Reviewer #2

answers in ***bold italic***

The manuscript by Alvera-Azcárate et al. explores the use of the DINEOF technique on satellite ocean colour data gathered from a coastal area between the North Sea and the English Channel. This technique was applied to achieve a gap-free, interpolated product with enhanced spatial resolution through the combination of Sentinel-2 and Sentinel-3 data. The topic is highly relevant, and the paper could contribute significantly to the field of multi-resolution products. The manuscript is well-organized, reasonably clear, and the quality of the English is good.

However, I have a few suggestions and comments that could improve the clarity and readability of the paper, address some issues/inconsistencies, and should be considered before publication:

Section 2.2: It is unclear what the satellite resolutions are and whether Sentinel-3A and Sentinel-3B (or Sentinel-2A and 2B) are treated as merged or separate sensors.

Lines 101–104: A brief summary of the switching method employed would be helpful.

We have added the following paragraph to provide an overview of the approach:

"To combine the two approaches, a comprehensive, region-independent, and pixel-based automatic switching scheme is required, along with a technique for achieving a seamless transition between the two algorithms. The C2RCC to ACOLITE/DSF pixel-based switching is performed by means of band comparison of the RRS560 and RRS865 products (defined as green-nir ratio) as provided by the C2RCC processor. The green-nir ratio can be modelled using a logarithmic regression curve which starts as linear for the smaller reflectance values, but bends at the point where the saturation of the most sensitive band (i.e. RRS560) occurs. C2RCC pixels which deviate from the logarithmic model are considered erroneous outputs. The ACOLITE/DSF processor has the ability to provide higher RRS ranges compared to C2RCC while being noisier for lower RRS values thus highlighting the complementary between the two approaches. The green-nir ratio value of 45 is selected as the transition point between C2RCC and ACOLITE/DSF products. To ensure a smooth transition between the different ACs, a weighted transition is applied between the green-nir ratio boundaries of 50 and 40 based on the method described by Novoa et al. (2017). The C2RCC to ACOLITE/DSF pixel-based switching is described in detail in Van der Zande et al. (2023). Compatibility between the Sentinel-2/MSI products..."

Figure 2 caption: Replace "atmospheric correct algorithms" to "atmospheric correction algorithms."

Done

Section 2.3: Why are multi-sensor satellite data described in a different section from "satellite data"? Aren't they also satellite data? Additionally, at this point in the manuscript, the purpose of the multi-sensor data is unclear, and it is not helpful to have to jump between different pages and sections to understand how these data are being used in the study. I recommend merging this section as a subsection of 2.2 (like the other satellite products) and explaining earlier in the manuscript how these data support the study's goals.

Thank you for pointing this out, which is indeed a misplaced subsection. It is now section 2.2.3, placed now within the Satellite data section. We have started section 2.2.3 explaining the aim of this dataset:

"In order to assess the small-scale information retained in the final DINEOF reconstructions, an additional test on a larger region has been done, using data at 1 km resolution. The aim is to create a degraded 5 km resolution dataset from the initial data and to compare the DINEOF results to the initial, non-degraded dataset. This scale assessment is described in section..."

4.3. Daily chlorophyll data at a spatial

Please provide a link or reference for the product "cmems_obs-oc_atl_bgc-plankton_my_l3-multi-1km_P1D" mentioned in the paper. This dataset typically covers from September 1997 until 8–10 days before the present day. The authors should clarify why only the data from 1 February 2022 to 1 November 2022 were chosen and why this particular period was selected.

We have added a link to the DOI of the dataset and an explanation of the choice of period in section 2.2.3:

"We have extracted data from 1 February 2022 to 1 November 2022 in order to have a long time series of data, avoiding January and December which have low-light conditions and prevent the calculation of ocean colour variables at the higher latitudes of the domain. The choice of the year was simply to avoid 2020 which is used in the other tests."

We have clarified the text and added a link to the data.

Lines 149–150: "The matchup validation protocol described by Bailey and Werdell (2006) was applied to remove erroneous matchups from the analysis." A brief summary of the criteria used to remove erroneous matchups would be useful.

We have added the following summary:

The matchup validation protocol described by Bailey and Werdell (2006) was applied to remove erroneous matchups from the analysis. Macro-pixels of 15x15 60m pixels for Sentinel-2/MSI and 3x3 300m for Sentinel-3/OLCI were extracted from the L2 products. This box allows for the evaluation of spatial stability, or homogeneity, at the validation point. For the satellite data it was required that at least 60% of the pixels in the defined box be valid (i.e. unflagged) to ensure statistical confidence in the mean values retrieved. The arithmetic mean and standard deviation of the non-masked pixels was determined enabling the computation of the coefficient of variation (standard deviation divided by the filtered mean,. Satellite retrievals with extreme variation between pixels in the defined box (CV > 0.15) were excluded from the matchup analysis.

Lines 150–151: "Macro-pixels of 3x3 60m pixels for Sentinel-2/MSI and 3x3 300m for Sentinel-3/OLCI were extracted from the L2 products." Given the different resolutions of these sensors, using the same 3x3 macro-pixel extraction leads to different spatial coverage, affecting spatial variability in macro-pixel extraction, especially in coastal zones. I recommend adjusting the number of pixels for macro-pixel extraction to account for each sensor's resolution to obtain comparable macro-area extractions for matchup analysis.

We have recalculated the Sentinel-2 matchup statistics for a 15x15 macro pixel to be comparable with the 3x3 macro pixel of Sentinel-3. The match-up metrics do not change significantly

compared to the 3x3 macro-pixel for S2 but we agree that this makes it so that the scatterplots for S2 and S3 can be compared directly.

Section 3.2: The authors mention the different spatial resolutions of the satellites, which may cause differences between the datasets. A quantitative analysis (using bias, RMSE, etc.) of these differences, when data from both Sentinel-2 and Sentinel-3 overlap, would be very useful.

The quantitative difference analysis between both data streams and the reconstructed data are done in section 4.2.

Line 197: Sentinel-2 data are available from 2017 and Sentinel-3 from 2016. Why did the authors only use data from 18 January 2020 to 17 December 2020? Why not a period such as August 2021 to March 2022 or any other period when data from both Sentinel-2 and Sentinel-3 are available?

For 2020 we had the largest unbroken timeseries of Hypernet data at the Oostende RT1 station allowing us to validate the super-resolution DINEOF product and its potential in capturing the coastal Turbidity dynamics. We have included this information in section 2.2.

What is the final output resolution? Is it 60m or 100m, as indicated in the caption of Figure 6 (for the first time in the manuscript)?

It is 60 m, which we considered to be enough for most coastal ocean applications. We have clarified this on the manuscript.

Section 4.1

line 210: Regarding Figure 4, I notice a peak in January, which is the highest point in the entire series. Why has this not been considered?

Light during mid-December to mid-January is very low in the North Sea, which does not allow to accurately derive ocean colour variables. The time series starts therefore on mid-January, and this does not allow to properly resolve the peak that appears to occur in January. We have however mentioned the January peak in the text.

Figure 6: Geographical coordinates in the maps are expressed in degrees and minutes, but in the bottom plot, latitude is in decimal degrees. Consistent units should be used across all figures (this applies to all figures where geographical coordinates are displayed). Also, the bottom plot shows latitude increasing from south to north, while in Figures 18, 19, and 20, latitude increases from north to south. Please ensure consistency in the presentation of geographical coordinates.

We have changed the axes to show degrees/minutes. On figures 18 and 19 the latitude is reversed so that it is easier to show the correspondence to the above plot. We understand this might be strange, so we have changed the x-axis to show longitude, which increases towards the right.

Section 4.2

Figures 10 and 11: It would be helpful to provide definitions and formulas for each of the metrics used.

Line 263: The authors refer to band 492, but Figure 10 shows the 490nm plot.

We have changed the text to match Figure 10

Line 267: The authors mention 179 matchup points for Sentinel-3, but Figure 11 shows matchups ranging between 168 and 179 points. Could this discrepancy be related to the criteria used for match-up quality flagging? Consider eliminating entire spectra when at least one band has quality issues.

The scatterplots have been redone removing all spectra that get flag in one or more bands

Line 270: The authors refer to bands 492 and 709, but Figure 11 shows 490nm and 704nm plots.

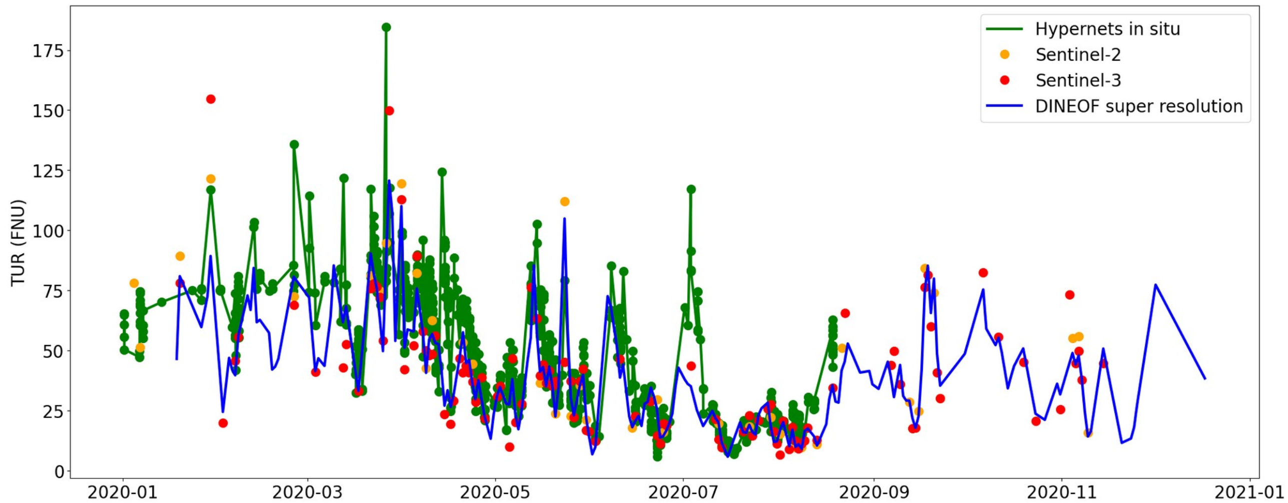
We have changed figure 11 to match the Sentinel-3 bands and we have changed the text.

Line 273: The authors discuss RMSE for Figure 12, but Figures 10 and 11 show RMSD or cRMSD. Please clarify. Additionally, the metrics in Figure 12 (slope, RMSE/RMSD, and MAPE) differ significantly from those in Figures 10 and 11. Figure 12 should present metrics in a comparable way to demonstrate consistency across the bands. Why are the metrics different?

Figure 12 indeed was using different metrics as used in figure 10 and 11. Figure 12 was removed since the other reviewer asked to reduce the figures.

Figure 13: The DINEOF (blue line) should have values for every day in 2020. The RT1 (green line) does not have daily values due to quality control or other reasons, but representing it as a continuous line makes it hard to distinguish actual data points. I suggest adding markers on the green line where RT1 measurements exist. I recommend the authors also indicate the number of valid RT1 data points in 2020. Furthermore, why does the green line only span from January to August 2020? Where are the data from September to December 2020, when RT1 data should be available from 2019 to 2023?

The Panthyr data, represented by the green line actually has multiple data points per day as it collects data every 20 minutes. This means that even after the quality control of the in situ data, it is still expected that the Panthyr RT1 data has the highest temporal frequency, especially compared to the satellite data. Please find the figure with a marker for each actual measurement which makes the plot less readable in our opinion so we suggest to leave it as it is. In total there were 980 observations available from the Panthyr system from January to August. There is no Panthyr data available from September-December as the system was dismounted from the platform for calibration reasons.



Lines 286–289: In this section, just before Figure 14, the authors repeat the technique used for the matchup analysis. However, I assume that the same method was also applied to Figure 13. If this is correct, I recommend moving this paragraph to precede the description of Figure 13. If not, please clarify the method used for the analysis in Figure 13.

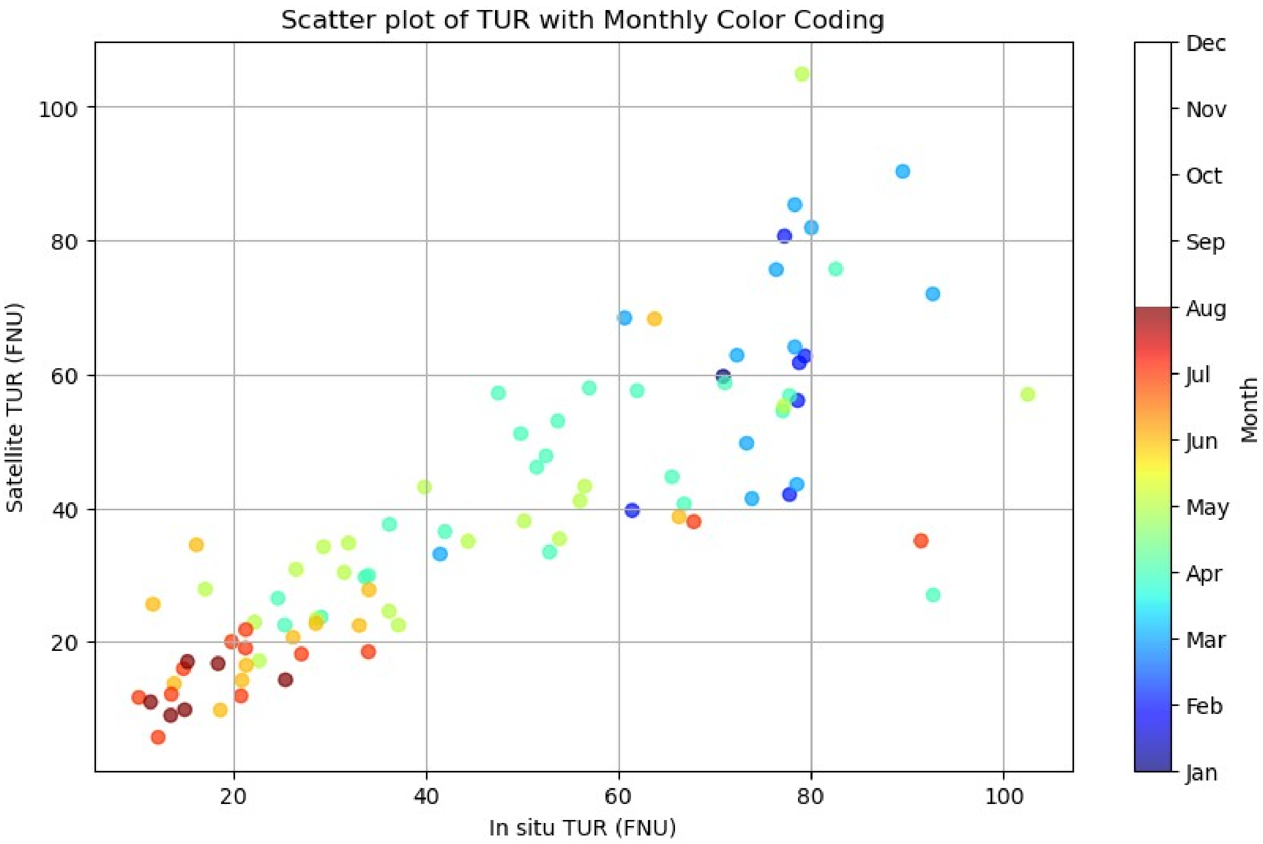
This is indeed the same method used (Bailey and Werdell, 2006). We are keeping the previous order of the text, as changing it reverses the two figures, which in our opinion makes less sense.

Figure 14: Is MAPD the same as MAPE? Also, as described in the text and shown in the figure, with DINEOF the number of matchup points increases from 67 to 90 for 2020. Therefore, can I assume that the maximum number of matchup points available from RT1 is 90? Please clarify this number of in-situ matchup points.

Yes, MAPD and MAPE are the same metric and we have changed figures 10 and 11 so that we consistently use MAPD throughout the manuscript. You are correct concerning the matchups, the 90 matchups are the result of the quality control where mainly we only allowed matchups with the in situ Panthyr data if the time difference with the satellite observations was smaller than 1 hour. While we have Panthyr data for most days, this does not mean that there is in situ data available within the allowed time difference window due to clouds.

Line 294–295: The authors, referring to Figure 14, mention an underestimation of DINEOF data for high TUR values. Could this underestimation be the same as the one observed in Figure 13 for January and February, possibly due to high cloud cover as indicated in the text a few lines above?

You are correct, when we replot the data with colours depending on the month of observation you can see that the underestimation mainly comes from the January-February period. We have mention this in the text.



Section 4.3: I assume that the plots in Figures 15, 16, and 17 are in log scale, but this isn't explicitly mentioned. Since the authors note that "the spatial distribution of chlorophyll is similar in all figures," it would be helpful to include percentage difference maps for each analyzed day between "DINEOF Super Resolution" and "DINEOF reference." Additionally, a performance analysis of "DINEOF Super Resolution" over the entire dataset adopted (February 2022–November 2022) using percentage difference maps, scatterplots, or density plots would provide a more comprehensive evaluation than focusing on just 2–3 single days. Expanding the dataset used for analysis is recommended to ensure a more robust evaluation.

The log units are mentioned in the caption of the figures. We have added a panel with percentage difference map for each of the examples. Doing a percentage difference overall, on the average for example, would not illustrate the detail that are gained through the DINEOF analysis. The evaluation is performed in the Validation section, in which all matchup data are used.

Section 5: In my opinion, the usefulness of this analysis is unclear. While the authors attempt to apply super-resolution interpolated data, this example may not be the most suitable due to high temporal variability, as acknowledged by the authors. It would be more useful if the authors could present the performance of the super-resolution interpolated data over all 210 days selected for 2020. This would help validate the positive results observed for the 2–3 days analyzed in the previous sections and demonstrate the potential utility of this technique for operational contexts or long-term application. Anyway, going into the paragraph, it is not mentioned that the data in all the figures are logarithmic. Moreover, including all the DINEOF data in the bottom panels of Figures 18 and 19 seems unnecessary, as superimposing data from other days makes it harder to interpret the trends for the specific day being analysed.

We have removed figure 20, which showed a Hövmuller diagram that did not add a lot to the discussion because the data lack the necessary high temporal resolution needed for that. The performance of the 120 days is difficult to show, as any averaging or aggregation of the data does not show the actual resolution of the data. We still believe this section shows how variable this region is and what the influence of sandbanks is on the turbidity.