

Rapport pour le comité d'accompagnement, troisième année de thèse

Guyliann Engels

02-12-2020

Avancement de la recherche

Titre provisoire de la thèse:

...

Résumé :

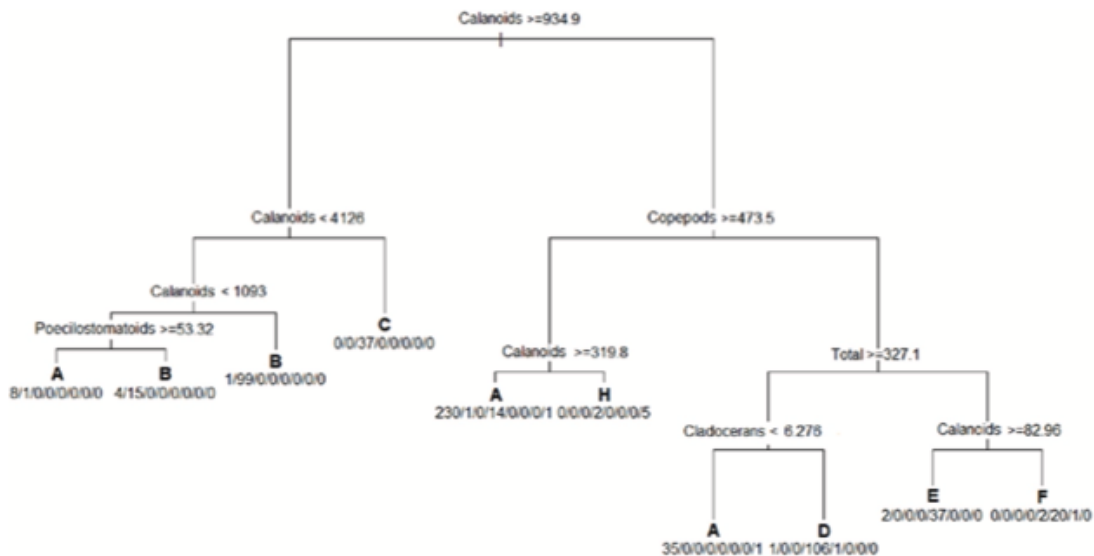
Le suivi de la dynamique du zooplancton est capital. Les organismes sont très intéressants comme indicateurs de changements.

...

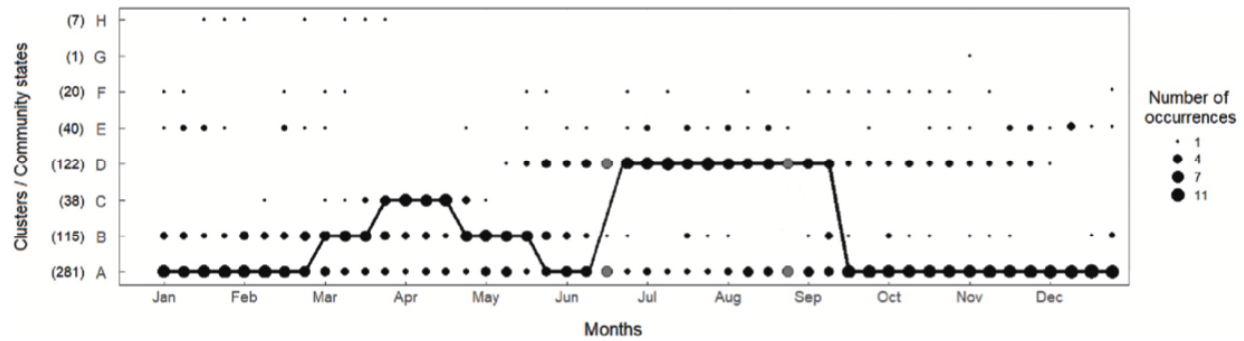
Chapitre 1 : Suivi de la dynamique zooplanctonique en subsurface au cours du temps en baie de Calvi Corse

Le suivi de la communauté planctonique est un élément majeur dans ...

L'originalité de cette étude est la définition de 8 stades qui représente l'état de la communauté planctonique. Ces stades sont déterminés avec un arbre de partitionnements.



Le nombre d'occurrence maximal des stades permet de définir un profil annuel type.



Avancement Un premier article est paru sur cette année sur cette thématique (Fullgrabe et al. 2020). Entre 2004 et 2016, 14 groupes taxonomiques ont été étudié. Néanmoins, cette première étude descriptive va être complétée avec des données collectées en 2017 et 2018. Le nombre de groupes étudié va également être augmenté. Enfin, de nouveaux outils vont être employés afin de déterminer une année typique.

Chapitre 2 : Comparaison entre les organismes planctoniques de subsurface et de profondeur en baie de Calvi Corse.

...

La méthode d'échantillonnage la plus utilisée est le trait de filet. Cette méthode est peu coûteuse et simple de mise en œuvre. Elle permet d'avoir une estimation de l'abondance planctonique.

On recense principalement 3 types de traits : les traits verticaux, les traits horizontaux et les traits obliques. Les traits verticaux sont les traits les plus utilisés.

La station de recherche de Calvi réalise des prélèvements via des traits de plancton horizontaux depuis 2003. Ils ont également débuté des prélèvements verticaux depuis 2012.

30 échantillons ont été sélectionnés afin de réaliser cette comparaison.

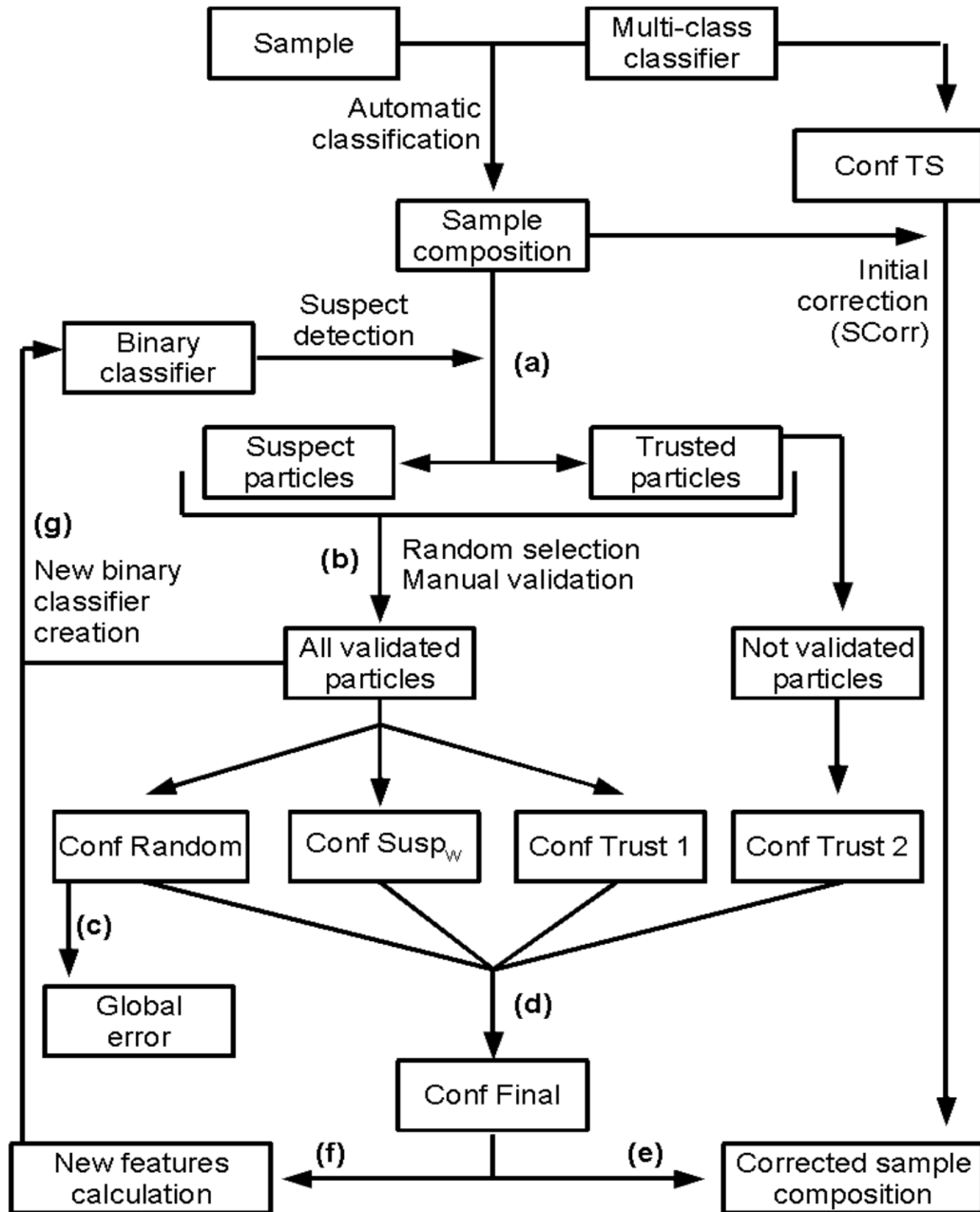
Avancement : Des recherches sont menées afin de proposer une comparaison entre les deux types de traits. Cette comparaison a pour objectif de comparer les valeurs obtenues avec les autres séries temporelles menées sur le plancton en Europe (ajouter le nombre de série en cours ...).

Chapitre 3 : Correction des erreurs due à la classification automatique via la validation de suspects.

La classification automatique est très employée actuellement afin de classer rapidement des images de plancton. La classification se base sur un training set et d'un algorithme qui vont utiliser ce set d'apprentissage afin de classer les items inconnus. Cette méthode seule comprend encore trop d'erreurs.

Différents outils ont été proposés afin d'améliorer la classification prédite par les algorithmes sur base du training set. La méthode la plus simple mais la plus longue est de réaliser une validation manuelle de chaque item. La classification automatique est employée comme un outil qui propose une classe uniquement. Elle requiert néanmoins d'avoir un expert en taxonomie qui valide en permanence les échantillons de plancton.

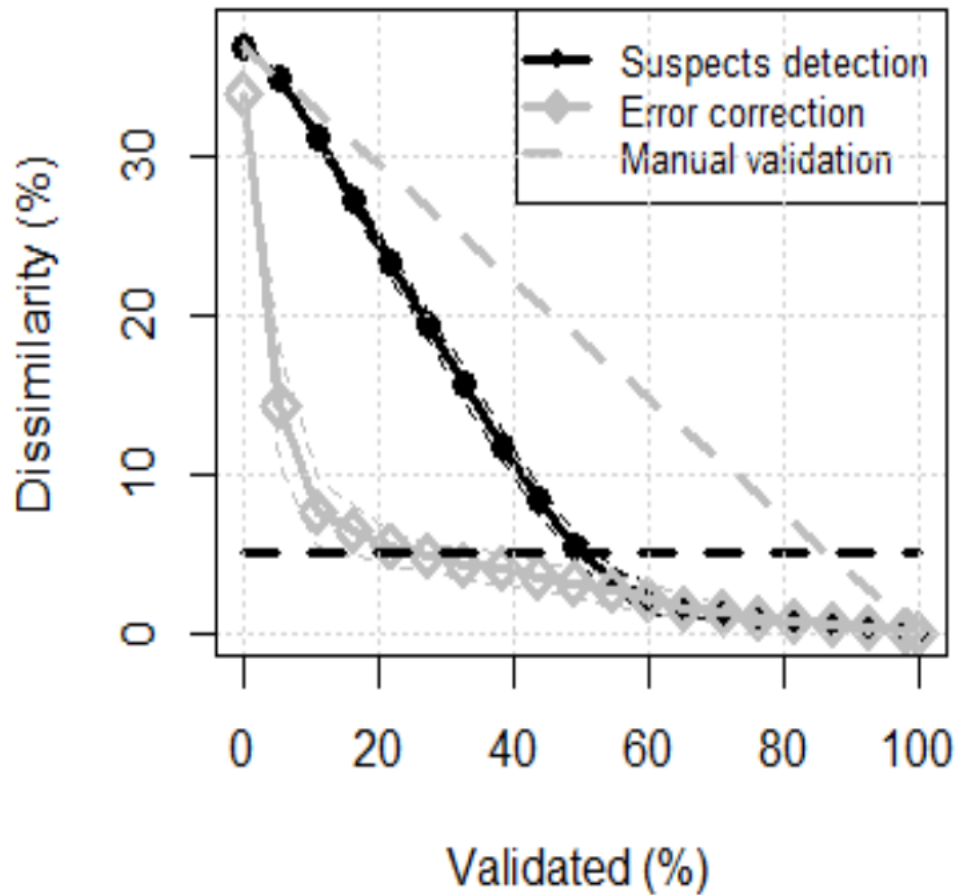
Une méthode de validation plus rapide est de déterminer des items suspects et de valider principalement ces items. Le schéma ci-dessous propose un schéma de la méthode.



On définit la composition d'un échantillon prélevé sur base d'outils de classifications automatiques. Les items peuvent être classés en items suspects ou en items de confiance (lors de la première détection, tous les items sont des suspects). Parmi ces items, une proportion de maximum 5% des images de l'échantillon sont à valider. Il y aura une faible proportion d'items aléatoires, une large proportion d'items suspects et une faible proportion d'items de confiances. Sur base des validations manuelles de nouvelles matrices de confusion sont calculées.

Ce processus est itératif. Les éléments validés sont employés pour améliorer la classification et la détection de nouveaux suspects à valider. Cette méthode permet donc de s'attarder principalement sur les items suspects d'un échantillon.

En partant du principe que la validation ne comprend aucune erreur, la validation manuelle classique va permettre de corriger les erreurs. Lorsque 100% de l'échantillon est validé, l'erreur tombe à 0%.



L'axe Y montre la dissimilarité entre les abondances prédites de chaque classe au sein de l'échantillon et les abondances réelles (obtenue en classant manuellement chaque items). L'axe X indique le pourcentage d'items validés manuellement.

La ligne en pointillé grisée indique la diminution linéaire du la dissimilarité lors de la validation manuelle de chaque item. Il faut dépasser 80% afin de tomber sous les 5% de dissimilarité.

...

Avancement :

Une publication est en cours d'écriture sur *Error correction of automatic classification of plankton digital images through partial validation of suspect items in Zoo/PhytoImage*.

Des tests vont débuter sur la correction d'erreur sur des échantillons de zooplancton.

Chapitre 4 : ...

...

Avancement : Une publication est en cours de finalisation sur *Assessment of active learning for semi-automatic classification of plankton digital images with Zoo/PhytoImage*.

Données disponibles

Les données suivantes sont disponibles et vont être employées dans le cadre de cette étude

- Série temporelle horizontale de la baie de Calvi : 335 échantillons
- Série temporelle verticale de la baie de Calvi : 76 échantillons

Activités valorisables pour la formation doctorale

L'ensemble des publications et communications scientifiques sont regroupés dans le fichier `publication_di_20201023.pdf`.

Activités supplémentaires valorisables pour la formation doctorale

- 2020
 - 2h de séminaire : ...
 - 2h de séminaire : RBZS Workshop Present your Research with Impact (1 décembre)
 - 4h (temps estimé) de formation sur OpenClassrooms : *Adoptez les API REST pour vos projets web*
 - 8h (temps estimé) de formation sur OpenClassrooms : *Réalisez des rapports statistiques clairs et impactants*
 - 8h (temps estimé) de formation sur OpenClassrooms : *Testez vos idées avec le lean prototyping*
 - 12h (temps estimé) de formation sur OpenClassrooms : *Initiez-vous à Python pour l'analyse de données*
 - 20h (temps estimé) de formation sur OpenClassrooms : *Initiez-vous à l'algèbre relationnelle avec le langage SQL*
 - 4h (temps estimé) de formation sur OpenClassrooms : *Mettez en place un système de veille informationnelle*
 - 12h (temps estimé) de formation sur OpenClassrooms : *Gérez votre temps efficacement*
 - 12h (temps estimé) de formation sur OpenClassrooms : *Initiez-vous à la statistique inférentielle*
 - 12h (temps estimé) de formation sur OpenClassrooms : *Initiez-vous au langage R pour analyser vos données*
 - 12h (temps estimé) de formation sur OpenClassrooms portant : *Utilisez Git et GitHub pour vos projets de développement*
- 2019
 - 1/2 journée de formation sur *Getting the most out of Git*
 - 1/2 journée de formation sur *R/exams: A One-For-All Exams Generator* lors useR 2019 à Toulouse
- 2018
 - Formation sur le deep learning. Formation d'1/2 journée sur le deep learning durant la septièmes rencontres R (Rennes, France) en 2018.
- 2017
 - C.I.B.I.M. (présence) Présence à la conférence annuelle du Centre Interuniversitaire de Biologie Marine à Louvain La Neuve 2017.
-
- 1 décembre 2020 (2h) : RBZS Workshop Present your Research with Impact
- 8 décembre 2020 (2h) : RBZS Workshop ...

Publications

Fullgrabe, Lovina, Philippe Grosjean, Sylvie Gobert, Pierre Lejeune, Michèle Leduc, Guyliann Engels, Patrick Dauby, Pierre Boissery, and Jonathan Richir. 2020. “Zooplankton dynamics in a changing environment: A 13-year survey in the northwestern Mediterranean Sea.” *Marine Environmental Research* 159: 104962. <https://doi.org/https://doi.org/10.1016/j.marenvres.2020.104962>.