

## Mémoire de Master ECD

# Téledétection de *Trichodesmium* dans la zone du Pacifique Sud-Ouest

Guillaume Rousset

soutenu le 29/06/2015

Encadrement : Florian De Boissieu(LOCEAN) ; Cécile Dupouy(MIO) ; Jérôme Lefèvre(MIO) ; Christophe Menkes(LOCEAN)

Localisation : Institut de Recherche pour le Développement, Nouméa, Nouvelle-Calédonie

### Résumé :

*Le Trichodesmium est une cyanobactérie diazotrophe présente dans les eaux du Pacifique Sud-Ouest. Selon plusieurs études, cette algue participerait à l'apport de nitrate dans les océans, et jouerait un véritable rôle de fertilisant dans les zones pauvres en nutriments. Elle serait à l'origine des concentrations exceptionnelles de chlorophylle détectées par satellite, dans la partie tropicale de l'océan Pacifique Sud, et notamment autour de la Nouvelle-Calédonie. Des travaux récents ont montré que les efflorescences de Trichodesmium pouvaient être détectées par satellite. Ces travaux se basent principalement sur les spécificités bio-optiques du Trichodesmium. Ce stage avait pour objectif d'implémenter et de tester ces algorithmes bio-optiques, de les adapter à la région du Pacifique Sud-Ouest, et de les comparer à des algorithmes statistiques encore très peu utilisés dans la reconnaissance du phytoplancton.*

### Abstract:

*Trichodesmium is a diazotrophic cyanobacteria present in oceans of the tropical belt. According to several studies, this algae would produce nitrate contributing this way to the fertilization of these waters especially poor in nutrients. Trichodesmium would also be responsible for the exceptional chlorophyll concentrations remotely sensed in the tropical part of the South Pacific Ocean, in particular around New Caledonia. Recent studies showed that the mats of Trichodesmium could be detected by ocean color remote sensing. These studies are mainly based on the bio-optical specificities of Trichodesmium. During this internship, my goals were the implementation and the evaluation of the existing bio-optical algorithms, their adaptation to the South-West Pacific region, and to compare them with statistical algorithms still poorly used in the recognition of the phytoplankton.*

## Table des matières

<b>1</b>	<b>Structure d'accueil : l'IRD de Nouvelle-Calédonie</b>	<b>1</b>
1.1	Département d'océanographie	1
1.1.1	MIO	1
1.1.2	LOCEAN	2
<b>2</b>	<b>Remerciements</b>	<b>3</b>
<b>3</b>	<b>Introduction</b>	<b>4</b>
<b>4</b>	<b>État de l'art</b>	<b>6</b>
4.1	Téledétection satellite de la couleur de l'eau	6
4.1.1	Principes	6
4.1.2	Capteurs satellites	8
4.1.3	Traitement des données satellites	9
4.2	<i>Trichodesmium</i>	11
4.3	Détection du <i>Trichodesmium</i>	12
4.3.1	Les propriétés optiques du <i>Trichodesmium</i>	13
4.3.2	Les algorithmes bio-optiques	13
4.4	Les algorithmes de classification	15
<b>5</b>	<b>Les données</b>	<b>16</b>
5.1	Données <i>in-situ</i>	16
5.2	Données satellites	16
5.3	traitement et visualisation des données	18
5.3.1	SeasDAS	18
5.3.2	Les développements techniques	18
5.3.3	Automatisation du processus	19
<b>6</b>	<b>Reconnaissance du <i>Trichodesmium</i> par des algorithmes bio-optique</b>	<b>21</b>
6.1	Les algorithmes de la littérature	21
6.1.1	Algorithme de McKinna	21
6.1.2	Algorithme de Hu	23
6.1.3	Les autres algorithmes	24
6.2	Développement d'un algorithme bio-optique	25
6.2.1	Utilisation de la correction atmosphérique	25
6.2.2	Critère de forme	25
6.2.3	Résultats	25
<b>7</b>	<b>Reconnaissance du <i>Trichodesmium</i> par méthode d'apprentissage supervisé</b>	<b>29</b>
7.1	Méthodes	29
7.1.1	Répartition des données et sélection des variables	29
7.1.2	Validation des algorithmes	30
7.2	résultats	30
<b>8</b>	<b>Discussion et conclusion</b>	<b>32</b>
<b>9</b>	<b>Plannification</b>	<b>35</b>
9.1	planning	35



# 1 Structure d'accueil : l'IRD de Nouvelle-Calédonie

L'Institut de Recherche pour le Développement (IRD) est un organisme français de recherche qui fut fondé il y a plus de 65 ans. Cet institut a pour but de contribuer au développement social, économique et culturel des pays du Sud, mais aussi d'étudier les relations entre l'homme et son environnement. Améliorer les conditions sanitaires, comprendre l'évolution des sociétés, préserver l'environnement et les ressources constituent les piliers de son action en Afrique, en Méditerranée, en Amérique latine, en Asie et dans l'Outre-Mer tropical français.

Grâce à son action de recherche, de formation et d'innovation en partenariat, il rayonne dans plus d'une cinquantaine de pays, en Afrique, sur le pourtour méditerranéen, en Asie, en Amérique latine et en Outre-Mer. Fondés sur l'interdisciplinarité, les projets menés conjointement traitent de questions cruciales pour les pays en développement : maladies tropicales et de civilisation, relations entre santé et environnement, changements climatiques, ressources en eau, sécurité alimentaire, écosystèmes tropicaux et méditerranéens, risques naturels, pauvreté, vulnérabilité et inégalités sociales, migrations, évolution du marché du travail, etc.

En Nouvelle-Calédonie le centre IRD de Nouméa est la principale implantation de l'Institut dans le Pacifique Sud. Pour répondre aux demandes de ses partenaires publics présent dans la région du Pacifique sud, les activités de recherche du centre ciblent particulièrement les thématiques suivantes :

- écosystèmes et ressources naturelles : biodiversité marine et terrestre
- Changement climatique et aléas naturels
- Enjeux sociétaux
- Santé

## 1.1 Département d'océanographie

L'IRD de Nouméa regroupe de nombreux chercheurs appartenant à des laboratoires différents. Le département d'océanographie regroupe plusieurs équipes entrant souvent en collaboration pour mener des recherches dans le Pacifique sud. Mes encadrants dépendent de laboratoires différents que sont : le Laboratoire d'Océanographie et du climat(LOCEAN) et l'Institut Méditerranéen d'Océanologie(MIO).

### 1.1.1 MIO

Mes encadrants rattachés à ce laboratoire sont : Cécile Dupouy et Jérôme Lefèvre.  
source : <https://www.mio.univ-amu.fr>

**Organisation** Le MIO est une unité mixte de recherche en Océanologie sous la tutelle du Centre National de la Recherche Scientifique (CNRS), de l'IRD, de l'Université d'Aix Marseille et de l'Université de Toulon. Ses objectifs sont de mieux comprendre le système océanique et son évolution en réponse au changement global. Il constitue un pôle de compétences en biologie, écologie, biodiversité, microbiologie, halieutique, physique, chimie et optique marine, biogéochimie et en sédimentologie marine. Ses cadres d'exercice sont l'océan mondial, ses interfaces avec le continent, l'atmosphère et le sédiment.

Le MIO regroupe plus de 200 personnes dont une centaine de chercheurs (CNRS, IRD) et d'enseignants-chercheurs des Universités, une soixantaine d'ingénieurs et techniciens, plus de 50 doctorants et une équipe d'administratifs au service de la recherche.

**Thèmes de recherche** Les scientifiques du MIO effectuent leurs recherches dans les domaines de la circulation océanique et atmosphérique, la compréhension des écosystèmes marins et de la biodiversité de la bactérie aux poissons, des fonctionnements biologiques dans les environnements extrêmes et de la pollution des océans. Ces travaux de recherches nécessitent des campagnes océanographiques, des stations d'observations, l'utilisation d'instruments immergés équipés de capteurs (bouées, instruments tractés, gliders ♦ ou drones sous marins), de radars installés sur la côte et d'analyses d'images satellitaires,

ainsi que des moyens analytiques et expérimentaux de laboratoire. Ces recherches font appel à l'analyse numérique et à la modélisation.

**En Nouvelle-Calédonie** Localement l'équipe de recherche du MIO étudie la circulation des masses d'eau à différentes échelles à l'aide de la modélisation et de la télédétection couleur de l'eau. Il s'agit notamment d'étudier les perturbations climatiques et anthropiques sur le fonctionnement de l'écosystème. D'autres travaux porte sur les cycles biogéochimiques en relation avec la biodiversité taxonomique et fonctionnelle des communautés planctoniques et leur structuration spatio-temporelle, son contrôle par les forçages physiques et la disponibilité nutritive.

### 1.1.2 LOCEAN

Mes encadrants rattachés à ce laboratoire sont : Florian De Boissieu et Christophe Menkes.  
source : <https://www.ipsl.fr/Organisation/Les-laboratoires/LOCEAN>

**Organisation** Le LOCEAN est une unité mixte de recherche, dont les tutelles sont l'Université Pierre et Marie Curie, le Muséum National d'Histoire Naturelle, le Centre National de la Recherche Scientifique et l'Institut de Recherche pour le Développement. Il compte environ 150 personnes, dont près des deux-tiers de chercheurs, enseignants-chercheurs, doctorants et post-doctorants et un tiers de techniciens et administratifs.

**Thèmes de recherche** L'activité centrale du LOCEAN est l'étude de l'océan et de la variabilité climatique. En ce qui concerne le premier volet, LOCEAN étudie un large spectre d'échelles qui va de celles du mélange vertical et des ondes internes à celles des mouvements planétaires, pour approfondir nos connaissances de la dynamique océanique et de son impact sur les grands compartiments chimiques (en particulier le carbone sous ses formes inorganique et organique) en liaison avec les cycles bio-géochimiques. Dans le second volet, le laboratoire s'intéresse plus spécifiquement au rôle de l'océan sur la variabilité du climat, en particulier aux échelles courtes qui vont de l'inter-annuel au siècle, aussi bien pour le climat passé, que présent et futur. Le laboratoire s'est par ailleurs engagé récemment dans des études sur les relations variabilité climatique-société en interaction avec d'autres communautés (agronomie, économie, santé).

**En Nouvelle-Calédonie** l'équipe de LOCEAN en Nouvelle-Calédonie effectue des recherches sur le climat, basées sur la modélisation de la dynamique des fluides océans et atmosphère, ainsi que sur l'étude du couplage océan-atmosphère. L'objectif est de comprendre comment le climat modifie les cyclones sur le présent (influence du phénomène El Niño/Oscillation australe ENSO) et dans le futur (pourquoi les cyclones ont tendance à diminuer dans le Pacifique sud-ouest).

## 2 Remerciements

Je tiens à remercier mes différents encadrants de stage, Florian De Boissieu, Cécile Dupouy, Jérôme Lefèvre et Christophe Menkes qui m'ont accueilli et aidé tout au long de ce stage de fin d'étude.

Je remercie ma collègue de bureau Chloé Martias avec qui j'ai pu effectuer quelques missions en mer.

Je remercie Guillaume Dirberg et Sandrine Bouisset dont les rapports m'ont aidé à mieux comprendre ce sujet de stage et à rédiger l'état de l'art de ce rapport.

### 3 Introduction

La région tropicale du Pacifique Sud-Ouest est connue pour ses faibles concentrations en sels nutritifs (Fer et Nitrate) limitant la floraison de la majorité des types de phytoplancton. Dans cette région dite oligotrophe (pauvre en éléments nutritifs), les diazotrophes, des phytoplanctons fixateurs de l'azote de l'air, sont privilégiés car moins susceptibles au manque de nitrate. Ces organismes jouent un rôle important dans l'océan tropical en participant à l'apport de nitrates nécessaires à la floraison d'autres espèces [Luo et al., 2012, Moisander et al., 2010] et pourraient de ce fait, contribuer à près de 50% de la production nouvelle dans la zone [Garcia et al., 2007, Biegala and Raimbault, 2008, Bonnet et al., 2009]. Parmi les diazotrophes, le *Trichodesmium* est l'espèce de phytoplancton qui nous intéresse plus particulièrement. Cette algue fleurit durant l'été austral et semble être à l'origine des concentrations de chlorophylle exceptionnelles observées par satellite dans les régions tropicales et en particulier dans les eaux de la Nouvelle Calédonie et du Vanuatu.

La cyanobactérie du genre *Trichodesmium* apparaît comme le principal fixateur de diazote dans le milieu océanique oligotrophe (milieu pauvre en éléments nutritifs) du Pacifique tropical Sud-Ouest. La télédétection des efflorescences de *Trichodesmium* est un sujet crucial pour la connaissance des cycles biogéochimiques de l'océan, car elle permettrait de quantifier la part de la production primaire induite par la diazotrophie et donc son importance dans le cycle global du carbone. L'hypothèse est que le *Trichodesmium* serait détectable sur les images satellites grâce à ses propriétés optiques originales, rétro-diffusion et absorption, qui influent sur les réflectances marines.

Plusieurs algorithmes ont été développés afin d'identifier de telles floraisons [Dupouy et al., 2011, Gower et al., 2014, Hu et al., 2010, McKinna et al., 2011, Westberry et al., 2005] à partir de la couleur de l'eau observée par satellite. Cependant ces algorithmes, souvent empiriques, ont été implémentés dans différentes régions avec des images provenant de différents capteurs. Nous les avons donc appliqués et adaptés à la région Pacifique Sud-Ouest et à l'imagerie satellite haute résolution (MERIS et MODIS). D'autre part nous avons appliqué des méthodes d'extraction de connaissances à la détection du *Trichodesmium*. Quelques travaux ont été publiés récemment sur l'utilisation de ce type de méthodes pour la reconnaissance automatique des grands types de phytoplanctons [de Boissieu et al., 2014, Prasad and Krishna, 2011]. Cependant, aucune application de ces méthodes d'extraction de connaissance n'a été développée pour la reconnaissance spécifique du *Trichodesmium*. Ces méthodes présentent pourtant un intérêt certain pour le traitement de données satellites de plus en plus nombreuses et de nature différentes.

Les méthodes de reconnaissance bio-optique se basent sur la physique de l'absorption et de la rétro-diffusion optique des différents éléments en présence (marins et atmosphériques). Ces méthodes se veulent ainsi généralisables indépendamment du capteur employé. Cependant, les méthodes bio-optiques impliquent une connaissance a priori ou des hypothèses sur les éléments en présence (eau, aérosols, fond marin, matière en suspension organique et non-organique, etc.) et leur influence sur le signal optique mesuré par le satellite. Malheureusement, les relations entre les éléments en présence et le signal optique sont complexes et difficiles à inverser. Les algorithmes bio-optiques restent donc souvent dépendant du capteur et des pré-traitements associés, ainsi que du contexte géographique dans lequel ils ont été développés. Et les relations empiriques doivent alors être adaptées manuellement.

Une autre option est celle des algorithmes de classification automatiques qui se basent sur les relations statistiques entre les variables d'influence (ici le signal optique et ses indices dérivés) et les éléments à reconnaître (ici les efflorescences de *Trichodesmium*). Ces méthodes bénéficient d'un apprentissage automatique permettant d'adapter les relations en fonction du jeu de données et donc du capteur et de ses pré-traitements. Cependant, cet apprentissage automatique présente également des inconvénients. Les estimations statistiques nécessitent une diversité d'exemples représentative de la situation analysée. De plus, les relations non-linéaires sont souvent difficiles à modéliser statistiquement.

Les objectifs de ce stage étaient donc de tester et éventuellement adapter à la zone d'étude les algorithmes bio-optiques existant pour la reconnaissance des efflorescences de *Trichodesmium*, et d'en comparer les performances aux algorithmes statistiques. La zone d'étude est centrée sur la région Pacifique Sud et plus

particulièrement autour de la Nouvelle-Calédonie et du Vanuatu. Ce stage fait parti du projet CNES-TOSCA SAVALEFER d'étude par satellite de l'influence du volcanisme vanuatais sur la fertilisation océanique de la région.



## 4 État de l'art

Première ressource en oxygène de la planète, l'océan est aussi une ressource alimentaire et économique pour de nombreux pays. Dans cet environnement complexe et fragile le phytoplancton, dont le *Trichodesmium*, joue un rôle très important dans le fonctionnement des écosystèmes marins. L'étude et le suivi de ces algues permettent une meilleure compréhension des mécanismes bio-géochimiques qui ont lieu dans l'océan. Pour effectuer ces observations, la détection par satellite est un outil particulièrement intéressant car il permet de caractériser la surface de l'océan avec des couvertures spatiale et temporelle réalisables avec les mesures de terrain seules.

### 4.1 Télédétection satellite de la couleur de l'eau

L'objectif de la télédétection optique marine est d'observer les caractéristiques spectrales de la mer (i.e. la « couleur de l'eau ») pour en déduire les différents constituants présents dans l'eau. En effet, chaque constituant a des caractéristiques optiques spécifiques d'absorption et de diffusion de la lumière qui altèrent les rayons lumineux reçus. La couleur de l'eau perçue au niveau du satellite correspond donc à un mélange des caractéristiques optiques des différents constituants (atmosphère, aérosols, eau, phytoplancton, matières dissoutes dans l'eau, etc.) rencontrés sur le chemin entre la source lumineuse (le soleil) et le capteur satellite. Le traitement de ces données satellites vise donc dans un premier temps à retrouver les caractéristiques spectrales au niveau de l'eau, et dans un second temps à séparer et quantifier les différentes sources marines d'altération de la lumière. Ces méthodes sont notamment utilisées pour étudier les différents procédés bio-géophysiques de l'océan, comme l'étude du cycle du carbone par exemple, à l'échelle régionale ou globale. Ces études permettent alors de quantifier l'impact écologique que peuvent avoir les différents facteurs environnementaux et/ou anthropiques sur la biogéochimie de l'océan.

Cette partie présente dans une première section les principes de l'optique marine, et dans une deuxième section les traitements standards nécessaires à l'obtention des caractéristiques spectrales au niveau de l'eau. Pour plus de précision sur les notions d'optique marine présentées ci-dessous, on se référera à [Kirk, 1994, Froidefond et Doxaran, 2004].

#### 4.1.1 Principes

Un rayon lumineux subit plusieurs altérations lors de son trajet avant de pouvoir être mesuré. Deux phénomènes principaux participent à son altération : l'absorption et la diffusion. L'absorption optique est le phénomène par lequel une molécule absorbe et convertit une partie de l'énergie lumineuse incidente en un autre type d'énergie. La diffusion est le phénomène par lequel un rayon incident sur une particule est dévié en une multitude de directions.

L'absorption varie selon la nature de la particule et de la longueur d'onde du rayon incident. La diffusion varie selon la longueur d'onde mais également selon la taille de la particule ou la rugosité de l'objet observé. Ainsi, sous une lumière blanche, les longueurs d'ondes les moins absorbées et/ou les plus diffusées sont celles constituant la couleur de l'objet observé. Les coefficients d'absorption et de diffusion constituent les propriétés optiques inhérentes (IOP) de l'objet observé. Inversement, la connaissance des IOP permettrait de déterminer la nature de l'objet observé.

Cependant le cheminement d'un rayon lumineux du soleil jusqu'au satellite où il est mesuré est complexe. Il intègre notamment les altérations des différents matériaux traversés (atmosphère, aérosols, surface de l'eau, etc.). La Figure 1 illustre de manière schématique le trajet d'un rayon lumineux ( $r$ ) issu du soleil jusqu'au satellite.

Du rayon incident ( $r$ ) qui a traversé l'atmosphère, une partie de la lumière réfléchit par la surface de l'eau ( $r_1$ ) tandis que l'autre partie pénètre dans l'eau ( $r_2$ ). Sous la surface, le rayon  $r_2$  est en partie absorbé ( $r_{2a}$ ) et en partie diffusé ( $r_{2b}$ ). Le rayon diffusé ( $r_{2b}$ ) peut se décomposer comme la somme de la diffusion descendante ( $r_{2bf}$ ) et la diffusion ascendante dite rétro-diffusion ( $r_{2bb}$ ). Sur les différents rayons issus de

la rétro-diffusion ( $r_{2bb}$ ), ceux qui ont un angle supérieur à l'angle limite sont réfléchis par la surface tandis que le reste arrive à émerger pour donner le rayon émergent ( $r_3$ ). Comme ( $r_1$ ), le rayon émergent ( $r_3$ ) traverse alors l'atmosphère et est mesuré par le capteur satellite.

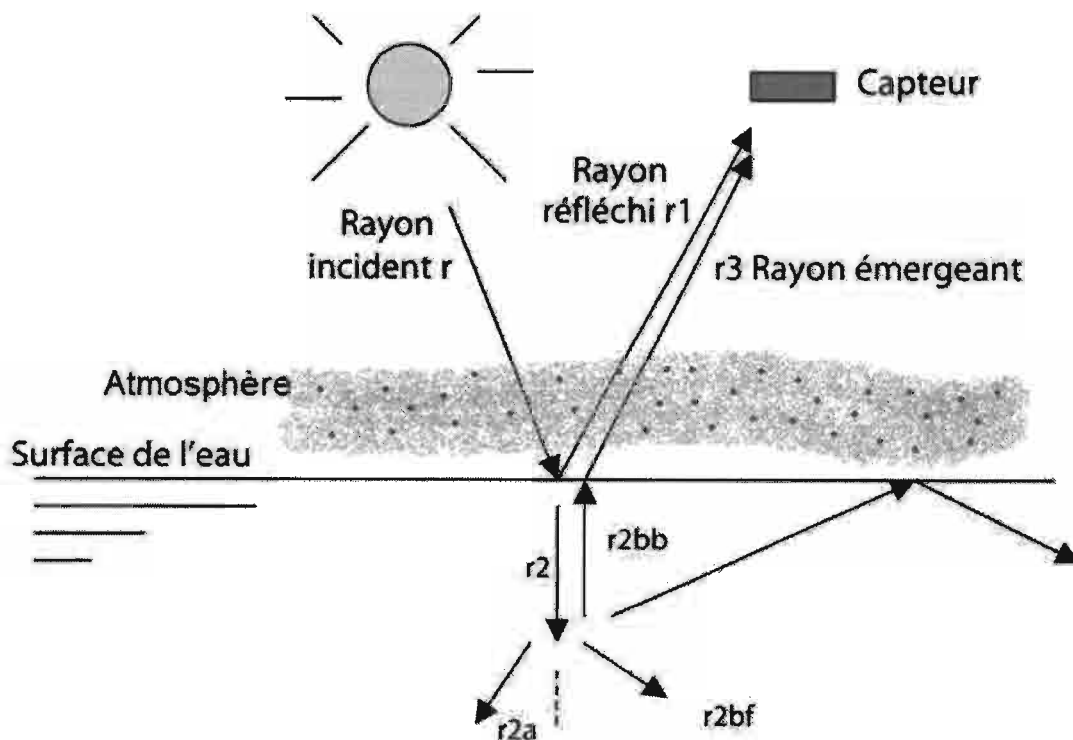


Figure 1 – Extrait de [Froidefond and Doxaran, 2004]. Trajets schématiques des rayons lumineux à l'interface atmosphérique/eau

Tous les phénomènes cités ci-dessus se produisent aussi au sein de l'atmosphère. Il existe donc de nombreux autres trajets que le rayon lumineux peut emprunter. Il est possible de distinguer la nature des différents parcours du rayon incident en mesurant la luminance (intensité lumineuse) de chaque rayon arrivant jusqu'au capteur.

Parmi ces parcours nous sommes particulièrement intéressés par le rayon émergent ( $r_3$ ) décrit plus tôt. En effet tout au long de son parcours ce rayon lumineux a subi plusieurs altérations dues au milieu qu'il a traversé : l'atmosphère et l'eau. Dans cette dernière, la lumière qui a été diffusée dépend des constituants moléculaires et particulaires présent dans l'eau. En étudiant le spectre du rayon ( $r_3$ ), on souhaite retrouver les constituants de l'eau.

On note  $a$ , le coefficient d'absorption exprimé en  $m^{-1}$  qui représente le rapport du flux radiatif perdu par absorption sur le flux radiatif incident. On note  $b$ , le coefficient de diffusion en  $m^{-1}$  qui est le rapport du flux radiatif perdu par diffusion sur le flux radiatif incident. Pour rappel le coefficient de diffusion peut être décomposé en deux parties, la partie qui suit la direction initiale du rayon incident (forward) qui sera notée  $b_f$  et la partie qui est rétro-diffusée vers la surface (backward) qui sera notée  $b_b$ .

$$b(\lambda) = b_b(\lambda) + b_f(\lambda) \quad (1)$$

$a$  et  $b$  sont des propriétés optiques inhérentes de la matière. Elles peuvent être décomposées de manière additive entre les différentes contributions des substances contenues dans l'eau. Ces substances sont l'eau pure, les matières organiques colorées dissoutes, les particules phytoplanctoniques et les particules minérales désignées respectivement par les indices  $w$ ,  $y$ ,  $ph$ , et  $s$ . Les valeurs d'absorption et de diffusion

de l'eau pure sont connues [Pope and Fry, 1997].

$$a(\lambda) = a_w(\lambda) + a_y(\lambda) + a_{ph}(\lambda) + a_s(\lambda) \quad (2)$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bph}(\lambda) + b_{bs}(\lambda) \quad (3)$$

Les capteurs optiques ne permettent pas d'accéder directement aux propriétés optiques inhérentes, mais au rapport d'intensités lumineuses appelées "réflectances". La réflectance est définie comme étant le rapport entre l'éclairement ascendant et l'éclairement descendant :

$$R(\lambda) = E_u(\lambda)/E_d(\lambda) \quad (4)$$

La réflectance peut être approximée à partir des propriétés optiques inhérentes selon l'équation suivante :

$$R(\lambda) = \frac{f \cdot b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (5)$$

Avec  $f = 0,33$ , une approximation communément admise, valable pour toutes les longueurs d'ondes du visible.

*Rrs* ou "Remote Sensing Reflectance" en  $sr^{-1}$  (stéradian : unité d'angle solide), est la réflectance télédéteectée par le capteur satellite. *Rrs* est définie de cette manière :

$$Rrs(\lambda) = \frac{nLw(\lambda)}{F0(\lambda)} \quad (6)$$

où  $nLw$  est la luminance normalisée de l'eau, une autre grandeur issue des données satellites, et  $F0$  l'éclairement solaire au sommet de l'atmosphère lorsque le soleil est au zénith. Avec l'équation 5 issue de la combinaison d'une série d'équations caractérisant un spectro-radiomètre [Froidefond and Doxaran, 2004] il est possible d'établir un lien direct entre le *Rrs* et les coefficients  $a$  et  $b$ .

$$Rrs(\lambda) = 1,57 \cdot R(\lambda) \frac{f}{Q_u} \quad (7)$$

Avec  $\frac{f}{Q_u}$  variant entre 0,075 et 0,095 suivant les différentes conditions optiques [Morel and Gentili, 1996].

#### 4.1.2 Capteurs satellites

La couleur de l'eau est observée par satellite depuis environ 50ans, avec le lancement successif de différents satellites (CZCS, SeaWifs,MODIS, MERIS, VIIRS, etc.). Au début des années 80, le spectro-radiomètre CZCS (Coastal Zone Color Scanner) sur NIMBUS a permis de mettre en évidence la présence d'une production primaire liée à *Trichodesmium* dans le Pacifique [Dupouy et al., 1988].

Lors de ce stage nous avons fait le choix de nous concentrer sur le capteur MODIS/Aqua (MODISA), dont les premiers enregistrements ont débutés en Juillet-2002 (sans interruption depuis) et pour lequel une partie de l'archive est disponible au Centre de Nouméa. Ce choix a été motivé tout d'abord par la résolution du capteur qui est de 250 m à 1 km selon les bandes spectrales et de l'abondance des données acquises sur la région depuis 2002. De plus, ce capteur est celui utilisé dans les études récentes de reconnaissance du *Trichodesmium* [McKinna et al., 2011, Hu et al., 2010]. Cela facilite la comparaison entre ces algorithmes et les algorithmes de détection que j'ai implémentés. Enfin, MODISA est toujours en activité, ce qui nous a permis d'étudier des scènes correspondantes à la campagne océanographique OUTPACE qui a eu lieu durant le stage.

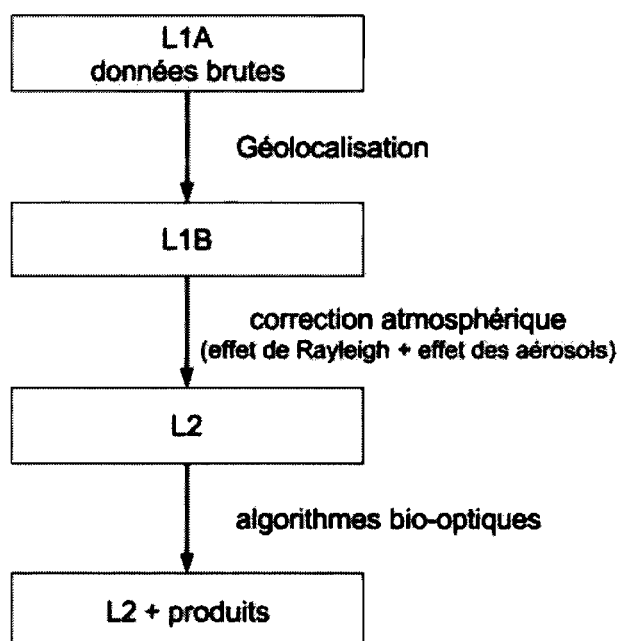
Les caractéristiques du satellite sont les suivantes <http://modis.gsfc.nasa.gov/about/specifications.php>. Le tableau ci-dessous décrit les bandes de longueurs d'ondes qui nous intéressent plus particulièrement.

Bande	Largeur de bandes	résolution (en m)
1	620 - 670	250
2	841 - 876	250
3	459 - 479	500
4	545 - 565	500
5	1230 - 1250	500
6	1628 - 1652	500
7	2105 - 2155	500
8	405 - 420	1000
9	438 - 448	1000
10	483 - 493	1000
11	526 - 536	1000
12	546 - 556	1000
13	662 - 672	1000
14	673 - 683	1000
15	743 - 753	1000
16	862 - 877	1000
17	890 - 920	1000
18	931 - 941	1000
19	915 - 965	1000

Table 1 – Tableau descriptif des différentes résolutions des bandes utilisées par MODIS

#### 4.1.3 Traitement des données satellites

Le capteur satellite fournit des données brutes de réflectance mesurée qui doivent être pré-traitées et normalisées avant utilisation. Parmi ces traitements il y a par exemple le redressement géodésique et la correction atmosphérique. On définit un niveau pour chaque étape de traitement, reflétant l'état et la constitution des données. SeasDAS <http://seadas.gsfc.nasa.gov/>, un logiciel créé par la NASA et enrichi par la communauté des utilisateurs, permet d'effectuer l'ensemble de ces traitements.



#### La chaîne de traitement

- L0 : ce sont les données brutes issues du capteur satellite.
- L1A : les données brutes sont structurées et annotées avec les informations radiométriques, la calibration géométrique leur correspondant, et les paramètres géographiques nécessaires. Mais les corrections correspondantes ne sont pas encore appliquées.
- L1B : ce sont les données L1A sur lesquelles ont été appliquées les calibrations radiométrique et instrumentale, et un géoréférencement.
- L2 : ce sont les données L1B corrigées des effets de l'atmosphère et des aérosols, puis combinées pour calculer une liste de produits tels que les réflectances normalisées, la concentration de chlorophylle dans l'eau et la température de surface de l'eau.

Figure 2 – Succession des étapes de production dans le logiciel SeasDAS.

- L3 : il s'agit de l'agrégation des données L2 sur une période donnée et subissant une reprojexion sur une grille régulière.

En suivant la chaîne de traitement de la Figure 2, les données peuvent être interprétées et exploitées dans les logiciels prévus à cet effet, qui permettent la visualisation et l'application d'algorithmes bio-optiques faisant apparaître de nouvelles informations. Comme expliqué ci-dessus, le traitement des données suit un enchainement précis :

**L1A vers L1B + GEO** A partir des L1A un premier traitement consiste à générer les fichiers GEO et L1B. Le fichier L1B contient les informations radiométriques et se décompose en 3 parties : une partie basse résolution(1km), une partie à moyenne résolution (500m) et une dernière partie à haute résolution (250m). Le produit GEO quant à lui contient toutes les caractéristiques géographiques nécessaires (associées aux 3 résolutions spatiales, 1km, 500 et 250m) permettant de géolocaliser proprement chaque pixel. Le fichier GEO est construit en utilisant le modèle de vol du satellite corrigé des effets de dérive et roulis moyennant l'utilisation des éphémérides (table comportant les positions observées du vaisseau).

**De L1B + GEO vers L2** Par la suite le produit L2 est formé à partir de ces fichiers intermédiaires (L1B + GEO). Nous pouvons alors calculer plusieurs produits à partir des algorithmes déjà implémentés dans SeasDAS. Par exemple il est possible de sortir le produit OC3, qui est le résultat de l'algorithme de Hu(2012) pour le calcul de la concentration chlorophyllienne. A ce niveau, l'utilisateur peut également ajouter de nouveaux produits dérivés, en faisant appel à ses propres algorithmes ou bien en ajustant à sa région d'étude les algorithmes proposés.

Nous souhaitons créer des produits standards contenant des produits dérivés utilisés fréquemment. De cette manière, on met à disposition une série de fichier L2 contenant des produits dérivés spécifiques. Pour ce stage on extraira 5 fichiers L2 contenant respectivement :

- Les Rrs pour les longueurs d'ondes allant de 412 à 1240nm
- les Rrc pour les longueurs d'ondes allant de 412 à 1240nm
- Les valeurs d'absorption pour les longueurs d'ondes allant de 412 à 678nm
- Les valeurs de rétro-diffusion pour les longueurs d'ondes allant de 412 à 678nm
- les valeurs de chlorophylle et de turbidité et les épaisseurs optiques

**La correction atmosphérique** Comme vu précédemment la traversée de l'atmosphère altère le rayon lumineux et pour obtenir le signal provenant de la mer, il nous faut corriger ce biais. Dans l'atmosphère deux types de diffusion de la lumière ont lieu. La première dite diffusion de Rayleigh, concerne les particules dont l'ordre de grandeur est inférieur à la longueur d'onde du rayon lumineux. La deuxième est la diffusion liée aux particules de taille supérieure à la longueur d'onde du rayon, typiquement celles qu'on nomme aérosols. On différencie ces deux cas car les modèles physiques qui les régissent sont différents. Dans la suite de ce rapport, on appellera Rrc les réflectances corrigées des effets Rayleigh mais pas des aérosols, et on appellera Rrs les réflectances corrigées des effets de Rayleigh et des aérosols. Il existe plusieurs corrections atmosphériques qui sont applicables sur les données brutes offertes par les satellites d'observation de la couleur de la mer. On utilisera la correction de Gordon and Wang (1994) pour corriger les réflectances des effets de Rayleigh.

Si la diffusion de Rayleigh est un phénomène bien connu dont l'influence sur les rayons lumineux peut être calculée, la diffusion des aérosols est plus difficile à modéliser en raison de leur nature et taille variable. La NASA a mis au point un logiciel permettant d'exploiter les données brutes délivrées par les satellites MODIS : SeasDAS. Actuellement il existe une douzaine de modèles de correction atmosphérique disponibles dans SeasDAS correspondant à différents modèles d'atmosphère. Ces modèles se basent sur des hypothèses quand à la nature des aérosols en présence (aidés par les mesures satellites) et leur utilisation dépend de la nature du milieu en lui-même (domaine côtier, plein océan, etc.). Au-dessus de

l'océan Pacifique Tropicale Sud, les observations se font le plus souvent au dessus d'un milieu océanique dit "pristine", subissant peu l'influence d'aérosols d'origine continentale. La composition des aérosols est dominée par les sels-marins.

D'une manière générale, les modèles de correction atmosphérique considérant l'océan comme un corps noir dans le proche infra-rouge : c'est à dire qu'il absorbe toute la lumière pour les longueurs d'onde supérieures à 670nm. Le signal satellite de l'infrarouge est donc interprété comme étant uniquement du à l'atmosphère (Rayleigh + aérosols). Les amas d'algues flottantes ne sont pas vraiment prévus dans les modèles de correction atmosphérique actuels. Comme le *Trichodesmium* qui dans le proche infra-rouge, la correction atmosphérique pourrait être erronée au dessus de mattes de trichomes. C'est justement cette particularité qui a été utilisée pour la détection des mattes de surface.

#### 4.2 *Trichodesmium*

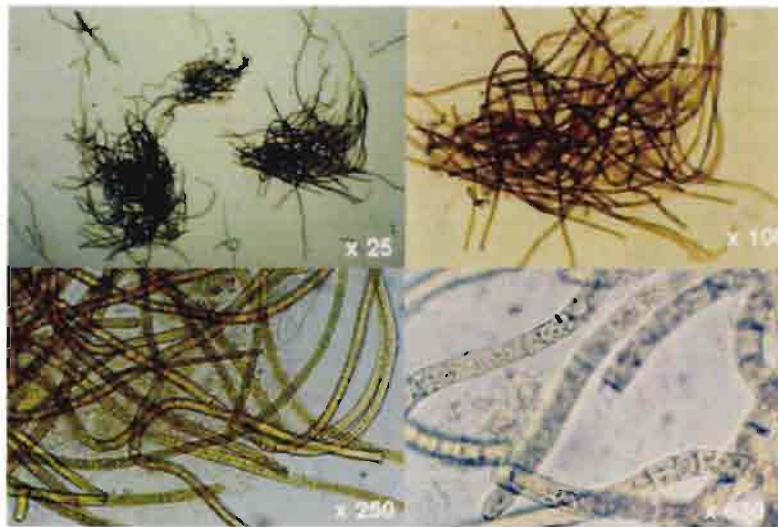


Figure 3 – Photo de C. Dupouy. Observation d'un groupe de trichomes au microscope



Figure 4 – Photo de prélèvement de *Trichodesmium* C. Dupouy

Elles seraient principalement constituées d'un type particulier de phytoplancton appelé *Trichodesmium* [Dupouy et al., 1988].

Les océans sont peuplés d'algues (ou phytoplanctons) qui est le premier maillon de la chaîne alimentaire marine. Pour sa croissance, le phytoplancton utilise la photosynthèse et a essentiellement besoin de lumière et de sels nutritifs dont les principaux sont le fer, l'azote sous forme de nitrate et le phosphate. Grâce à un brassage vertical généré par les courants marins, ces nutriments sont habituellement distribués dans toute la colonne d'eau, et notamment dans la couche éclairée (dite zone euphotique) où s'effectue la photosynthèse. Cependant, dans les eaux chaudes, la température élevée en surface engendre une stratification de la colonne d'eau empêchant les eaux riches en nutriments de remonter vers la couche éclairée. Pour cette raison, les eaux de la ceinture tropicale sont réputées pour être oligotrophes (pauvres en nutriments) et plus particulièrement en nitrates. Cependant, des accumulations d'algues beiges apparaissent régulièrement dans ces eaux pauvres en nutriments. Ces efflorescences, visibles sur les images satellites, ont notamment été observées autour des archipels mélanésien [Dupouy et al., 2011, Shiozaki et al., 2014] et au large des côtes australiennes [Kuchler and Jupp, 1988, McKinna et al., 2011].

Le *Trichodesmium* est un type de phytoplancton de la famille des cyanobactéries. Il possède les caractéristiques d'une bactérie (absence de noyau et d'organites intracellulaires) et les fonctions de photosynthèse des algues [ref]. Parmi les cyanobactéries, le *Trichodesmium* appartient à la famille des diazotrophes, car il a la particularité de pouvoir fixer l'azote atmosphérique pour s'alimenter en azote. Ainsi, contrairement aux autres types de phytoplancton, la diazotrophie permet au *Trichodesmium* de se développer dans les eaux pauvres en nitrate. Grâce à la particularité de cette niche écologique, le *Trichodesmium* est considéré comme le principal fixateur de di-azote dans les eaux oligotrophes tropicales et sub-tropicales [Paerl, 1999]. Des études récentes tendent à montrer que par ses particularités de diazotrophie et de taille millimétrique, le *Trichodesmium* pourrait alimenter les eaux en nitrates et contribuer à leur fertilisation par d'autres types de phytoplancton [Berthelot et al., 2015]. Le *Trichodesmium* jouerait donc un rôle important dans le cycle de l'azote et du carbone, rôle largement sous-estimé jusqu'à récemment [Capone, 1997, Gruber and Sarmiento, 1997, Karl et al., 2002]

Au niveau morphologique, le *Trichodesmium* se présente sous la forme d'un filament constitué de plusieurs dizaines de cellules et appelé trichome. Lors de conditions météorologiques favorables (houle et vent de surface faibles, température de surface élevée) [Capone, 1997], plusieurs centaines de trichomes peuvent s'agglomérer en colonies (Figure 3) et former des "mattes". Les mattes de *Trichodesmium* peuvent mesurer jusqu'à plusieurs dizaines de kilomètres de long pour quelques centaines de mètres de large [Kuchler and Jupp, 1988]. Lors d'efflorescences importantes, elles sont visibles à la surface de l'eau sous la forme de longues bandes étroites beiges (Figure 4).

Enfin, comme tout type d'algue, le *Trichodesmium* contient une série de pigments qui lui permettent la photosynthèse. Le plus connu est la Chlorophylle commune à toutes les espèces de phytoplancton. D'autres types de pigments dits accessoires sont spécifiques à certaines espèces de phytoplancton. Le pigment spécifique des cyanobactéries est la phycoérythrine [Subramaniam et al., 1999a, Subramaniam et al., 1999b]. Outre leur rôle dans la photosynthèse, les pigments sont également responsables de la couleur des algues. La chlorophylle et la phycoérythrine pigmentent ainsi le *Trichodesmium*, lui donnant cette couleur beige caractéristique.

L'observation des structures de *Trichodesmium* par imagerie satellite permettrait de mieux suivre la distribution spatiale et temporelle de ce type de phytoplancton, et de resituer son rôle dans les cycles de l'azote et du carbone.

### 4.3 Détection du *Trichodesmium*

Ces 20 dernières années plusieurs algorithmes bio-optiques ont été développés pour détecter les efflorescences de *Trichodesmium* [Hu et al., 2010, Westberry et al., 2005, Subramaniam et al., 2002, McKinna et al., 2011, Dupouy et al., 2011, Gower et al., 2014]. Pour distinguer le *Trichodesmium* des autres composants de l'eau, ces algorithmes exploitent majoritairement les propriétés optiques particulières du *Trichodesmium*, dont les pigments mais également la structure physiologique et la distribution spatiale sont essentiellement responsables. Ces algorithmes sont dits bio-optiques. Les sections suivantes décrivent successivement les propriétés optiques du *Trichodesmium*, puis les deux principaux algorithmes de la littérature implémentés lors de mon stage.

Pour détecter le *Trichodesmium* il faut réunir plusieurs conditions : des conditions hydroclimatiques favorables au développement de la cyanobactérie et l'absence de couverture nuageuse afin de disposer d'images satellites exploitables. Si ces conditions sont rencontrées il existe de nombreux autres biais dont il faut tenir compte afin de pouvoir détecter le *Trichodesmium*. Outre la correction atmosphérique, d'autres paramètres peuvent altérer le signal comme la présence d'autres algues mélangées au *Trichodesmium* ou encore la présence de hauts fonds.

### 4.3.1 Les propriétés optiques du *Trichodesmium*

Les propriétés optiques du phytoplancton ont été extensivement étudiées [Stramski and Mobley, 1997, Stramski et al., 2001]. Il apparaît que plus le phytoplancton est petit, plus il rétro-diffuse, et que les bactéries sont encore plus rétro-diffusantes [Morel and Maritorena, 2001].

Malgré sa grande taille, le *Trichodesmium* se caractérise par une forte rétro-diffusion qui serait due à la présence de vésicules gazeuses [Subramaniam et al., 1999a, Subramaniam et al., 1999b]. Il serait également remarquable par son absorption et sa fluorescence du fait de la présence de phycoérythrine [Subramaniam et al., 1999a]. Selon [Subramaniam et al., 1999b], lors de fortes efflorescences de *Trichodesmium* ( $[CHL] > 1g.L^{-1}$ ), la signature optique du *Trichodesmium* serait suffisamment caractéristique pour être distinguée des autres types de phytoplancton en présence.

### 4.3.2 Les algorithmes bio-optiques

Parmi les algorithmes bio-optiques de détection des efflorescences de *Trichodesmium* cités ci-dessus, seuls les algorithmes de [McKinna et al., 2011] et de [Hu et al., 2010] ont été implémentés et sont présentés ci-dessous. Les autres algorithmes utilisent la concentration de chlorophylle pour reconnaître le *Trichodesmium* [Dupouy et al., 2011, Gower et al., 2014, Westberry et al., 2005]. Or il a été montré que l'estimation de la concentration de chlorophylle sur des mattes de *Trichodesmium* pouvait être fortement sous-estimée [Subramaniam et al., 2002, Hu et al., 2010].

**McKinna** L'algorithme de détection des efflorescences de *Trichodesmium* présenté dans [McKinna et al., 2011]

se base sur l'analyse du spectre de réflectance d'une efflorescence de *Trichodesmium* mesuré en mer, tout

à fait semblable à celui mesuré sur des mattes de *Trichodesmium* sur filtre [Dupouy et al., 2008]. Ce spectre de réflectance est présenté dans la Figure 5.

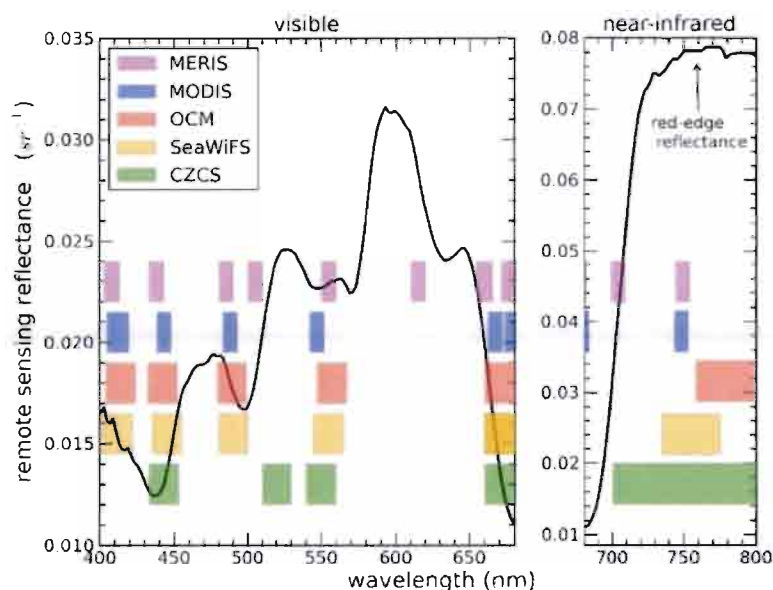


Figure 5 – Extrait de [McKinna et al., 2011]. Spectre du *Trichodesmium*. Les barres colorées représentent les bandes spectrales des satellites généralement utilisés dans les algorithmes bio-optiques de détection du *Trichodesmium*.

Sur ce spectre dans la partie « NIR-IR » (Near-Infra-Red, Infra-Red= Proche infra-rouge, Infra-rouge), on note une réflectance proche de zéro pour la longueur d'onde à 678nm, et des réflectances très fortes pour les longueurs d'onde proches de l'infrarouge. Outre ces caractéristiques, le domaine visible du spectre peut varier en fonction de la concentration et du taux d'agglomération du *Trichodesmium*



[McKinna, 2015]. La réflectance à 678nm semble rester la plus faible du spectre.

L'algorithme de McKinna [McKinna et al., 2011] utilise ces caractéristiques spectrales pour définir 4 critères de détection du *Trichodesmium* applicables au spectre de réflectance satellite (Rrs). Les trois premiers critères se basent sur la forme sur le spectre et sont données par les équations 8, 9 et 10. Le dernier critère écarte tout pixel dont la luminance normalisée est négative 11. Lorsque ces 4 critères sont respectés le pixel est identifié comme appartenant à une efflorescence de *Trichodesmium*.

$$Rrs(859) > c1.Rrs(678) \quad (8)$$

$$Rrs(645) > Rrs(678) \quad (9)$$

$$Rrs(555) > Rrs(678) \quad (10)$$

$$Rrs(555), Rrs(645), Rrs(678), \text{ ou } Rrs(859) < 0 \quad (11)$$

Dans [McKinna et al., 2011], l'algorithme est appliqué à des images MODISA L2 de la zone côtière Nord-Est australienne. Les données spectrales ont été interpolées à une résolution spatiale de 250m afin de pouvoir détecter et identifier les structures caractéristiques du *Trichodesmium*. En plein océan, la nature et la concentration des aérosols sont généralement différentes de celles sur les terres. La côte se trouvant à la frontière entre ces deux milieux, une correction standard adaptée au milieu océanique n'est pas forcément la plus adaptée. [McKinna et al., 2011] choisit d'appliquer une correction atmosphérique du type NIR-"SWIR" (Shortwave infrared) selon l'algorithme de Wang, 1999. Le NIR-SWIR permet de choisir entre deux types de correction atmosphérique en fonction du milieu en présence, on utilisera NIR pour la partie océanique, et le SWIR pour les côtes.

**Hu** Un autre algorithme de détection des efflorescences de *Trichodesmium* a été développé par Hu et al. (2010) sur des images MODIS Aqua de la côte ouest de la Floride et du golfe du Mexique. Cet algorithme se décompose en deux étapes : 1) la détection de mattes de phytoplancton avec un indice d'algue flottante (FAI) ; 2) la discrimination du *Trichodesmium* des autres types d'algues flottantes (e.g. *Sargassum*) grâce à des caractéristiques spectrales dans le visible. Pour cela [Hu et al., 2010] met à profit à la fois les bandes spectrales MODIS Aqua destinées à l'étude de l'océan mais également celles destinées au terrestre (dans le domaine NIR).

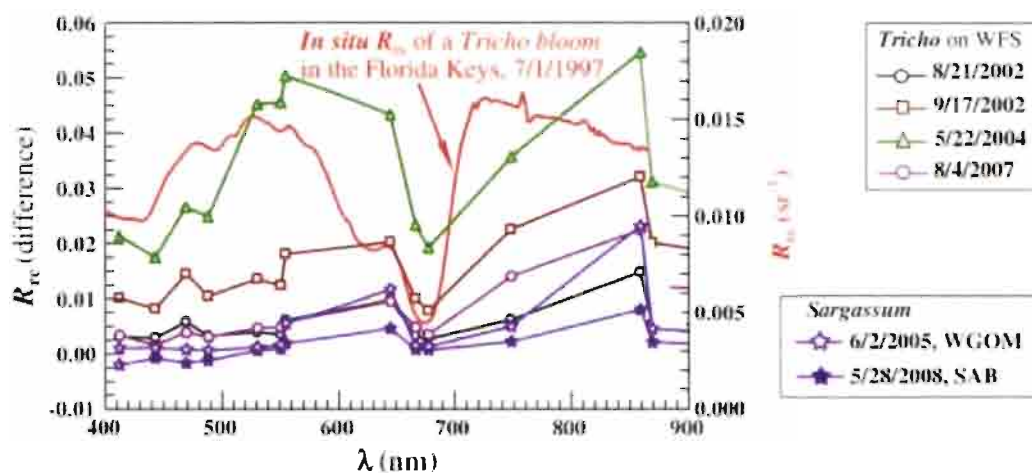


Figure 6 -

Le FAI est un indice développé par [Hu, 2009] dans le but de détecter les fortes réflectances dans l'infrarouge (red-edge) caractéristiques des agglomérations d'algues en surface de l'océan. En effet, dans ces conditions, les amas d'algues ont des propriétés optiques similaires à celles de la végétation terrestre,

qui se manifestent par une forte réflexion dans le domaine proche infrarouge. Afin d'éviter une sur-correction atmosphérique liée à l'effet red-edge des algues flottantes organisées en amas [Hu, 2009], le calcul de cet indice se base sur des réflectances corrigées des effets de Rayleigh uniquement ( $Rrc$ ). Cette correction permet de s'affranchir de la majeure partie de la couleur de l'atmosphère selon [Hu et al., 2010]. Le FAI est alors défini comme la différence entre une  $Rrc$  du proche infrarouge (859 nm pour MODIS) et une réflectance de référence ( $Rrc'$ ) calculée par interpolation linéaire entre le rouge et le proche infrarouge, respectivement 667 nm et 1240 nm pour le capteur MODIS. Selon [Hu et al., 2010], la différence entre  $Rrc$  et  $Rrc'$  permet de s'affranchir de la majorité des effets des aérosols qui ont une forme spectrale quasi-linéaire entre 667nm et 1240nm. Ce calcul est présenté dans l'équation 13.

$$FAI = Rrc(859) - Rrc'(859) \quad (12)$$

$$Rrc'(859) = Rrc(667) + \frac{Rrc(1240) - Rrc(667)}{1240 - 667} \quad (13)$$

La seconde étape consiste à identifier l'espèce d'algue identifiée mis en valeur par le FAI grâce à la forme du spectre dans le visible. De manière à corriger le biais induit dans la partie visible du spectre par la présence éventuelle d'aérosols, [Hu et al., 2010] propose d'appliquer aux pixels présentant une forte valeur de FAI, la correction d'une zone voisine sans efflorescence. Cette démarche étant très coûteuse en temps de calcul, elle est substituée par une simple différence entre le spectre  $Rrc$  des pixels suspects et celui d'une zone voisine sans efflorescence.

Après comparaison avec d'autres types d'agrégations d'algues flottantes (*Sargassum*, *Ulva prolifera* et *Microcystis Aeruginosa*) le spectre de différence de  $Rrc$  du *Trichodesmium* présente un motif qui semble lui être spécifique : un schéma du type haut-bas-haut-bas-haut pour les longueurs d'onde 469-488-531-551-555 nm, tel qu'illustré par la figure 6.

#### 4.4 Les algorithmes de classification

Il existe de nombreuses techniques de classification. Un algorithme à base de Support Vecteur Machine (SVM) a été appliqué à la reconnaissance d'un groupe d'algues dont fait partie le *Trichodesmium* [de Boissieu et al., 2014]. De même pour les réseaux de neurones [Smits et al., 1992]. Pour le *Trichodesmium* la limite d'application de ces méthodes réside dans le fait qu'il nous faut un nombre suffisant de mesures fiables à la fois satellite et sur le terrain pour pouvoir constituer les jeux d'apprentissage.

## 5 Les données

Afin de développer et d'appliquer les algorithmes de reconnaissance du *Trichodesmium*, il est nécessaire de constituer une base d'observations satellites en coïncidence avec les observations in-situ. Celle-ci servira de base d'analyse et d'apprentissage pour le réglage des algorithmes de reconnaissance (bio-optiques et statistiques) et permettra à travers des outils statistiques de juger de la pertinence des résultats de la détection.

### 5.1 Données in-situ

Deux jeux de données de terrain étaient disponibles. Le premier contient des observations visuelles de formation ou non de mattes de *Trichodesmium* à la surface de l'eau. Ces données géolocalisées s'appuient sur des photos aériennes et des observations consignées par la marine nationale.

Le second jeu de données in-situ est une compilation de bases de données internationales de concentration de cellules de *Trichodesmium* ou de Trichomes (DIAPALIS [Tenório, 2006, Neveux et al., 2006], MARE-DAT [Luo et al., 2012], PANDORA [Bonnet et al., 2015]). Cependant il n'indique pas si l'algue formait des mattes ou non. Les concentrations de cellules ou de Trichomes sont transformées en concentration de *Chl - a* spécifique au *Trichodesmium* et comparées à la concentration de *Chl - a* totale pour estimer la part de *Trichodesmium* par rapport aux autres espèces de phytoplancton. Les facteurs de conversion ont été définis selon la littérature avec respectivement 100 cell./Trichome et 0.2 ng *Chl-a*/Trichome [Carpenter et al., 2004, Finzi, 2006]. Dans les deux jeux de données nous disposons de la date et de la localisation géographique du point de prélèvement. Chaque jeu de données compte environ une centaine d'observations depuis le 04 juillet 2002, début de la période d'acquisition du satellite MODIS.

### 5.2 Données satellites

Les images satellites utilisées proviennent de deux capteurs sur le satellite MODIS, MODIS Terra et MODIS Aqua, respectivement mis en service en 2000 et 2002. Les données sont en libre accès sur les serveurs ftp de la NASA. La base de données satellite a été constituée à partir des images coïncidentes aux mesures in-situ. Étant donné une couverture nuageuse importante dans la région tropicale du Pacifique Sud, seulement 5% à 10% des scènes s'avèrent exploitables, i.e. en conservant les scènes comportant une fraction nuageuse inférieure à 60%. A cela s'ajoute les dynamiques marines et atmosphériques qui déplace les efflorescences entre le moment de la mesure in-situ et la prise de vue satellite.

Dans ce contexte, deux jeux de données ont été constitués. Le premier jeu de données, dit jeu de données « restreint » contient les pixels coïncidents à une zone  $\pm 0.005$  deg autour des points de mesure in-situ. Ce jeu de données est donc appuyé par une vérité terrain mesurée. Il contient 1238 pixels dont 556 identifiés comme du *Trichodesmium*.

Le deuxième jeu de données, dit jeu de données élargi, vise à identifier des zones d'efflorescences de *Trichodesmium* ou des zones vierges par photo-interprétation. Ce jeu de données est donc appuyé par une vérité terrain experte. Afin d'augmenter les chances d'observer des efflorescences de *Trichodesmium* sur les images satellites, le jeu de données élargi a été étendu au voisinage de  $\pm 3$  jours autour de la date de chaque mesure. Et pour limiter le temps d'analyse et de traitement des scènes la zone d'intérêt a été réduite à  $\pm 1$  deg autour de la localisation géographique de chaque mesure. La taille de cette zone d'intérêt se justifie en prenant pour hypothèse que la vitesse des courants marins ne dépasse généralement pas  $10 \text{ cm.s}^{-1}$  soit environ  $0.1 \text{ deg.jour}^{-1}$  pour les latitudes tropicales, ce qui est le cas dans la zone. Pour chaque mesure in-situ la base de données contient donc 7 scènes de 2 deg x 2 deg (soit environ  $1000 \times 1000$  pixels). Parmi l'ensemble des scènes, une dizaine ont été analysées afin d'y sélectionner des zones annotées avec la présence ou non d'efflorescences de *Trichodesmium* visible sur les images vraie-couleur et les concentration de chlorophylle. La base de données de *Trichodesmium* élargie contient environ 340296 pixels dont 9908 identifiés comme du *Trichodesmium*.

Ces deux jeux de données comprennent les variables de réflectance *Rrc* (correction des effets Rayleigh seulement) et *Rrs* (correction atmosphérique standard : Rayleigh + aérosols) pour chaque longueur d'onde MODIS entre 412nm à 1240nm. A cela ont été ajoutés les coefficients d'absorption et de rétrodiffusion

estimés à partir des données satellites pour les longueurs d'onde du visible (412nm à 678nm). Au total, les jeux de données sont donc constitués de 48 variables.

Parmi les scènes de la base de données, nous avons sélectionné 3 scènes sur lesquelles des efflorescences de *Trichodesmium* sont bien visibles. Ces scènes m'ont notamment servi de test pour le développement et l'ajustement des algorithmes de reconnaissance.

La première scène sélectionnée que l'on nommera scène AUSTEST est l'une de celle utilisée dans [McKinna et al., 2011]. Cette scène date de 17 octobre 2007 et se situe au niveau du canal du Capricorne, entre la côte Est australienne et la Grande Barrière de Corail (granule A2007290.0355). Cette scène présente plusieurs intérêts : l'acquisition satellite de cette scène s'est déroulée juste après un fort événement pluvieux apportant beaucoup de nutriments dans le canal. Cette fertilisation de la zone a conduit à un développement massif de *Trichodesmium*. Bien qu'il n'y ait pas eu de mesure scientifique lors de ce phénomène, les structures filamentaires observées sur l'image vraie couleur (Figure 7A) sont significatives d'une algue flottante selon [McKinna et al., 2011]. McKinna et al. s'appuient également sur une présence régulière d'efflorescences de *Trichodesmium* dans cette région, tel que signalé par [Kuchler and Jupp, 1988]. Par contre, les efflorescences visibles dans cette scène se trouvent en partie au-dessus de haut-fonds. Cela constitue une difficulté supplémentaire pour la reconnaissance, car ils perturbent le signal optique des efflorescences. Cette scène présente également une faible couverture nuageuse. Enfin, cette scène m'a permis de tester mon implémentation de l'algorithme bio-optique de McKinna et de le comparer aux autres algorithmes implémentés durant mon stage.

La deuxième, que l'on nommera scène SPOT (SPOT= South Pacific Time Series, ayant fait la couverture du site NASA EarthExplorer <http://earthobservatory.nasa.gov/IOTD/view.php?id=85073>), est une scène recouvrant un point de mesure de la campagne en mer SPOT8 (Figure 7B). Elle a été prise le 20 décembre 2014 au large de la côte Est de Nouvelle-Calédonie (granule A2014353.0240). Elle présente une très forte efflorescence de *Trichodesmium* confirmé visuellement lors de la campagne. Mais malheureusement, l'analyse des concentrations en *Trichodesmium* n'a pas encore pu être effectuée.

Enfin la troisième scène, qu'on nommera NCEST du fait qu'elle est située à l'Est de la Nouvelle-Calédonie (Figure 7C). Elle a été prise le 27 janvier 2004 et a pour identifiant de granule : A2004027.0255. Encore une fois une très forte concentration de *Trichodesmium* a été mesurée et les propriétés optiques des mattes mesurées. Contrairement à AUSTEST et SPOT, les données de la scène NCEST n'ont pas été incluses dans les jeux de données d'apprentissage. Cette scène constitue donc un exemple pour juger de la généralisation des algorithmes de reconnaissance des efflorescences de *Trichodesmium* à de nouvelles scènes.

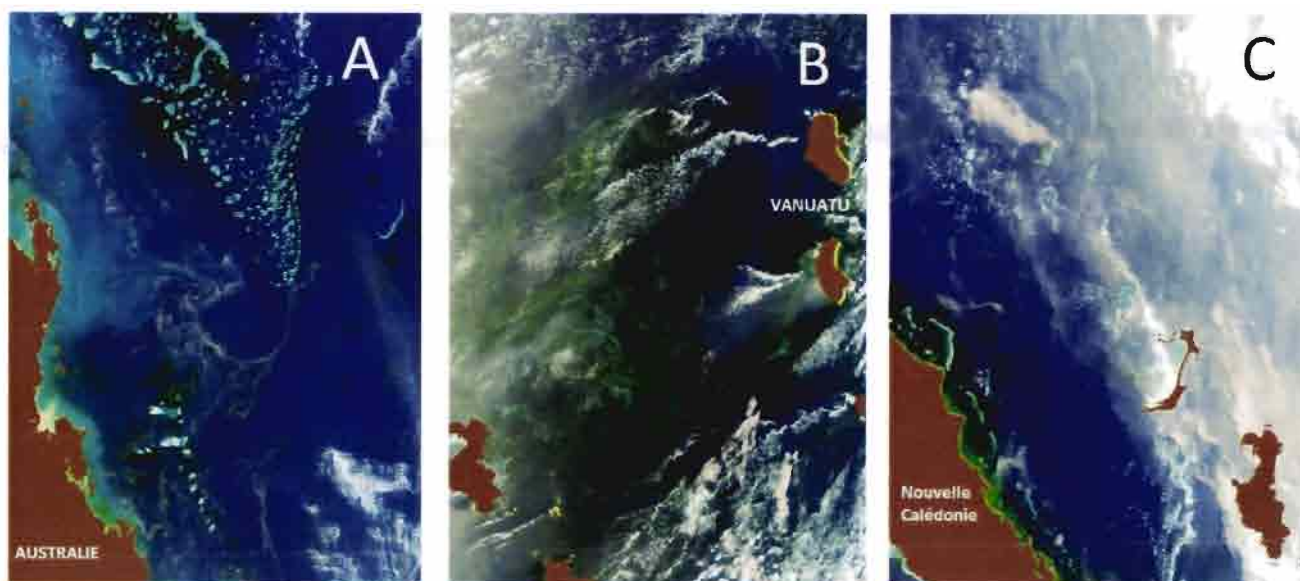


Figure 7 – Images True Color des scènes AUSTEST(A) SPOT(B) et NCEST(C)

### 5.3 SeasDAS

SeasDAS est un logiciel complexe dont de nombreuses fonctionnalités ont été ajoutées au cours du temps. Une interface graphique est intégrée au logiciel, mais il est possible d'écrire directement par des scripts pour effectuer certaines tâches. Lors de mon stage, j'ai utilisé deux versions de SeasDAS, la SeasDAS 6.4 et SeasDAS 7.1. SeasDAS 7.1 a été utilisé pour les traitements du niveau L1A au L2 car cette version dispose d'algorithmes à jour et plus performants que les versions précédentes. SeasDAS 6.4 permet une implémentation plus facile d'algorithmes à partir de script IDL. De plus, cette version de SeasDAS ne présente pas de problème de visualisation des scènes traitées contrairement à la version 7.1 toujours en développement sur ce plan. Cette version a donc été utilisée pour l'implémentation des algorithmes bio-optique, pour l'intégration des produits développés au fichier de produits MODIS, et pour la visualisation des résultats.

## 6 Reconnaissance du *Trichodesmium* par des algorithmes bio-optique

Afin d'évaluer les algorithmes bio-optiques de la littérature sur une région/image différente de celle du développement, j'ai tout d'abord implémenté et appliqué ceux de [McKinna et al., 2011] et [Hu et al., 2010] aux trois scènes d'études. Au vu des résultats, j'ai essayé d'adapter ces algorithmes afin d'améliorer la reconnaissance sur les 3 scènes d'étude. Enfin, en essayant de combiner les avantages des deux algorithmes j'ai abouti au développement de mon propre algorithme bio-optique. Dans une première section, je présente dans une première section l'adaptation et l'évaluation des algorithmes de [McKinna et al., 2011, Hu et al., 2010]. Dans une deuxième section, le nouvel algorithme que j'ai développé et ses résultats sont exposés dans la deuxième section de cette partie.

### 6.1 Les algorithmes de la littérature

Au cours de ce stage, j'ai implémenté les algorithmes de [McKinna et al., 2011] et de [Hu et al., 2010] tels que mentionnés dans la partie État de l'Art. Ces algorithmes ont été appliqués à la scène AUSTEST (Figure 7A). La suite de cette section présente les résultats obtenus et leur analyse.

#### 6.1.1 Algorithme de McKinna

Tout d'abord, l'algorithme de McKinna est appliqué de manière identique à [McKinna et al., 2011], i.e. avec une correction atmosphérique NIR-SWIR des données satellites. Le résultat sur la scène AUSTEST est présenté en Figure 8.

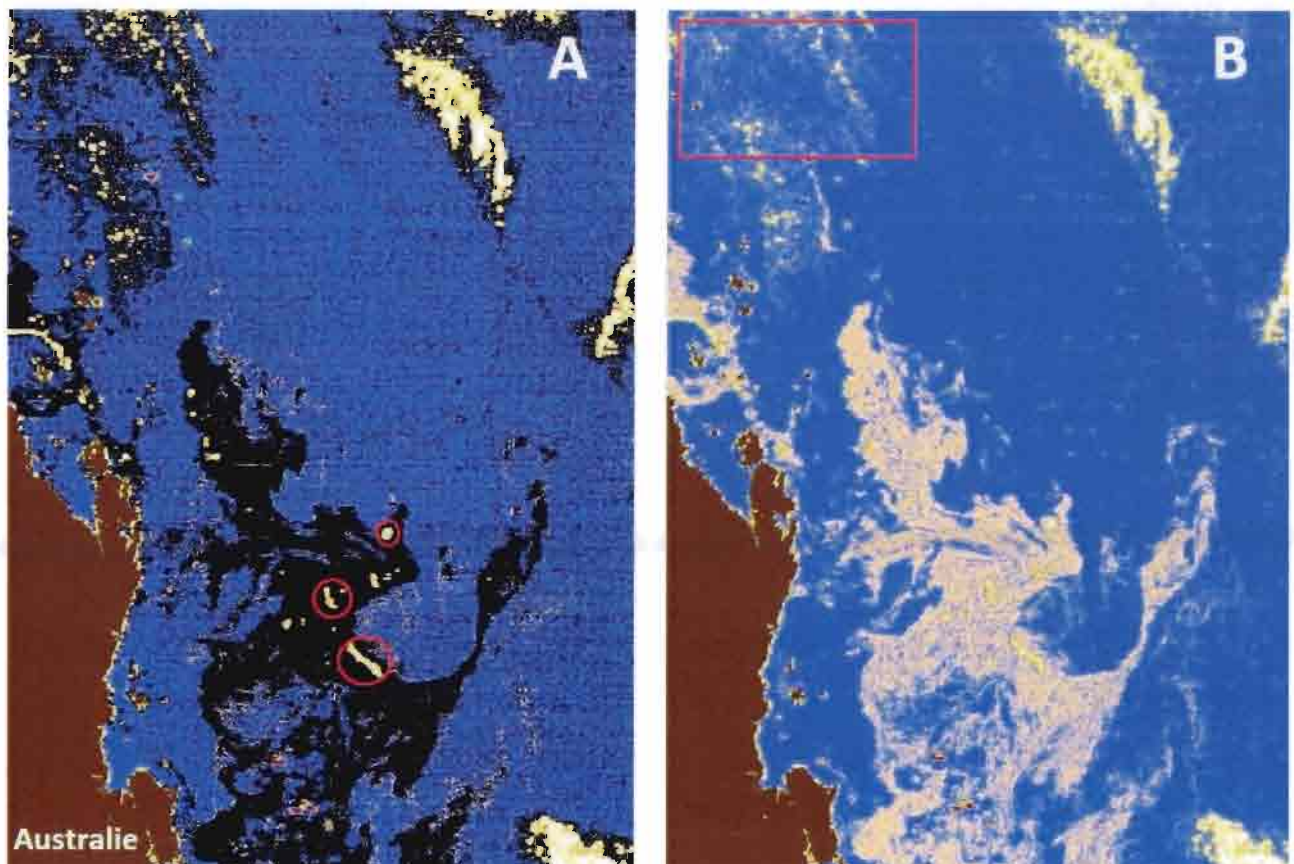


Figure 8 – Détection de *Trichodesmium* à l'aide d'un algorithme de détection (McKinna et al., 2011) sur la scène MODIS-Aqua prise le 17 Octobre : (A) Avec une correction atmosphérique Nir-Swir. (B) Avec la correction atmosphérique Nir-Swir et sans utiliser le critère supprimant les réflectances négatives (Équation 11).

On peut noter que la majeure partie des pixels identifiés comme du *Trichodesmium* dans [McKinna et al., 2011] comportent des valeurs de réflectances négatives (en noir sur 8A) et sont écartés par l'équation 11 de l'algorithme en utilisant la nouvelle chaîne de traitement basée sur SeaDAS 7. On note également que certaines zones au centre des efflorescences sont masquées car identifiées comme des nuages (encadrées en rouge sur la Figure fig :resultMKA) alors que l'image RGB (Figure 7A) n'indique pas la présence de nuages dans cette zone. L'implémentation de l'algorithme en lui-même n'étant pas en cause, nous avons attribué ces différences, par rapport aux résultats de [McKinna et al., 2011], à une modification des algorithmes de correction atmosphérique lors du passage de SeasDAS 6 à SeasDAS 7. En effet, lorsque le masque des réflectances négatives est retiré, le résultat (figure 8B) correspond plus à celui de [McKinna et al., 2011] où on peut observer une très grande quantité de *Trichodesmium* (pixels beiges). De même, lorsque l'algorithme est appliqué aux réflectances non corrigées atmosphériquement (Rrc), le masque de réflectances négatives n'intervient presque plus dans la zone d'intérêt (Figure absente). Enfin, tout autour de la scène on peut constater d'autres faux positifs induit par l'algorithme lui-même. Ces pixels se distinguent par leur répartition spatiale ne correspondant pas à ce qu'on pourrait attendre d'une matte de *Trichodesmium* et peuvent donc être éliminés (zone entourée en rouge dans la Figure 8B).

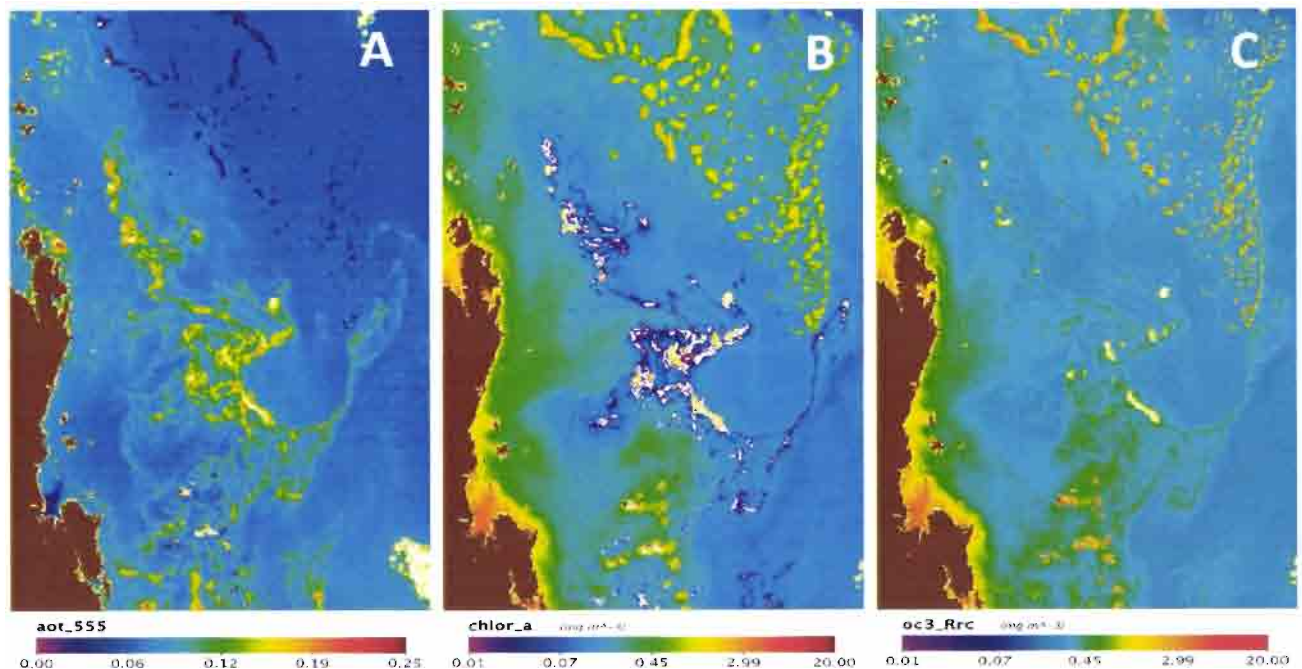


Figure 9 – Scène MODIS-Aqua A2007290.0355 montrant : (A) l'épaisseur optique mesurée à la longueur d'onde 555nm, (B) une diminution de concentration de la chlorophylle autour des mattes de *Trichodesmium*, (C) la chlorophylle recalculée sans la correction des aérosols, à partir des Rrc.

La sur-corrrection atmosphérique est un problème général aux fortes concentrations d'algues constaté à plusieurs reprises [Hu et al., 2010]. Les fortes réflectances dans l'infrarouge dues au fait que la partie chlorophyllienne (red-edge) des mattes est située au-dessus de l'eau sont interprétées par l'algorithme de correction atmosphérique comme provenant d'aérosols puisqu'en principe dans le NIR l'océan est un corps noir. Cette correction aboutit à des valeurs de réflectances exagérément diminuées voire négatives. Dans le cas extrême d'agglomérats très denses d'algues flottantes, l'algorithme de correction atmosphérique les identifie comme des nuages.

Ce problème de correction atmosphérique a des conséquences sur l'ensemble des produits dérivés L2, tous calculés à partir des réflectances corrigées (Rrs). Ce phénomène est notamment visible sur le produit standard de concentration en chlorophylle (Figure 9B) évalué selon l'algorithme OC3 de [Hu et al., 2012]. Sur cette figure la concentration en chlorophylle tombe à zéro juste au niveau des efflorescences de *Trichodesmium*, alors que la concentration réelle est sûrement plus élevée dans les efflorescences qu'en dehors

(Tenorio 2006, y mesure jusqu'à  $200 \text{ mg.m}^{-3}$ ). De plus l'indice d'épaisseur optique, significatif de la présence ou non d'aérosol, Figure 9A, indicateur de la présente exactement les mêmes structures filamenteuses que la chlorophylle. Or il est peu vraisemblable que des structures d'aérosols soient aussi précises. Celles-ci se rapportent plutôt à des structures marines. La correction atmosphérique NIR-SWIR employée dans [McKinna et al., 2011] pour éviter cette sur-correction ne semble pas suffisante.

Sachant que la correction atmosphérique est erronée à cause d'une mauvaise estimation des aérosols, comme préconisé par Hu et al., 2011, un nouveau produit chlorophylle a été calculé en ne corrigeant que les effets de Rayleigh, Figure 9C. On distingue toujours la même structure marine mais cette fois-ci accompagnée d'une concentration en chlorophylle plus forte.

Enfin, tout autour de la scène on peut constater d'autres erreurs de reconnaissance induites par la présence de nuages. Ces pixels se distinguent par leur répartition spatial ne correspondant pas à ce qu'on pourrait attendre d'une matte de *Trichodesmium* (zone entourée en rouge dans la Figure 8A).

### 6.1.2 Algorithme de Hu

L'algorithme de Hu a aussi été appliqué à la scène AUSTEST. Pour éviter les problèmes de sur-correction atmosphérique, Hu et al. (2010) ont fait le choix de ne corriger les réflectances que de la diffusion atmosphérique de Rayleigh. Le calcul du FAI avec ces données atteste de l'efficacité de cette approche : sur la Figure 10A, les zones comprenant d'importantes concentrations de *Trichodesmium* correspondent à de fortes valeurs de FAI. Cependant, on retrouve également de fortes valeurs de FAI aux abords des nuages (faux positifs). La condition d'un fort FAI n'est donc pas suffisante pour permettre l'identification des mattes de *Trichodesmium* sans ambiguïté.

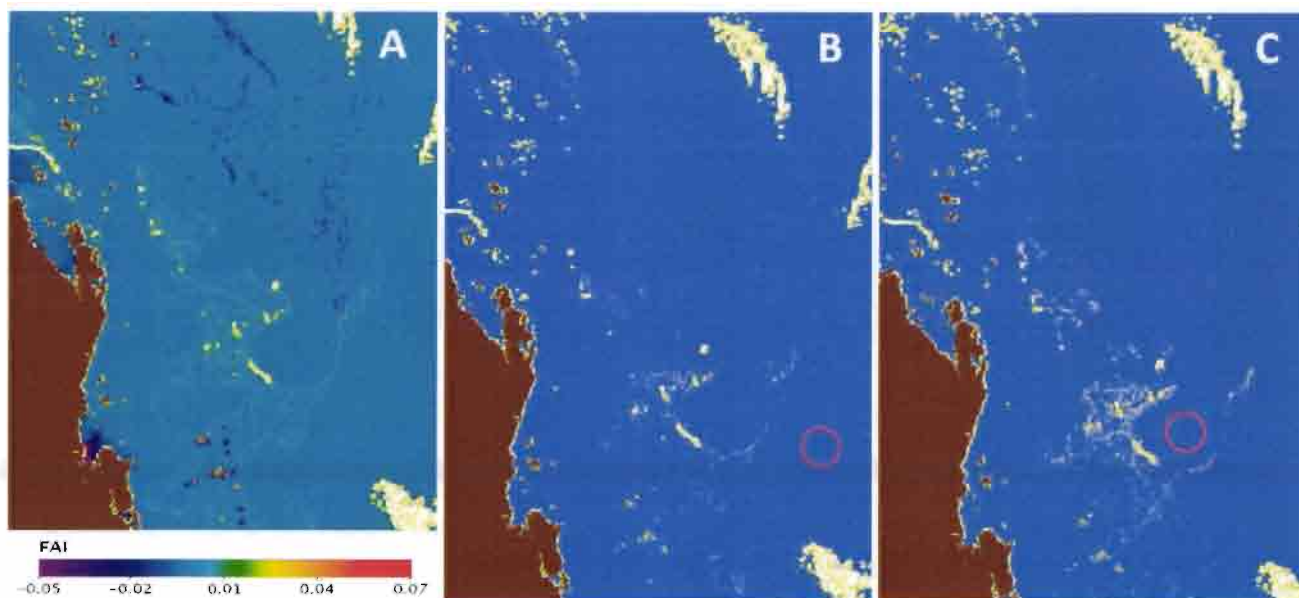


Figure 10 – Scène MODIS-Aqua A2007290.0355 montrant (A) la distribution de l'indice FAI (Hu et al. 2010) mettant en évidence les mattes de *Trichodesmium* également repérées par McKinna et al. (2013). (B) et (C) l'influence de la zone de référence sur les résultats de l'algorithme de Hu.

La seconde partie de l'algorithme de Hu a pour objectif d'identifier la nature des anomalies de FAI (mattes de *Trichodesmium* ou non). La méthode simplifiée, i.e. la différence de Rrc avec une zone de référence, est employée pour effectuer la correction d'éventuels aérosols sur la zone. La zone de référence pour cette correction se trouve entourée en rouge sur les figures 10B et 10C.

Ces figures présentent le résultat de classification obtenu sur la scène AUSTEST. Les mattes de *Trichodesmium* détectées (pixels beiges) par l'algorithme ne correspondent qu'à une petite partie des structures observées sur l'image True Color (Figure 7A). En effet, les 4 critères sur la forme du spectre entre 468nm



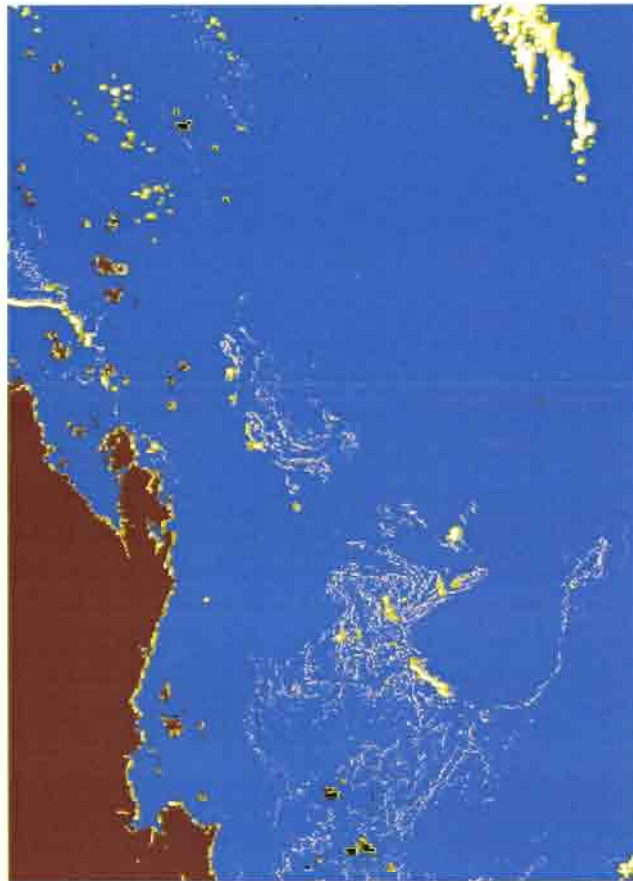


Figure 11 – Adaptation de l’algorithme de Hu par suppression de certains critères de forme

et 555nm sont rarement respectés tous en même temps. Enfin, on peut constater par la différence de résultats entre la figure 10B et 10C, que la méthode de correction des aérosols s’avère très sensible au choix de la zone de référence. De plus les différences de réflectances utilisées comme critère de forme sont faibles, cette correction peut faire varier énormément le résultat. Sur la Figure 11, le critère de forme concernant les longueurs d’ondes 531nm et 555nm a été supprimé. A partir du support visuel (Figure 7A) nous constatons que l’algorithme parvient à mieux détecter la structure de *Trichodesmium* sans ajouter de faux positifs notables. Cependant, encore une fois ce résultat dépend de la zone de référence choisie en premier lieu, et l’adaptation des critères variera en conséquence aussi.

Ainsi, le FAI nous permet d’obtenir des résultats convenables pour la détection de mattes d’algues flottantes à condition d’avoir une couverture nuageuse quasi-nulle. En revanche, la deuxième partie de l’algorithme de Hu ne nous donne pas des résultats satisfaisants. Cette méthode manuelle dont le résultat dépend trop du choix de la zone de référence n’a donc pas été utilisée par la suite.

### 6.1.3 Les autres algorithmes

Comme la plupart des auteurs, nous avons bien constaté de très fortes réflectances dans l’infrarouge au-dessus des mattes de *Trichodesmium*, perturbant ainsi la correction atmosphérique avec des produits satellites, dont la chlorophylle, se retrouvant avec des valeurs biaisées, et affectant par la même occasion d’autres algorithmes basés sur ces produits de type bio-optique. Nous n’implémenterons donc pas les algorithmes de Gowen, Westberry et Dupouy [Dupouy et al., 2011, Gower et al., 2014, Westberry et al., 2005] dont une partie de la méthode est basée sur des concentrations de chlorophylle et de plus adapté à SeaWiFS et non MODIS.

L’algorithme de Subramaniam [Subramaniam et al., 2002] développé sur le satellite Seawifs et antérieur

à l'algorithme de Westberry, pourrait être adapté au satellite Modis. Les deux satellites possèdent des canaux proches et l'algorithme consiste en une série simple de critères binaires. Cependant un précédent stage de Master, réalisé par Sandrine Bouisset, a démontré que cet algorithme n'était pas adapté à la zone du Pacifique Sud-Ouest.

## 6.2 Développement d'un algorithme bio-optique

En sortie des algorithmes de détection du *Trichodesmium* de Hu et McKinna, on constate un très grand nombre de détections à proximité et à distance des nuages. Ces points sont souvent issues d'une fausse détection par les algorithmes bio-optiques, mais nous n'avons pas les moyens de différencier systématiquement les bonnes des mauvaises détections. De plus la zone du Pacifique Sud subit une très importante couverture nuageuse rendant inexploitable la plupart des scènes capturées par satellite. Nous souhaitons donc améliorer la détection du *Trichodesmium* en limitant le nombre de faux positifs et ainsi exploiter un maximum de scènes, en rendant fiable les zones comme les trouées nuageuses. L'algorithme de McKinna ayant montré des résultats prometteurs, il servira de base à notre nouvel algorithme.

### 6.2.1 Utilisation de la correction atmosphérique

A l'aide du spectre typique du *Trichodesmium* de la littérature, nous savons que la valeur en *Rrs* pour la longueur d'onde à 678nm est toujours très faible approchant zéro. Associée à une trop forte correction atmosphérique induit par la présence de *Trichodesmium* cette valeur peut se retrouver négative. La première partie de notre algorithme consistera à écarter toutes les valeurs positives de *Rrs*(678), mettant ainsi en évidence toutes les mattes de *Trichodesmium* ayant subi une trop forte correction atmosphérique. Ensuite, nous avons constaté que les plus fortes agglomérations de *Trichodesmium* correspondaient aux endroits avec la plus forte correction atmosphérique. Ainsi en prenant l'inverse de *Rrs*(678) nous obtiendrons un domaine continu qui pourrait dénoter la concentration d'une matre de *Trichodesmium* relativement aux autres mattes détectées.

$$Rrs678 < 0 \quad (14)$$

### 6.2.2 Critère de forme

Par la suite on applique des critères de forme telle que dans l'algorithme de McKinna pour en extraire des caractéristiques plus fines du *Trichodesmium*. Pour déterminer ces critères, nous prenons des zones identifiées comme étant du *Trichodesmium* que l'on compare directement au spectre de faux positifs détectés. Par une analyse des spectres on essaie d'extraire les caractéristiques principales permettant de distinguer les deux spectres. Nous obtenons les deux critères suivants :

$$Rrc(748) < Rrc(859) \quad (15)$$

$$Rrc(645) < Rrc(531) \quad (16)$$

### 6.2.3 Résultats

Nous utilisons de nouveau la scène AUSTEST pour tester cet algorithme. Nous obtenons les mêmes structures spatiales paternes de *Trichodesmium* que pour l'algorithme de McKinna. Cependant là où l'algorithme de McKinna offrait une très vaste étendue de *Trichodesmium* dont on ne pouvait distinguer la structure interne, nous pouvons à présent distinguer la répartition filamentaire des mattes de *Trichodesmium*.

Les zones supplémentaires détectées par l'algorithme de McKinna et identifiées comme faux positifs, ne sont plus détectées par notre algorithme et cela même au plus près des nuages. Notre algorithme semble beaucoup plus strict dans la détection du *Trichodesmium*, mais limite le nombre de fausses détections. Afin de tester plus en avant notre algorithme, nous l'avons appliqué à nos autres scènes de références

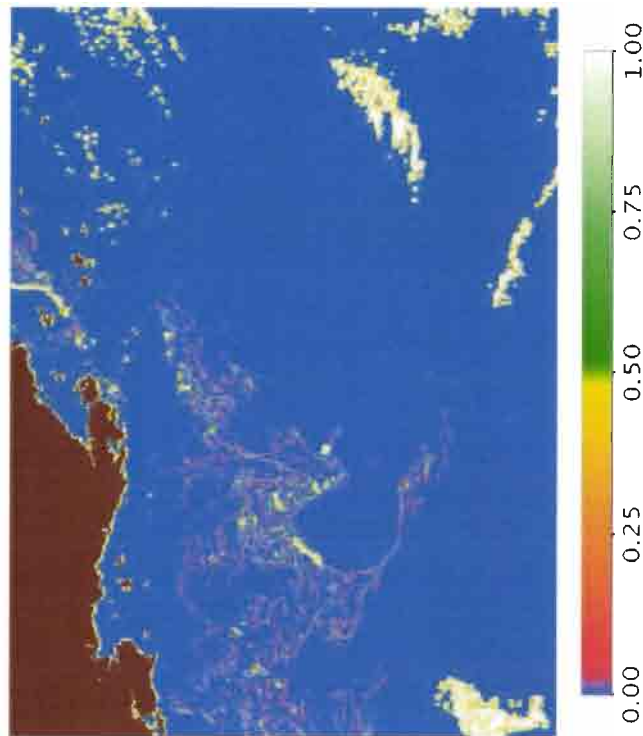


Figure 12 – Résultat du nouveau algorithme bio-optique développé sur la scène AUSTEST

Comme dit précédemment, en retenant le FAI, l'algorithme de McKinna et notre propre algorithme, on souhaite appliquer nos algorithmes à d'autres zones du Pacifique. Pour cela nous avons identifié deux zones proches de la Nouvelle-Calédonie avec une très forte concentration en *Trichodesmium*, les deux scènes SPOT et NCEST.

Sur la première scène, NCEST, nous arrivons avec l'ensemble de nos algorithmes à détecter des formes de *Trichodesmium*. Ces algorithmes peuvent donc fonctionner dans cette partie de l'océan. Cependant avec notre deuxième scène, SPOT, on ne détecte qu'une partie infime du *Trichodesmium* qui devrait être détectée car visible avec l'image True Color 7B.

En observant le produit de chlorophylle de cette scène on constate que ses valeurs sont relativement hautes près du *Trichodesmium*, or on avait fait le constat inverse avec la scène de McKinna, ce qui indique que les mattes sont moins concentrées et que la correction atmosphérique est donc correcte sur une plus grande partie de l'efflorescence. Donc, en écartant l'hypothèse d'une espèce de *Trichodesmium* différente, il ne reste que le cas où le *Trichodesmium* ne forme pas de mattes partout. Nous pouvons avoir une très forte concentration d'algues, mais il est possible qu'avec des conditions météorologiques océaniques telles que beaucoup de vent, que le *Trichodesmium* se soit dispersé et mélangé verticalement. Le *Trichodesmium* de s'agglomérant pas dans ce cas, nous perdons notre fort signal dans l'infrarouge.

Nous sommes alors en présence d'une très forte efflorescence de *Trichodesmium* que nous ne pouvons détecter entièrement avec nos algorithmes bio-optiques. Ceux-ci sont performants pour localiser les mattes de cette algue, mais ne parviennent pas à détecter celle-ci lorsqu'elle est dispersée sous la surface de l'eau même à de très importantes concentrations. Il nous faudrait alors un nouvel algorithme permettant de détecter ce nouveau cas de figure afin de pouvoir par exemple estimer la fixation d'azote de l'ensemble de l'efflorescence.

En extrayant les spectres de notre *Trichodesmium* dispersé, il ne nous a pas été possible d'identifier distinctement des caractéristiques permettant de dissocier le *Trichodesmium* du reste, comme constaté par

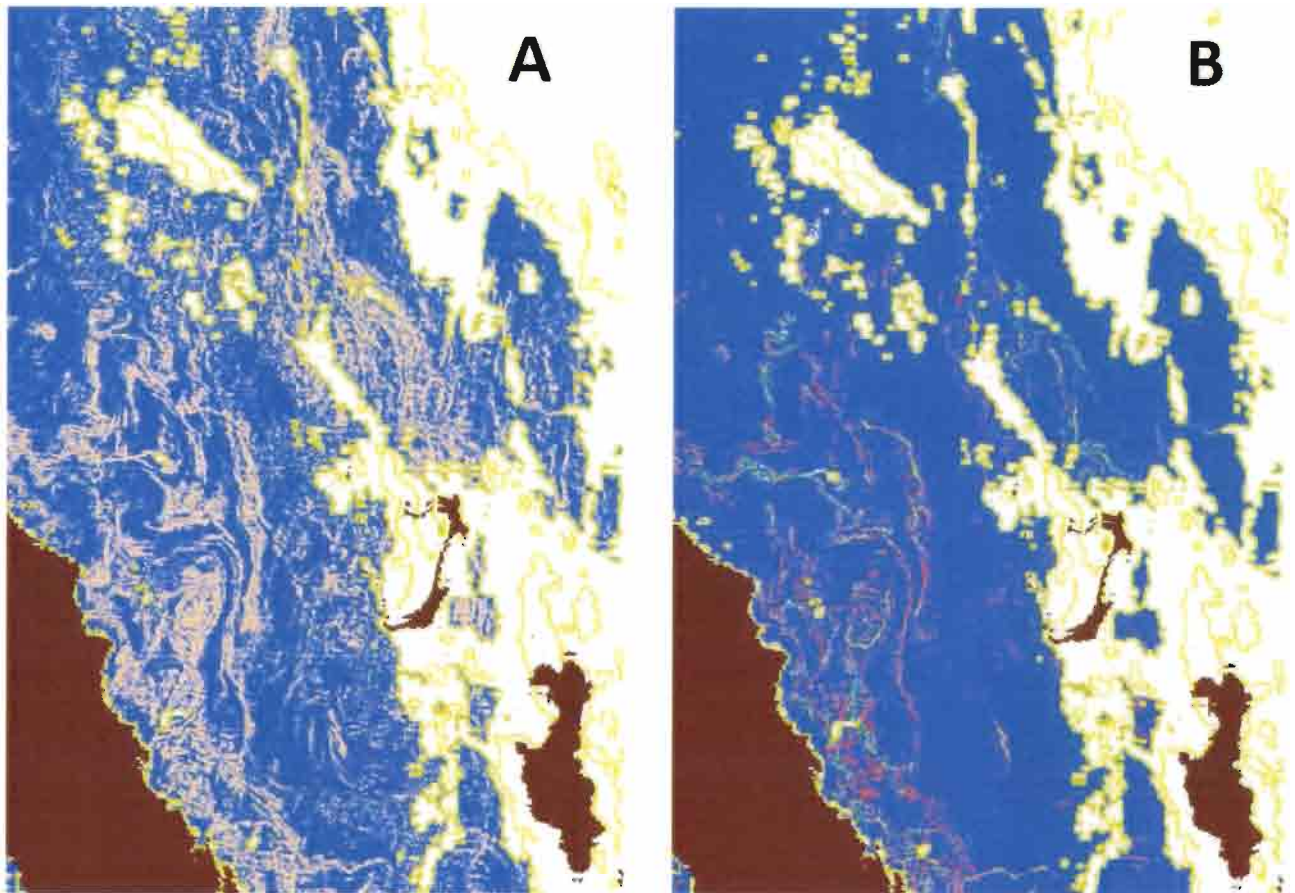


Figure 13 – Application de l’algorithme de Mckinna (A) et de notre algorithme (B) à la scène NCEST

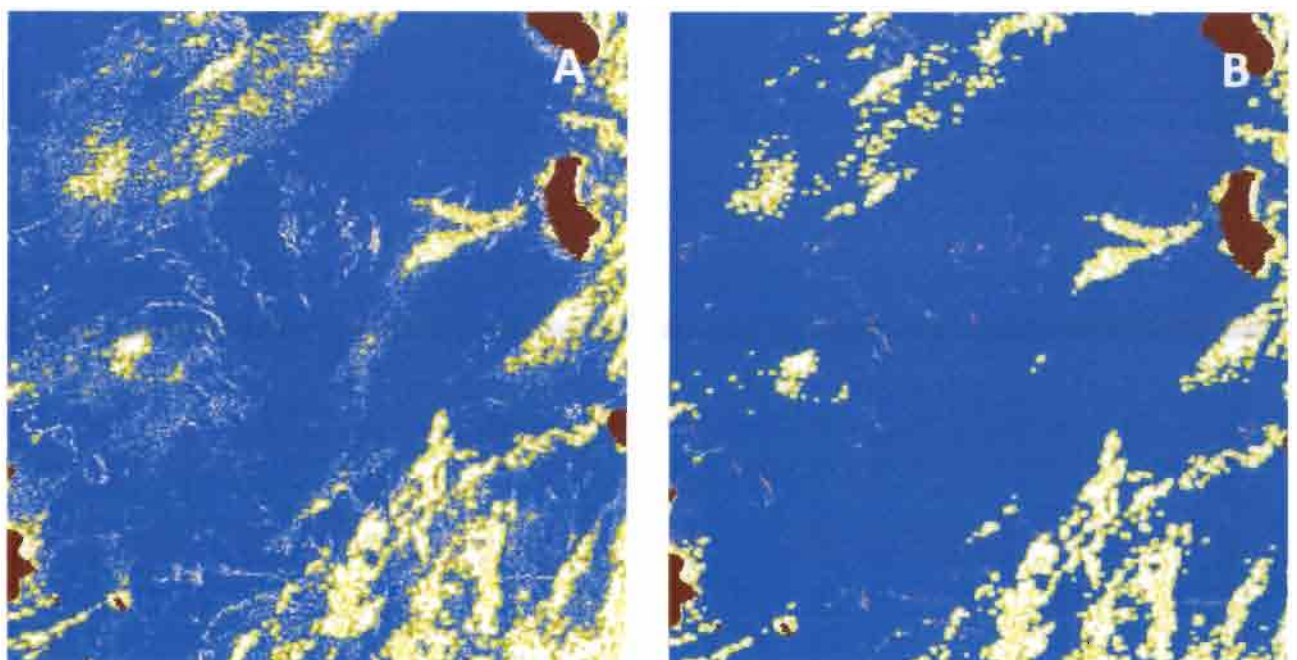


Figure 14 – Application de l’algorithme de Mckinna (A) et de notre algorithme (B) à la scène SPOT

Dupout et al., 2008. Même en ajoutant des variables telles que les valeurs de rétro-diffusion et d’absorption, il n’a pas été possible d’établir un lien direct avec les caractéristiques du *Trichodesmium*. Nous nous

retrouvons avec un nombre conséquent de variables avec un problème complexe qui, avec nos connaissances actuelles, ne nous permettent pas de créer un nouvel algorithme bio-optique pour identifier le *Trichodesmium* sous la surface de l'eau.

## 7 Reconnaissance du *Trichodesmium* par méthode d'apprentissage supervisé

Les méthodes statistiques de classification, telles que les réseaux de neurones, les SVM, les réseaux bayésiens, etc., présentent l'intérêt d'apprendre automatiquement les relations statistiques permettant de relier les variables d'entrées à la classe de sortie. Lorsque l'apprentissage utilise des données annotées avec la classe de sortie, l'apprentissage est dit supervisé. Dans le cas inverse (classes non-connues à l'avance) on parle d'apprentissage non-supervisé (ou clustering en anglais). A partir des deux bases de données présentées dans la partie 6 de ce rapport, nous avons souhaité évaluer le potentiel des méthodes statistiques de classification supervisée pour la détection du *Trichodesmium*. Parmi les différentes méthodes existantes et au vu du temps de stage restant, nous avons choisi de nous concentrer sur une méthode de classification SVM.

### 7.1 Méthodes

La classification SVM est une méthode de classification supervisée reconnue pour ses bonnes performances, en particulier pour des applications en télédétection [Mountrakis et al., 2011]. Cette méthode est adaptée en particulier à la séparation non-linéaire de classes et aboutit le plus souvent à une bonne généralisation de la classification sur des données nouvelles, même avec un jeu de données d'apprentissages limité. De plus, contrairement à d'autres méthodes de classification, les SVM n'ont pas besoin d'une connaissance a priori de la distribution des données. Enfin, dans une étude récente [de Boissieu et al., 2014], cette méthode a notamment montré de bons résultats pour la reconnaissance des types fonctionnels de phytoplancton les plus communément rencontrés sur le globe (*Diatomées*, *Nanoecaryotes*, *Synechococcus* et associées, *Prochlorococcus*). La classification SVM, comme l'ensemble des méthodes de classification supervisée, nécessite un jeu de données d'apprentissage représentatif des caractéristiques à reconnaître. Dans le cadre de mon stage, j'ai appliqué la classification SVM selon les deux approches définies par les deux jeux de données décrits dans la partie 6. Avec le jeu de données restreint, nous souhaitons évaluer le potentiel de la classification SVM avec un petit nombre de données mais pour lesquels nous avons la certitude, grâce aux mesures in-situ, que des efflorescences de *Trichodesmium* avaient lieu. Avec le jeu de données élargi, construit par photo-interprétation, nous souhaitons évaluer la reconnaissance SVM des efflorescences typiques observées par satellite dans la région Nouvelle-Calédonie/Manuatu assimilées à du *Trichodesmium*. L'algorithme de classification SVM mis en place se base sur les fonctions de la bibliothèque `e1071` de R, elle-même issue de la bibliothèque LIBSVM [Chang and Lin, 2011]. Plusieurs paramètres, tel que la fonction noyau et le coefficient de régularisation  $C$  sont réglables. Cependant, nous nous sommes limités à la configuration par défaut qui donne généralement de bons résultats. La fonction noyau est une Gaussienne à base radiale dont la variance est paramétrée grâce au coefficient  $\gamma = 1/d_x$  avec  $d_x$  le nombre de variables d'entrée. Le paramètre  $C$  est fixé à 1.

#### 7.1.1 Répartition des données et sélection des variables

Comme pour beaucoup d'algorithmes d'apprentissage statistique, l'algorithme SVM nécessite un jeu de données d'apprentissage équi-réparti entre les classes à reconnaître. Dans le cas contraire, le modèle tend à reconnaître préférentiellement les classes ayant le plus grand nombre d'exemples d'apprentissage. Pour le jeu de données restreint, le nombre de données d'apprentissage est fixé à 500 individus par classe. Pour le jeu de données élargi, le nombre de données d'apprentissage est fixé à 1000 individus par classe. Les données restantes sont utilisées pour l'évaluation de généralisation des performances à des données nouvelles.

L'apprentissage peut être également biaisé si le nombre d'individus n'est pas suffisant au vu du nombre de paramètres du modèle à entraîner. Dans ce cas, le modèle aura de biens meilleurs résultats de reconnaissance sur des données d'apprentissage que pour des données nouvelles, et biaisant ainsi l'évaluation des performances de classification. Ce phénomène est appelé le sur-apprentissage. De plus, certaines variables peuvent s'avérer redondantes. Ainsi, pour le jeu de données restreint une sélection de variables est effectuée avec un algorithme `glouton`. Partant d'un jeu de variable vide, cet algorithme sélectionne

séquentiellement les variables permettant d'augmenter le plus les performances de reconnaissance. La sélection est arrêtée lorsque les performances n'augmentent plus et/ou lorsque le sur-apprentissage intervient.

### 7.1.2 Validation des algorithmes

Pour évaluer les performances des modèles nous avons utilisé la méthode de k-fold cross-validation. Cette méthode consiste à séparer aléatoirement le jeu de données d'apprentissage en k parties égales et à utiliser successivement chaque partie pour l'évaluation de la généralisation de l'algorithme tandis que les k-1 parties restantes servent à l'entraînement des paramètres. Les performances sont ensuite moyennées sur l'ensemble de ces combinaisons. Cette méthode permet d'optimiser l'utilisation du jeu de données d'apprentissage pour l'évaluation du comportement de l'algorithme face à de nouvelles données. Afin d'évaluer le potentiel de l'algorithme indépendamment de la sélection aléatoire des jeux d'apprentissage, en particulier pour le jeu de données élargi est qui fortement déséquilibré entre les deux classes, la sélection des données d'apprentissage ainsi que la k-fold cross-validation est renouvelée 10 fois, donnant lieu à 10 expériences sur lesquels les performances sont moyennées. L'évaluation des performances de l'algorithme est effectuée à travers la matrice de confusion qui récapitule, sous la forme d'un tableau, la distribution des résultats de classification en fonction de la cible des individus. A partir de la matrice de confusion il est possible de calculer plusieurs indices qui nous informent sur la qualité du modèle. Parmi ceux-ci, nous avons retenu l'exactitude (ou accuracy ACC), la précision et le rappel, respectivement calculés selon les équations 17, 18 et 19.

$$ACC = \frac{T \text{ richodesmium prédit correctement} + N \text{ on Trichodesmium prédit correctement}}{\text{Somme Trichodesmium} + \text{Somme Non Trichodesmium}} \quad (17)$$

$$\text{précision} = \frac{T \text{ richodesmium prédit correctement}}{\text{Somme Trichodesmium prédit}} \quad (18)$$

$$\text{rappel} = \frac{T \text{ richodesmium prédit correctement}}{\text{Somme Trichodesmium}} \quad (19)$$

Cependant le support visuel reste très important pour juger des performances des algorithmes et notamment de la cohérence spatiale des résultats. Par exemple, la présence de faux positifs aux abords des nuages est facilement repérable visuellement sur les résultats, mais difficile à quantifier statistiquement à moins d'annoter chaque pixel de la scène analysée. Les modèles sélectionnés sont donc appliqués aux scènes AUSTEST, SPOT et NCEST. Comme précisé en partie 6, la scène NCEST n'étant pas intégrée au jeu de données d'apprentissage, elle constitue un exemple indépendant pour évaluer la généralisation des performances de classification à de nouvelles scènes.

## 7.2 résultats

Le résultat de la sélection séquentielle des variables est présenté en figure 15. Avec trois variables uniquement, l'algorithme permet d'obtenir une ACC proche de 95%. Les performances ensuite augmentent faiblement pour atteindre un plateau à partir de 6 variables. Étant donné que les performances de classification sont déjà satisfaisantes, nous choisirons les 3 premières variables sélectionnées pour notre modèle : Rrs(412), Rrs(531) et Rrc(869). Pour le jeu de données étendu nous n'avons pas effectué cette pré-sélection et avons utilisé l'ensemble des 48 variables disponibles pour construire notre modèle.

Suite aux 10 expériences de 10-fold cross-validation, nous obtenons les matrices de confusion suivantes :

La table 7.2 correspondant au modèle « restreint ♦ montre une précision de 94,5% et un rappel de 93,6%. Le modèle « étendu ♦ permet d'atteindre une précision de 98% et un rappel de 99,9%. La classification SVM présente donc de très bonnes performances de reconnaissance sur les deux jeux de données. Une fois les modèles construits nous les avons appliqués à nos scènes de références. Pour le jeu d'apprentissage issu des données restreintes nous avons pu l'appliquer à l'ensemble des trois scènes dont on peut voir les résultats Figure 16, 17B et 17B. Si avec la scène AUSTEST on retrouve bien les motifs des

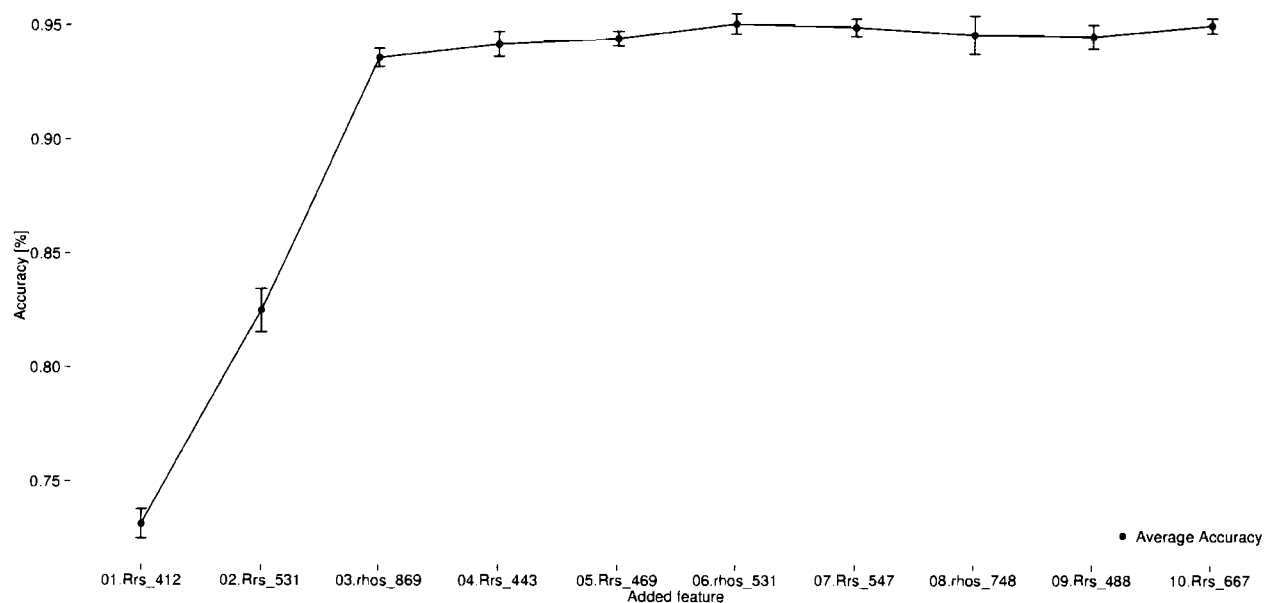


Figure 15 – Scène MODIS-Aqua A2007290.0355 montrant la distribution de l'indice FAI (Hu et al. 2010) mettant en évidence les mattes de *Trichodesmium* également repérées par McKinna et al. (2013).

		Réal	
		Tricho	Non Tricho
Prédit	Tricho	42127	2440
	Non Tricho	2873	42560

Table 2 – Matrice de confusion issu de l'apprentissage de notre modèle SVM avec les données restreintes

efflorescences de *Trichodesmium* visibles sur l'image vraie-couleur, on peut noter que toute une zone ne présentant pas d'efflorescence a été classé comme *Trichodesmium*. Pour la scène SPOT aucune détection des principaux motifs de *Trichodesmium* ne sont détectés tandis que des zones sans efflorescences sont classées *Trichodesmium*. La généralisation des performances à de nouvelles scènes n'est donc pas très bonne. Pour le jeu de données étendu, nous avons utilisé les scènes AUSTEST et SPOT lors de l'apprentissage. La généralisation à de nouvelle scènes est donc évaluée à partir des résultats obtenus sur la scène NCEST, Figure 17C. On peut noter que l'on retrouve les efflorescences de *Trichodesmium* détectées par notre algorithme bio-optique 17A.

		Réal	
		Tricho	Non Tricho
Prédit	Tricho	88269	1731
	Non Tricho	50	88269

Table 3 – Matrice de confusion issu de l'apprentissage de notre modèle SVM avec les données étendues



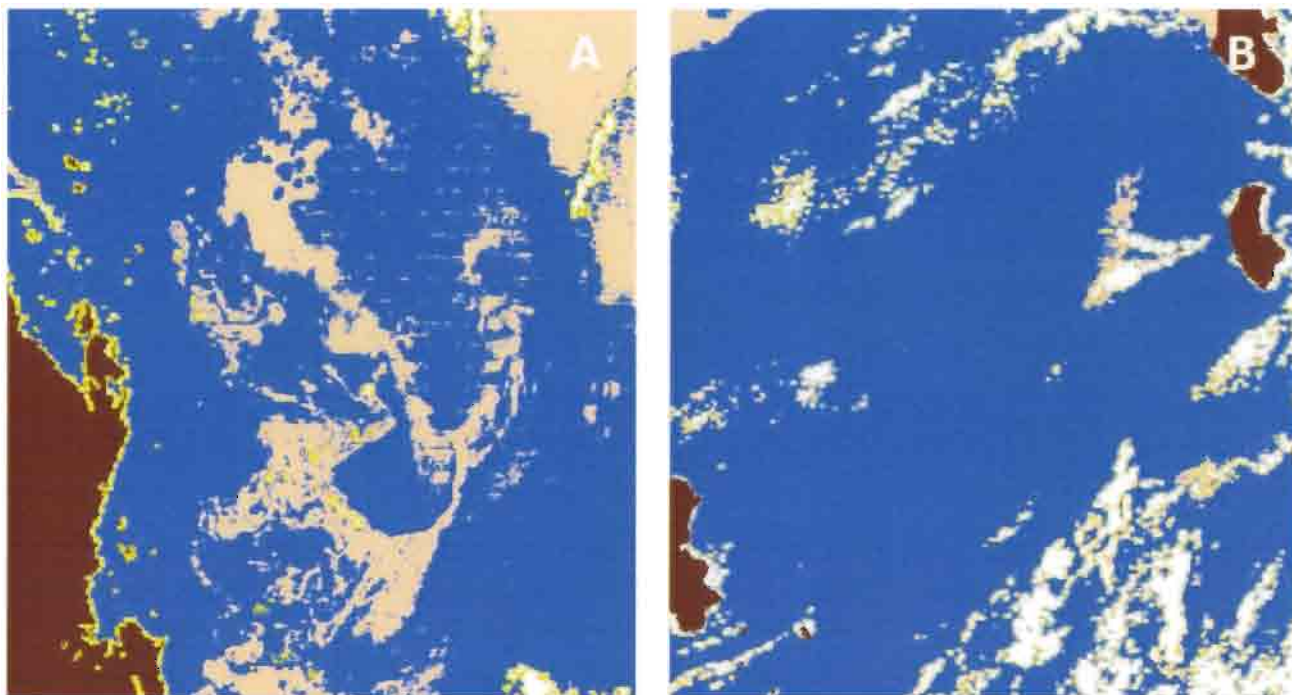


Figure 16 – Application de modèle SVM utilisant les données restreintes sur les scènes AUSTEST (A) et SPOT (B)

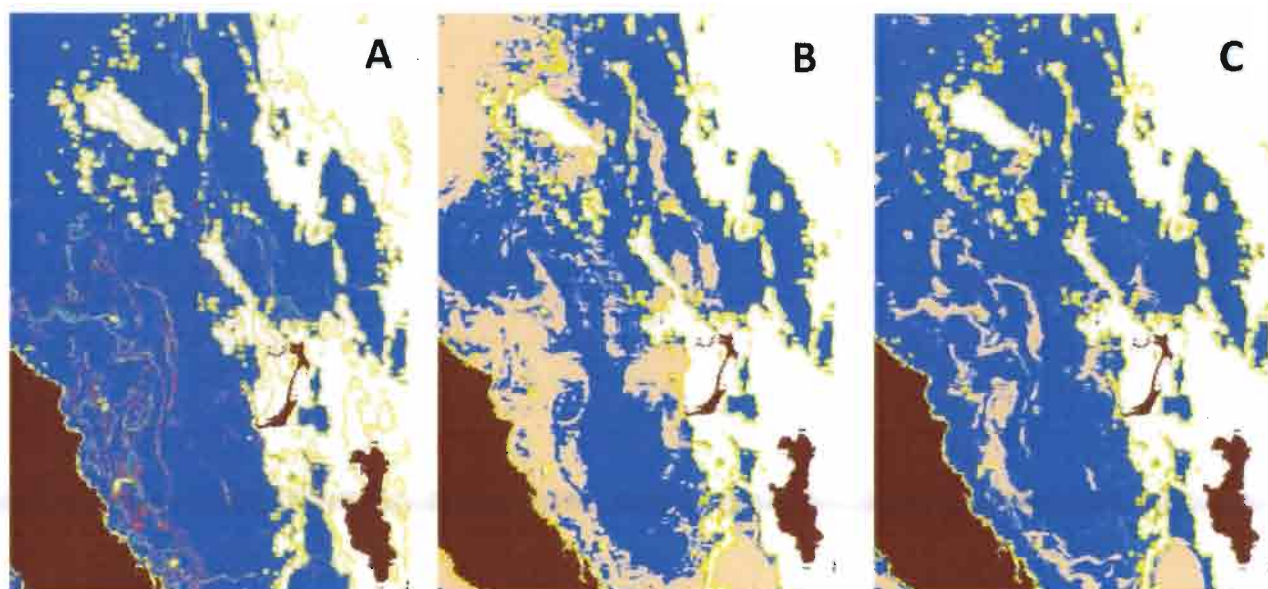


Figure 17 – Comparaison de notre algorithme bio-optique (A) avec les modèles SVM crée à partir du jeu de données restreint (B) et du jeu de données étendu (C) sur la scène NCEST

## 8 Discussion et conclusion

L'étude des algorithmes bio-optiques pour la détection du *Trichodesmium* issus de la littérature nous ont permis de mettre en évidence la mauvaise correction atmosphérique en présence de fortes concentration de *Trichodesmium* à la surface. Cette mauvaise correction est due à une erreur d'identification du signal infrarouge considéré comme provenant d'aérosols au lieu des algues. Elle impacte directement la mesure des réflectances satellites et indirectement tous les produits calculés à partir des réflectances, dont la concentration en chlorophylle souvent utilisée pour évaluer la production primaire. De nombreux

algorithmes de détection exploitant cette information ressortent alors des résultats biaisés. Les problèmes de détection liés aux fortes concentrations d'algues sont connus, et certains algorithmes comme celui de Hu, s'en affranchissent par d'autres méthodes.

Parmi les algorithmes répertoriés dans la littérature, les algorithmes de [Hu et al., 2010] et [McKinna et al., 2011] ont été implémentés au cours de ce stage. L'algorithme de McKinna implémenté selon [McKinna et al., 2011] ne semble pas fonctionner correctement et masque la majeure partie de la scène analysée. Cependant, en enlevant le masque sur les valeurs de réflectance négatives, et malgré les problèmes de correction atmosphérique, la détection des mattes de *Trichodesmium* est fortement améliorée. En revanche, les résultats montrent un certain nombre de fausses détections, notamment à proximité des nuages. L'algorithme de reconnaissance de Hu passe par l'évaluation d'un indice d'algue flottante, qui détecte bien les mattes de *Trichodesmium* visibles sur les images satellites. Cependant, cet indice n'est pas suffisant pour l'identification de l'espèce d'algue en présence. De plus, le FAI est également sensible à la présence de nuages. La seconde partie de l'algorithme de Hu qui doit permettre l'identification du *Trichodesmium* détecte correctement les mattes les plus denses et permet d'éviter la majorité des fausses détections aux abords des nuages. Cependant, dans cette partie de l'algorithme de Hu, la sensibilité de détection des efflorescences reste très dépendante de la zone de référence nécessaire à la normalisation des réflectances. Cet algorithme a donc été écarté pour la suite des expériences.

En prenant compte des erreurs de correction atmosphérique et des enseignements tirés des algorithmes de Hu et McKinna, j'ai créé mon propre algorithme bio-optique. Comparé à l'algorithme de McKinna, cet algorithme limite le nombre de faux positifs, même au plus près des nuages, mais ne semble détecter que les mattes de *Trichodesmium* les plus denses.

Lors de l'application de la comparaison des deux algorithmes (McKinna et l'algorithme que j'ai développé) à la scène SPOT 14, nous avons remarqué que seule une petite partie de l'efflorescence de *Trichodesmium* clairement visible sur l'image vraie-couleur était détectée. Pourtant des mattes de *Trichodesmium* avaient été repérées dans cette zone quelques jours auparavant lors de la mission en mer SPOT8. Plusieurs éléments nous laissent penser que les efflorescences de *Trichodesmium* n'étaient à ce moment plus en surface. Notamment, la correction atmosphérique standard, habituellement biaisée par la présence de *Trichodesmium*, ne présente pas d'anomalie particulière sur les efflorescences de cette scène. Il se pourrait donc que les efflorescences se trouvent sous la surface supposément suite à l'action du vent et/ou de la dynamique océanique.

Afin d'élargir la détection à ces efflorescences sous-marines nous développons un algorithme d'apprentissage supervisé basé sur les Machines à Supports Vectoriels (SVM).

Un des avantages des algorithmes bio-optiques est d'être théoriquement indépendant du capteur satellite tant que celui-ci comprend des bandes spectrales similaires. En pratique, les algorithmes de calibration des capteurs et de correction atmosphérique sont régulièrement mis à jour et changent les mesures de manière significative. Pour cette raison et afin d'élargir la détection à ces efflorescences sous-marines, nous avons souhaité évaluer le potentiel d'une méthode de classification supervisée basée sur des Machines à Support Vectoriel (SVM). L'apprentissage supervisé de cet algorithme a été réalisé sur deux bases de données satellites correspondant à des mesures in-situ coïncidentes. Le jeu de données satellites « restreint » a été construit par coïncidence géographique et temporelle exacte avec les mesures in-situ (500m et même jour), et correspond à une vérité terrain mesurée. Le jeu de données « étendu » a été construit par photo-interprétation des zones voisines des mesures (analyse à 1 deg et 7 jours du point de mesure). Les résultats de la classification SVM entraînée avec le jeu de données restreint sont excellents selon les chiffres, avec environ 95% de précision et de rappel. Cependant, lors de l'application des modèles à de nouvelles scènes, on retrouve certaines structures d'efflorescence observées sur les images vraie-couleur, mais globalement la classification n'est pas satisfaisante : des zones sans efflorescence apparente sont identifiées comme *Trichodesmium* tandis que les zones d'efflorescence ne sont pour la plupart pas détectées.

Nous expliquons ce phénomène par un problème de diversité des données d'apprentissage. En effet les prélèvements en Océanie s'effectuent très souvent dans les mêmes zones ce qui limite l'apprentissage à la diversité de situation rencontrée dans celle-ci. Le décalage temporel entre la mesure in-situ et la prise

de vue satellite ne permet pas d'assurer la présence d'efflorescence à l'endroit où elles ont été observées. Enfin, la couverture nuageuse limite le nombre de données coïncidentes.

C'est notamment pour ces raisons qu'une classification a été effectuée sur le jeu de données étendu basé sur la photo-interprétation des efflorescences. Ce jeu de données étendue se base sur le jugement expert sur l'image satellite pour détecter les mattes de *Trichodesmium* les plus évidentes, et identifier les différentes zones d'intérêts possédant des propriétés optiques originales (hauts fonds, barrière de corail, domaine côtier, etc.). Comme précédemment, les résultats selon les chiffres sont excellent une précision et un rappel de l'ordre de 99%, mais appliqués à une nouvelle scène les résultats ne semblent pas aussi bon que les algorithmes bio-optiques. La reconnaissance reste tout de même bien meilleure qu'avec l'apprentissage sur le jeu de données restreint. En alimentant le jeu d'apprentissage avec plus de cas de figure, atteindre de meilleurs résultats et de pouvoir au final détecter de façon systématique le *Trichodesmium* même immergé.

Dans l'ensemble, il reste donc difficile de pouvoir comparer directement ces algorithmes et déterminer le meilleur d'entre eux. La limite de ces algorithmes semble notamment provenir du trop faible nombre de données in-situ. Et à défaut d'une vérité terrain étendue et coïncidente stricto-sensu avec des données satellite, le développement, l'amélioration et la validation des algorithmes est avant tout basée sur l'interprétation visuelle des images. Toutefois, plusieurs pistes restent à explorer. Tout d'abord, il serait nécessaire d'approfondir la classification statistique qui n'a pu être abordée qu'à la fin du stage. Notamment, d'autres types d'algorithmes de classification tels que les réseaux de neurones, les forêts aléatoires ou les réseaux bayésiens pourraient aboutir à de meilleurs résultats. Au niveau du jeu de données, les mesures in-situ utilisées ont été faites majoritairement en haute mer. Les points originaux (côtiers, barrière de corail, etc.) sont marginaux dans la base de données. Une solution pour palier à ce problème pourrait être de pondérer l'apprentissage pour revaloriser les points originaux. Il serait également intéressant d'étendre les algorithmes de reconnaissance à d'autres satellites. Si les algorithmes bio-optiques sont dépendant des satellites ou plus précisément des canaux disponibles, les méthodes d'apprentissage supervisé peuvent être implémentées sur n'importe lequel d'entre eux. Cette couverture satellite nouvelle nous permettrait d'obtenir une meilleure couverture géographique et temporelle des zones d'études, la possibilité d'inter-calibrer les données des différents capteurs et ainsi pouvoir exploiter le maximum de point de mesures in-situ. Une autre aspect que nous n'avons pas eu le temps de traiter durant ce stage est l'exploitation du contexte spatial et temporel autour du *Trichodesmium*. En effet, la détection des structures filamenteuses particulières du *Trichodesmium* devraient permettre d'améliorer la reconnaissance. La relation spatiale et temporelle à d'autres types de phytoplancton détectés par satellite apporterait également des informations sur la probabilité de formation de mattes.

## Références

- [Berthelot et al., 2015] Berthelot, H., Moutin, T., L'Helguen, S., Leblanc, K., Hélias, S., Grosso, O., Leblond, N., Charrière, B., and Bonnet, S. (2015). Dinitrogen fixation and dissolved organic nitrogen fueled primary production and particulate export during the VAHINE mesocosms experiment (New Caledonia lagoon). *Biogeosciences Discussions*, 12(5) :4273–4313.
- [Biegala and Raimbault, 2008] Biegala, I. C. and Raimbault, P. (2008). High abundance of diazotrophic picocyanobacteria ( $>3 \mu\text{m}$ ) in a Southwest Pacific coral lagoon. *Aquatic Microbial Ecology*, 51(1) :45–53.
- [Bonnet et al., 2009] Bonnet, S., Biegala, I. C., Dutrieux, P., Slemmons, L. O., and Capone, D. G. (2009). Nitrogen fixation in the western equatorial Pacific : Rates, diazotrophic cyanobacterial size class distribution, and biogeochemical significance. *Global Biogeochemical Cycles*, 23(3) :1–13.
- [Bonnet et al., 2015] Bonnet, S., Rodier, M., Turk-Kubo, K., Germaineaud, C., Menkes, C., Ganachaud, A., Cravatte, S., Raimbault, P., Campbell, E., Quéroué, F., Sarthou, G., Desnues, A., Maes, C., and Eldin, G. (2015). Contrasted geographical distribution of  $\text{N}_2$  fixation rates and nifH phylotypes in the Coral and Solomon Seas (South-Western Pacific) during austral winter conditions. *Global biogeochemical cycles*.
- [Capone, 1997] Capone, D. G. (1997). Trichodesmium, a Globally Significant Marine Cyanobacterium.
- [Carpenter et al., 2004] Carpenter, E. J., Subramaniam, A., and Capone, D. G. (2004). Biomass and primary productivity of the cyanobacterium Trichodesmium spp. in the tropical N Atlantic ocean. *Deep-Sea Research Part I : Oceanographic Research Papers*, 51(2) :173–203.
- [Chang and Lin, 2011] Chang, C.-C. and Lin, C.-J. (2011). LIBSVM.
- [de Boissieu et al., 2014] de Boissieu, F., Menkes, C., Dupouy, C., Rodier, M., Bonnet, S., Mangeas, M., and Frouin, R. J. (2014). Phytoplankton global mapping from space with a support vector machine algorithm. 9261 :92611R.
- [Dupouy et al., 2011] Dupouy, C., Benielli-Gary, D., Neveux, J., Dandonneau, Y., and Westberry, T. K. (2011). An algorithm for detecting Trichodesmium surface blooms in the South Western Tropical Pacific. *Biogeosciences*, 8(12) :3631–3647.
- [Dupouy et al., 2008] Dupouy, C., Minghelli-Roman, A., Despinoy, M., Röttgers, R., Neveux, J., Pinazo, C., and Petit, M. a. (2008). MODIS/Aqua chlorophyll monitoring of the New Caledonia lagoon during the 2008 La Nina event. *Proceedings of SPIE*, 7150 :715014–715014–8.
- [Dupouy et al., 1988] Dupouy, C., Petit, M., and Dandonneau, Y. (1988). Satellite detected cyanobacteria bloom in the southwestern tropical Pacific. *Remote Sensing*, 9 :389–396.
- [Finzi, 2006] Finzi, J. A. (2006). Photophysiological Parameters for  $\text{CO}_2$  and  $\text{N}_2$  of Trichodesmium spp. in Natural Populations and in Culture Nutrient Limitation Experiments.
- [Froidefond and Doxaran, 2004] Froidefond, J. M. and Doxaran, D. (2004). Télédétection optique appliquée à l'étude des eaux côtières. *Télédétection*, 4(2) :579–597.
- [Garcia et al., 2007] Garcia, N., Raimbault, P., and Sandroni, V. (2007). Seasonal nitrogen fixation and primary production in the Southwest Pacific : Nanoplankton diazotrophy and transfer of nitrogen to picoplankton organisms. *Marine Ecology Progress Series*, 343(Paerl 1990) :25–33.
- [Gower et al., 2014] Gower, J., King, S., and Young, E. (2014). Global remote sensing of Trichodesmium. *International Journal of Remote Sensing*, 35(14) :5459–5466.
- [Gruber and Sarmiento, 1997] Gruber, N. and Sarmiento, J. L. (1997). Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochemical Cycles*, 11(2) :235.
- [Hu, 2009] Hu, C. (2009). A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, 113(10) :2118–2129.
- [Hu et al., 2010] Hu, C., Cannizzaro, J., Carder, K. L., Muller-Karger, F. E., and Hardy, R. (2010). Remote detection of Trichodesmium blooms in optically complex coastal waters : Examples with MODIS full-spectral data. *Remote Sensing of Environment*, 114(9) :2048–2058.

- [Hu et al., 2012] Hu, C., Lee, Z., and Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans : A novel approach based on three-band reflectance difference. *Journal of Geophysical Research : Oceans*, 117(1) :1–25.
- [Karl et al., 2002] Karl, D., Michaels, a., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D., and Stal, L. (2002). Dinitrogen fixation in the world's oceans. *Biogeochemistry*, 57-58 :47–98.
- [Kirk, 1994] Kirk, J. T. O. (1994). *Light and Photosynthesis in Aquatic Ecosystems*. publisherName-Cambridge University Press.
- [Kuchler and Jupp, 1988] Kuchler, D. A. and Jupp, D. L. B. (1988). Shuttle photograph captures massive phytoplankton bloom in the Great Barrier. *Remote sensing*.
- [Luo et al., 2012] Luo, Y.-W., Doney, S. C., Anderson, L. a., Benavides, M., Berman-Frank, I., Bode, a., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, a., Foster, R. a., Furuya, K., Gómez, F., Gundersen, K., Hynes, a. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy, D. J., Moisander, P. H., Moore, C. M., Mouriño Carballido, B., Mulholland, M. R., Needoba, J. a., Orcutt, K. M., Poulton, a. J., Rahav, E., Raimbault, P., Rees, a. P., Riemann, L., Shiozaki, T., Subramaniam, a., Tyrrell, T., Turk-Kubo, K. a., Varela, M., Villareal, T. a., Webb, E. a., White, a. E., Wu, J., and Zehr, J. P. (2012). Database of diazotrophs in global ocean : abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, 4(1) :47–73.
- [McKinna et al., 2011] McKinna, L., Furnas, M., and Ridd, P. (2011). A simple, binary classification algorithm for the detection of *Trichodesmium* spp. within the Great Barrier Reef using MODIS imagery. *Limnology and Oceanography : Methods*, 9 :50–66.
- [Mckinna, 2015] Mckinna, L. I. W. (2015). Three decades of ocean-color remote-sensing *Trichodesmium* spp. in the World's oceans : A review. *Progress in Oceanography*, 131 :177–199.
- [Moisander et al., 2010] Moisander, P. H., Beinart, R. a., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. a., Montoya, J. P., and Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic N<sub>2</sub> fixation domain. *Science (New York, N.Y.)*, 327(5972) :1512–1514.
- [Morel and Gentili, 1996] Morel, a. and Gentili, B. (1996). Diffuse reflectance of oceanic waters. III. Implication of bidirectionality for the remote-sensing problem. *Applied optics*, 35(24) :4850–4862.
- [Morel and Maritorena, 2001] Morel, A. and Maritorena, S. (2001). Bio-optical properties of oceanic waters : A reappraisal. *Journal of Geophysical Research*, 106(C4) :7163.
- [Mountrakis et al., 2011] Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3) :247–259.
- [Neveux et al., 2006] Neveux, J., Tenorio, M. M. B., Dupouy, C., and Villareal, T. A. (2006). Spectral diversity of phycoerythrins and diazotroph abundance in tropical waters. 51(4) :1689–1698.
- [Paerl, 1999] Paerl, H. W. (1999). Physical-chemical constraints on cyanobacterial growth in the Oceans. *Bulletin de l'Institut océanographique*, pages 621–622.
- [Pope and Fry, 1997] Pope, R. M. and Fry, E. S. (1997). Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements. *Applied optics*, 36(33) :8710–8723.
- [Prasad and Krishna, 2011] Prasad, S. V. S. and Krishna, I. V. M. (2011). Classification of Multispectral Satellite Images using Clustering with SVM Classifier. *International Journal*, 35(5) :32–44.
- [Shiozaki et al., 2014] Shiozaki, T., Kodama, T., and Furuya, K. (2014). Large scale impact of the island mass effect through nitrogen fixation in the western South Pacific Ocean. *Geophysical Research . . .*, pages 1–7.
- [Smits et al., 1992] Smits, J., Breedveld, L., Derksen, M., Kateman, G., Balfoort, H., Snoek, J., and Hofstraat, J. (1992). Pattern classification with artificial neural networks : classification of algae, based upon flow cytometer data. *Analytica Chimica Acta*, 258(1) :11–25.

- [Stramski et al., 2001] Stramski, D., Bricaud, a., and Morel, a. (2001). Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Applied optics*, 40(18) :2929–2945.
- [Stramski and Mobley, 1997] Stramski, D. and Mobley, C. D. (1997). Effects of microbial particles on oceanic optics : A database of single-particle optical properties. *Limnology and Oceanography*, 42(3) :538–549.
- [Subramaniam et al., 2002] Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J., and Capone, D. G. (2002). Detecting Trichodesmium blooms in SeaWiFS imagery. *Deep-Sea Research Part II : Topical Studies in Oceanography*, 49(1-3) :107–121.
- [Subramaniam et al., 1999a] Subramaniam, A., Carpenter, E. J., and Falkowski, P. G. (1999a). Bio-optical properties of the marine diazotrophic cyanobacteria Trichodesmium spp. I. Absorption and photosynthetic action spectra.
- [Subramaniam et al., 1999b] Subramaniam, A., Carpenter, E. J., and Falkowski, P. G. (1999b). Bio-optical properties of the marine diazotrophic cyanobacteria Trichodesmium spp. II. A reflectance model for remote-sensing.
- [Tenório, 2006] Tenório, M. M. B. (2006). Les cyanobactéries pélagiques en milieu tropical oligotrophe : occurrence, distribution et dynamique.
- [Westberry et al., 2005] Westberry, T. K., Siegel, D. A., and Subramaniam, A. (2005). An improved bio-optical model for the remote sensing of Trichodesmium spp. blooms. *Journal of Geophysical Research : Oceans*, 110(6) :1–11.