An intercomparison of bio-optical techniques for detecting phytoplankton functional types from space

Robert J.W. Brewin, School of Earth Ocean and Environmental Science (SEOES), University of Plymouth, UK, PL4 8AA, robert.brewin@plymouth.ac.uk.
Samantha Lavender, SEOES, University of Plymouth, UK, PL4 8AA. ARGANS Limited, Tamar Science Park, Plymouth, UK, PL6 8BX
Nick J. Hardman-Mountford, Plymouth Marine Laboratory, The Hoe, Plymouth, UK. PL1 3DH.
Rosa Barciela, Fitxroy Road, Met Office, Exeter, UK, EX1 3PB
Dionysios Raitsos-Exparchopoulos, HCMR, Institute of Oceanography, P.O. Box 712 Anavissos 19013, Attika, Greece.
Takafumi Hirata. Plymouth Marine Laboratory, The Hoe, Plymouth, UK. PL1 3DH.

ABSTRACT

Satellite remote sensing of ocean colour is currently the only way of measuring synoptically wide-area ocean properties such as phytoplankton abundance, and recent bio-optical and ecological methods have been established that use satellite data to differentiate between certain phytoplankton functional groups (PFTs). These techniques have been applied to a long-term ocean colour data series from the SeaWiFS satellite sensor in the Atlantic Ocean, and validated using in situ data from two different sources in order to test their ability to detect three phytoplankton size classes (micro-, nano-, and picoplankton). In case 1 regions all the published algorithms were seen to perform fairly accurately. Results from AMT diagnostic pigment analysis (DPA) from 1997-2005 indicate that an unpublished model outlined in this study and referred to as BLM1, and the chlorophyll-a approach developed by Hirata et al. (2008) seem to be functioning the most accurately, closely followed by Uitz et al. (2006), and Devred et al. (2006) (microplankton only). Results from CPR data in the North Atlantic (1997-2003), which comprised both case 1 and case 2 regions, indicated the size class approach of Uitz et al. (2006) was functioning the most accurately, though in all cases differences were found in accuracy between different size classes. Potential biases are highlighted in this study, which indicate the results are indicative and not definitive.

INTRODUCTION

In recent years a variety of bio-optical and ecological methods have been established that use satellite data to differentiate between PFTs in full, or phytoplankton size classes (PSCs) (e.g. Sathyendranath et al. (2004); Alvain et al. (2005, 2008); Uitz et al. (2006); Devred et al. (2006); Ciotti and Bricaud (2006); Aiken et al. (2007); Raitsos et al. (2008); and Hirata et al. (2008)). Platt et al. (2006) describe this as one of the major problems of the day in ocean optics, which is further complicated by the sparseness of in situ data that is required to validate these algorithms.

Through implementing these satellite techniques to long-term ocean colour data series such as the GlobColour (see www.globcolour.info) or SeaWiFS data series, and comparing the results to extensive composite in situ data, a better understanding as to how these algorithms are performing can be made. Furthermore, where and why they are performing accurately and incorrectly is important information needed to improve these algorithms. Accurate PFT measurements from space could be incorporated into primary production Earth Observation models, which would result in a greater understanding of the contribution of different PFTs to global ocean primary production. The measurements could also be
used to validate PFT ecological models, as the lack of availability of \textit{in situ} PFT data is regarded as a major problem in PFT ecosystem modelling. This study aims to take the first step toward this goal, by comparing different PFT satellite algorithms to three different sources of \textit{in situ} data in the Atlantic Ocean, spanning from 1997-2005, in order to assess the ability of the various satellite approaches at detecting three size classes of phytoplankton in the Atlantic Ocean.

\textbf{DATA}

Nair \textit{et al.} (2008) highlight that the use of any one \textit{in situ} method in isolation would imply identifications of phytoplankton that may not be entirely dependable, and hence incorporating different types of \textit{in situ} methodologies would lead to a more accurate diagnosis of the phytoplankton groups. For the purpose of this study two different \textit{in situ} datasets were used: Atlantic Meridional Transect (AMT) HPLC pigment data from 1997-2005 (AMT 5-17), and Continuous Plankton Recorder (CPR) data from the North Atlantic (1997-2003).

AMT HPLC pigment data was quality assured by statistical methods according to Aiken \textit{et al.} (In Press). This data was then classified using Diagnostic Pigment Analysis (DPA) according to Vidussi \textit{et al.} (2000), Uitz \textit{et al.} (2006) and Hirata \textit{et al.} (2008). The dominant size class was established based on where a size class (pico-, nano- or microplankton) had a DPA value greater than 45%. From AMT 5-17, 3132 HPLC measurements were originally utilised. This was reduced to 1369 measurements by only including data taken in the top 10m of the water column. Where there were two or more measurements in a vertical profile of the 10m surface layer, either the dominant size class closest to the surface was prioritised, or if there were multiple measurements the most frequent dominant size class was prioritised. The data was then matched to Level 3 SeaWiFS daily products, at 9km resolution, + or – one pixel. Mean data for nLw(412, 443, 490, 510, 670), chlorophyll-a (Chl-a), and optical aerosol thickness (T865) as well as the associated standard deviations were calculated across the 9 pixels, and the IOP model of Lee \textit{et al.} (2002) was used to calculate the phytoplankton absorption coefficient ($a_{ph}(\lambda)$) at these data points. This resulted in 380 daily match-ups spanning from 1997-2005, shown in Figure 1.

Raitsos \textit{et al.} (2006) analysed measurements of phytoplankton abundance (cell counts) from the CPR in the North Atlantic from 1997-2003. Samples were collected by a high speed plankton recorder towed behind ‘ships of opportunity’ in the surface layer of the ocean ~ 6-10 m deep, from which plankton were filtered on a constantly moving band of silk. Raitsos \textit{et al.} (2006) calculated the total number of species per sample for four PFTs (diatoms, dinoflagellates, silicoflagellates, coccolithophores), and determined the dominant phytoplankton type for each sample using the Z factor standardized method, with the largest Z values for each group categorized as the dominant species. Where the CPR data indicated that there were no dominant cells a category of ‘no dominance’ was used (note, the CPR silk filter only filters plankton larger than 2µm (unless caught between silk filaments), and therefore this \textit{in situ} dataset focused on the nano and micro size ranges). Raitsos \textit{et al.} (2006) matched the CPR data to Level 3 SeaWiFS 8 day Chl-a products at 9km resolution, which resulted in 17,061 matching data points, shown in Figure 1.
Figure 1: In situ data points used for the study. Red points indicate CPR data, and blue points indicate AMT data.

METHODS

The following PFT and size class algorithms were implemented for the satellite in situ match-up data: Alvain et al. (2005, 2008) PHYSAT method; Uitz et al. (2006) statistical size class approach; Hirata et al. (2008) $a_{ph}(443)$ size class method referred to in this study as PML ($a_{ph}$) and the Hirata et al. (2008) Chl-a size class method referred to in this study as PML (Chl); Devred et al. (2006) size class method; and a basic unpublished size class model developed by the author, which shall be referred to as BLM (Brewin, Lavender, and Mountford) (versions 1 and 2 as described later). BLM involved HPLC DPA of the NOMAD dataset (new release (22/02/2007) (Werdell and Bailey (2005))) and the BENCAL02 dataset (Fishwick et al. (2006); Aiken et al. (2007)) to distinguish three size classes, which were matched to in situ absorption measurements of $a_{ph}(\lambda)$, whereby the spectrum was normalised to $a_{ph}(443)$, and distinct differences were found between the dominant size classes of micro-, nano-, and picoplankton between $a_{ph}(510)/a_{ph}(443)$ and $a_{ph}(555)/a_{ph}(443)$ (see Figure 2).
A basic power law relationship was created between Chl-a and $a_{ph}^*(\lambda)$ in the NOMAD bio-optical dataset, similar to that of Bricaud et al. (1995) (see Figure 3), and satellite pixels were classified as a certain size class based on mean square error differences between the satellite $a_{ph}^*(\lambda)/a_{ph}(443)$ values between 510-555nm, and the mean $a_{ph}(\lambda)/a_{ph}(443)$ values of the three size classes in the in situ results, this method is denoted BLM1 (Note: $a_{ph}^*(\lambda)/a_{ph}(443)$ and $a_{ph}(\lambda)/a_{ph}(443)$ are theoretically identical). Lee et al. (2002) IOP model was also used to derive the satellite based $a_{ph}(\lambda)$ normalised spectrum from which the same method was applied, and shall be denoted as BLM2.
according to where the percentage of a particular size class was greater than 45% (in the few pixels it was not, the highest percentage of the three size classes was allocated as dominant). The techniques of PML ($a_{ph}$), PML (Chl), BLM1 and BLM2 already give a dominance size class so they could be directly applied to the data. The model of Devred et al. (2006) calculates the percentage contribution of micro- and combined nano- and picoplankton to $a^*_{ph}(\lambda)$ at a given wavelength based on the Chl-a concentration. For this study, the wavelength of 443nm was adopted as differences in $a^*_{ph}(\lambda)$ between size classes have been shown to be greatest at this wavelength (see Ciotti et al. (2002)). The Devred et al. (2006) method was focused only on the microplankton size class, as the mixing model developed separated $a^*_{ph}(\lambda)$ into microplankton contribution, and combined nano- and picoplankton contribution, as opposed to attempting to separate the smaller size classes as in the case of the previous models. Where the percentage of microplankton or combined nano- and picoplankton was greater than 45%, the pixel was allocated to the respective size class. All methods were applied to the AMT dataset, however, as the CPR data was only matched to Chl-a products only the methods that rely on Chl-a as an input were applied to the dataset, i.e. BLM1, PML (Chl-a), Uitz et al. (2006) and Devred et al. (2006).

Two methods were adopted to analyse the in situ and satellite match-ups. Method 1 was designed to assess the percentage of satellite and in situ match-ups in terms of where they both predict a size class as well as where they both do not predict a size class. The in situ data was divided into three groups representing the size classes, and within each group the data was divided into data dominated, and data not dominated by a size class. This data was then matched to the mean satellite predictions of each technique. The percentage accuracy of each method was then calculated by dividing the number of data points that both the in situ and satellite method observed as the same, by the number of data points the in situ data observed, for both dominated and non-dominated pixels of a given size class, before being averaged and converted into a percentage.

According to this method, however, if one was to guess the size class it would be likely to be accurate to 50%. For instance, if one was to assume all pixels were microplankton, where the in situ data was microplankton the method would agree 100%, where the in situ data was not dominated by microplankton it would be 0% an average of 50%. As for the nano- and picoplankton it would show 0% where the in situ data predicted the size class, and 100% where it did not, again showing an average of 50%. Therefore, in order to be confident the satellite algorithms are performing better than guesswork, a judged percentage accuracy was calculated according to the percentage accuracy minus 50% and then doubled. This gave a percentage where 0% indicted the satellite algorithm was as good as guesswork, and 100% indicted the algorithm was functioning perfectly. A flow chart of method 1 is shown in Figure 4.
Method 2 was adapted from the validation method used in Hirata et al. (2008). This method based the validation on a scoring technique, with a direct match indicating 2 points, a near match 1 point, and no match indicating 0. In this study for each size class, a direct match (2 points) was assigned when either the mean, or the mean + or – the standard deviation of the 9 pixels matched with the dominant size class shown by the in situ data. For the near match criteria, the in situ data was re-analyzed to assess a more mixed environment. For the HPLC diagnostic pigment data, where the dominant size had a DPA ratio greater than 45%, the data was also assessed to find if another size class had a DPA ratio of greater than 40% at the same point, and if so, a second size class was recorded. For CPR cell counts, if a second size class had a Z factor of less than 0.2 smaller than the dominant size class it was recorded as a second size class. Then if the mean, or the mean + or – the standard deviation of the 9 pixels was matched to the second size class a near match (1 point) was recorded, and where there were no matches of either the dominant or where applicable the second size class a no match (0 points) was recorded. The results were then converted into a percentage for each size class, to easily assess how the algorithms were performing.

POTENTIAL BIASES

In this study the in situ data is essentially deemed to be the truth, which is not correct. There could clearly be instances where a satellite algorithm is correct and the in situ is not. The AMT DPA data, as highlighted by Vidussi et al. (2001) and Uitz et al. (2006), does not strictly reflect the true size of phytoplankton. Diatoms for example have been observed in the nanosize range, whereas in this procedure they are categorized as only microplankton. Furthermore, some taxonomic pigments might be shared by various phytoplankton groups, such as fucoxanthin (the main indicator of diatoms) may also be found in some prymnesiophytes. The cell count approach used for the CPR data is a more definitive approach. However, 8-day SeaWiFS mean values that are matched to the cell count data cannot account for the mesoscale variability associated with dynamic zones, and therefore are only really applicable to areas with low heterogeneity. Furthermore, due to the CPR data having a mesh size of 2 µm, it is a much more effective tool at filtering microplankton and less effective at filtering nanoplankton, which can slip between the silk mesh.
With regards to the satellite algorithms, each method is very different in its approach and it is thus very difficult to make a quantitative comparison with the *in situ* data. This study has focused primarily on size class, whereas the PHYSAT approach looks at specific taxonomic groups. PHYSAT does not attempt to account for all the taxonomic groups within a size class that this study is assuming (though is does predict the more abundant groups). Furthermore, this study assesses dominance, and some of the approaches have been adapted to fit this criterion. In light of this, it would therefore be sensible to assume the results of this study are indicative and not definitive.

**RESULTS AND DISCUSSION**

Figure 5 shows the results from method 1. The CPR data was collected in both case 1 and case 2 regions, and with regards to microplankton, all the algorithms used on this dataset were shown to perform greater than 15% implying they are better than guesswork (Figure 5 (A)). The algorithms were shown not to perform as accurately with regards to nanoplanckton. This could be due to the algorithms themselves not predicting this size class accurately, but is also likely to be due to the CPR dataset not accurately capturing the dominance of this size class, as previously highlighted. Uitz et al. (2006) algorithm was shown to perform with most accuracy according to this criteria (23% micro- and 20% nanoplanckton), followed by BLM1 (20% micro-, 8% nanoplanckton) and Devred et al. (2006) (19% microplankton). These values are still quite low, and this is also thought to be associated with inaccuracies of the SeaWiFS OC4-v4 Chl-a algorithm in the case 2 locations.

![Figure 5: Histograms showing results from Method 1, (A) CPR data, and (B) AMT data. Microplankton are shown in red, nanoplanckton in green, and picoplankton in blue.](image)

The AMT data was collected in primarily case 1 regions (see Figure 1) where the satellite signal is essentially affected by only optical properties of phytoplankton. It is the best of the two datasets for comparing algorithms, as all the algorithms in this study are tested on this dataset (CPR dataset only 4), and it is case 1 areas where the algorithms were developed. This is highlighted in Figure 5 (B), as all the algorithms are shown to improve considerably in their performance when compared to the CPR dataset. The BLM1 algorithm is shown to perform the most accurately according to this criterion with 57% micro-, 30% nano- and 57% picoplankton, this is closely followed by the Hirata et al. (2008) Chl-a algorithm with 36% micro-, 37% nano-, and 54% picoplankton. What is interesting with both these algorithms is that the Chl-a based algorithm of PML (Chl) is shown to perform better than the IOP PML (*a*<sub>ph</sub>) based algorithm, and BLM1 which uses Chl-a to calculate *a*<sub>ph</sub>(λ) is seen to perform better than BLM2 which uses IOP *a*<sub>ph</sub>(λ). This could be due to the IOP model used in this study not
performing as accurately as is required; Chl-a algorithms such as OC4v4 are empirical approaches, but relatively robust in case 1 waters (O’Reilly et al. 1998). Running different IOP models in this study and comparing them could result in a more definitive answer (e.g. Smyth et al. (2006) and Maritorena et al. (2002) (GSM)). Uitz et al. (2006) was seen to perform accurately with microplankton (50%), but more poorly with nano- (6%) and picoplankton (21%). Devred et al. (2006) was shown to do well at predicting microplankton (54%), and the PHYSAT method was seen to perform less well 8% micro-, 16% nano-, and 6% picoplankton, which is thought to be possibly due to reasons previously highlighted, i.e. it is not a size class model.

Figure 6: Histograms showing results from Method 2, (A) CPR data, and (B) AMT data. Microplankton are shown in red, nanoplanckton in green, and picoplankton in blue.

The CPR data in Figure 6 (A) is shown to have similar results to Figure 5 (A) with regards to how the algorithms are performing in relation to each other. Uitz et al. (2006) algorithm was shown to perform with most accuracy according to this criteria (59% micro- and 74% nanoplanckton), followed by BLM1 (62% micro-, 55% nanoplanckton), PML (Chl) with (40% micro- and 70% nanoplanckton), and Devred et al. (2006) (51% microplankton). Figure 6 (B) shows the results from the AMT data, and what is striking is that the algorithms of BLM1, BLM2, PML (a_ph), and PML (Chl) all perform very accurately at predicting picoplankton, all with greater than 92% accuracy. Similar to the results in Figure 5 (B), the BLM1 algorithm was seen to perform the most accurately with 67% micro-, 51% nano-, and 94% picoplankton; this was closely followed by PML (Chl) algorithm with 47% micro-, 59% nano-, and 94% picoplankton. Similarities between these algorithms could be linked to the fact they are both developed using the NOMAD in situ dataset. Uitz et al. (2006) was found to more accurately predict micro- (61%), and nanoplanckton (76%), than picoplankton (42%). Devred et al. (2006) was shown to perform as accurately as any method at predicting microplankton (65%).

CONCLUSION

Comparing and validating different PFT satellite algorithms in both case 1 and case 2 regions using in situ data is a critical issue to address with regards to improving synoptic estimates of PFTs needed to validate ecosystem models, and improve our understanding of the biogeochemical interactions between phytoplankton and our environment. In case 1 regions (AMT data) all the algorithms were shown to perform with a higher accuracy than combined case 1 and case 2 regions (CPR data). In both the AMT DPA comparison and the CPR North Atlantic comparison the BLM1 algorithm, the PML (Chl) algorithm and the Uitz et al. (2006) algorithm were shown to perform the most accurately over the three phytoplankton size classes. The technique of Devred et al. (2006) was also seen to perform as
accurately at predicting dominant microplankton. However, potential biases were highlighted which indicate the results are indicative and not definitive.

Future work will incorporate the satellite PFT algorithms of Ciotti and Brinaud (2006), and Raitos et al. (2008), as well as output from a variety of biogeochemical models such as Diat-HadOCC at the UK Met Office and ERSEM at PML. In order to gain a more quantitative assessment of how the different satellite algorithms are performing, more in situ data needs to be gathered, over larger spatial scales. This will aim to include cell count time series data from the L4 station in the English Channel, BATS HPLC data from the Bermuda, and HOTS HPLC data in Hawaii. This will ultimately result in understanding where and why certain algorithms are functioning better or worse, which could lead to an improved approach through either, combining the output of several methods, improving an existing technique or developing a modified version. Future work will also aim to compare CO₂ flux variation hindcasts produced by the FOAM-HadOCC 3D physical model with coupled biology, in order to better understand the contribution of different PFTs to global CO₂ flux variability.

ACKNOWLEDGEMENTS – We thank present and past staff of SAHFOS who have contributed to the maintenance of the CPR time series and the ocean colour data used in this study were produced by the SeaWiFS Project at the NASA Goddard Space Flight Center. This work was also supported by the Natural Environment Research Council, UK, through the Centre for observations of Air-Sea Interactions and fluxes (CASIX). Thanks would also like to be given to both Sèverine Alvain and Julia Uitz for their useful collaboration, and Josephine Ras and Andre Morel for the BENCAL02 absorption data.

REFERENCES:


