SIMULATING HYPERSPECTRAL DATACUBES FOR OCEAN COLOR: A FOUNDATION FOR PHYTOPLANKTON RECOGNITION IN SPECTRAL SIGNATURES

Oudijk, A.E.[1], *Johansen, T. A.*[1], *Johnsen, G.*[2], *Alver, M. O.*[1]

[1]Dept. Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Norway
[2]Dept. Biology, Norwegian University of Science and Technology (NTNU), Norway

ABSTRACT

This paper focuses on the first steps in setting up simulations of oceanic hyperspectral datacubes, to simulate hyperspectral images of coastal areas as obtained with a pushbroom hyperspectral imager. The purpose is to learn more about recognizing and estimating water constituents in spectral signatures, thereby facilitating the development of better algorithms for processing of hyperspectral data that is becoming more available from new satellites. In this work we show that we can distinguish diatoms and flagellates from observations of the remote sensing reflectance (R_{rs}). We do this by using in-vivo spectral reflectance settings from either diatoms or flagellates in a light model (Ecolight-S) that models how incident light on the ocean surface propagates in the water column and produces estimates of R_{rs}. Ecolight-S is also applied to diatom and flagellate concentrations from numerical ocean simulations using the SINMOD model. The estimated R_{rs} values are converted to a hyperspectral datacube of an area that an image from a hyperspectral satellite such as HYPSO-1 could cover. By this method we can make large datasets with variable ocean constituents, and simulated datacubes corresponding to these. In future research, it should be studied how we can distinguish mixed concentrations of diatoms and flagellates in the remote sensing reflectance. In addition, the influence of atmospheric effects on the spectral signature should be implemented in the model. Last, it should be studied how effective the modelled, hyperspectral datacubes perform as labeled datasets to train machine learning algorithms in classifying phytoplankton in hyperspectral datacubes from satellite imaging.

Index Terms— Ocean Simulations, Light Modelling, Marine Biology, Remote Sensing, Hyperspectral Imaging

1. INTRODUCTION

Ocean observations are important to learn more about hydrodynamic and ecological phenomena in the ocean. Imaging the oceans from satellite is commonly done to map physical variables like sea surface temperature, and also ecological phenomena, for example chlorophyll-a (chla). Hyperspectral ocean data will be more often collected in the future with upcoming, hyperspectral imagers on small satellites about 500-800 km above sea-level, e.g. [1, 2]. These smaller satellites can typically be controlled to track selected target areas, which allows for a lower revisit time. Moreover, these small satellites allow for on-demand imaging of locations where there is a sudden event such as a harmful algae bloom. Hyperspectral data resulting from pushbroom imaging typically consists of a datacube of information that contains a 3D dataset with two spatial axes (cross-track, along-track) and one spectral axis, see the illustration in Fig. 1. The spectral axis contains the continuous light spectrum.

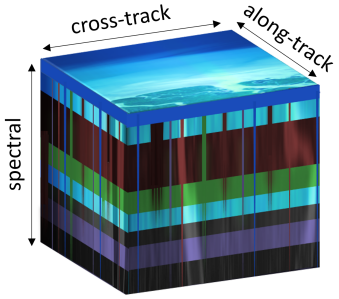


Fig. 1: Hyperspectral datacube illustration.

The light that the hyperspectral imagers capture is the light that leaves the oceans after interacting with the water. These spectra contain information about what is inside the top layer of the water column. Important terms that are often used to describe these spectra are scalar irradiance and remote sensing reflectance (R_{rs}). The scalar irradiance is used to measure the photon flux and is the equally weighted irradiance from all directions on one point, this term is often used because phytoplankton are randomly oriented in measurements of coastal areas. The R_{rs} is the surface reflectance per steradian. The surface reflectance tells how much of the incident light is reflected back. The spectral absorption is then how much light was absorbed by the watercolumn as the light passes through it, where the absorption differs for each wavelength depending on the water constituents. The light that backscatters is mostly caused by scattering, and not due to the absorption of the water column. Therefore, backscatter-

This research was funded by the research council of Norway and NTNU through the HYPSCI project number 325961. The simulations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway

Table 1: Summary of abbreviated model variables and parameters.

Symbol	Type	Unit/Value	Description
R_{rs}	Output from Ecolight-S	sr^{-1}	Remote sensing reflectance
P_d	Variable	mg/m^3	Diatom concentration
P_f	Variable	mg/m^3	Flagellate concentration
r_f	Parameter as assumed by SINMOD	7.6	Redfield ratio
Ac_d	Variable	m^{-1}	Diatoms in-vivo chla specific absorption coefficient
Ac_f	Variable	m^{-1}	Flagellates in-vivo chla specific absorption coefficient
As_d	Parameter from [3]	$m^2(mg\ chla)^{-1}$	Diatoms in-vivo chla specific absorption spectrum
As_f	Parameter from [3]	$m^2(mg\ chla)^{-1}$	Flagellates in-vivo chla specific absorption spectrum
$chla_d$	Parameter as assumed by SINMOD	$0.03\ mg/m^3$	Estimated chla concentration in diatoms
$chla_f$	Parameter as assumed by SINMOD	$0.012\ mg/m^3$	Estimated chla concentration in flagellates
$chla_{d_t}$	Variable	mg/m^3	Total chla concentration in diatoms in one layer
$chla_{f_t}$	Variable	mg/m^3	Total chla concentration in flagellates in one layer
Ac_t	Variable	m^{-1}	Total absorption coefficient
σ	Parameter	1	Exponent for calculating of Ac_d from chla

ing usually changes more slowly with wavelength than absorption. The spectral absorption and backscattering coefficients of phytoplankton are species specific, and are necessary to use in light modelling as inherent optical properties ($IOPs$). With these inputs, the radiative transfer equations can be solved, for example by using a model set-up of Ecolight-S [4] or SINMOD light [5].

In [6] hyperspectral differentiation of phytoplankton taxonomic groups is performed, where using the R_{rs} is compared to using absorption spectra for distinguishing phytoplankton types. The absorption spectra are more thrustworthy, however, the study indicates that there is a potential for using the R_{rs} to distinguish different phytoplankton types. The R_{rs} is also modelled in for example [7], where remote sensing of zooplankton is performed, mainly by finding chla peaks in satellite data and combining this with in-situ data. In this study the R_{rs} of the area was modelled, where the IOP values were measured in-situ, however, the modeled R_{rs} still deviated from the satellite R_{rs}, in some regions of the spectrum. The authors assume that this is because the IOP values were not precise enough and only achieved from small volumes, showing how important setting correctly achieved IOP values in modelling approaches is for realistic modelling. In [8] Empirical Orthogonal Functions (EOF) were applied to R_{rs} data, to be able to predict which species would be present in R_{rs} data. The limitation of this technique is that the pigments only can be predicted if they were in every collocated samples. In [9], a model is presented to go from R_{rs} to phytoplankton pigment composition. In [9] the same limitation occurs that the model is only limited to the dataset that it was developed for. These papers made use of the fact that each water constituent has its unique spectral signature. In theory, the hyperspectral ocean data should be useful as a data-source to distinguish between water constituents. However, to the best of our knowledge, there is no trustworthy processing method of hyperspectral data found that is able to distinguish different types of algae applied to different datasets. The problem is that there can be different concentrations and types of colored dissolved organic matter (cDOM) and algae in the ocean water columns, including vertical variations, whereas the hyperspectral image capture by satellites only provides one spectrum coming from each water column composition.

One way to understand the composition of hyperspectral datacubes captured with the small satellites, is by taking in-situ samples. Although sampling is a necessity for validation of findings, doing so over the full area that can be covered by hyperspectral satellite images is a tedious task, and only provides small, labeled datasets that do not cover each water column imaged by satellites. Therefore, it is important to set-up missions where in-situ data and hyperspectral data are collected simultaneously, as for example in [10, 11, 12]. The question is how to combine these different datasets of different spatial and temporal resolutions, because in both [8, 9] the differences in spatial and temporal resolutions of in-situ data gathering and remote sensing observations make it difficult to overcome the limitations. Because of the low spatial and temporal coverage of in-situ data it is difficult to get large, labeled datasets.

One possible way to exploit the in-situ datasets and satellites observations is by using ocean simulation models. Ocean simulation models are used to model physical and ecological variables, as for example in the Sintef Ocean model (SINMOD) set-up [13, 14]. In recent works, as for example in [15] it is shown that improvements are made to make modelling outputs even more realistic by implementing in-situ and satellite measurements with Kalman filtering in the model. This makes numerical modelling an interesting candidate to provide for a exteneded dataset.

In this paper we study how two of the outputs, the diatom and flagellate concentrations over the water columns, of these ocean models can be used as an input for ocean light models, such as Ecolight-S [4] or SINMOD light [5]. This way, we can create labeled ocean color datasets. This means that we can make a large number of datasets with different

ocean constituents, and the corresponding, expected hyperspectral datacube that belongs to these. In future work we can use this model to set up lots of different variations in water constituents and train e.g. neural networks to predict the relationship between the hyperspectral datacube data and the water column that includes different concentrations and types of phytoplankton. The purpose of this study is to see if we can recognize and distinguish between different algae groups, diatoms and flagellates, from their pigment groups, by using their remote sensing reflectance.

First, we are setting up a simplified model chain of ocean simulations with SINMOD and light modelling with Ecolight-S, which will be briefly described in section 2. Then, the results of modelling diatom and flagellates in this set-up are shown and discussed in section 3, where it can be seen in this simplified set-up that the R_{rs} differs for diatoms and flagellates, and finally the set-up is also done for a part of the Norwegian coastal area that is regularly imaged with HYPSO-1 hyperspectral satellite. In section 4 the conclusions and future work are described.

2. METHODS

2.1. SINMOD for Ocean Simulations

SINMOD is a numerical ocean model that can be used to perform spatio-temporal simulations of physical and biological properties of the ocean. The model is run in a nested setup where the largest scale grid covers the Greenland, Iceland and Norwegian seas with a horizontal resolution of 20 km. The top level domain provides boundary conditions for a domain with 4 km resolution covering the Norwegian Sea, Barents Sea and a part of the Arctic. More details about the physical set-up of the model can be found in [14] and details about the ecological model used for the Norwegian sea set up is described in [13].

The model set-up used in this paper covers the coastal line of mid-Norway, including the area of Frohavet and Trondheimsfjorden. The model has a horizontal spatial resolution of 800 meters and a temporal resolution of 120 seconds, in the current set-up the storage interval for biology was 24 hours. The set up is run for the period in the spring of 2022, for in total 118 days. The outputs from SINMOD include hydrodynamical and biological values. For this work we specifically focus on the output of diatoms and flagellates concentrations, and to use these as input concentrations for Ecolight-S.

An example output of the diatom and flagellate concentrations in the top surface, at one moment in time in the middle of Norway set-up can be seen in Fig. 2.

2.2. Ecolight-S for light modelling

Ecolight-S is a model that is developed to to solve radiative transfer problems in systems that include physical and biological effects [4]. The model is set-up as a subroutine, and can therefore be easily combined and run with separate user

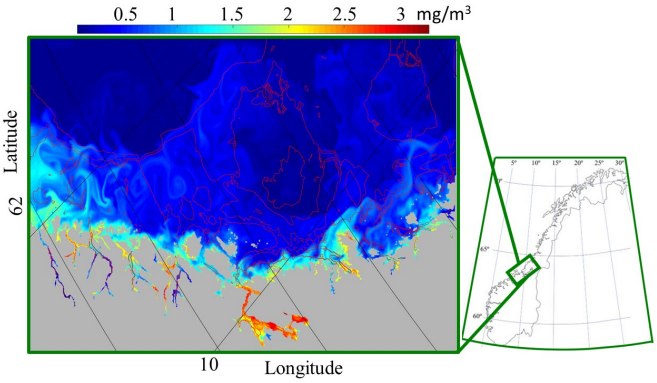


Fig. 2: The diatom and flagellate concentrations in the coast of Norway from SINMOD. The red contours refer to the depth contours, and the gray areas are land. The map of Norway is just for geographical context and not the model domain it is nested from

codes. Ecolight-S is mainly focused on fast run times by instead of solving the radiative transfer equations for every layer depth, the user can set a parameter to a value F_0 that determines that the radiative transfer equation will be solved up to the depth where the irradiance has decreased to F_0 of its surface value. The further values of the scalar irradiances down and up are extrapolated from this dataset. In the set-up that was used in this model, the value of F_0 is set to 0.1.

In Ecolight-S the bottom reflectance properties should be chosen depending on the type of bottom. In our model set-up we chose to set the bottom boundary condition for infinitely deep water, therefore we did not need to define bottom reflection properties.

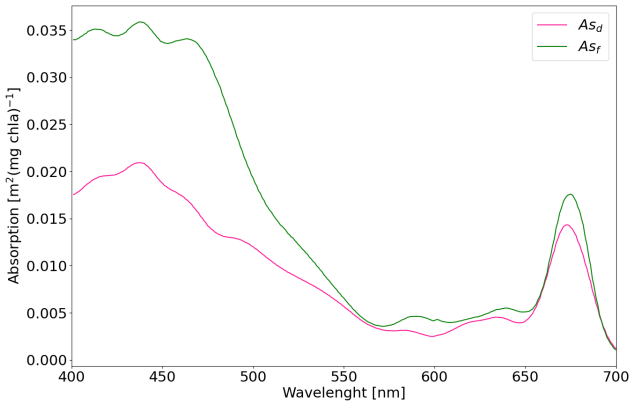


Fig. 3: The in-vivo spectral chla specific absorption spectra of diatoms and flagellates from [3].

For the watercolumn depths we used the same 15 layers and layer depths for all inputs and model runs, including while implementing the previously described concentrations of diatoms and flagellates coming from SINMOD. When the layer depths of SINMOD were containing less than these lay-

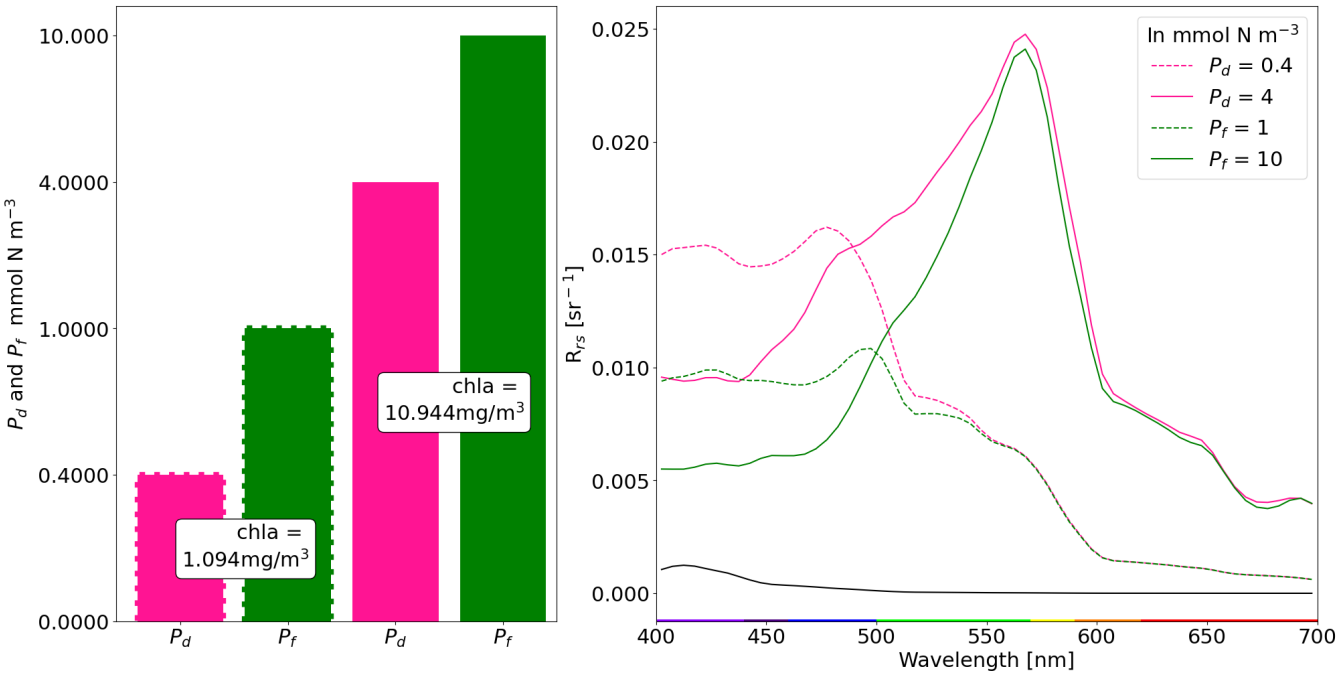


Fig. 4: The chla concentration of 10.094 mg/m^3 (1.094 mg/m^3) is set up by using only diatoms in water (P_d) in pink (dots). The R_{rs} is shown right in pink (dotted) lines. Similarly, only P_f in water was given as an input, in green (dotted) lines. The black line indicates the R_{rs} of only pure water (without P_f and P_d).

ers, the concentrations would be set to 0. In a model set-up in SINMOD the bathymetry and layer depths belonging to the water columns are known, and in a future set-up of this model combination this bathymetry should be used correctly in the Ecolight-S simulations.

The diatoms and flagellates that we want to model in the watercolumn are described with their individual IOP values. In the next section it is described how the $IOPs$ are determined.

2.3. IOP values

The Ecolight-S set-up needs as an input the IOP values. The IOP of interest in this specific set up is the absorption coefficients specific for flagellates and the absorption coefficients for diatoms. First the concentration of diatoms and flagellates (P_d and P_f, see table 1 for the most used abbreviated model variables and parameters) over the depth is given as an input. These concentrations can be generated or coming from SINMOD. As SINMOD in this simulations runs a non-spectral light model, no absorption spectra are actually used in SINMOD. Therefore, as a second input the absorption spectrum of diatoms As_f and flagellates As_f is used. Different options are available for using absorption spectra, as for example shown in [16] for different functional groups. Recenty in [17], a model is developed for calculating IOP's, for different species, which could be an interesting option when setting up model runs with a mixture of more species than the diatoms and flagellates. For this work we had access to use the mean in-vivo chla-specific absorption spectrum of 6 diatoms species, and of 8 bloom forming prymnophycean species (prymno) as determined by [3], which is a verified and extensive data set. The spectra are normalized for 1 mg/m^3 chla, and more information about these absorption spectra can be found in [3]. The mean spectra can be seen in Fig. 3. The data is available with a spectral resolution of 1 nm, and for using the spectral resolution of 0.5 nm the As_f and As_f coefficients are interpolated according to the wavelength absorption spectra.

Based on the SINMOD states and the model's assumptions, it is estimated that the chla concentration in diatoms is $chla_d = 0.03$ mg/m^3 and in flagellates the chla concentration is $chla_f = 0.012$ mg/m^3. To determine the total chla concentration in the flagellates, this chla concentration estimation is multiplied by the atomic weight of carbon (12) and the Redfield ratio (see table 1). The $chla_d$, $chla_f$, and r_f are assumption of the SINMOD model. The total chla concentration ($chla_{d_t}$) is estimated as follows for diatoms;

$$chla_{d_t} = P_d \cdot chla_d \cdot 12 \cdot r_f. \quad (1)$$

The absorption coefficient for the diatoms (Ac_d) and flagellates are calculated by multiplying (for each wavelength) the absorption spectrum by the chla concentration,

$$Ac_d = As_d \cdot chla_{d_t}^{\sigma}. \quad (2)$$

In this set-up the same simplification is made as in [5], where the value of 1 has been used for σ. For flagellates; the

same result can be found when substituting the index d with f in equations (1) and (2).

Finally, for each layer and wavelength the total absorption coefficients are determined by adding up all different contributions:

$$Ac_t = Ac_w + Ac_f + Ac_d + Ac_{cdom} \tag{3}$$

where Ac_w is the absorption coefficient for pure water and Ac_{cdom} is the absorption coefficient of cDOM.

3. RESULTS AND DISCUSSION

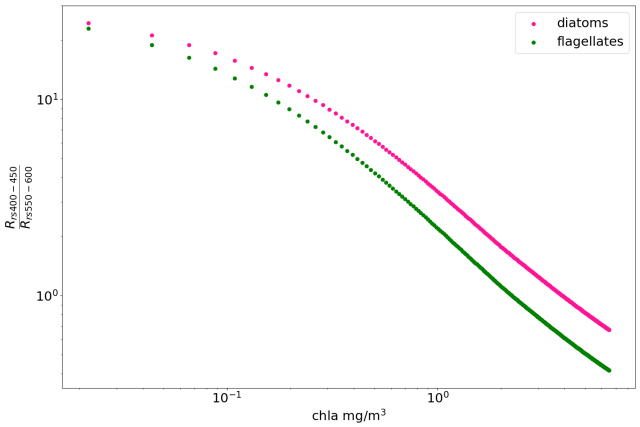


Fig. 5: For different chla levels the mean value of R_{rs} in the region of 400-450 nm is divided by the mean value of R_{rs} at 550-600 nm. This is done for the set-up that only contains diatoms (pink) and the set-up that only contains flagellates (green).

3.1. Varying diatom and flagellate concentrations

The Ecolight-S model was run for different set-ups of varying diatom and flagellate concentrations. First, two datasets are generated for different diatom concentrations. The P_d is chosen such that the chla content, which is calculated accordingly with equation (1), has one low (1.094 mg/m^3) and one higher value (10.944 mg/m^3). Then, *mutatis mutandis* two datasets are generated for different flagellate concentrations, using the same chla concentration values. The values of P_d and P_f, and their resulting chla densities can be seen in Fig. 4. Each dataset is given as input values to Ecolight-S, and the resulting R_{rs} can be seen in the right side of Fig. 4. In Fig. 4 the pink graphs represent diatom results, and green flagellate results. The dotted results are for chla densities of 1.094 mg/m^3, and the solid values represent chla densities of 10.944 mg/m^3. From Fig. 4, we can see a difference in the R_{rs} for the P_d and P_f, even though the amount of chla in the set-up is the same. The shape of the R_{rs}, changes significantly for different concentrations of the phytoplankton, making it complex to find a method that determines whether

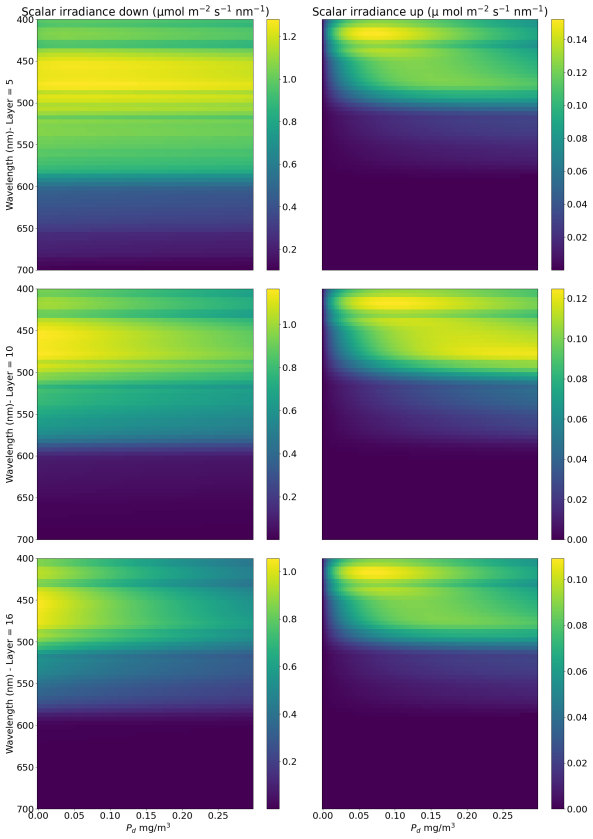
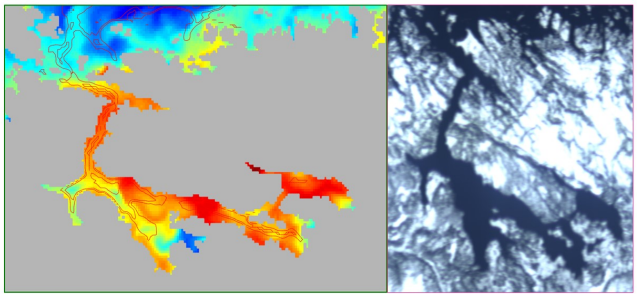


Fig. 6: The scalar irradiance down and up for an increasing P_d, while the P_f and P_{cdom} remain zero.

the R_{rs} corresponds to diatoms or flagellates. However, it seems that the R_{rs} over the wavelengths 440-450 nm for diatoms is larger than for flagellates for low and high concentrations of chla. In the region of wavelengths of 550-600 nm, the R_{rs} seems slightly higher for diatoms, but mostly similar for both diatoms and flagellates at low and high chla. In Fig. 5, the ratio of the average R_{rs} over the wavelengths 440-450 nm ($R_{rs400-450}$) and the average R_{rs} over the wavelengths of 550-600 nm ($R_{rs550-600}$), so ($\frac{R_{rs400-450}}{R_{rs550-600}}$), is shown.

For lower P_d and P_f (so lower chla concentrations), it is harder to distinguish the different species based on their R_{rs} ratios. This is as expected, as for lower concentrations the contribution of the phytoplankton to the R_{rs}, is naturally smaller. However, for higher concentrations, it can be seen that $\frac{R_{rs400-450}}{R_{rs550-600}}$ is consistently larger for diatoms than for flagellates.

From comparing the R_{rs} to the absorption spectra of the phytoplankton from Fig. 3, it can be seen that higher absorption peaks of flagellates in 400-500 nm of the spectrum, has as a consequence that the R_{rs} peaks in this part of the spectrum is lower than the diatom R_{rs}, this holds for the low and high chla concentrations. For both the diatoms and the flagellates there is a small effect visible of the second peak at 675 nm in



(a) (b)

Fig. 7: The diatom and flagellate concentrations Trondheimfjord from a SINMOD model run simulating 04-04-2022 at time 00:00:00 in (a). For the geographical location and the color scaling please see Fig. 2. (b) Shows a cropped RGB render of a hyperspectral image taken of Trondheimfjord with the HYPSO-1 satellite on 27-03-2023.

the absorption spectrum in Fig. 3, as it can be seen that for this absorption peak the reflectance has a small dip when for the high chla values. However, for lower chla densities (the dotted lines), this effect cannot be observed in Fig. 4, this is because in the wavelength region of 600-700 nm there is a high absorbance of pure water.

When we look into more detail on the effects of the diatom and flagellate concentrations on how the light propagates in the water column, a dataset is generated with increasing diatom concentrations, while the flagellate and CDOM concentrations are kept constant at zero. The scalar irradiance down and up can be seen in Fig. 6, for three different layers of depth. It can be seen that the model behaves as expected, the scalar irradiance up is low for higher concentrations, and decreases for lower depths. The scalar irradiance down is not decreasing as fast for lower depths as the scalar irradiance up, and is higher for lower concentrations of diatoms. Similar effects are found for setting up the model with only flagellates.

Because of the high absorbance of pure water in the wavelength region of 600-700 nm it can be seen in the top left plot of Fig. 6 that for the fifth layer the scalar irradiance down from 600 nm and higher is only a third or less of the scalar irradiance in the region of 450 nm. For different scaling of the scalar irradiance values over the spectrum it could be tried to set other σ values. As explained in section 2 and shown in table 1, the value is now equal for all wavelengths and has value one. By setting this in a more accurate way, this could potentially be improved. For example, if one would choose the same wavelength spectrum of values as in [18], the spectra values would be relatively higher in the 600-700 nm region.

3.2. Simulating a hyperspectral datacube

For the region of the middle of Norway the SINMOD model was run with the set-up described in section 2.1. The region

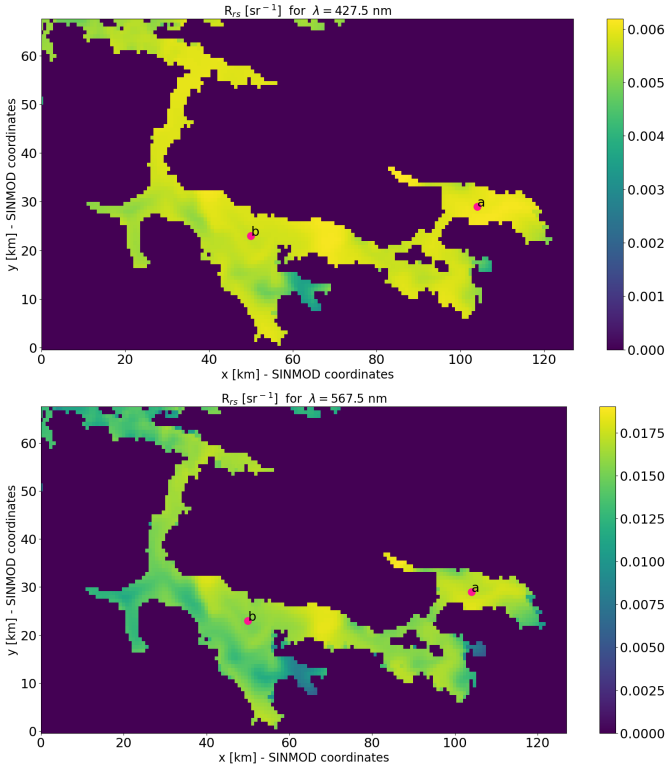


Fig. 8: The R_{rs} distribution for λ = 427.5 nm (top) and λ = 567.5 nm (bottom) over the area of Trondheimsfjorden. a and b denote the locations of which the R_{rs} is plotted in Fig. 9. The data is generated from a SINMOD model run simulation at timestep 04-04-2022 at 00:00:00.

of Trondheimsfjorden is regularly imaged with the HYPSO-1 satellite, and an example of a RGB render of a HYPSO-1 image can be seen in Fig. 7b, along with a detailed map of the flagellate and diatom concentrations in the fjord estimated by SINMOD in Fig. 7a. The P_d and P_f of this area were used as an input in the Ecolight-S model, and for each cell the R_{rs} is determined. For two example locations the P_d and P_f, and their resulting chla concentrations are shown respectively in the two left plots of Fig. 9. The R_{rs} is shown in the right plot in Fig. 9. For two different wavelengths (λ = 427.5 nm and λ = 576.5 nm) the complete map of every cell can be seen in Fig. 8.

In Fig. 8, it can be seen that for the whole area of Trondheimfjord the R_{rs} at 427.5 nm is lower than at 567.5 nm, which is also expected from R_{rs} simulations in two different model set ups shown in Fig. 4. For the more shallow areas, for example nearby land, it can be seen that a higher R_{rs} is not related to high diatom and flagellate concentrations as can be seen in Fig. 7. This can be explained by the approximation that was made that in Ecolight-S, where for now the depths are all considered constant, and therefore, more shallow areas have instead of a earlier bottom reflection now just as a flagellate and diatom concentration 0.

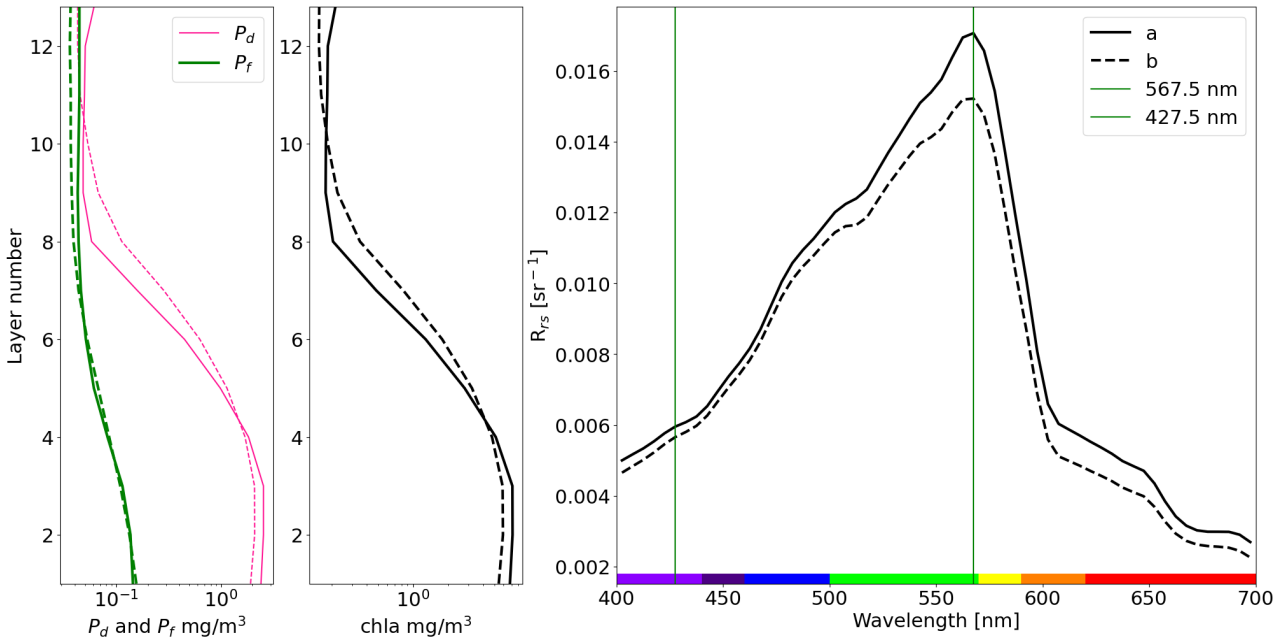


Fig. 9: The R_{rs} of two different input cells of SINMOD. The lines show cell a of Fig. 8, and the dotted lines cell b. The vertical lines represent the wavelengths that were mapped in Fig. 8. The two plots on the left denote the P_d and P_f, and their resulting chla concentrations from cell a and b, for every layer. In this plot layer 0 is the surface layer.

4. CONCLUSION

The R_{rs} resulting from model set-ups only containing diatoms, can be distinguished from model set-ups only containing flagellates, while each set-up has the same amount of total chla. This is shown by taking ratio of the average R_{rs} over the wavelengths 440-450 nm ($R_{rs400-450}$) and the average R_{rs} over the wavelengths of 550-600 nm ($R_{rs550-600}$), so ($\frac{R_{rs400-450}}{R_{rs550-600}}$). It was found that this ratio is consistently larger for diatoms than for flagellates over a wide range of tested chla levels. In future work algorithms should be tested in determining ratio's of flagellates and diatoms based on the R_{rs} in mixed water columns.

We have shown that we can use the diatom and flagellates concentrations in the watercolumn, coming from numerical ocean simulation modelling (using SINMOD) as an input for light interaction modelling (using Ecolight-S). This way we can create a hyperspectral data based on realistic ocean model simulations. One application of this tool is determining whether hyperspectral imaging allows the detection of harmful algal blooms.

In future research, the influence of atmospheric effects on the spectral signature should be implemented in the model. For example, by performing data-assimilation with in-situ data, and comparing this to hyperspectral images taken with the satellite (at the same time as the in-situ samples are collected). Then, it should be studied how effective the modelled, hyperspectral datacubes are as labeled datasets to train machine learning algorithms in labeling.

5. REFERENCES

[1] Sivert Bakken, Marie B Henriksen, Roger Birkeland, Dennis D Langer, Adriënne E Oudijk, Simen Berg, Yeshi Pursley, Joseph L Garrett, Fredrik Gran-Jansen, Evelyn Honoré-Livermore, et al., "Hypso-1 cubesat: first images and in-orbit characterization," *Remote Sensing*, vol. 15, no. 3, pp. 755, 2023.

[2] Darryl J Keith, Blake A Schaeffer, Ross S Lunetta, Richard W Gould Jr, Kenneth Rocha, and Donald J Cobb, "Remote sensing of selected water-quality indicators with the hyperspectral imager for the coastal ocean (HICO) sensor," *International Journal of Remote Sensing*, vol. 35, no. 9, pp. 2927–2962, 2014.

[3] Geir Johnsen and Egil Sakshaug, "Biooptical characteristics of psii and psi in 33 species (13 pigment groups) of marine phytoplankton, and the relevance for pulse-amplitude-modulated and fast-repetition-rate fluorometry 1," *Journal of Phycology*, vol. 43, no. 6, pp. 1236–1251, 2007.

[4] Curtis D Mobley, "Fast light calculations for ocean ecosystem and inverse models," *Optics express*, vol. 19, no. 20, pp. 18927–18944, 2011.

[5] Morten O Alver, Kasper Hancke, Egil Sakshaug, and Dag Slagstad, "A spectrally-resolved light propagation model for aquatic systems: Steps toward parameterizing primary production," *Journal of Marine Systems*, vol. 130, pp. 134–146, 2014.

[6] Hongyan Xi, Martin Hieronymi, Rüdiger Röttgers, Hajo Krasemann, and Zhongfeng Qiu, "Hyperspectral differentiation of phytoplankton taxonomic groups: A comparison between using remote sensing reflectance and absorption spectra," *Remote Sensing*, vol. 7, no. 11, pp. 14781–14805, 2015.

[7] Sünnje L Basedow, David McKee, Ina Lefering, Astthor Gislason, Malin Daase, Emilia Trudnowska, Einar Skarstad Egeland, Marvin Choquet, and Stig Falk-Petersen, "Remote sensing of zooplankton swarms," *Scientific reports*, vol. 9, no. 1, pp. 686, 2019.

[8] Astrid Bracher, MH Taylor, Bettina Taylor, Tilman Dinter, Ruediger Roettgers, and Francois Steinmetz, "Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations," *Ocean Science*, vol. 11, no. 1, pp. 139–158, 2015.

[9] Sasha J Kramer, David A Siegel, Stéphane Maritorena, and Dylan Catlett, "Modeling surface ocean phytoplankton pigments from hyperspectral remote sensing reflectance on global scales," *Remote Sensing of Environment*, vol. 270, pp. 112879, 2022.

[10] Sina Keller, Philipp M Maier, Felix M Riese, Stefan Norra, Andreas Holbach, Nicolas Börsig, Andre Wilhelms, Christian Moldaenke, André Zaake, and Stefan Hinz, "Hyperspectral data and machine learning for estimating cdom, chlorophyll a, diatoms, green algae and turbidity," *International journal of environmental research and public health*, vol. 15, no. 9, pp. 1881, 2018.

[11] A. E. Oudijk, O. Hasler, H. Øveraas, S. Marty, D. Williamson, T. Svendsen, S. Berg, R. Birkeland, D. Halvorsen, S. Bakken, M. B. Henriksen, M. Alver, G. Johnsen, T. A. Johansen, A. Stahl, P. Kvaløy, A. Dallolio, S. Majaneva, G. Fragoso, and J. Garrett, "Campaign for hyperspectral data validation in north atlantic coastal waters," in *12th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Rome*, 2022.

[12] David Roddan Williamson, Glaucia Moreira Fragoso, Sanna Kristiina Majaneva, Alberto Dallolio, Daniel Ørnes Halvorsen, Oliver Kevin Hasler, Adriënne Esmeralda Oudijk, Dennis David Langer, Tor Arne Johansen, Geir Johnsen, et al., "Monitoring algal blooms with complementary sensors on multiple spatial and temporal scales," *Oceanography*, vol. 36, 2023.

[13] P. Wassmann, D. Slagstad, C.W. Riser, and M. Reigstad, "Modelling the ecosystem dynamics of the barents sea including the marginal ice zone: Ii. Carbon flux and interannual variability," *Journal of Marine Systems*, vol. 59, no. 1-2, pp. 1–24, 2006.

[14] D. Slagstad and T.A. McClimans, "Modeling the ecosystem dynamics of the barents sea including the marginal ice zone: I. Physical and chemical oceanography," *Journal of Marine Systems*, vol. 58, no. 1-2, pp. 1–18, 2005.

[15] Daniel Ørnes Halvorsen, Alberto Dallolio, and Morten Omholt Alver, "Assimilation of heterogeneous measurements at different spatial scales in the arctic ocean and norwegian sea," in *OCEANS 2022, Hampton Roads*. IEEE, 2022, pp. 1–10.

[16] Shubha Sathyendranath and Trevor Platt, "Spectral effects in bio-optical control on the ocean system," *Oceanologia*, vol. 49, no. 1, 2007.

[17] Shun Bi, Martin Hieronymi, and Rüdiger Röttgers, "Bio-geo-optical modelling of natural waters," *Frontiers in Marine Science*, vol. 10, pp. 1196352, 2023.

[18] Annick Bricaud, André Morel, Marcel Babin, Karima Allali, and Hervé Claustre, "Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models," *Journal of Geophysical Research: Oceans*, vol. 103, no. C13, pp. 31033–31044, 1998.