# Reviewer 1

**We would like to thank the reviewer for their comments and thorough review. The main critique seems to be the absence of validation against observations. Initially that part of the work was omitted to streamline the paper and keep the focus on the novelty of the set up, and to avoid retreading work that was done in the previous study of Ford 2022. However in response to the reviews it is clear that this weakens the paper and its message, which we have corrected in the revised manuscript.**

**Specific comments:**

l.13: "features of interest, namely chlorophyll and oxygen": the later ones are variables not features. From my understanding, the feature of interest in this paper is the spring bloom associated with oxygen minimum and chlorophyll-a maximum. **Fixed**

L.27: making our observations more efficient **and** through the use of low-carbon autonomous platforms, such as ocean gliders (Testor et al. (2019)): "and" can be removed. **Fixed**

l.49: Three issues in previous study by Ford et al. are mentionned as improved in this paper but I did not find where the reduction of "biases in the observational source" is addressed in this paper. **We will add more validation into the revised manuscript**

l.140: ...from satellite the physics observations... -> ...from satellite, the physical observations... **Fixed**

l.175: Does NEMO and ERSEM shares the same spatial resolution at 1.5 km in the AMM15 configuration? **Yes, they are on the same grid. We have added 'which operates on the same numerical grid as NEMO' to clarify**

l.223: I would add for clarity some information in the sentence, even if mentioned later:
...transect based sampling **of the feature of interest** ...
disregarding **forecast** uncertainty ... **Fixed**

l.299: Does the other in-situ observations than the glider ones are also DT quality in AMM15-DT? **No, only the glider and satellite observations went through extra processing. Text added to clarify**

l.303-312: In this section the impact of glider data assimilation is discussed but the simulations that are compared to infer it are AMM15-DT and AMM15-NoG-"NRT". Those simulations differ also in processing level of the assimilated data sets, in addition to the glider assimilation. The comparison should be computed against the AMM15-NRT so the differences could solely be attributed to glider data assimilation. **The different runs are compared against the 'best possible' representation. Comparing both simulations to this still highlights the impact of glider assimilation.**

Figure 4: “depth average”: Can you specify over which range of depth the average is computed? **Changed depth to water column to clarify**

l.351: the impact of model differences on the glider: Do you mean on the glider **path**? **Yes, fixed**

l.369-372: Can you also interpret the different paths of the gliders in AMM15-NRT and AMM15-NoG by looking at the Chl-a maximum and Oxygen minimum locations in the 2 simulations? Which of the two path is the better sampling the different extrema? **Given these paths are hypothetical, by definition the path will target the extrema of the simulation it is calculated on. The NRT run should be closer to reality due to the assimilation of glider data so will better sample the extrema of reality**

l.378 - 382: Only differences between the simulation is shown. No diagnostic in the paper supports that the “inclusion of in situ glider observations improves the data assimilation products and subsequently the predictive skill of the model”. The model analysis and forecasts are not evaluated in term of realism to represent the real ocean, only between each other’s. **It is common to compare forecasts from several days ahead to the analysis solution to represent forecast predictability rather than realism, with the realism evaluated through the validation. In the absence of validation in the paper we appreciate that we weren’t clear enough in the demonstration of the benefits, which we have hopefully addressed.**

l.387: The improved realism of the simulation with the increased resolution is not shown in the paper. Only differences between the simulations are analysed. **See previous response.**

l.396: I do not find where the impact on glider path of using DT mode observations instead of NRT ones is shown. Figure 10 shows the different paths for the simulation with/without glider NRT observations assimilated. T**he comparison between DT and NRT path planning was removed to streamline the results, as the difference with just highlight the difference in simulations.**

**Technical corrections**

l.10: to address**es Fixed**

I would suggest checking the format of references in the text. For example, line 24: “… within marine autonomy Ford et al. (2022).” could be changed to ” … within marine autonomy (Ford et al., 2022).” **Fixed**

l.68: assoiated -> asso**c**iated **Fixed**

# Reviewer 2

**The authors would like to thank the review for their thorough review of the manuscript. We appreciate the lack of validation included in the paper. This decision was to try and keep the focus on the DTO system and to avoid repeating the work from the previous study of Ford 2022. However from the reviewer response it is clear that the paper would benefit from including this element of the research, which we have added in for the revision. Thank you also for highlighting the impact of the change in path planning strategy. Originally there was going to be a separate manuscript covering elements of the path planning and keeping the focus here on the modelling, but for completeness discussion for that aspect be added here as well.**

**Specific comments:**

Lines 18-21: This is correct but very generic, and we miss the scientific motivation of this. It will be useful to mention here why DTOs are needed specifically for coastal forecasting. **Will add in some coastal focus justification**

Line 26: Here you have to justify why it is essential. Could you explain why autonomy without DTO feedback is insufficient for the coastal system? **Added**

Line 26: This is a secondary aspect, and it is not related to the work. You either make it shorter or remove it. The authors should focus more on the scientific return and not on the cost efficiency. **The economical and environmental benefits are a key justification for the work, with only this single sentence covering it. The authors feel it does not detract from the scientific return.**

Line 28: This statement is too broad and risks overstating the maturity of the field. **Statement adjusted**

Line 43: The limitations have been listed but biogeochemistry is not explained, especially why the 7 km resolution model and single glider mission fail to capture chlorophyll and oxygen features. **This is covered in Section 3.3, we will adjust the introduction to reflect this.**

Lines 51-53 and 61-63: The choice of Karenia mikimotoi is scientifically justified, but the chlorophyll from the glider is only a proxy, and the species-specific detection remains unresolved, which is why the adaptive and multivariable sampling is needed. However, the author needs to clarify here if the oxygen is treated as a direct indicator for the bloom or as an independent parameter. **It is an independent parameter, this has been edited to make clearer in the revised text.**

Line 66: This sentence needs to be rephrased as "overselling outcomes," given that the bloom did not enter the study domain, as it is not guaranteed the bloom tracking, but surely indicates the system testing capability. **Text rephrased**

Line 74: It shifts the reader's focus away. **Removed**

Line 78: It will be beneficial if the authors specify what the manuscript evaluates here. **Added in**

Line 84: This statement is vague. It will be important for the authors to describe the frequency of the DTO system. Is it aligned with the real-time observations or performed daily? Furthermore, the authors should specify the quality level of the data incorporated into the model. Is quality control performed in real-time observations? **This is already mentioned in the individual sections for the different components, and we have added some more detail to the initial description.**

Line 89: “AI derived” is vague and somewhat fashionable. I would recommend the author here to use accurate terminology. **A description of the path planning model is given in Section 2.4, we have added in more detail to the general description here**

Lines: 94-95. Further information is needed for the horizontal and vertical resolution? Furthermore, at which depth range were the gliders operated during the experiment? Are shallow Slocum gliders or deep? **Added**

Line 97: The authors well stated the issue with the oxygen sensor, which is something that can happen during an experiment. However, the paper should discuss the impact of data assimilation if the oxygen variable was used as an indication of the bloom. **Oxygen was not used as an indication of the bloom, it was used to try and identify areas of hypoxia. This has been made clearer in the text**

Line 99: Did you focus on the oxygen hysteresis between downcast and upcast profiles? Please describe the handling of the oxygen data in NRT and delay mode. **More detail has been added**

Line 100: A table that describes each sensor type, frequency, model, and glider is needed for clarity. **Detail added**

Line 110: Could you provide us more details regarding the manual calibration that was performed in the glider chlorophyll observation? Is it about dark counts, correction factor, or something else? **Added**

Line 115: Could you please specify the correction methods that were performed? **Added**

Line 134: Is there any specific reason why the SLSTR SST was not assimilated, particularly given the value of these observations near the coast? **The data assimilation system we were using was already set up and configured to assimilate**

**GHRSST fields, as mentioned in section 2.2.3 as that is what is used operationally by the met office. This has been made clearer in the text**

Line 183: Could you shortly justify the use of median values rather than mean values, especially on the biogeochemical data that are skewed? **The median reduces the error in the system (See: Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model, Ford 2012). It is what is currently done operationally. Some text has been added to justify.**

Line 209: Why are 6 and 7 days key for chlorophyll-A and dissolved oxygen? Do you have the examination for forecasting the decorrelation scales? **Day 6/7 is from the start of the hindcast, starting several days before the current day. This has been made clearer**

Line 211: I assume you refer to the maximum of chlorophyll and minimum of oxygen. **Correct, this section was written to be independent of the variables being measured. The text has been made clearer**

Line 221: Please link this to the classic adaptive sampling strategy that was used before the DTOs. **Text added**

Line 260: The trajectories of gliders should be shown in Figure 1 as part of a zoomed-in map. **Added**

Line 279: How do you verify that the Karenia bloom that was detected in the Celtic Sea occurred? Please confirm if you carried out independent measurements. **The bloom was captured by satellite measurements, text has been added to clarify**

Line 281: Please indicate and report the periods that the system was fully operational. This sentence is very vague. **This has been made clearer**

Line 285: The reader until now has seen an extensive discussion of the DTO architecture and path planning, but with the scope in this manuscript, it is too late. I restructure the manuscript suggested for clarity. **Thank you for the suggestion. We have added some validation and made the story clearer**

Lines 295-301: The names of experiments, AMM15-NRT, AMM15-DT, etc., are introduced without any name convention. **We have adjusted to make the text clearer**

Lines 303-304: This is an assumption of the reference, not a validation. Please indicate on figure 3 the period of the time-average RMSD. **For reference it was the length of the mission, this has been added to the revised manuscript, along with validation against observations and making it clearer in the text what is validation and what is comparison against a reference.**

Line 315. I assume the spatial-average RMSD in Figure 4 is for the whole domain, but what about the results on the small red box where the gliders are operating? Did you

notice any influence on assimilation from the number of vertical profiles that assimilated in the red box? **It is from the reduced domain. This has been made clearer in the text. The plots show that the impact of the vertical profiles extends significantly beyond the operational area**

Line: 321-322: The gliders mentioned here target 'event states; it will be useful for the author to mention here the glider limitations regarding this. The glider is progressing ~21 km/day. The thresholds are reasonable, but what kind of criteria have you used to define them for the study area? **Text added with justification for the values used.**

Lines 331-335: It is indicated that the variances are caused only by the varying resolutions of the models, rather than any variations in configuration? In addition, it is mentioned that the difference is significant due to the assimilation, yet you do not provide a numerical value for that difference. Is the satellite chlorophyll assimilation behaving differently across resolutions? **This is demonstrated in figure 6, showing that at the time of assimilation the difference in chlorophyll is small, whereas after 3 days forecast that error has grown. We have made this clearer in the text.**

Line 336: It has been mentioned that in the bottom oxygen the difference is large, but it has been discussed how the models are so sensitive with the resolution. Are there other mechanisms that can affect such things as the timescale of ventilation? **Unlike chlorophyll, the only assimilation of oxygen come from the gliders which is not enough to reduce the differences between models of different resolutions. Given the oxygen model is identical between the runs, changes must come from the either the resolution or differences in physics which also are mostly resolution differences.**

Lines 344-345: That shows only the variability between the models and does not validate the models. What is the variability in the glider oxygen observations in the red box compared to the simulations for the same area? **Validation will be included in the revised manuscript**

Lines 362-364: In the section, the authors should compare the post-mission paths with the actual operation paths flown during the mission, making clear how assimilation alters their trajectory with what the glider should have done in practice. **This was done originally, but we felt it wasn't as interesting as the difference in paths between the different post mission runs. We will look at this again when revising.**

Lines 370-372. The authors should consider that typical glider progress is ~21 km/day and explicitly interpret path divergences of similar magnitude as effectively a one-day navigation error, with direct consequences for missed or delayed event sampling. **Thank you for the suggestion, we have the following in the text: "up to around 20km, meaning if the minimum zone**

**is a localised event the glider would be almost a full day of travel away."**

In section 4: The conclusion would benefit from a focused summary that synthesizes the main scientific findings rather than restating the framework or future potential. In particular, the authors should explicitly summarize the demonstrated impacts of multi-glider assimilation and model resolution. **We will adjust the conclusion to focus on the scientific findings**