Quick Overview: Reviewer comments and the addressed line no.s from the revised manuscript

Reviewer comments	Line no.	Comments
RC1 - 01	-	-
RC1 - 02	467	Fig F1 and Table E1 have been added.
RC1 - 03	143, 595, 533	Fig G1 has been added.
RC1 - 04	161	
RC1 - 05	169	
RC1 - 06	778	Appendix H with Fig H1 has been added.
RC1 - 07	405	Fig 5,6, and 7 have been corrected.
RC1 - 08	-	The captions of Fig 6 and 7 are updated.
RC1 - 09	504	
RC1 - 10	-	
RC2 - 01	213-224	
RC2 - 02	194	
RC2 - 03	299	
RC2 - 04	595	
RC2 - 05	19	
RC2 - 06	130	
RC2 - 07	738, 557	
RC2 - 08	545	
RC2 - 09	564	

Quick Overview: List of Figures and Tables

Original manuscript figure/table no.	Revised manuscript figure/table no.	Comments
1	1	No change.
2	2	No change.
3	3	No change in the figure, but the caption is updated.
4	4	No change.
5	5	The figure has been reproduced with the year 2010 corrected. (See RC1 07)
6	6	The figure has been reproduced with the year 2010 corrected. (See RC1 07)
7	7	The figure has been reproduced with the year 2010 corrected. (See RC1 07)
-	8	A new figure in the revised manuscript. (See RC1 02)
A1	A1	No change.
B1	B1	No change.
C1	C1	No change.
D1	D1	No change.
-	E1	A new table in the revised manuscript. (See RC1 02)
-	F1	A new figure in the revised manuscript. (See RC1 02)
-	G1	A new figure in the revised manuscript. (See RC1 03)
-	H1	A new figure in the revised manuscript. (See RC1 06)

Detailed Response to Reviewer 1

RC1 01: *The authors employ machine learning techniques in order to identify and analyze marine carbon biomes over space and time. They apply the tool using a global biogeochemistry model and identify 7 unique biomes globally. Analysis of these biomes and the drivers of each allow for conclusions about seasonal variation at regional scales and how these climatic patterns are shifting over time. This tool will be publicly accessible, providing an important resource for future research to improve analysis of ocean carbon and carbon cycle projections. The paper provides an important scientific tool, but I do have some edits recommended before publication-ready.*

AR 01: We appreciate your constructive feedback and your time in providing details on the segments that can be improved. Please find our responses below.

RC1 02: *My largest comment is that in the abstract, the authors mention observation of a 10% expansion of the subtropical biome in the North Atlantic over time, and a 10% expansion of the subpolar biome in the Southern Ocean. These are very interesting results, worthy of highlighting in the abstract, but I felt they weren't actually expounded upon enough in the results section. We have one paragraph at the end discussing it, but I was curious about putting it in context a bit more with the impacts of climate change and what these changing biomes could imply for the future. Additionally, I felt there was no real visual representation of these shifts. Is there a way to emphasize or include it more clearly in figure 6, or perhaps even mention it in the figure caption, to allow the reader to absorb this information better?*

AR 02: We agree with the Reviewer that the effect of climate change on the biome coverage should be given more attention. We took the following steps to address this comment:

- We now compute the trends in biome coverage between 1970 and 2018 over the North Atlantic Ocean and Southern Ocean for all months (not only January and July). In **Table E1**, we show the values of the linear trends and their statistical significance. The table will be included in the appendix as Table E1.
- To better visualize the trends in biome coverage over the global ocean, we added a figure in the main manuscript (**Fig. 8**) showing the spatial patterns of coverage change for five biomes between 1970-1979 and 2009-2018.
- To interpret the trends in biome coverage in the context of the changing climate, we computed SST changes between 1970-1979 and 2009-2018 in both the model and the observation-based data set EN4.2.2 (Good et al., 2013) and

plotted them for January and July (**Fig. F1**). This figure will be added to the appendix as Figure F1.

Based on the above-mentioned new analysis, we found the following:

- ➔ In the North Atlantic, the percentage coverages of the SUBTR I and SUBTR II biomes are negatively correlated and show opposing linear trends between 1970-2018 for each month of the year (**Table E1**). While the correlations are always statistically significant ($p<0.04$), the linear trends are statistically significant only between January and May. Over these months, SUBTR II (the most thermally driven biome) expanded at a rate of ~1-2% per decade, while the SUBTR I biome contracted at a similar rate. Locally, the trends in biome coverage may reach even higher values, i.e., the changes may exceed 50% between the decade 1970-1979 and the decade 2009-2018, as shown in **Fig. 8**. Besides, **Fig. 8** also shows that SUBTR II expanded in all basins in the winter hemisphere (January in the Northern Hemisphere, July in the Southern Hemisphere) at the expense of SUBTR I, which instead expanded in the summer hemisphere and the Tropics. The expansion of the SUBTR II biome could be related to a concomitant increase in SST over the North Atlantic and subtropical gyres (**Fig. F1**), in agreement with the observational data set EN4.2.2 (Good et al., 2013). This might have enhanced the thermal control of SST on fCO_2, thereby favoring the thermally-driven SUBTR II biome. We can further speculate that the expansion of SUBTR II is particularly strong in winter, which may be related to the low phytoplankton carbon uptake in these months, a characteristic that enhances the thermally-driven character of the biome.
- ➔ In the Southern Ocean, the percentage coverages of the SUBP+UP II and SUBP+UP III biomes are also negatively correlated on interannual time scales and mostly show opposing linear trends between 1970-2018 (**table E1**). The statistically-significant trends are here found in the summer months (December-March), with an expansion of the non-thermally-driven SUBP+UP III biome of 0.5-1% and a contraction of the SUBP+UP II biome. Again, locally the trends in biome coverage may reach values exceeding 50% between the decade 1970-1979 and the decade 2009-2018. The expansion of the non-thermally-driven SUBP+UP III biome might be related to the concomitant increase in Southern Hemisphere westerly winds (Swart et al. 2015), which has created more favorable conditions for DIC upwelling and therefore enhanced the non-thermal control on fCO_2 (Gruber et al. 2023). The increased upwelling caused a negative trend in SST (**Fig. 4**), which in the model is particularly strong in the austral summer months, and much less pronounced in the austral winter months. This

might explain why the expansion of SUBP+UP III was stronger in the austral summer months. However, since this winter-summer asymmetry is not visible in the observation-based EN4.2.2 data set, it remains uncertain whether the model might underestimate the SUBP+UP III expansion in winter.

[Updated text has been added at **line no—467** in the revised manuscript.]

RC1 03: *My second comment has to do with clarification of the input data: this was all done using one single model and it's output, correct? I think some supplementary discussion of the model itself's strengths and weaknesses could be included—I know, for example, some models have unrealistic mixed layer depths when compared with observations. How would something like this impact these biome patterns? Could there even be a supplementary figure comparing some of this with observations? For example, the figure 7 showing the SST, SSS, and MLD for each biome—how well does this match observations that are for roughly the same geographical region as defined by the machine learning biomes? I believe the paper could benefit from a little added discussion about how this method is employed within a model, and how that applies to future research--does it need to be regenerated with selected observations (if so, what are the base requirements for the obs) or someone's own model to usefully apply the biomes, or can they use your defined biomes explored here, and how does that affect research decisions?*

AR 03: We deconstruct this comment in three parts:

1) In reply to: *"I think some supplementary discussion of the model itself's strengths and weaknesses could be included—I know, for example, some models have unrealistic mixed layer depths when compared with observations. How would something like this impact these biome patterns? Could there even be a supplementary figure comparing some of this with observations?*"

We thank the Reviewer for pointing out the importance of a careful evaluation of the model output with respect to observations. We added a figure (**Fig. G1**) showing the comparison of climatological SST, SSS, MLD, surface pre-industrial DIC and surface alkalinity with observation-based data sets.

Based on this new analysis, we found the following:

The magnitude and spatial patterns of SST, SSS and winter MLD are reasonably simulated with respect to observational data sets (Good et al., 2013; Sallée et al., 2021). This is not surprising, since the physical ocean model is forced by observed reanalysis data and additionally contains a weak surface salinity restoring in ice-free regions (1

year time scale over 50 m depth). In comparison to the GLODAPv2 data set (Lauvset et al. 2022), pre-industrial DIC and alkalinity show reasonable spatial patterns but a shift in their mean values. This is due to the model adjustment during the 250-year spin-up, which causes biogeochemical properties to deviate from the GLODAPv2 initial conditions. We argue, however, that the bias in mean properties should not significantly affect our results since 1) spatial gradients are reasonably simulated, 2) the biomes are built on spatial relationships between fCO_2 and its drivers so that a shift in mean values of DIC and ALK likely doesn't play a substantial role.

[Updated text has been added at **line no—143** in the revised manuscript.]

2) In reply to: *"For example, the figure 7 showing the SST, SSS, and MLD for each biome—how well does this match observations that are for roughly the same geographical region as defined by the machine learning biomes?"*. The model-based biome coverage changes from month to month and from year to year. As seen from **Fig. 5**, the biomes are defined as a "probability" that a particular biome will be found at one particular location. The biomes are thus not defined as static masks over which observational data sets (which are often provided as averages over longer periods) can be averaged. There could be two ways of addressing the Reviewer comment, but both are - in our view - not optimal: i) Since SSS and SST in the EN4.2.2 data set are defined for each month and year, we could in theory subsample them in the temporally-changing model-based biomes. However, we believe this would not be an accurate calculation, since the biomes have been constructed based on the model output. ii) The best solution would be to run the detection and tracking tool directly on observational data sets. The issue here is that observational data sets do not have the spatial resolution necessary to build meaningful target-driver relationships over 2º boxes (please see next reply).

[The comment is not added in the manuscript. This comment was for answering the reviewer here.]

3) In reply to:*" …does it need to be regenerated with selected observations (if so, what are the base requirements for the obs) or someone's own model to usefully apply the biomes, or can they use your defined biomes explored here, and how does that affect research decisions"*. The tool (detection and tracking) is transferable. Still, the exact biome locations and model weights and parameters (linear regression, clustering, neural networks) are not, as they are specific to the ocean model experiments. Ideally, the tool should be run from scratch based on a specific model output or observational data set. Requirements for this include saving surface CO_2 fugacity or partial pressure, DIC, alkalinity and temperature at a resolution sufficient to resolve the seasonal cycle. It

should be noted that if the data set has a coarse spatial resolution, bigger boxes will be needed to construct the spatial target-driver relationships, with the drawback of achieving a less refined picture, of cutting through sharp current, or of averaging out different regimes.

[Updated text has been added at **line no—595** in the revised manuscript.]

This study does not provide the exact locations of static biomes, but rather - due to the temporally-varying nature of the ocean - a probability that a certain biome will be found in a certain place (**Fig. 5**). We acknowledge that it would be practical for model evaluation or intercomparison studies to define a set of static biomes over which different model diagnostics can be averaged (as done in e.g., DeVries et al., 2023). However, the usage of static biomes comes with disadvantages owing to time-varying and model-specific locations of ocean fronts (Fay and McKinley, 2014). We suggest that future research could use the detection and tracking tool proposed here across different data classes to test the spatial homogeneity of the carbon biomes as well as to better understand the specific dynamics of each model.

[Updated text has been added at **line no—533** in the revised manuscript.]

RC1 04: *Line 146+: The authors mention for both fCO2 and DIC, they use natural components rather than contemporary. How are these separated? Also, the note 'they are substantially similar when using contemporary DIC/fCO2"…does this imply that the influence of anthropogenic carbon is not impacting the biomes? I feel this could be explored with a sentence or two here*

AR 04: The natural components of DIC and fCO_2 were obtained by running the model under constant pre-industrial values of the atmospheric CO_2 mixing ratio equal to 284.32 ppm.

In our previously published work (Mohanty et al. 2023, 10.1145/3609956.3609973 , Fig. 4), we built an online tool to facilitate marine scientists to detect carbon biomes. In the tool, we provide an option to construct the biomes from anthropogenic and pre-industrial CO_2 uptake. We observed that the biomes detected in both cases look spatially almost identical, indicating that anthropogenic carbon does not significantly impact carbon biomes built upon the target-driver relationship.

[Updated text has been added at **line no—161** in the revised manuscript.]

RC1 05: *Line 151: The authors note they decided to build biomes on target-driver relationships rather than drivers themselves, because it's better for the methodology. Did they test this, or how do they know this is better?*

AR 05: We agree that the word 'better' is not a good choice here. Thank you for challenging it. The text has been rephrased to: "We decided to build biomes on target-driver relationships rather than on the drivers since we aim to capture regionally specific relationships between fCO2 and its drivers."

We have not conducted separate experiments for building regimes based only on drivers.

[Updated text has been added at **line no—169** in the revised manuscript.]

RC1 06: *Line 249+: The authors select January 2009 as the training date for the FNN. They do address the sensitivity of this month selection, and acknowledge the caveats in the discussion, which is both good and necessary. However, they don't really explain why January 2009 is chosen. What about the year 2009—it's not in the middle of the analyzed time range, in fact it's near the end. In addition, why the month of January? I think in the methods, this could be explained with more detail and justification.*

AR 06: We agree that the choice of the specific month and year used to identify the labels should be discussed more thoroughly. The short answer to why we selected January 2009 is that the choice is random. However, we now better justify why the random choice is not expected to make a difference in the biome outcome.

We included a new figure (**Fig. H1**) showing kernel density distribution of the regression coefficients (RCs) for some random years and the mean and standard deviation of RCs corresponding to each tracked biome computed over the 12 months between 1958 and 2018. The new analysis indicates that the RC kernel density distribution shows only subtle changes from year to year. The RCs within each biome (except for the strongly variable ICE biomes) are substantially stable overall in months and years. This stable behavior of the RCs over single biomes suggests that the specific month and year selected to build labels through hierarchical clustering should not affect the results. Instead, we argue that the leading subjective choice is the choice of the parameters used to cut the dendrogram resulting from the hierarchical clustering. The parameter choice yields a different degree of fragmentation or aggregation of the resulting clusters (not shown). We don't consider this a caveat because a different amount of aggregation/fragmentation may be desired depending on the user-defined application.

Once the labels are defined, however, we argue that the neural network can track the different RC combinations in each detected biome consistently.

[Updated text has been added at **line no—778 (Appendix H)** in the revised manuscript.]

RC1 07: *Line 370: "Only a couple of years were found to be inconsistent with overall pattern" while looking at the figures, those years were pretty significantly outside the expected pattern. Any theories on why that might be? What was going on in those years? How did it bounce back so quickly, with no longer-term shifts on the biomes?*

AR 07: After careful inspection, we realized that we had made a mistake in handling the year 2010 and that several ocean points were missing from the analysis. We have now corrected the mistake and updated figures **5, 6, and 7** accordingly.

This error did not apply to any other year. Therefore the occasional abrupt shifts in biome coverage (e.g January 1969 and March 1958 in the North Atlantic, and September 1969 in the Southern Ocean) are to be considered as climate-driven transient features.

[Figures 5,6, and 7 have been corrected. The updated text has been added at **line no—405** in the revised manuscript.]

RC1 08: *Figures 6&7: While I know the white box was labeled in a figure, I'd appreciate latitudinal/longitudinal boundaries for the NA and SO regions in both these figure captions*

AR 08: Added.

[Figure captions of 6 and 7 have been updated.]

RC1 09: *Line 438: "instead of directly environmental parameters," I believe might be missing a word in this line*

AR 09: **Updated text:** instead of directly feeding in the environmental parameters to the clustering method

[Updated text has been added at **line no—504** in the revised manuscript.]

RC1 10: *Line 484: Should be an extra line space between paragraphs*

AR 10: Added.

[We have added extra spaces between all paragraphs.]

References

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J.D., Bopp, L., Carroll, D., Carter, B., Chau, T.T.T. and Doney, S.C., 2023. Magnitude, trends, and variability of the global ocean carbon sink from 1985 to 2018. *Global Biogeochemical Cycles*, *37*(10), p.e2023GB007780.

Good, S.A., Martin, M.J. and Rayner, N.A., 2013. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), pp.6704-6716.

Gruber, N., Bakker, D.C., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., McKinley, G.A. and Müller, J.D., 2023. Trends and variability in the ocean carbon sink. *Nature Reviews Earth & Environment*, *4*(2), pp.119-134.

Lauvset, S.K., Lange, N., Tanhua, T., Bittig, H.C., Olsen, A., Kozyr, A., Alin, S., Álvarez, M., Azetsu-Scott, K., Barbero, L. and Becker, S., 2022. GLODAPv2. 2022: the latest version of the global interior ocean biogeochemical data product. *Earth System Science Data*, *14*(12), pp.5543-5572.

Mohanty, S., Kazempour, D., Patara, L. and Kröger, P., 2023, August. Interactive Detection and Visualization of Ocean Carbon Regimes. In *Proceedings of the 18th International Symposium on Spatial and Temporal Data* (pp. 171-174).

Sallée, J.B., Pellichero, V., Akhoudas, C., Pauthenet, E., Vignes, L., Schmidtko, S., Garabato, A.N., Sutherland, P. and Kuusela, M., 2021. Summertime increases in upper-ocean stratification and mixed-layer depth. *Nature*, *591*(7851), pp.592-598.

Swart, N.C., Fyfe, J.C., Gillett, N. and Marshall, G.J., 2015. Comparing trends in the southern annular mode and surface westerly jet. *Journal of Climate*, *28*(22), pp.8840-8859.

Detailed Response to Reviewer 2

RC2 01: *1) The method section - and in particular section 2.3 onward are difficult to read especially for folks that are not familiar with machine learning. This is the result of the many specific terms used (e.g. "merging at a higher height", "Ward variance", "Euclidian Distance", "Ward Linkage", ...). To make the methods section more accessible to the wider audience of the journal, I would suggest to provide less technical text in the main section and add the required detail and terminology to the appendix.*

AR 01: Indeed, the accessibility and comprehensiveness of the text shall be ensured. For that purpose, we rephrased the terms in such a way that they become more intuitive.

- The phrase 'merging at a higher height' implies that clusters are grouped together on the dendrogram's vertical axis.
- Euclidean distance is the distance between two points in space
- Ward variance aims to group data points so that the variance is minimized.
- Ward linkage implies the link generated after clustering data points after the application/ calculation of Ward variance.

The text in the revised manuscript will be updated accordingly.

[Updated text has been added between **line no—213 to 224** in the revised manuscript.]

RC2 02: *2) I dont find the argument about the choice of an MLR that convincing. Figure A1 clearly shows that the relationships are not "perfectly linear". Furthermore, the arguments provided on lines 178-180 that the MLR is faster and easier interpretable are only to a certain extend true. Using e.g. a simple single layer FFN instead of the MLR could account for the slight divergence from linearity without compromising on speed. For me, the main argument is interpretability. The single weights of the MLR are easier to interpret and process than the more complex weight Matrixes of a FFN.*

AR 02: Thank you for challenging the argument about the choice of MLR. Your objection that runtime performance is not a valid/convincing argument for relying on MLR instead of an FFN is a strong point. As you suggested, we redirected our focus to the interpretability of the weights of MLRs compared to FFN weight matrices.

[Updated text has been added at **line no—194** in the revised manuscript.]

RC2 03: *3) This may be a misunderstanding on my end, but I am still puzzled why you need a FFN for the time variation in the biomes. I fully understand the approach and I*

endorse it, but would you not also get changing biomes by doing the MLR followed by the hierarchical clustering for each month/year separately? Thought he changing HC relationships, you would also get changes in the biomes, no?

AR 03: We thank you for bringing up this point. The challenge of how to track the carbon biomes, in the absence of an initial set of labeled clusters, is one to which we devoted considerable thought. Since our approach is fully unsupervised and clusters do not have pre-defined labels, performing a one-to-one mapping between biomes over months and years becomes difficult. Indeed, we initially also used the approach that you suggest. In Mohanty et al. (2023) we tested whether the carbon biomes could be tracked only based on the changes of the feature-driver relationships over time. Specifically, we ran the clustering algorithm for every month of each year and then calculated the Frobenius Norm, a distance metric, between the RCs of each cluster between one time step and the next. We subsequently matched those clusters exhibiting the lowest Frobenius distance. However, this approach was not optimal, since the cluster number was not fixed and the Frobenius Norm failed to match all clusters over time.

Hence, for this manuscript we decided to use a more flexible tracking method. We selected a random month/year to define labels (please see also reply #6 to Reviewer 1 for more details on this particular aspect) and trained a neural network to learn the associations between labels and the underlying regression coefficients (RCs). This allowed us to predict the locations of the same labels over time in an effective way. As shown in **Fig. H1**, the RCs within each biome are (with the exception of the strongly variable ICE biomes) substantially stable over all months and years.

This aspect is now explained more thoroughly in the revised manuscript.

[Updated text has been added at **line no—299** in the revised manuscript.]

RC2 04: *A more general question I had that was not answered in the paper: is your approached that was designed from a single model easily adoptable for other models?*

AR 04: The tool (detection and tracking) is transferable to other ocean models. Still, the exact biome locations and model weights and parameters (linear regression, clustering, neural networks) are not, as they are specific to the ocean model experiments. Ideally, the tool should be run from scratch based on a specific model output or observational data set. Requirements for this include saving surface CO_2 fugacity or partial pressure, DIC, alkalinity and temperature at a resolution sufficient to resolve the seasonal cycle. It should be noted that if the data set has a coarse resolution, bigger boxes will be needed

to construct the spatial target-driver relationships, with the drawback of achieving a less refined picture, of cutting through sharp current, or of averaging out different regimes.

This study does not provide the exact locations of static biomes, but rather - due to the temporally-varying nature of the ocean - a probability that a certain biome will be found in a certain place (**Fig. 5**). We acknowledge that it is practical for model evaluation or intercomparison studies to define a set of static biomes over which different model diagnostics can be averaged (as done in e.g., DeVries et al., 2023). However, the usage of static biomes comes with disadvantages owing to time-varying and model-specific locations of ocean fronts (Fay and McKinley, 2014). We suggest that future research could use the detection and tracking tool proposed here across different data classes to test the spatial homogeneity of the carbon biomes as well as to better understand the specific dynamics of each model.

[Updated text has been added at **line no—595** in the revised manuscript.]

RC2 05: *line 20: please add "annual" to the 25% (the number refers to the present day uptake rate - historically, over the industrial period, the ocean uptake was larger)*

AC 05: Added.

[Updated text has been added at **line no—19** in the revised manuscript.]

RC2 06: *line 130: Please provide more detail how the outlier removal was done*

AC 06: We removed data points with pre-industrial DIC below 1500 micro-mol/kg, alkalinity below 1700 micro-mol/kg, and pre-industrial fCO_2 above 500 micro-atm.

[Updated text has been added at **line no—130** in the revised manuscript.]

RC2 07: *lines 255-270: The architecture of the NN are provided by no justification to why. Have you done some optimalization testing (e.g. on the optimal number of neutrons), or are these subjective choices?*

AR 07: We did not conduct a defined set of tests to choose the number of layers and neurons (which affect the model complexity), and we tried a very small batch of NN architectures. We focused rather on optimizing different hyperparameters of a selected FFN architecture

In the revised manuscript, we now provide more detail on the optimization testing, as detailed below.

- First, we fixed an architecture with the number of layers to six (a relatively small number given the number of input features, i.e., three slopes) and no. of neurons per layer with 64, 128, 256, etc. (a standard practice in ML community to choose neurons as a power of 2) to conduct sensitivity analysis for the hyperparameters (optimizer, learning rate, batch size, no. of epoch, etc.) used in the neural network architecture. We have utilized a 5-cross validation technique for this analysis.
    - This can be found here: https://github.com/swemoh/Detection-and-Tracking-of-Carbon-Biomes/tree/main/jupyter_notebooks/output_files (the link for Gihub is available under Code Availability in the revised manuscript.)
    - We chose the parameters for which we received lower test loss, higher and comparable accuracy, precision, and recall and then trained the model to track the carbon regimes.
- Second, to understand how the model complexity could affect the outcome, we kept the hyperparameters constant and attempted to alter the neural network architecture by removing/adding a hidden layer, which resulted in a decrease/not significant increase of test loss and accuracy (not shown).

We could only check an NN architecture's accuracy, precision, etc., with different hyperparameters. Experimenting with different non-linear functions, individual weights, and biases for every layer of NN could be incredibly time-consuming. Neural networks can be opaque, making it hard to interpret precisely what each neuron has learned. This is especially true for deep networks with many layers. Besides, interpretability techniques like Layer-wise Relevance Propagation (LRP), SHAP, and LIME could have been applied to understand which features are essential and how neurons or layers contribute to the model's prediction capabilities. We suggest that this could be an outlook to be addressed in future research.

[Updated text has been added at **line no—738, 557** in the revised manuscript.]

RC2 08: *line 471: "personality" is an odd choice of wording*

AR 08: Updated to 'nature'

[Updated text has been added at **line no—545** in the revised manuscript.]

RC2 09: *lines 485-490 are a repeat from the introduction and can be removed in my view*

AR 09: In the revised manuscript, we rewrote the sentences to make it more compact.

[Updated text has been added at **line no—564** in the revised manuscript.]

References

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J.D., Bopp, L., Carroll, D., Carter, B., Chau, T.T.T. and Doney, S.C., 2023. Magnitude, trends, and variability of the global ocean carbon sink from 1985 to 2018. *Global Biogeochemical Cycles*, *37*(10), p.e2023GB007780.

Fay, A.R. and McKinley, G.A., 2014. Global open-ocean biomes: mean and temporal variability. Earth System Science Data, 6(2), pp.273-284.

Mohanty, S., Kazempour, D., Patara, L. and Kröger P., "Detection and Tracking of Dynamic Ocean Carbon Uptake Regimes Built Upon Spatial Target-Driver Relationships via Adaptive Hierarchical Clustering," 2023 IEEE 19th International Conference on e-Science (e-Science), Limassol, Cyprus, 2023, pp. 1-10, doi:10.1109/e-Science58273.2023.10254820.