**  Sélection et affichage du répertoire courant (ici, "\\Users\Ben\Documents\Phyto"). Ce répertoire contient les fichiers compressés disponibles au format ZIDB (ce sont les échantillons SANS parenthèses).  
**Définir un chemin standard par défaut (le plus simple possible) dans lequel l'utilisateur DOIT ranger ses échantillons (afin d'éviter que des échantillons à traiter ne se retrouvent dans de multiples répertoires sur l'ordinateur) ???**
- 2**  Sélection du répertoire contenant les fichiers bruts issus du FlowCAM. Si des échantillons n'ont pas encore été importés (traitement d'images + création des ZIDB), ceux-ci sont notés ENTRE parenthèses.
- 5** Ces données brutes sont utilisées pour le traitement d'image et la création du ZIDB effectués par ZooPhytoImage (après clic droit sur l'échantillon ciblé et sélection de "**Import**").
- 5** Cependant, si de nouvelles mesures doivent être réalisées sur des échantillons déjà importés, ces données brutes seront utiles pour retraiter les images et recréer les ZIDB (après clic droit sur l'échantillon ciblé et sélection de "**Import**").
- 3**  Recherche les échantillons traités ou non traités dans les sous-répertoires du répertoire courant.
- 4**  Importation AUTOMATIQUE des données (traitement des images et création du ZIDB à partir des données brutes du FlowCAM) si simple clic souris sur un échantillon non traité.
- 5**  Barre de progression pour l'importation des données (traitement d'image et création du fichier ZIDB EN TACHES DE FOND).

Ce module permet également de pouvoir explorer les données des échantillons sélectionnés grâce aux différentes fenêtres interactives. En effet, une fois l'échantillon importé dans ZooPhytoImage (c'est-à-dire les images traitées et le fichier compressé ZIDB créé), des graphes ainsi que les vignettes et le tableau de mesures associé sont affichées dans la zone de travail.

#### AVANTAGES

- Importation en tâche de fond (traitement d'images et création du ZIDB)
  - Pas d'utilisation directe d'ImageJ (pour le traitement d'images)
  - Paramétrage automatique grâce aux fichiers bruts issus du FlowCAM
  - Pas de création manuelle du fichier ZIDB
- Exploration interactive aisée des données avec visualisation des vignettes et des mesures grâce à des graphes, etc.

### 3.3. Création d'un training/test set

Ce second module concerne la création d'un training set ou d'un test set.

- **Zone d'arrivée - Groups**

Contient tous les groupes définis dans le training set ou le test set visé. L'arborescence des différents groupes est ici respectée. Pour la création d'un training/test set, aucun classifieur ne doit être sélectionné (cf. figure 3.3-1).

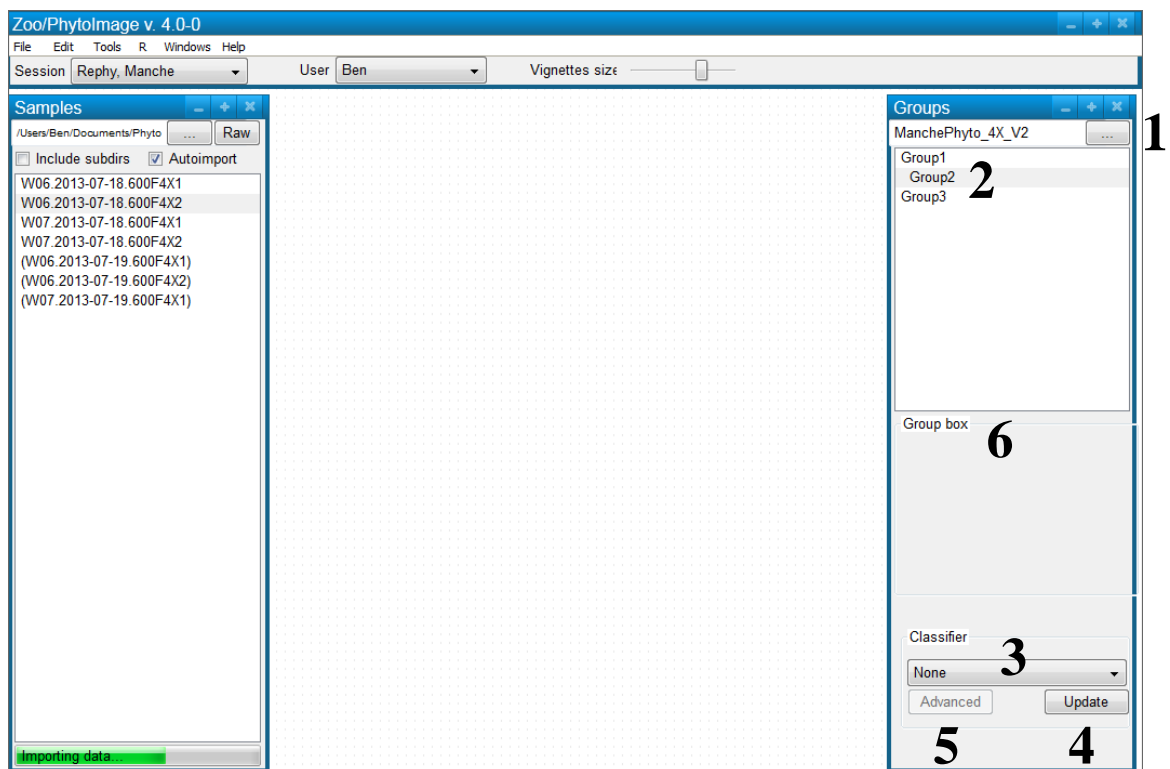

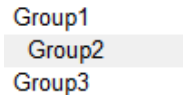

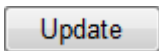


Figure 3.3-1 : Sélection du training/test set

- 1  Sélection et affichage du set d'apprentissage ou du set de test à compléter (ici, "ManchePhyto\_4X\_V2"). Ce répertoire peut contenir des vignettes déjà triées par l'utilisateur.
- 2  Arborescence des groupes du set sélectionné
- 3  **Classifier** : Sélection d'un classifieur pour la création d'un outil de reconnaissance.  
Dans ce contexte, aucun classifieur n'est sélectionné : *None* → classification manuelle active.
- 4  **Classifier** : Mise à jour des performances de l'outil de reconnaissance et affichage des résultats (vignettes, graphes, etc.).

**5** **Advanced**

**Classifier** : Optimisation des performances de l'outil de reconnaissance, réglages des paramètres du classifieur (paramètres optionnels, nombre de folds pour la validation croisée, etc.), visualisation des performances par outils spécifiques, ...

Ce bouton n'est disponible que pour un mode avancé d'utilisation (mode utilisateur "Avancé").

**6** **Group box**

Affichage des vignettes caractéristiques du groupe sélectionné dans la liste.

• **Zone de travail - Unclassified items**

Contient toutes les vignettes non triées de l'échantillon sélectionné. Cette zone permet également de naviguer dans le tableau de mesures, de visualiser des scatterplots, mais également d'interagir entre ces différents outils d'exploration de données (figure 3.3-2).

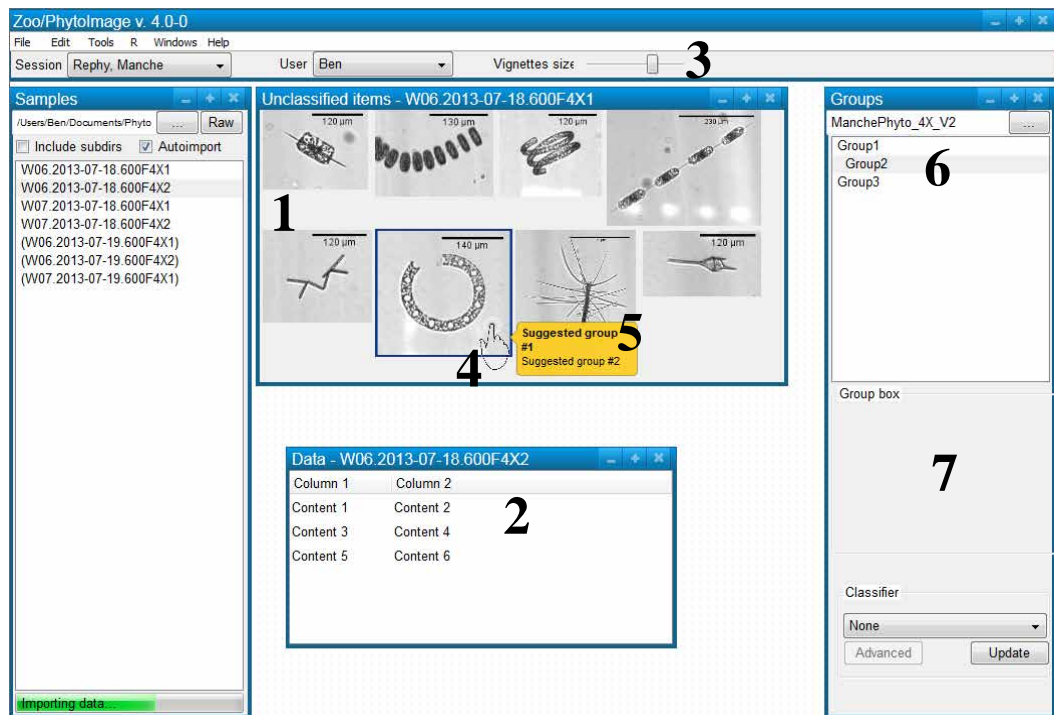



Figure 3.3-2 : Classification manuelle des vignettes dans le training/test set

- 1** **Unclassified items - W06.2013-07-18.600F4X1** Affichage des vignettes de l'échantillon sélectionné
- 2** **Data - W06.2013-07-18.600F4X2** Affichage du tableau de mesures associé aux vignettes
- 3** **Vignettes size**  Curseur permettant de régler la taille des vignettes

### Classification manuelle de vignettes pour la création d'un training/test set

- Classifier → None
- Passage du curseur de la souris sur une vignette
  - 4** ➤ Affichage de suggestions (sur la base d'une mesure de similarité par rapport aux vignettes préalablement rangées dans les groupes du set d'apprentissage) après un court laps de temps.
    - Groupe de plus haute probabilité d'appartenance :
      - 5** ▪ caractères gras dans la liste des suggestions
      - 6** ▪ sur-brillance dans la liste des groupes du set sélectionné
      - 7** ▪ affichage de vignettes caractéristiques du groupe dans la fenêtre "*Group box*"
        - si OK → ENTREE, sinon → déplacement manuel de la vignette (drag and drop).

**Remarque** : il est possible d'interagir entre les différents outils proposés dans cette zone de travail. En effet, en cliquant sur un objet (vignette, point dans l'espace de projection, ligne du tableau de mesures, ...) dans une fenêtre, l'objet correspondant dans les autres fenêtres sera mis en avant. Ceci pourra permettre de renseigner l'utilisateur sur les objets les plus suspects et ainsi comprendre les erreurs commises par l'outil de reconnaissance (objets aux frontières de deux groupes, vignettes peu lisibles, etc.).

#### **AVANTAGES**

- Classification intuitive (aide à la classification manuelle sur base de suggestions)
  - Suggestions pour chacune des vignettes
  - Affichage de vignettes caractéristiques pour le groupe sélectionné
- Classification manuelle effectuée sous R (et non via un logiciel indépendant).

### 3.4. Création d'un outil de reconnaissance

Ce second module concerne la création d'un classifieur sur base d'un training set.

- **Zone d'arrivée - Groups**

Contient tous les groupes définis dans le training set sélectionné. L'arborescence des différents groupes est ici respectée. Pour la création d'un outil de reconnaissance, six algorithmes sont disponibles (cf. figure 3.4-1).

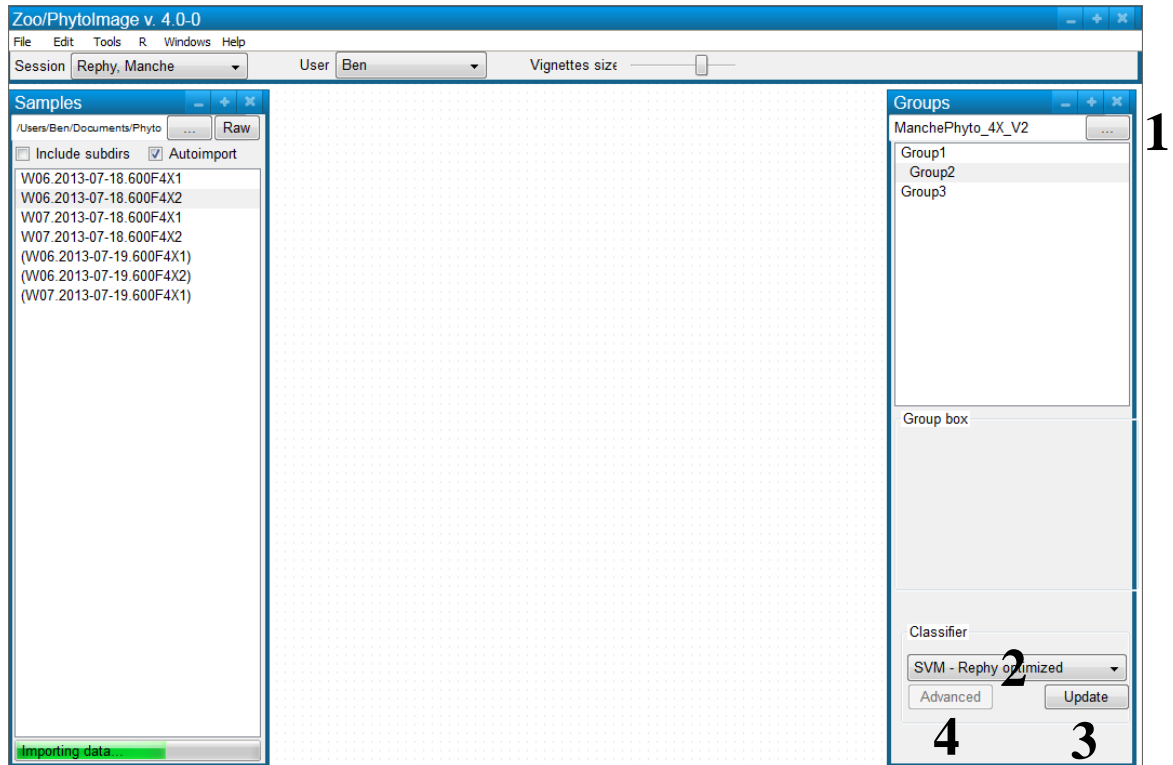


Figure 3.4-1 : Sélection du classifieur

- 1  Sélection du set d'apprentissage (ici, "ManchePhyto\_4X\_V2").
- 2  **Classifier** : Sélection d'un classifieur pour la création d'un outil de reconnaissance. Dans ce contexte, sélection d'un classifieur : 6 algorithmes disponibles (LDA, Tree, KNN, LVQ, Neural Network, Random Forest).
- 3  **Classifier** : Mise à jour des performances de l'outil de reconnaissance et affichage des résultats (vignettes, graphes, etc.).
- 4  **Classifier** : Optimisation des performances de l'outil de reconnaissance, réglages des paramètres du classifieur (paramètres optionnels, nombre de folds pour la validation croisée, etc.), visualisation des performances par outils spécifiques, ...  
Ce bouton n'est disponible que pour un mode avancé d'utilisation (mode utilisateur "Avancé").

- **Zone de travail - Unclassified items**

Présente un aperçu des vignettes classées dans chacun des groupes définis dans le set d'apprentissage par l'outil de reconnaissance sur base d'une validation croisée. Cette zone permet également d'afficher des outils d'évaluation des performances du classifieur (matrice de confusion, erreur par groupes, etc.).

### AVANTAGES

- Affichage de l'évolution des scores de performance au fur et à mesure des changements dans le training set (fusion/séparation de groupes, ajout/suppression de vignettes, etc.)
- Navigation aisée et passage rapide entre les différents algorithmes.

## 3.5. Reconnaissance automatique

Ce second module concerne la reconnaissance automatique.

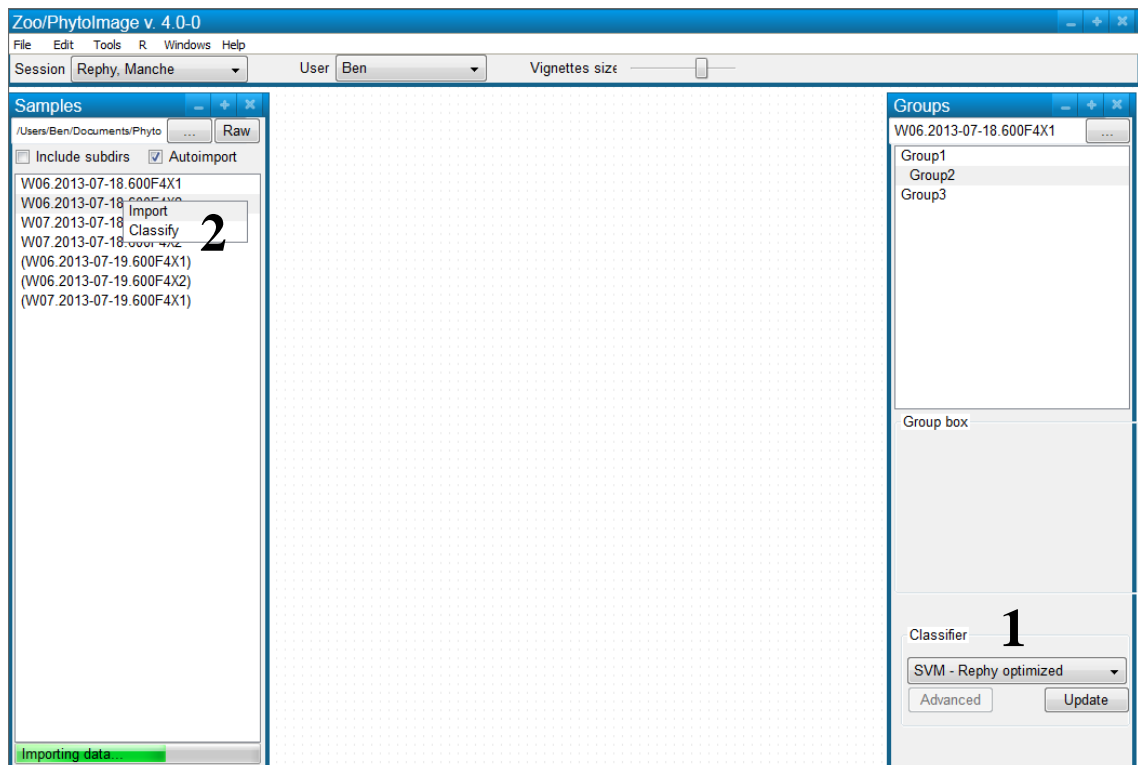


Figure 3.5-1 : Lancement du module de reconnaissance automatique

Pour démarrer l'étape de reconnaissance automatique (cf. figure 3.5-1) :

- sélection d'un outil de reconnaissance
- clic droit sur l'échantillon souhaité puis sélection de l'option "Classify".



- **Zone d'arrivée - Classes**

Contient tous les groupes connus du classifieur. L'arborescence des différents groupes est ici respectée. Les vignettes classées automatiquement sont rangées dans ces groupes (cf. figure 3.5-2).

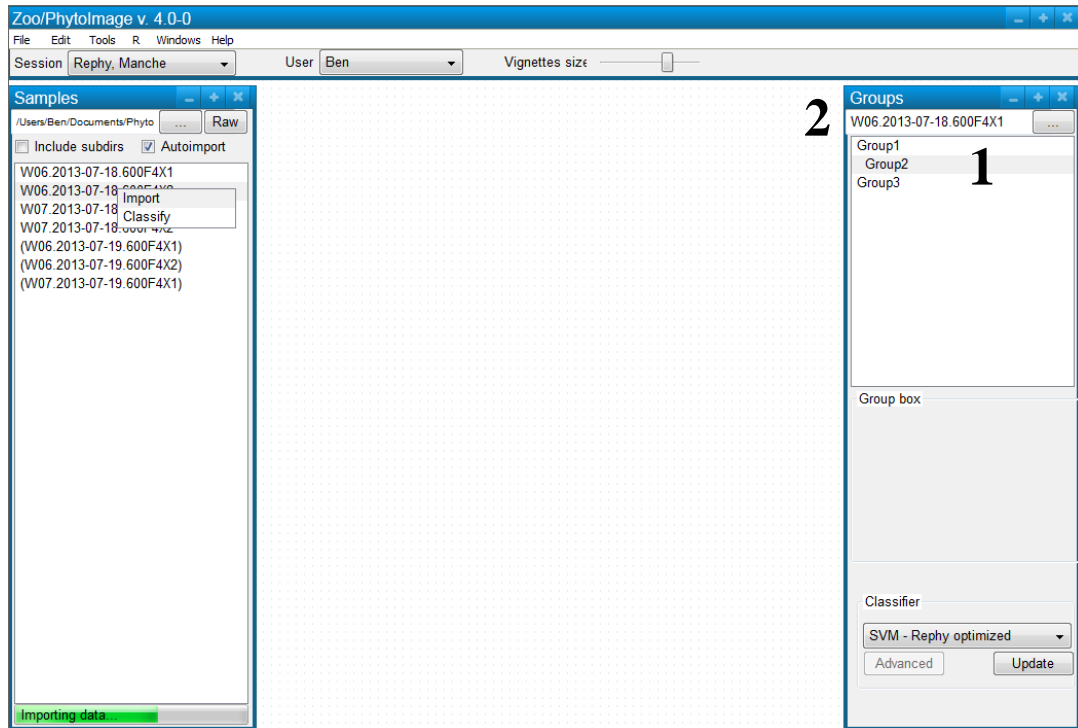


Figure 3.5-2 : Reconnaissance automatique des vignettes

- |  |   |
|--|---|
| <p><b>1</b> Group1<br/>Group2<br/>Group3</p> | Arborescence des groupes dans lesquels sont rangées les vignettes classées automatiquement. |
| <p><b>2</b> W06.2013-07-18.600F4X1</p>       | Nom de l'échantillon classé automatiquement.  |

- **Zone de travail - Suspect items**

### Classification automatique de vignettes sur la base d'un classifieur

- Sélection d'un classifieur (ex : *SVM - Rephy optimized*)
- Sélection d'un (des) échantillon(s) à traiter.
- Affichage des résultats de classification (cf. figure 3.5-3).

- 1** ➤ Sélection d'un groupe de vignettes classées dans la liste
- 2** ▪ affichage des vignettes du groupes dans la fenêtre "Group box"
- 3** ▪ affichage des mesures pour chacune des vignettes du groupe
- 4** ▪ affichage des estimations d'abondance, de biomasse et des spectres de taille.
- 5** ➤ Affichage des vignettes les plus suspectes.

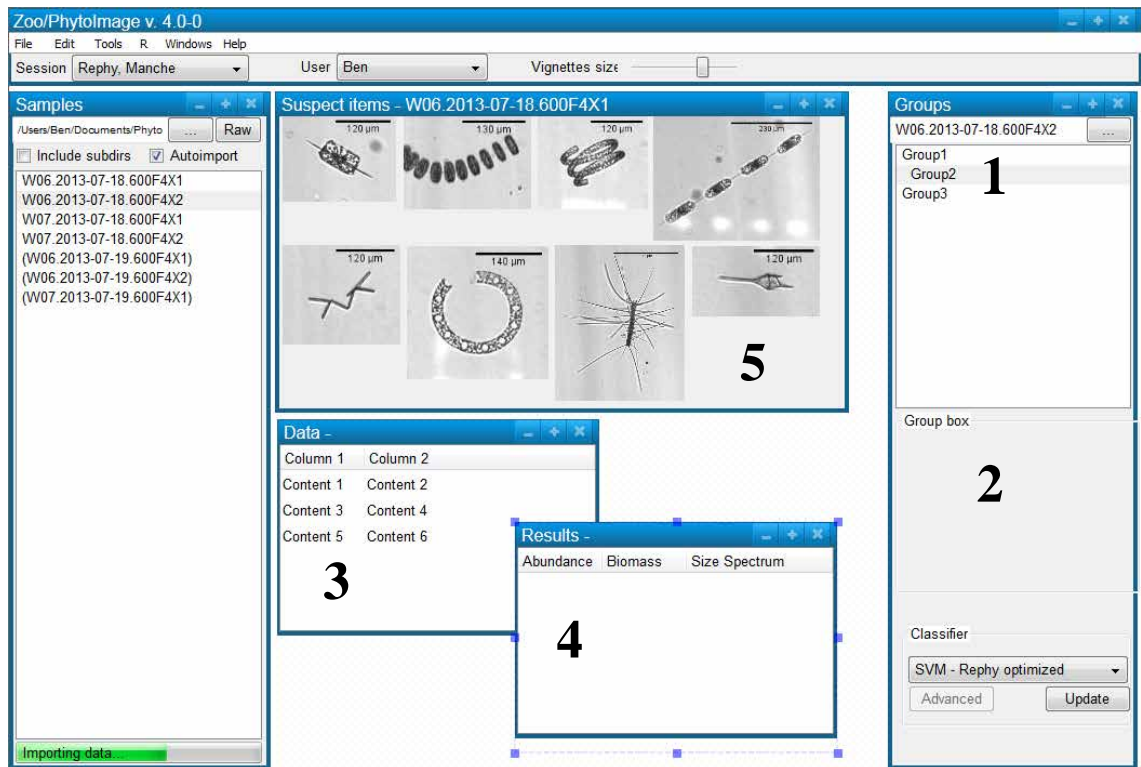


Figure 3.5-3 : Affichage des résultats de la reconnaissance automatique

**Remarque :** de la même manière que pour les modules précédents, il est possible d'interagir entre les différents outils proposés dans cette zone de travail.

### AVANTAGES

- Visualisation des vignettes classées par groupe
- Navigation aisée et passage rapide entre les différents algorithmes.

### 3.6. Validation (correction statistique de l'erreur)

Ce dernier module concerne la validation des résultats de l'identification automatique et la correction statistique de l'erreur. Après affichage des résultats de classification automatique, l'utilisateur peut choisir de corriger la prédiction d'appartenance d'une vignette à un groupe en déplaçant manuellement cette vignette dans son nouveau groupe (drag and drop). Ce déplacement active alors le module de correction de l'erreur (cf. figure 3.6-1).

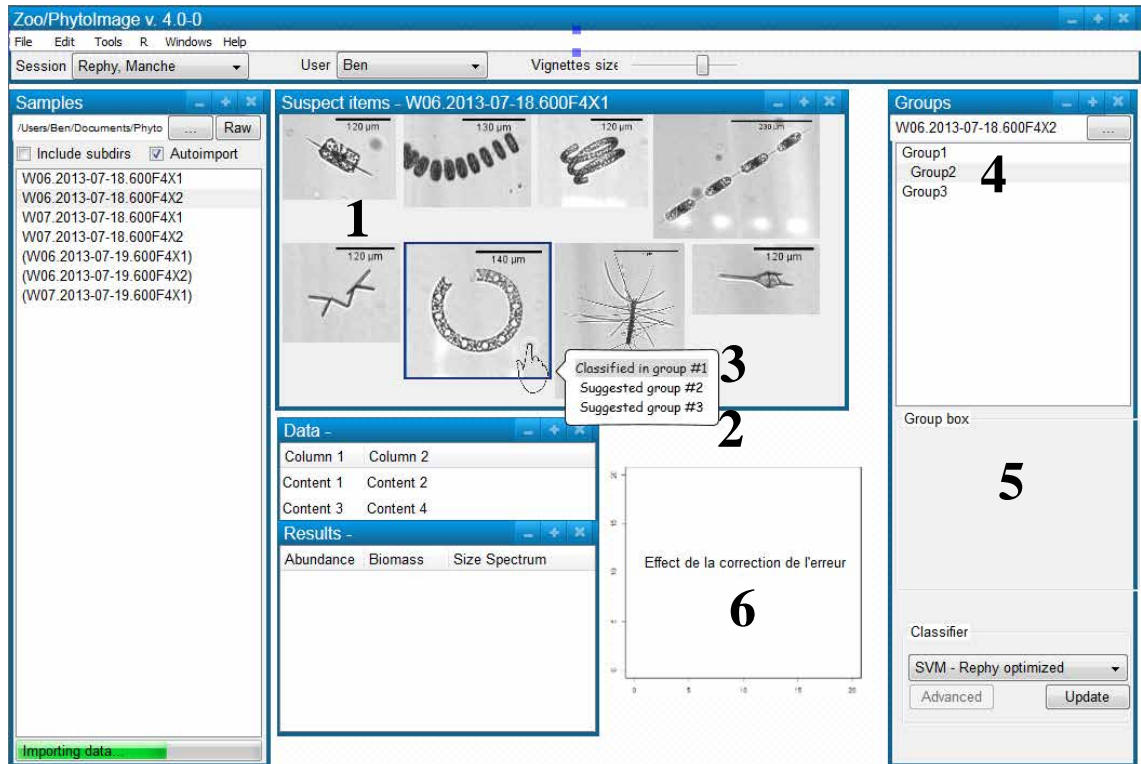


Figure 3.6-1 : Module de correction statistique de l'erreur

- **Zone de travail - Suspect items**

1 Présente les vignettes les plus suspectes au sens d'un critère (basé sur des critères algorithmiques ou bio-écologiques tels que la probabilité de présence d'une espèce selon la saison, la zone géographique, etc.). Cette zone permet également de naviguer dans le tableau de mesures, de visualiser des scatterplots, etc.

2 De la même manière que pour la construction d'un set d'apprentissage (classification manuelle), le passage du curseur de la souris sur une vignette permet d'obtenir une liste composée du groupe dans lequel la vignette a été classée automatiquement mais également de suggestions concernant les groupes d'appartenance possibles.

## Classification manuelle de vignettes pour la correction de l'erreur

- Passage du curseur de la souris sur une vignette
  - 2 ○ Affichage de suggestions (sur la base d'une mesure de similarité par rapport aux vignettes préalablement rangées dans les groupes du set d'apprentissage) après un court laps de temps.
    - Groupe dans lequel la vignette a été classée automatiquement :
      - 3 ▪ caractères gras dans la liste des suggestions
      - 4 ▪ sur-brillance dans la liste des groupes du set sélectionné
      - 5 ▪ affichage de vignettes caractéristiques du groupe dans la fenêtre "*Group box*"
        - si OK → ENTREE, sinon → déplacement manuel de la vignette (drag and drop).

- 6 Lors de cette étape de validation, l'utilisateur peut visualiser l'effet de la correction de l'erreur. Pour cela, un graphe présentant l'évolution de l'erreur de prédiction est proposé. L'utilisateur peut également évaluer la pertinence de la correction de l'erreur grâce à un critère subjectif basé sur la visualisation des vignettes les plus suspectes (sous forme de collage) à chaque itération.

## 4. Outils complémentaires

### 4.1. Fichier de configuration

Un fichier INI (fichier de configuration) peut être créé afin de sauvegarder la configuration des préférences pour chaque utilisateur. En effet, des paramètres tels que les options d'importation des données, les fenêtres à afficher, la position de ces différentes fenêtres, la taille des vignettes, etc. peuvent être différents d'un utilisateur à un autre. C'est pourquoi, dans un souci d'ergonomie, d'aisance et de rapidité, il est souhaitable que la configuration de chaque utilisateur soit sauvegardée à la fin de chaque session. Ceci pourrait alors permettre un chargement automatique des préférences à chaque réouverture de session pour un même utilisateur.

### 4.2. Historique des actions

Après chaque modification apportée durant le traitement d'un échantillon (déplacement de vignettes, suppression de vignettes, nouveaux paramétrage d'un algorithme, etc.), il serait judicieux de pouvoir annuler cette modification via un bouton "*Undo*" (comme illustré sur la figure 4.2-1).

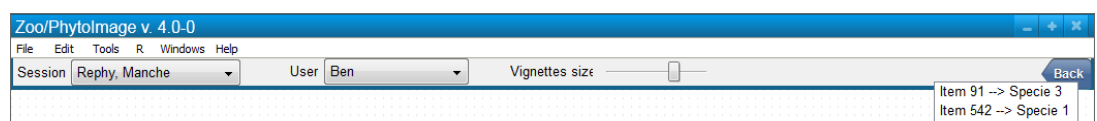


Figure 4.2-1 : Exemple de bouton "*Undo*"

Un clic souris sur ce bouton permet alors d'afficher une liste réduite aux 10 (?) dernières modifications. L'utilisateur peut alors choisir l'événement à annuler.

De plus, après chaque modification, un événement est créé dans un journal. Ce dernier (communément appelé "log file" ou "log") répertorie alors l'ensemble des actions affectant le processus classées par ordre chronologique.

Cette technique de journalisation peut s'organiser comme suit :

- une ligne par événement,
- moment exact (date, heure, minute, seconde) de l'événement.

L'utilisateur peut alors avoir accès à tout moment à un historique détaillé et classé chronologiquement de ses actions. Ce fichier peut également être utilisé comme outil de restauration grâce aux fonctions "**Undo**" (annulation du dernier événement) et "**Redo**" (annulation du dernier "**Undo**"), comme illustré sur la figure 4.2-1. Cette fenêtre pourra être disponible dans le menu "**Tools**" de la barre de menu principal.

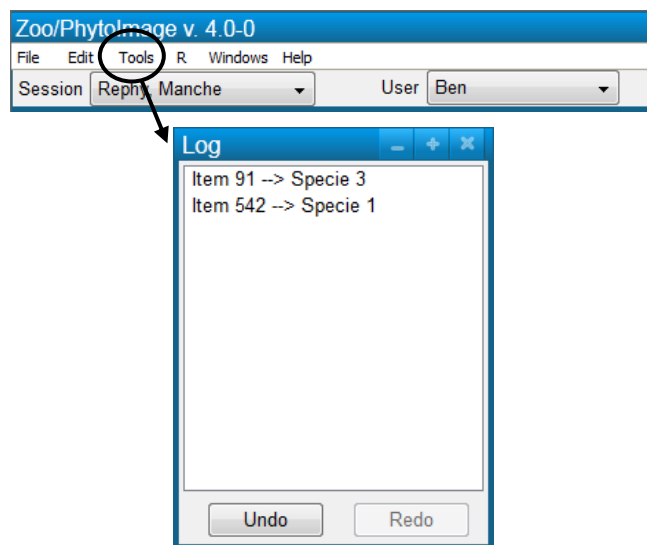


Figure 4.2-2 : Exemple de fichier Log

### 4.3. Dénombrement des cellules en colonie

Les cellules agencées en colonies ne sont pas encore identifiées individuellement par le logiciel ZooPhytoImage. Une colonie est donc considérée comme un objet unique. Or, l'ensemble des estimateurs de la biomasse sont calibrés essentiellement sur l'abondance en termes de cellules par unité de volume et non en termes de colonies. Il n'est, de ce fait, pas possible de comparer les comptages cellulaires avec les comptages automatiques. Les espèces coloniales étant fréquentes dans les échantillons observés et les mesures d'abondance en termes de cellules représentant une information capitale, ce travail est donc nécessaire afin de pouvoir comparer les résultats obtenus par le système couplé FlowCAM/PhytoImage avec ceux des dénombrements REPHY.

Question : intégration d'un module de dénombrement interactif dans ZooPhytoImage ???

- Affichage de la vignette et proposition automatique d'une estimation du nombre de cellules dans la colonie, basée sur un abaque (par exemple, pour *Phaeocystis*) ou méthodes de segmentation (par exemple, pour *Lauderia*), comme illustrée sur la figure 4.3-1) ???
- Correction manuelle (par clics souris) ou validation (cf. figure 4.3-2) ???

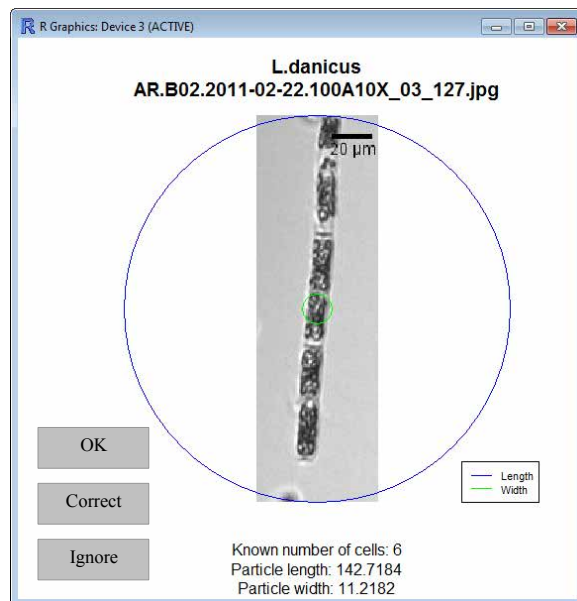


Figure 4.3-1 : Estimation du nombre de cellules

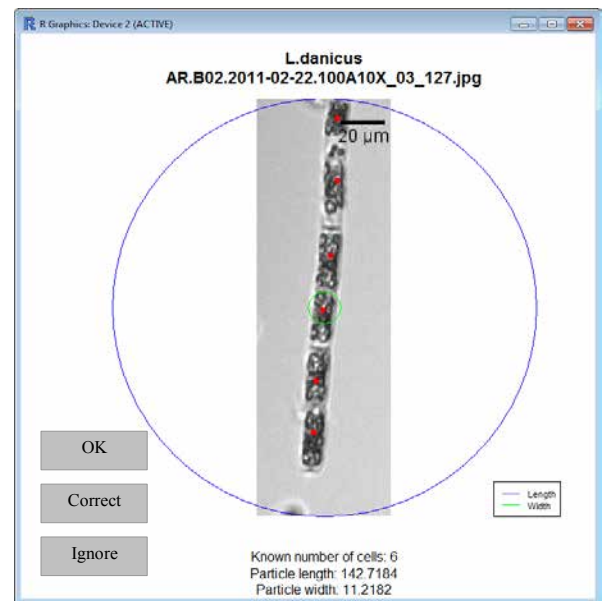


Figure 4.3-2 : Correction manuelle

## Bibliographie indicative

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

Dereume-Hancart F., 2013. Correction statistique de l'erreur dans le cadre de la classification automatique du plancton. Mémoire de fin d'étude. Université de Mons Hainault, UMONS - Service d'Ecologie Numérique des Ecosystèmes Aquatiques : 71p.

Grosjean P. & Denis K., 2010. Rapport Université de Mons – ZooPhytoImage – Optimisation du set d'apprentissage REPHY en 2010.

Tunin-Ley A. & Maurer D., 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. Rapport RST/LER/AR/11/002

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.

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.