1) Rev.: My major criticism is the apparent lack of a clear detection of the expected spectral signature of Trichodesmium slicks, which show a strong red edge, as described in section 3.3. A strong red edge is shown in Gower et al., 2014, Figure 6 derived from MERIS, and this should also appear in MODIS data. Spectra in Figure 4 show little sign of a red edge.

Resp: The values shown Figure 4 is the average of 600 differents values taken from different Trichodesmium mat density. The majority of them present a weak response in the red-edge, but are nevertheless Trichodesmium algae agglomeration. The densest mats have a stronger reflectance in the NIR. We can see these difference with the in situ spectrum taken by McKinna et al. (2015), Figure 3.

McKinna, L., Furnas, M., and Ridd, P.: 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, https://doi.org/10.4319/lom.2011.9.50, 2011.

2) Rev.: The paper needs work. English is poor. There are several typos.

Resp: A read-through of the article have been carried out to limit typos.

3) Rev.: Line 27 of page 7 mentions the need to distinguish Sargassum in this area. How realistic is this? Has Sargassum ever been observed there in quantities large enough to cause confusion?

Resp: To today, no important Sargassum bloom have been observed in the Southwestern Pacific (Payri and Richer de Forges, 2000). If the second step of the algorithm of Hu et al. (2010) resolve the ambiguity between Trichodesmium and Sargassum, this distinction have not been studied with the new algorithm.

Payri C. and B. Richer de Forges, 2000. Compendium of marine species from New Caledonia. ISSN 1297-9635, Second edition, N°117, ORSTOM editions, IRD Center of Noumea, 480 pages.

Hu, C., Cannizzaro, J., Carder, K. L., Muller-Karger, F. E. and Hardy, R.: Remote detection of Trichodesmium blooms in optically complex coastal waters: Examples with MODIS full-spectral data, Remote Sens. Environ., 114(9), 2048–2058, 2010.

4) Rev.: A possible daily cycle in Trichodesmium observability is mentioned on page 9. This could be important and needs more discussion. How big a variation has been observed? How consistently?

According to Villareal and Carpenter, *Trichodesmium* colonies migrate vertically in the water column. To our knowledge, no study of the abundance variation of *Trichodesmium* between these cycles have been carried out. Page 9, the main point was around the vertical migration of *Trichodesmium* because of natural or environmental causes that may affect the detection of the algorithm.

Villareal, T. A. and Carpenter, E. J.: Buoyancy Regulation and the Potential for Vertical Migration in the Oceanic Cyanobacterium Trichodesmium, Microbial Ecology, 45, 1–10, http://www.jstor.org/stable/4287673, 2003.

5) Rev.: The MCI time sequence in Gower et al., 2014 shows monthly composite images over this area for the period 2002 to 2012 (Figure2 in that paper). Do any of the high signals shown there correspond to in-situ data considered here?

Resp: For the area considered in Gower et al. and the period 2002-2012, only 80 observations remain. 45% of these observations are coincident with the monthly composite images of Gower et al., 2014. The *Trichodesmium* blooms during the period Jan-Feb 2003, Feb 2004, Mar-Apr and Dec 2010 are especially interesting and match with 27 of our observation. For the observations not detected, 30 were observed in April and July 2003.

6) Rev.: Line 5. The new algorithm discussed in section 3.5 uses negative Rrs(678) to measure a red-edge signal. Using the absolute value of Rs(678) will result in errors if Rs(678) is ever positive. The negative value alone should be used.

Resp: The reviewer is correct, actually the detection does not use the magnitude of Rrs(678). The absolute value is only used when negative. The sentence has been reformulated to remove the confusion.

Minor points:

Page 1.

7) Rev.: Line 1 "is the major dinitrogen-fixing organisms." Should be "organism." Do we need the "di"? Are there "mono-nitrogen" fixers?

Resp: "dinitrogen-fixing" have been changed into "nitrogen-fixing"

8) Rev.: Line 15, "fails to detect sub-surface booms." This should be "blooms." But why criticise the algorithm for not being able to do something which must be fundamentally impossible? There can be no signal from (deeper) sub-surface blooms.

Resp: This sentence states the limit of the be detection for reader which would not be familiar with this kind of remote sensing algorithm. The sentence has been deleted nevertheless.

Page 3

9) Rev.: Lines 11,12, "blooming period (Nov to Mar) coincides with the South Pacific Convergence Zone." A period has to coincide with another period, or Zone with another Zone. I'm not sure what point you are making here.

Resp: The South Pacific Convergence Zone (SPCZ) is an area of convergence with an important cloud cover. The SPCZ does not have a specific location and move between seasons. Here, during the period from November to Mars, the SPCZ intersect our zone of study, the WTSP. Formulation corrected according to comment.

10) Rev.: Line 32, "three datasets intersecting." Maybe you mean "three datasets providing data in the acquisition period"

Resp: The reviewer is correct. Formulation corrected according to comment.

Page 4

11) Rev.: Lines 17,18 "SOB database and the OUTPACE campaign are used. So REMMOA is not used? Why not? You discuss it above and call it "most favorable for satellite data validation" (line 6).

Resp: As discussed before, in section 2.1, the REMMOA dataset is included in the SOB database . As it did not seem clear that SOB integrated REMMOA, the paragraph was reformulated to make it clearer.

Page 10

12) Rev.: Lines 15, 16. Where in Figure 7 is the area you discuss?

Resp: The area is just around the stations near the longitude 180° at the South of Fidji of the OUTPACE campaign. This information has been added to the text.

Page 11

13) Rev.: Line 5. "Two adjacent wavelengths 645 and 748." The two adjacent wavelengths are 665 and 748. Why use 645 in Equation 4?

Resp: Indeed the wording was not correct, as 645 is not adjacent to 678. The choice of 645 instead of the adjacent wavelength allows to estimate the trough at 678 nm in the spectrum (Figure 4A) which would not be as clear when using the adjacent wavelength at 665 nm. The wording have been corrected.

14) Rev.: What is the criterion in equation 4? Must this expression be positive? Or must it be greater than some threshold value?

Resp: You're right, there was a missing criteria in the version you had for equation 4. It has been corrected in the new version (Equation 4 must be greater than 0).

15) Rev.: Line 15,16 Which "former one"?

Resp: The algorithm proposed section 3.5. The text have been changed accordingly

Page 12

16) Rev.: Line 6. What is sub-mesoscale? By one convention mesoscale is about 300km, but you must mean a much smaller value than this.

Resp: Lévy et al. 2012 name "submesoscales" scales between 1 and 10 km at most and "mesoscales" is used for scales of the order of the first baroclinic radius of deformation. Typically at 20° S these scales are ~50 km (e.g. http://www-po.coas.oregonstate.edu/research/po/research/rossby_radius/fig2.html). We wanted to emphasize that visual observations at sea typically show sub mesoscales structures that cannot be fully described by > 12.5km FSLEs as used by Rousselet et al (2018).

Lévy, M., R. Ferrari, P. J. S. Franks, A. P. Martin, and P. Rivière (2012), Bringing physics to life at the submesoscale, *Geophys. Res. Lett.*, 39, L14602, doi: 10.1029/2012GL052756.

Page 13

17) Rev.: Line 21. "citetMcKinna2011" should be "cited McKinna (2011)"

Resp: corrected

Page 14

18) Rev.: Line 7 "floating algae organized in a heap" is a strange description. Perhaps you mean "floating algae in mats of varying surface concentration"?

Resp: The description has been changed, we meant "organized in dense mats".

Figure 2

19) Rev.: Panels B and C captions are interchanged

Resp: To our understanding the captions between panel B and C are not interchanged. Panel B is the spectra of Trichodesmium mats at 1 km resolution with R_{rc} reflectances. Panel C is Trichodesmium mats at 250 m resolution with R_{rs} reflectances.

Figure 4

20) Rev.: Why are points not plotted at 748 and 870nm in panels A, C and E?

Resp: Because of the atmospheric correction used in SeaDas, these two wavelength are set to 0. Details Section 3.3.

Figure 5

21) Rev.: The geographic area needs to be better described. I assume this is the Great Barrier Reef, with land at bottom left and cloud in grey? It would be useful to outline the area of the reef itself.

Resp: The label of the Great Barrier Reef have been added Figure 2 and Figure 5, and the caption of the figure have been completed to describe the grey pixel as clouds.

Figure 6

22) Rev.: The grey areas appear to be clouds. This needs to be made clear.

Resp: The caption of the figure have been completed to describe the grey pixel as clouds.

Figure 7

23) Rev.: The Figure needs better description. Presumably, A is the top panel, B the bottom one? What are the units of the two color bars? At present the left color bar is referred to as fixation rate, the central bar is undefined. Text describing the colored dots in the lower panel is ambiguous: they are said to be blue for absence, red for presence, but the last sentence says they are only colored for presence.

Resp: The figures have been labelled, the central bar have been defined, and the caption have been made clearer.

Figure 8

24) Rev.: "Only a few pixels are remaining." "Only a few pixels remain" is better English. In fact, I would say that a significant number remain, or "The main feature is well-detected at the lower resolution."

Resp: The sentence have been corrected

Abstract. *Trichodesmium* is the major ~~organism~~~~dinitrogen-fixing species~~ nitrogen-fixing species in the Western Tropical South Pacific (WTSP) region, a hotspot for diazotrophy. Due to the paucity of in situ observations, remote-sensing methods for detecting *Trichodesmium* presence on a large scale have been investigated to assess the regional-to-global impact of this~~these~~ organism~~s~~ on primary production and carbon cycling. A number of algorithms have been developed to identify *Trichodesmium* surface blooms from space, but determining with confidence their accuracy has been difficult, chiefly because of the scarcity of sea-truth information at time of satellite overpass. Here, we use a series of new cruises as well as airborne surveys over the WTSP to evaluate their ability to detect~~accuracy in detecting~~ *Trichodesmium* surface blooms in the satellite imagery. The evaluation, performed on MODIS data at 250 m and 1 km resolution acquired over the region shows limitations due to spatial resolution, clouds, and atmospheric correction. A new satellite-based algorithm is designed to alleviate some of these limitations, by exploiting optimally spectral features in the atmospherically corrected reflectance at 531, 645, 678, 748, and 869 nm. This algorithm outperforms former ones near clouds, limiting false positive detection and allowing regional scale automation. ~~Compared with observations, 80 % of the detected mats are within a 2 km range, demonstrating the good statistical skill of the new algorithm.~~ Application to MODIS imagery acquired during the February-March 2015 OUTPACE campaign reveals the presence of surface blooms Northwest and East of New Caledonia and near 20°S-172°W in qualitative agreement with measured nitrogen fixation rates. ~~The new algorithm, however, fails to detect sub-surface blooms evidenced in trichome quantification by molecular methods.~~ Improving *Trichodesmium* detection requires measuring ocean color at higher spectral and spatial (< 250 m) resolution than MODIS, taking into account environment properties (e.g., wind, sea surface temperature), fluorescence, and spatial structure of filaments, and a better understanding of *Trichodesmium* dynamics, including aggregation processes to generate surface mats. Such sub-mesoscale~~s~~ aggregation processes for *Trichodesmium* are yet to be understood.

1 Introduction

The Western Tropical South Pacific (WTSP) is a Low Nutrient Low Chlorophyll (LNLC) region, harboring surface nitrate concentrations close to detection limits of standard analytical methods, and limiting for the growth of the majority of phytoplankton species (Le Borgne et al., 2011). This lack of inorganic nitrogen favors the growth of dinitrogen (N2)-fixing organisms (or diazotrophs), which have the ability to use the inexhaustible pool of N2 dissolved in seawater and convert it into bioavailable ammonia. Several studies have reported high N2 fixation rates in the WTSP (Berthelot et al., 2017; Bonnet et al., 2009, 2015; Garcia et al., 2007), which has recently been identified as a hot spot of N2 fixation (Bonnet et al., 2017). During austral summer conditions, N2 fixation supports nearly all new primary production and organic matter export (Caffin et al.,2018; Knapp et al., 2018) as nitrate diffusion across the thermocline and atmospheric sources of N are < 10 % of new N inputs. The cyanobacterium *Trichodesmium* is one of the most abundant diazotrophs in our oceans (Capone, 1997; Luo et al., 2012) and in the WTSP in particular (Tenorio et al., 2018; Stenegren et al., 2018). C~~Based on c~~ell-specific N2 fixation measurements recently conducted in the WTSP have revealed~~,~~ *Trichodesmium* ~~has recently been identified, based on cell-specific N2 fixation measurement,~~ as the major N2-fixing organism, accounting for > 60 % of total N2 fixation (Bonnet et al., 2018). One of the characteristics of *Trichodesmium* is the presence of gas vesicles, which provide buoyancy (van Baalen and Brown, 1969; Villareal and Carpenter, 2003) and help maintain this cyanobacterium in the upper ocean surface. *Trichodesmium* cells are aggregated and form long chains called trichomes. Trichomes then can gather into colonies called "puffs" or "tuffs," and these colonies can aggregate at the surface of the water and form large mats that can extend for miles. They had already been observed during James Cook and Charles Darwin's expeditions. During the southern austral summer, *Trichodesmium* blooms have long been detected by satellite in the region, mostly around New Caledonia and Vanuatu (Dupouy et al., 2000, 2011), and later confirmed by microscopic enumerations (Shiozaki et al., 2014).

Identifying the occurrence and the spatial distribution of *Trichodesmium* blooms and mats is of primary importance to assess their regional contribution to primary production and biogeochemical cycles regionally. However because of their paucity, scientific cruises alone are not sufficient to achieve such goal, and remote sensing completed by sea observations of mats appear as the unique alternative for assessing its global impact. By using *Trichodesmium* spectral characteristics, among which pigment absorption due to phycoerythrin (PE) between 490 and 570 nm, absorption/scattering increase in the red-Near-Infrared (NIR), and particle backscattering (Subramaniam et al., 1999a, b; Hu et al., 2010), several bio-optical algorithms have been developed to detect *Trichodesmium* blooms in real time from various satellite sensors. They are extensively documented in Blondeau-Patissier et al. (2014) and Mckinna (2015). Former algorithms are classification schemes using thresholds applied to reflectances in the blue-green (440-550) range, like those proposed by Subramaniam et al. (2002) and Dupouy et al. (2011). They were designedderived using Sea-viewing Wide Field-of-view Sensor (SeaWiFS) observations. Using spectral features in Moderate Resolution Imaging Spectroradiometer (MODIS)--Aqua observations at 250 m in the red-NIR, McKinna et al. (2011) elaborated a simple reflectance classification algorithm to detect dense *Trichodesmium* surface aggregation. By providing an operational Floating Algae Index (FAI), Hu et al. (2010) also demonstrated the potential for using the red-edge effect, that is the increasing absorption in the red (620-700) and scattering in the NIR (beyond 700) region of the spectrum due to floating algae. Using Medium Resolution Imaging Spectrometer (MERIS) observations in the red-NIR band, Gower et al. (2014) provided a similar index of *Trichodesmium* surface slicks. In their *Trichodesmium* bio-optical model, Westberry et al. (2005) used specific inherent optical properties (*Trichodesmium* specific absorption and backscattering parameters) to estimate *Trichodesmium* biomass from SeaWiFS reflectances.

TBut the application of these algorithms to MODIS imagery revealed several issues, some of which were raised and discussed in the aforementioned articles. For example, the red-edge over *Trichodesmium* mats can lead to controversial results, since atmospheric correction for aerosols isare based on information at similar wavelengths (Hu et al., 2010). Sampling effects are also exacerbated due to the occurrence of clouds in the WTSP, since the blooming period of *Trichodesmium* (mainly November to March (Dupouy et al., 2011)) coincides with the cloudiest period of South Pacific Convergence Zone, i.e. heavy cloudiness making difficult the identification of coincident in-situ mats in satellite imagery. Moreover, bBecause *Trichodesmium* mats are narrow (~ 50 m typically), it raises the question of the suitability of MODIS visible-NIR spatial resolution to detect such surface aggregations is questionable.

The aim of this study is to provide a systematic detection of *Trichodesmium* blooms in the vast WTSP Ocean between latitudes 26° and 10°S and longitudes between 155° and 190° E, building on previously published algorithms and using marine reflectances measured by MODIS onboard/ AquaQUA. To evaluate the detection performance, a large database of historical mat observations in this region was created and , updated with recent datasets and particularly during the March-April 2015 “Oligotrophy from Ultra-oligoTrophy PACific Experiment” (OUTPACE) cruise ofin March-April 2015 (Moutin et al., 2017). As a consequence of improvements in MODIS Collection 6 calibrations and algorithm updates for aerosol and cloud screening (Casey et al., 2017), *Trichodesmium* detection algorithms developed with previous collections had to be adapted whenever possible. From this experience, a new algorithm less prone to contamination by clouds emerged, combining methods to detect *Trichodesmium* blooms from algorithms by McKinna et al. (2011) and Hu et al. (2010) and was evaluated using high resolution MODIS imagery. The paper is organized as follows.: In Section 2 in-situ and satellite data used in this study are presented. In Section 3 methods to extract *Trichodesmium* spectral signature and their limitations are described, and details are provided about the former detection algorithms of Hu et al. (2010) and McKinna et al. (2011), adapted for this study, as well as the newly developed algorithm. In Section 4 presents these algorithms are compared, and the proposed algorithm is evaluated along the OUTPACE cruise transect. In Section 5 the new algorithm performances is discussed. In Section 6, finally, the conclusions of the study are drawn and perspectives are provided for future work.

2 Material

2.1 In situ observations

The in-situ databasebase used to train and test the *Trichodesmium* detection algorithmmodel is a combination of twothree datasets, providing data in the acquisition period of intersecting MODIS acquisition period of Aqua or Terra missions (March 2000-present). The first dataset includes It includes the *Trichodesmium* mat observations published in Dupouy et al. (2011). These observations were madedone between 1998 and 2010, from aircraft, French Navy ships, research vessels (e.g. R/V Alis) and ships of opportunity. Some of these visual observations were confirmed by water samples analyzed with photomicrographs confirming the abundant presence of abundant *Trichodesmium* (Dupouy et al., 2011). The second dataset includes aAirborne visual observations were also gathered in December 2014 in the vicinity of New Caledonia during the "REcensement des Mammifères marins et autre Mégafaune pélagique par Observation Aérienne" (REMMOA) program (Laran et al., 2016). This second dataset provides a large number of *Trichodesmium* mat observations along numerous and repetitive transects, which is most favorable for satellite data validation. In total, the database created from the compilation of these two datasets open ocean observations contains 507 open ocean observations in the region 15°S-25°S, 155°E-180°E (Figure 1). It is referred to as the Simple Observation Base *Trichodesmium*(SOBT) in the following and used mainly for algorithm training.

A third dataset In addition to SOB, was obtained along a latitudinal transect around 20°S was carried out during the OUTPACE scientific cruise (Moutin et al., 2017) incovering the region 160°E-160°W from 23 February 23 th to 01 April 2015. Seawater samples were collected for *Trichodesmium* identification and estimation by quantitative PCR, a molecular method described in Stenegren et al. (2018), microscopic counts at selected stations (Caffin, personal communication, 2017), as well as N2 fixation rates as described in Bonnet et al. (2018). Moreover, *Trichodesmium* abundance from the Underwater Vision Profiler 5 (UVP5; (Picheral et al., 2010)), calibrated on trichome concentration from pigment algorithms and on visual counts of surface samples at every stations, allowed one to describe the *Trichodesmium* distribution along the transect (Dupouy et al., 2018).

2.2 Satellite data

Marine reflectances from MODIS_-Aqua and -Terra Collection 6 coinciding in time and space to the SOB database and the OUTPACE campaign are used in this study. Level-1A observations were downloaded from NASA's Goddard Space Flight Center (http://oceandata.sci.gsfc.nasa.gov) and processed with SeaDAS v7.0.2 to produce Level-2 reflectances at 250 m and 1 km resolution. SeaDAS v7.0.2 default values were applied during this processing for atmospheric correction failure, land, sun glint, very high or saturated radiance, sensor view zenith angle exceeding threshold, stray light contamination, and cloud contamination. Satellite observations with a proportion of valid pixels lower than 40 % within the 0.5° searching radius around each in-situ observation were screened out.

MODIS atmospherically corrected (aerosol + Rayleigh) remote-sensing reflectances (R_{rs}) in the visible, NIR, and short wavelength infrared (SWIR) were used at different resolutions: 250 m resolution for bands 1 (645 nm) and 2 (859 nm), 500 m resolution (bands 3-7, visible and SWIR land/clouds dedicated bands) and 1 km resolution (bands 8-16). Bands 8 to 16 are dedicated to ocean color (Table 1), but the information in high-resolution bands located in the visible-NIR region is also used to track floating blooms. To evaluate the influence of resolution on detection performances, Level-2 remote sensing data was produced at both 250 m and 1 km resolutions, with interpolation of 500 m and 1 km data and aggregation of 250 m and 500 m resolution data, respectively. The consequences of these processing are discussed in Section 5.

The aerosol correction was carried out with the standard Gordon and Wang (1994) method since the study was conducted in the open ocean (Case-1 waters). For reasons detailed in next section, this type of correction is not appropriate in the presence of strong floating algae concentrations. Therefore, only Rayleigh-corrected only reflectances (R_{rc}) were computed in addition to R_{rs} in Level-2 outputs.

3 Methods

3.1 Motivation in using Rayleigh-corrected reflectance (R_{rc}) in bio-optic algorithms for floating blooms

The atmospheric correction schemeaerosol model of Gordon and Wang (1994) utilizes a pair of NIR bands at 748 and 869 nm. Spectral differences in the two NIR bands are used to select the most probable aerosol model, allowing an estimation of the aerosol effects in the visible. The aerosol contamination is then removed from the TOA signal after Rayleigh scattering correction to yield water-leaving radiance, therefore R_{rs}, in the visibleand water-leaving radiances are corrected from aerosol contamination to give Rrs. But to retrieve the aerosol model over Case-1 waters, the Gordon and Wang (1994) method relies on the assumption that optical constituents have negligible contribution to water-leaving radiances in the NIR region. Over floating mats, this assumption does not hold due to red edge effects, and the atmospheric correction may give erroneous (i.e., too low) and even negative) values of R_{rs} as detailed in Hu et al. (2010). For this reason, R_{rc} was preferred in floating algal bloom concentration detection algorithms, as done with the Floating Algal Index (FAI) in Hu et al. (2010) or with the Maximum Chlorophyll Index (MCI) in Gower et al. (2008).

This issue is illustrated in Figure 2, which presents a MODIS-Aqua image of the Australian coast acquired just after a period of heavy rain that led to a massive *Trichodesmium* bloom. Fortunately, this bloom was also documented through field studies (McKinna et al., 2011). Figure 2A shows the “true-color” image obtained by combination of R_{rc} at 469, 555 and 645 nm. On this image, large visible *Trichodesmium* mats distributed over a vast area can be seen. Figure 2B displays the SeaDAS--derived aerosol optical thickness (AOT) at 555 nm, an indicator of the aerosol load in the atmosphere. The high AOT values match the filamentous spatial structure noticed in the “true color” image. However, this structure is more likely to be floating material as the patterns are very thin and do not seem to be driven by wind. Moreover, several patches within this structure are flagged as cloud in the center (grey patches on figure 2B), although the “true color” image does not indicate the presence of clouds in this particular area. The misclassification of marine to cloudy pixel is a failure of the cloud detection algorithm, resulting from high water-leaving signal in the NIR bands exceeding the albedo threshold of 2.7 % (Wang and Shi, 2005). Figure 2C shows the chlorophyll concentration (Chl_a) estimated according to the OC3 algorithm when Chl_achlorophyll retrievals are above 0.2 mg.m^{-3} and to the OCI algorithm when they are below (Hu et al., 2012). Chl_a decreases systematically, even falling to zero, in the vicinity of the *Trichodesmium* distribution patterns, although the real concentrations are certainly larger at the core of the mats than at their periphery. In addition to biomass underestimation, the spectral miscorrection may reduce the performance of reflectance classification schemes, like those based on simple reflectance thresholds or using reflectance ratios in the blue-green range. Prior to using such classification schemes, the spectral signature of *Trichodesmium* mats using MODIS R_{rs} and R_{rc} need to be further investigated (see below).

3.2 Extraction of the spectral signature of mats

With the persistent cloud cover of the region, the number of strict coincidences of in-situ observations and cloud free MODIS pixels with visible *Trichodesmium* mats is small. Therefore, the search for coincidences has been extended in space and time. To extract the *Trichodesmium* spectral signature corresponding to the OBT, 6 tiles have been specifically selected (Table 2) and are used later on to test the different bio-optical algorithms designed to detect the *Trichodesmium* presence. These images have been chosen because they are relatively free of clouds, *Trichodesmium* mats are visible in the "true color", and they contain many "ground-truth" observations (Figure 3).

The NASA Ocean Biology Processing Group method (Bailey and Werdell, 2006) to select match-ups, i.e., average or nearest pixel, was used to find coincidences between in situ observations and clear MODIS satellite pixels. A total of 468 coincident MODIS -Aqua pixels and SOB observations were found. After data screening, only 50 pairs remained, indicating that approximately 90 % of in-situ observations were not usable, primarily because of cloud cover. Once inspected, 19 MODIS-Aqua spectra out of the 50 pixels exhibited *Trichodesmium* features similar to those documented in Hu et al. (2010) and McKinna et al. (2011). In order to increase the number of useful observations, the searching window was extended both in time (up to +/- 4 days), with the hypothesis that a bloom can last for ~one week (e.g., Hegde et al., 2008), and in space (extend spatiallyup to over +/- 50 km, i.e., ~200 pixels at 250 m resolution). This spatial relaxation up to +/- 50km is motivated by considering a maximum drifting speed of ~10 cm.s^{-1} for algae mats compatible with the observed surface surface drifts in that region (Rousselet et al., 2018; Cravatte et al., 2015). Also, some in-situ observations close spatially and temporarily (in the same tile and at intervals of +/- 4 days) increased our degree of confidence in identifying the filamentous patterns as due to *Trichodesmium*.

After extending the searching windowrelaxing thresholds in our search algorithm, a total of 1200 spectra were extracted. Spectra were labelled, based on information provided by FAI (Hu, 2009) and the presence/absence of visible mats in the "true color" image. When the pixel is coinciding with a visible mat on the "true color" image and is characterized with high FAI but low Chl_a, it iswas labeled Type i: it is the signature of high algae concentration. Pixels in close proximity to *Trichodesmium* mats with low FAI and undetected mats in "true color" image arewere labeled Type ii. This second type of pixels is expected to have a high amount of free *Trichodesmium* colonies although not aggregated, offering an opportunity to detect spectral features due to the presence of *Trichodesmium* pigments other than Chl_a.

3.3 Robust spectral features over and near *Trichodesmium* mats

Figure 4B shows Type i average R_{rc} spectrum for 1km resolution images. Compared to blue water spectrum (see blue curve) no *Trichodesmium* spectral feature could be detected in the visible and NIR spectral domain. The increase of R_{rc} beyond 700 nm indicates that, at low resolution, the red edge effect is weakly discernible over floating algae, contrasting with robust signals detected by McKinna et al. (2011) and Hu et al. (2010) using both hyperspectral and MODIS 250x250 m observations over *Trichodesmium* mats. As suggested by these authors, spectral sensors with 1 km resolution are inappropriate to capture floating *Trichodesmium* due to negligible signal from discrete mats once mixed with adjacent water in 1x1 km cells. Middle and bottom rows of Figure 4 present R_{rc} and R_{rs} spectra interpolated to 250 m and partitioned in Tttype i and Tttype ii (second row and third row respectively). For Type i, it is noteworthy that the red edge signal is well captured in the NIR region at 859 nm, a MODIS land band with true 250 m resolution , while in the 758 and 896 nm bands the derived signals using bilinear interpolation are still low. One can notice that, after application of the NIR atmospheric correction, the shape of the mean R_{rs} spectra remains similar to that of R_{rc}, minus an offset in the visible region, corresponding to subtraction of the flawed aerosol contribution (compare Figure 4C withand Figure 4D). Derived R_{rs} at 748 and 869 nm are set to zero, as a result of the black pixel approximation.

For Type ii, the red-edge effect totally disappears from the true 250 m band at 859 nm and the derived baseline 758-8696 nm appears above the R_{rc} (and R_{rs}) at 859 nm. Hu (2009) had already noticed this issue for pixels with strong spatial gradients in the NIR part of the spectrum. It is interpreted as a result of the bilinear interpolation of 758 and 8~~9~~69 nm bands from 1 km to 250 m, using information from adjacent 1 km pixels.

Unfortunately, in the visible region, there is little useful information to capture *Trichodesmium* with R_{rc} (or R_{rs}) spectra interpolated to 250 m, both in Type i and Type ii situations. In *Trichodesmium* blooms, ~~it was observed~~ a release of dissolved colored substance was observed, as suggested by Hu et al. (2010) for coastal waters. For oligotrophic waters studied here, there was no such absorption in the blue region (412 nm) due to possible dissolved colored substances. Other~~s~~ characteristic features due to absorption maxima around 495 and 550 nm caused by~~due to~~ the presence of phycourobilin and phycoerythrobilin respectively (see Figure 5 in Hu et al. (2010)~~, their figure 5~~), are not seen using the 600 spectra-composite spectrum. Finally, the only ~~one~~ robust spectral feature in the visible range is a minimum in R_{rc} (and R_{rs}) occurring at 678 nm, due to increasing reflectance beyond 700 nm.

Interestingly, comparison between R_{rc} and R_{rs} shows that standard deviation error bars are much smaller for R_{rc} reflectances while the range of magnitudes between wavelengths is larger. This is a significant argument for using R_{rc} instead of R_{rs}~~-~~, as it would lead to a better discrimination of *Trichodesmium* mat spectra against other spectra.

3.4 Two published algorithms

Among the existing *Trichodesmium* mat detection algorithms, only the McKinna et al. (2011) and Hu et al. (2010) ones~~algorithms~~ were designed for the MODIS sensor and were tested in this study. The other algorithms were ignored as they were ~~either~~ not adaptable to MODIS sensor or ~~they~~ would give erroneous results due to Chl_a misestimation over *Trichodesmium* mats.

The *Trichodesmium* mat detection algorithm of McKinna et al. (2011) is based on 4 criteria relative to the shape of R_{rs} (~~see criteria~~~~definition~~given in the Appendix A). When applied to the same MODIS image (new collection) as the one used by McKinna et al. (2011), ~~the detection results of this algorithm showed~~ more pixels were discarded~~regarded pixels because~~ due to~~of~~ the fourth criterion, which ~~(~~eliminat~~ed~~ing pixels with~~which have a~~ negative ~~magnitude of~~ nLw at wavelength 555, 645, 678 or 859 nm). Indeed, the test of a negative R_{rs} value at 678 nm due to aerosol overcorrection excludes many pixels. Skipping the fourth criterion of the algorithm allowed ~~to~~ matching the results of McKinna et al. (2011). Therefore, this modification was adopted for the present~~this~~ study and the algorithm is called “McKinna modified” in the following.

The *Trichodesmium* detection algorithm presented in Hu et al. (2010) is based on a two steps analysis of R_{rc} spectra. The first step captures floating algae with FAI. The second step ~~is to~~ resolves the ambiguities between *Trichodesmium* and *Sargassum*, by analyzing spectral features in the blue-green region. To overcome possible spectral influence due to inappropriate atmospheric correction, Hu et al. (2010) proposed a simple correction method based on the difference of R_{rc} spectra between bloom and nearby algae free region. After several tries~~y~~ on the data presented above, the second step was found to be sensitive to the choice of the algae free region (not shown). Thus, the detection was made only by applying thresholds on FAI. After tuning, best results were obtained when FAI was between 0 and 0.04.

3.5 New algorithm criteria

Our criteria for detecting *Trichodesmium* mats were defined based on spectral characteristics of R_{rs} and R_{rc} (Figure 4). Similarly to the algorithms of Hu et al. (2010) and McKinna et al. (2011) (see Appendix), three criteria given in equations 1-3 were defined to extract the typical spectrum shape of *Trichodesmium*. The first criterium (Eq. 1) takes advantage of~~Indeed,~~ the systematic negative R_{rs} values at 678 nm over strong *Trichodesmium* mat concentrations~~ is taken as an advantage here~~. All pixels with negative R_{rs} value at this

wavelength have a high probability to be floating algae and thus *Trichodesmium* in this region. Moreover, this criterium has been found useful can also be used to detect some artifact, e.g., sun glint.
The absolute value of $R_{rs}(678)$ is actually used as an index of mats concentration, and can also be used to detect some artifact, e.g., sun glint (Eq. 1).
The second criterium (Eq. 2) is based onSimilarly to the algorithms of Hu et al. (2010) and McKinna et al. (2011) (appendix A), three criteria were defined to extract the typical spectrum shape of *Trichodesmium*. They are based on the use 1) $R_{rs}(678)$ since the valuespectrum shape may be affected by the aerosol miscorrection of SeaDAS standard atmospheric correction algorithm in the presence of mats (Eq. 1); 2) $R_{rc}(748)$ and $R_{rc}(859)$ to detect the spectral shape in the NIR presence of the red-edge associated with the surface *Trichodesmium* mats., which is one of the main criteria (Eq. 2); and 3) The third criterium (Eq. 3) uses $R_{rc}(645)$ and $R_{rc}(531)$ are used to resolve ambiguities between *Trichodesmium* mats and incorrectly seldetected pixels after processing with the previous criteria, the misdetections occurring mostly in near clouds neighbourhood (Eq. 3). To summarize, wWe usehave:

$R_{rs}(678) < 0$ (1)
$R_{rc}(748) < R_{rc}(859)$ (2)
$R_{rc}(645) < R_{rc}(531)$ (3)

4. Results
4.1 Algorithm application

TheAn attempt to detection resultscomparisone of the efficiency of the three *Trichodesmium* detection algorithms areis illustrated in Figure 5 on the MODIS tile A2007290.0355, used in McKinna et al. (2011). The McKinna modified algorithm shows the same detection patterns as the ones found in McKinna et al. (2011). It is a vast *Trichodesmium* area within which the filamentous structures cannot be distinguished. The new algorithm and the threshold FAI detectshow thin filamentous structures more similar to *Trichodesmium* mat structures observed on airborne photographs. When negative, the absolute value of $R_{rs}(678)$ can also be used as an index of mats concentration as values increase when getting to the core of patches. TIn the figure, the amplitudes of the Rrs(678) (stored in negative values) negative correction (at 678 nm) are displayed as colour gradient Figure 5Cfor the new algorithm., which could be inferred as an estimate of the different “thickness” within the mat: the more negative the value of Rrs(678) is, the “thicker” is the mat of the pixel.
Compared with both former algorithms, the new algorithm performs much better near clouds. Figure 6 is a zoom of the area delineated by the red rectangle inof Figure 5. This area presents a cloud patch where McKinna modified algorithm and FAI algorithms both detect *Trichodesmium* pixels. These pixels were identified as false positives as their spatial distribution is sparse and only in the vicinity of clouds. This conclusion is also supported by the “true color” composition (Figure 2) where the only *Trichodesmium* mats appear located at the bottom of the image. The robustness of the new detection algorithm to exclude cloud-contaminated pixels discard clouds while keeping accurate *Trichodesmium* mat detection is an important improvement for regions with high cloudiness covering, such as the WTSP.

4.2 Algorithm performance against in-situ mat observations

The exact coincidence in time and space between in-situ *Trichodesmium* mats observations and satellite mat detection is quite difficult to reach in general. One of the main reasons is by far the cloud cover, which eliminates a large quantity of the possible comparisons (90%). A second reason is the elapsed time between in-situ observations and the corresponding satellite pass during which the floating algae could have drifted at sea surface and/or migrated vertically depending on sea conditions (temperature, wind, etc.). For example,

the abundance of *Trichodesmium* at the sea surface may vary with the time of day, as a daily cycle of rising and sinking colonies in the water column is often observed as a result of cell ballasting (Villareal and Carpenter., 2003). Moreover, as *Trichodesmium* is as buoyant particle, it can be advected by surface currents.
To circumvent that problem and perform~~present~~ a more statistically robust comparison of the detection results of our algorithm with in-~~-~~situ data, we used the following strategy. An analysis of the spatio-temporal distance between the ~~closest~~ in-situ observations and the nearest detected mats was conducted. As explained previously, for each day in a range of +/- 4 days around the date of observation~~s~~, the spatial distance between the position of the observation and the nearest detected mat was computed. Overall, 80% of the observed mats have a corresponding mat detection within less than 5 km range independent of the detection algorithm used. These results demonstrate the statistical capability of the new algorithm to retrieve a mat near a point of observation.

4.3 Algorithm application for the OUTPACE cruise

The new algorithm was applied to MODIS data at the OUTPACE cruise time. A total of 140 tiles at 250 m resolution were covering the time period ~~15~~ February 15 - ~~07~~ April 7, 2015 and the spatial area of the cruise. Due to high cloud cover during the cruise, only a few tiles were exploitable. *Trichodesmium* mats were detected from 12 MODIS-Aqua and 3 MODIS-Terra tiles. Figure 7B shows the detected mats over these tiles (in blue). Note that the OUTPACE cruise actually crossed a number of *Trichodesmium* satellite detections. In order to further illustrate the results, a crude qualitative presence/absence scheme wa~~i~~s applied to better visualize which OUTPACE stations were coincident with the algorithm positive detection. We selected areas within 50 km of each OUTPACE station~~s~~ and labeled the station as presence when there was at least one pixel detected as positive in the satellite algorithm. In figure 7B, red points indicate presence and blue points absence.
Trichodesmium mats were mostly observed visually northeast of New Caledonia one week before the cruise and during the first days of the cruise (on board observation) by video and photographs. No other mats were observed during the rest of the cruise, but there was not any dedicated observer that would actually permit such visual observation, unlike during the ~~"REcensement des Mammifères marins et autre Mégafaune pélagique par Observation Aérienne" (~~REMMOA~~)~~ campaigns. Nevertheless, Underwater Vision Profiler 5 (UVP5) counts of colonies, phycoerythrin and trichome concentrations along the transect show that *Trichodesmium* contribution was maximum in the Melanesian Archipelago, the Western part of the transect (Dupouy et al., 2018), where slicks are numerous, and then fairly well related to *Trichodesmium* concentrations in the upper layer. The other high spot of mats is at LDB (station name in OUTPACE), where no slick was observed but where *Trichodesmium* was in high concentration, although mixed with a high abundance of picoplanktonic cyanobacteria (Dupouy et al., 2018).
Bonnet et al (2018) reported a significant ($p<0.05$) correlation between N2 fixation rates and *Trichodesmium* abundances during OUTPACE. Based on bulk and cell-specific N2-based isotopic measurements indicated that *Trichodesmium* accounted for 50 to >80 % of N2 fixation rates in this region at the time of the cruise. Such a high correlation between *Trichodesmium* biomass (here phycoerythrin) was also measured in New-Caledonia waters (Tenorio et al, 2018). Hence the in situ N2 fixation rate measured during the cruise (Figure 7A) was used as a robust proxy of the *Trichodesmium* concentration to further evaluate accuracy of satellite detections. A qualitative comparison between Figures 7A and 7B indicates that when significant fixation rates were observed, *Trichodesmium* presence was detected by satellite and when N2 fixation rates were low, *Trichodesmium* absence was stated. Although qualitative, this successful validation gives confidence in using our algorithm for *Trichodesmium* detection.

5 Discussion

5.1 Algorithm limitations

The proposed algorithm was designed to detect strong concentrations of floating *Trichodesmium* mats and limit wrong detections. However, floating *Trichodesmium* mats are occurring when sea surface is little agitated since they tend to sink and disperse in rough conditions (Cecile Dupouy, pers. comm.). In such a case, because of the low penetration depth of NIR irradiance (below 1 m), our algorithm does not allowfailed to detect sinking *Trichodesmium* mats even in strong concentration because of the low penetration depth of NIR irradiance (below 1 m). This situation occurred during the OUTPACE cruise, where measurements reveal high *Trichodesmium* abundances near the Fiji island (Stenegren et al., 2018), while our algorithm was unable to detect *Trichodesmium* mats (Figure 7, stations adjacent 180°E).

Another limitation concerns the validity of the new algorithm for future MODIS collections and other regions. The first criterion is taking advantage of the aerosol overcorrection as an index of floating algae. The aerosol correction algorithm will certainly be adapted in the future. Thus, first criterion would have to be replaced by another floating algae index; the FAI could be a good solution. However, it will always be possible to use the current (imperfect) aerosol correction. Moreover, the zero threshold onfor R_{rs}(678) , under which he pixel is labelled as a Trichodesmium matfloating algae is considered in high concentration, would have to be tested and tuned in other situations, e.g., in the presence of aerosols and other floating material. This study was carried out in the WTSP area, where the observed spectra seem slightly different from the ones reported in McKinna et al. (2011) and Hu et al. (2010). More specifically, the spectral oscillations of reflectance observed by Hu et al. (2010) in the range [412, 678] nm were not noticed in the present study. Therefore, the robustness of this algorithm in the presence of other floating algae (e.g., *Sargassum*), as well as in other regions of the world, would have to be tested to make it more general.

This study was limited to the processing of MODIS images, mainly because of the availability of images corresponding to field measurements (2014 and before). However, it would be worth extending the detection algorithm to other sensors such as MERIS, e.g., for comparison with the Gower et al. (2014) algorithm, and the recently launched Ocean and Land Colour Instrument (OLCI). This would require further investigations to adapt the algorithm to the specific bands of these sensors and evaluate the results. However, since these sensors cover the same spectral range, one can expect a quite similar behavior.

5.2 A different algorithm

Aware of these limitations, one may consider using a second algorithm is proposed that uses R_{rc} only, since R_{rs} is sensitive to the accuracy of the atmospheric correction. To emulate an index of mats concentration, Equation (4) is proposed in place of Equation (1). The magnitude of the trough at 678 nm is an indication of the mats concentration, which can beis retrieved fromusing the difference between the observed R_{rc}(678) and the result of the linear R_{rc} interpolation between the two adjacent wavelengths 645 and 748 nm. In the following, Tosimilar to the form criteria of McKinna et al. (2011) (Aappendix A), Equations (5-7) confirm the trough at 678 nm, check the spectral form typical to *Trichodesmium* (Figure 4), and assess the red-edge of the signal, equations (5-7), similar to the form criteria of McKinna et al. (2011), could be used:.

(4)

$$R_{rc}(678) < R_{rc}(859) \quad (5)$$

$$R_{rc}(678) < R_{rc}(748) \quad (6)$$

$$R_{rc}(678) < R_{rc}(667) \quad (7)$$

This alternative algorithm has the advantage to not to use absolute values (threshold of Eq. 1), and to be easily adapted of being easily adaptable to other sensors with similar spectral bands as (e.g. MODIS). Moreover, theis procedurealgorithm is free of atmospheric overcorrection and is while still exploitings the red-edge effect. However in its application, a large part of pixels detected as *Trichodesmium* by the former algorithm (i.e., described in Section 3.5) is discarded. The different criteria cannot be relaxed without a false positive detection increase. Therefore this alternative algorithm is more restrictive, and its suitability and performance require further examination.~~proposed only as an alternative to the former one~~, ~~but more restrictive.~~

5.2~~3~~ Spatial resolution impact

As indicated previously, only few spectral bands (land channels) have a high resolution (250 m or 500 m), while the rest have a nominal resolution at 1 km. To investigate the influence of resolution on the spectral signature of *Trichodesmium* mats the spectral analysis was also conducted at a 1 km resolution. Dense groups of extended mats are still well detected at 1 km resolution. However, thinner mats with a weaker signal visible at 250 m resolution are lost at 1 km resolution. Figure 8 illustrates this behavior on MODIS data.

The spatial structure of *Trichodesmium* aggregates is complex. When mats are present, *Trichodesmium* have a tendency to form a filamentous pattern much narrower than 250 m (50 m at most, according to visual detections), and thus the satellite sensor at 250 m resolution can only detect the largest ones (Figures 8 and 9). Hence there is a scale mismatch between the exact form of the thin filaments and the actual detection by the current satellite data, which must average in a way the thin and strongest filaments into signals detectable at 250 m. Understanding the shape of the filaments, and their physical characteristics (e.g width) will require much higher resolution satellite date (at least 50 m) which are available at present but without repetitive coverage. Figure 9 additionally illustrates that the *Trichodesmium* filaments are but a tiny part of the chlorophyll tongues and are inserted into the much wider chlorophyll patterns. There can exist, within a chlorophyll tongue such as Figure 9, several thin elongated filaments.

One would also intuitively believe that the filaments illustrate the presence of dynamical fronts where convergent dynamics can maintain and participate in the mat aggregations. A natural dynamical criteria allowing to characterize the presence of the filaments could be found in the finite-size Lyapunov exponents (FSLE) methods (Rousselet et al., 2018).~~,~~ A relationship between FSLE only and in situ chlorophyll edges "fronts" was found during OUTPACE with a 25 % correlation score. The same kind of relationship was expected for floating algae filaments. However, the relationship between FSLE and the organisation of the filaments could not be proven within this study.~~but we could not associate the presence of the FSLE with the presence of the filaments, for instance on Figure 9 (not shown). Rousselet et al. (2018) discuss the fact that FSLE only matched in situ chlorophyll "fronts" during OUTPACE with a 25 % correlation but we have seen that our filaments are present at a scale finer that the chlorophyll scale detected by the satellite during OUTPACE (see also Figure 9). Our filaments are typically present at the sub-mesoscales, and~~ W~~w~~e believe that it is here also a question of scale. T~~unlikely that t~~he present calculation of FSLE~~,~~ is using 12.5 km satellite data at best (Rousselet et al., 2018), thus it is difficult to think that such a resolution could catch dynamics Trichodesmium filaments wide of a few hundred meters. A~~can in fact be used to understand the filament dynamics. If FSLE are the right tools to understand filament formation, they must be calculated using a much higher spatial resolution than presently available. Hence, we lack the tools at present with which to understand the organization of the detected filaments and~~ dedicated in situ experiments wou~~i~~l~~l~~d have to be specifically undertaken to resolve the~~at~~ question of filament organization.

6 Conclusions and perspectives

At present, previously published algorithms detecting *Trichodesmium* using the current MODIS data archive (Hu et al., 2010; McKinna et al., 2011), cannot be directly used in the South Pacific as they either miss the mats due to algorithms failures (Section 3.4) and/or do not eliminate numerous false positive in the presence of clouds. In our study, we have devised a new algorithm building on the previous ones, which allows a cleaner and more robust detection of those mats. Validation was accomplished using a new, updated database of mats in the South Pacific. This algorithm can however detect only the densest slick but achieves the goal of limiting the detection of false positive due to clouds. During the OUTPACE cruise, satellite detections confirmed the presence of *Trichodesmium* slicks at much wider spatial range than what is possible to observe from the ship. This illustrates the important contribution and complementary nature of satellite observations to in situ measurements. Yet, the new detection algorithm was developed and evaluated for the WTSP region. Hence, future prospects will be to extend the evaluation to other regions, especially in the presence of other floating algae such as *Sargassum*.

During the same campaign Dupouy et al. (2018) found that ocean color measured with a Satlantic UV-VIS radiometer at greenish blue and yellowish green wavelengths were not totally linked to Chl_a and~~. Principal component analysis of Satlantic UV-VIS radiance showed~~ that an extra factor independent of Chl_a, may be related to colony backscattering or fluorescence, governs part of the variability. This extra factor~~,~~ is not observed during BIOSOPE (Biogeochemistry and Optics South Pacific Experiment), a cruise in the tropical Southern Pacific gyre (Claustre et al., 2008), where the radiance is~~measured with the same instrument and where *Trichodesmium* were absent, are~~ correlated to Chl_a only. Further investigations have to be conducted to confirm that such a signal is produced by *Trichodesmium* and can~~could~~ be detected from space.

MODIS-Terra and MODIS-Aqua satellite sensors are acquiring data since 2000 and 2002 respectively. However, the data quality of these sensors is becoming more and more uncertain with time going by, as their lifetime was not expected to last more than 6 years. The new algorithm could be adapted to other satellite instruments with similar spectral bands, for example Visible Infrared Imaging Radiometer Suite (VIIRS) onboard NPP and NOAA-20 (1 km resolution) and OLCI onboard Sentinel-3 (300 m spatial sampling), but the spatial resolution remains a problem as we observed that 250 m was already too coarse a resolution to understand the thinner mat dynamics. A study with a better spectral and spatial resolution may lead to better performances and to a new and better algorithm.~~,~~ It~~and this~~ may be possible, at least regarding spatial resolution, with Multispectral Instrument (MSI) onboard the Sentinel-2 series (10 to 60 m resolution).

It has previously been documented that near dense *Trichodesmium* mats some satellite product like the satellite chlorophyll concentration are inaccurate. In order to better constrain the contribution of *Trichodesmium* to nitrogen and carbon biogeochemical cycles, this algorithm must be corrected. Using R_{rc} instead of the R_{rs} is possible (section 5.1~~2~~), but some adjustments and comparisons with in-situ measurements must first be carried out. Globally such algorithm would allow one to estimate the *Trichodesmium* aggregated in sea surface mats. The next step is to understand the quantitative aspect linking the *Trichodesmium* abundances to N2 fixation rates, including their vertical distribution even when *Trichodesmium* filaments/colonies are spread out in the water column. Another important field of interest is to be able to understand phytoplankton functional types using satellites including *Trichodesmium* (de Boissieu et al., 2014). At present, we do not know any such study that included *Trichodesmium* but we have hopes that with our new in situ database and our understanding of the mat shapes detected in the present study, and the development of performing statistical methods such as machine learning, advances can be made in ~~that~~ that regard, a perspective for future work.

Finally Dutheil et al. (2018) explore the regional and seasonal budget of the N2 fixation due to *Trichodesmium* in a numerical model based on physical and biogeochemical properties that does not take into consideration the spatial characteristic of *Trichodesmium* colonies to aggregate in mats~~part of Trichodesmium that aggregates in mats~~. One interesting aspect will be to find a way to integrate our results in such model to better estimate the regional effects of that species.

Appendix A: McKinna et al. (2011) algorithm

The McKinna et al. (2011) algorithm is based on the analysis of the above-water reflectance spectrum of a moderate *Trichodesmium* mat, similar to the one measured on colonies in a small dish with an Ocean Optics spectroradiometer (Dupouy et al., 2008). It uses typical spectral characteristics of the normalized water-leaving radiance (nLw) after atmospheric correction to define 4 *Trichodesmium* detection criteria. The first three criteria relate to the shape of the spectrum and the last criteria discards any pixel with negative nLw. When these 4 criteria are respected the pixel is identified as revealing the presence of *Trichodesmium*:

$$nLw(859) > c\,1\,nLw(678) \quad (A1)$$
$$nLw(645) > nLw(678) \quad (A2)$$
$$nLw(555) > nLw(678) \quad (A3)$$
$$nLw(555)\,,\, nLw(645)\,,\, nLw(678)\,,\, nLw(859) < 0 \quad (A4)$$

Appendix B: Hu et al. (2010) algorithm

TheAnother detection algorithm, originally developed by (Hu, 2009) for floating algae, can be applied to *Trichodesmium* mats, as demonstrated by Hu et al. (2010) on MODIS-Aqua images of the west coast of Florida and the Gulf of Mexico, even though the *Trichodesmium* mats occurred in Case 2 waters. This algorithm can be decomposed into two steps: 1) detection of floating algae (FAI, Floating Algal Index), and 2) test-of-form criteria of the radiance spectrum.
The FAI aims at detecting the strong reflectance in the infrared (red-edge) of the algal agglomerate. To avoid the atmospheric overcorrection linked to the red-edge effect of the floating algae organized in dense matsa heap (Hu, 2009), the calculation of this index is applied to reflectance corrected only for the effects of Rayleigh scattering (R_{rc}). This correction accounts for the major part of the color of the atmosphere if aerosols are not too abundant (i.e., small optical thickness). The FAI is then defined as the difference between R_{rc} of the infrared domain (859 nm for MODIS) and a reference reflectance ($R_{rc,0}$) calculated by linear interpolation between the red and shortwave infrared, i.e., 667 nm and 1240 nm for MODIS:

(B1)

$$\lambda_{RED} = 645\text{ nm}\,,\ \lambda_{NIR} = 859\text{ nm}\,,\ \lambda_{SWIR} = 1240\text{ nm} \quad (B2)$$

where RED = 645 nm, NIR = 859 nm, and SWIR = 1240 nm. According to Hu et al. (2010), the difference between R_{rc} and $R_{rc,0}$ (the second term of Equation B18) allows one to deal with the majority of the atmospheric effect which has a quasi-linear spectral shape between 667nm and 1240nm.The second step of the algorithm consists in identifying the mats emphasized by the FAI thanks to the shape of the spectrum in the visible domain. So as to correct the bias inferred in the visible part of the spectrum by the possible presence of mats, Hu et al. (2010) suggests applying to the pixels presenting a strong FAI value the correction of an area situated immediately next to this pixel and without bloom. This approach being very computer expensive, it is substituted by a simple difference between the R_{rc} spectrum of the pixels considered (i.e., eventually with *Trichodesmium*) and that of a nearby zone without mat. The spectrum of this R_{rc} difference presents a pattern (spectral signature) that seems to be specific to *Trichodesmium* presence, i.e., a succession of high type low-top-low-top for the wavelengths 469-488-531-551-555 nm.

FIGURE LEGENDS

Figure 1: Map of in-situ (visual) observations of *Trichodesmium* mats gathered on the studied region.

Figure 2: MODIS-Aqua tile A2007290.0355 used in McKinna et al. (2011): (A) RGB image from R_{rc} generated with R = 645nm, G = 555nm and B = 469nm computed using the formulae 0.29319407 + 0.45585 × atan(50 × (R_{rc}) − 0.015)); (B) aerosol optical thickness at 555nm derived using the NIR atmospheric correction algorithm by Gordon and Wang (1994); (C) Chlorophyll concentration product computed from R_{rs} using OC3 and OCI (Hu et al., 2012). Cloud pixels in grey and land pixels in brown.

Figure 3: "True color" image of the 17th December "2014A2014351.0255" for which in situ observations (red crosses) exist in the SOBT database, and used for the test and adjustment of the different bio-optical algorithms.

Figure 4: MODIS Spectra of *Trichodesmium* mats (A, B, C and D) and adjacent areas (E and F) normalized by the maximum spectral value of all wavelengths, with pixels resolution at 1 km (A and B) and 250m (C, D, E and F). A, C and E are R_{rs} reflectances. B, D and F are R_{rc} reflectances. Average is red line and error bars are the standard deviation. The average of the water signal is the blue line.

Figure 5: *Trichodesmium* mats detection results on MODIS tile "A2007290.0355": (A) pixels detected pixels with McKinna modified algorithm in red, (B) pixels detected with FAI value of detected pixels with FAI algorithm, (C) pixels detected with the new algorithm, color represents showing values of R_{rs}(678) (absolute value of the detected pixels). Cloud pixels in grey.

Figure 6: MODIS tile "A2007290.0355" zoomed of the red square area (24°S, Figure 5). Pixels resulting from a false positive detection of (A) the McKinna modified algorithm, (B) FAI and (C) the new algorithm for the area in red on the Figure 5. Cloud pixels in grey.

Figure 7: (A) Background is the mMonthly composite of satellite chlorophyll for March 2015, with the corresponding chlorophyll-a concentrations given by the colorbar underneath the map. Superimposed colored dots aretogether with in- situ fixation rates measured superimposed on the OUTPACE track, as colored dots (with values given byon the left colorbar). (B) In the background, cyan pixels are pixelsPoints detected as *Trichodesmium* (bluecyan dots) by the ourpresent algorithm. The colored dots superimposedtogether with show thea summary of absence (blue) or/ presence (red) of detected pixels within a radius of 50 km around OUTPACE stations.denoted as colored blue (absence) and red (presence) points along the OUTPACE track. A point at the OUTPACE station is colored when the algorithm shows presence within a 50km radius off the station.

Figure 8: FAI application to the MODIS tile "A2007290.0355": zoomed of the green square area (22.5 °S, Figure 5). (A) Results at 250 m resolution, (B) the same scene at 1 km resolution. Only a few pixels are remaining corresponding to the densest part of the surface mat, showing the loss of detected mats.

Figure 9: MODIS-Aqua image at 250 m taken on March, 6th, 2015 to the south of the OUTPACE cruise illustrating the structure of the chlorophyll (colors) together with the filaments of *Trichodesmium* detected by our algorithm (in black). Cloud pixels in white.

TABLE LEGEND

Table 1: Specification of the first 1336 MODIS channels, including primary use, central wavelength, bandwidth and spatial resolution. http://eoweb.dlr.de:8080_guide/D-MODIS.html

Table 2: Satellite image with in-situ observations used to analyze *Trichodesmium* mat and adjacent spectra.