We thank RC2 for their constructive and thorough feedback. Please find below the comments from RC2 and our responses in red italic text.

RC2

Review of M. Chatting et al. : "Improving Marine Sediment Carbon Stock Estimates: The Role of Dry Bulk Density and Predictor Adjustments", egusphere-2025-661

Summary

The present manuscript by Chatting et al. describes an improved modelling approach for calculating OC stock for marine shelf areas by upscaling point observations on OC content. Two approaches are compared: (1) The traditional one uses available globally resolved datasets for predictor variables, including dry bulk density as calculated from sediment porosity and combines it with point observations of OC content in the Irish Sea to derive the local OC stock using a random forest approach. (2) The improved one first performs a bias adjustment, which transforms global data to better represent local point observations. Dry bulk density is then extrapolated using a random forest model based on these transformed predictor data. A second random forest model is applied for the OC content and lastly dry bulk density and OC content are combined into a new OC stock estimate. The new approach shows improved performance (i.e. agreement with in situ data) in predicting dry bulk density and lowers the OC stock estimate for the region by around one third.

General Comments

The approach is methodologically sound. Bias adjustment has so far been mostly applied in climate models and incorporating it into marine OC stock estimation is a novel but timely application. The other modelling approach, random forest models, is well established in geospatial modelling and is an appropriate choice for the present case. The manuscript convincingly shows that bias adjustment as well as including refined DBD estimates can substantially improve our OC stock assessments.

However, the MS is in some section hard to follow, especially in the methods section. A more clear description of the modelling workflow (supported by Fig 1), a plain language summary of what bias adjustment (a central method of this study) entails, and how the success of these measures is actually measured should appear in the introduction or early in the methods sections. This way the reader can be more effectively guided through the novel data handling approach.

Another point of concern which needs to be clarified is the choice of study area. It does not become clear what the advantages of choosing the Irish Sea are, although certainly there were some. The authors mention that only 3% of DBD data (the most important predictor variable!) are available for the study area. Either expanding the scope of the modelling to the entire NW European shelf, which is the source for this DBD data, or making a good case for limiting it to the Irish Sea are needed.

Despite these points the present MS provides a useful and novel blueprint study which can help improving OC stock modelling globally, a topic with general relevance for climate science as well as biogeochemistry.

Specific Comments

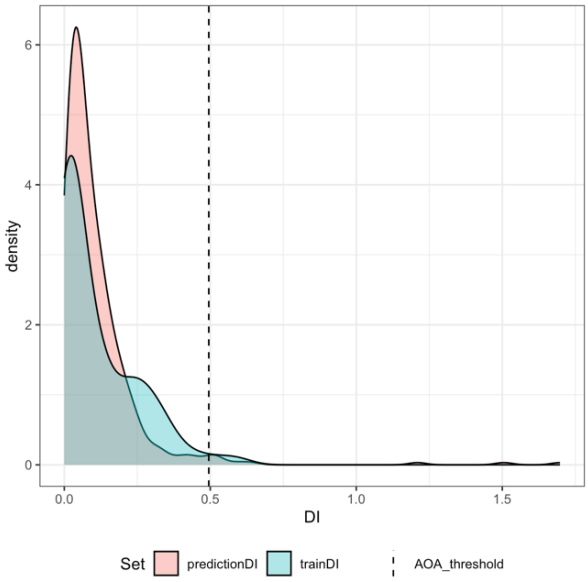
Clarify methods: The manuscript would benefit from some more plain language step-by-step guidance throughout the methods section. E.g. in L66-72 or later, a comprehensible description of what QQ mapping entails could help familiarizing the reader with the approach.
Choice of study area: The authors should better justify their focus on the Irish Sea (e.g. in L22-23), especially as DBD seems to be sparsely available here. The reader gets the impression the entire shelf area might have been a better focus (600+ data points for DBD). Certainly there are arguments for this tighter spatial focus, which could be presented here.
Generalizability: The authors should briefly state how the findings are transferable to other geographical regions, and what it would mean for existing OC stock assessments. Is this only applicable in areas with dense available data and therefore limited to well-studied zones, or can we improve global estimates? How, e.g., would the results of Atwood et al. (2020) which are cited in the text change considering the findings of the MS? Adding a global relevance section to the discussion can help the reader better grasp the implications of the novel modelling framework.

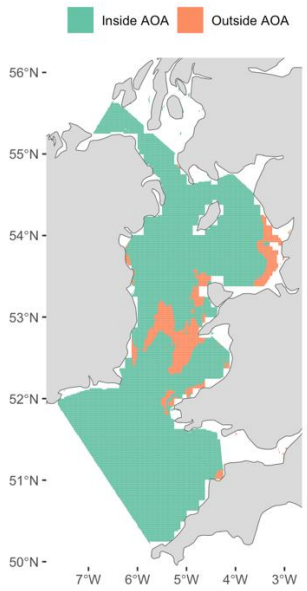
1) A more clear description of the modelling workflow (supported by Fig 1), a plain language summary of what bias adjustment (a central method of this study) entails, and how the success of these measures is actually measured should appear in the introduction or early in the methods sections.
    - *We appreciate this comment and agree that clarity in the methods section is critical. In response, we have rewritten large sections of the manuscript for clarity, especially within the methods. This includes shortening sentences, using more intuitive terminology for predictors, and introducing a summary paragraph at the beginning of the methods section as suggested by the reviewer. We have also clarified the bias adjustment process. The simple overview we added is at the start of the methods between lines 140 and 154 in the revised manuscript and reads:*
        i. *"To estimate organic carbon (OC) stocks in surficial sediments, we developed and compared two modelling workflows. Each workflow involved predicting OC content (%) and dry bulk density (DBD), which were then combined to calculate OC stock (kg m^{-2}). The key difference between the two workflows was the way environmental input data (predictors) were treated. The first approach used unadjusted, commonly available predictors and a standard DBD estimation method, while the second approach used bias-adjusted predictors, which were corrected using observational data, combined with a machine learning model to estimate DBD. A schematic overview of the workflow is provided in Fig. 1. Briefly, the process of bias-adjusting shifts the distribution of predictor data based on observational data in*

an effort to align predictor data with in situ observations. We evaluated the success of these improvements in two ways. First, we tested whether bias-adjusted predictors more closely matched local measurements, using an error metric (Root Mean Squared Error; RMSE) which measured how far predictions deviated from in situ observations. Second, we assessed whether these improved predictors led to more accurate predictions of OC content and DBD using machine learning models, using cross-validation and RMSE. The assumption underpinning this study is that predictors that better align with in situ data would produce more reliable predictions of OC content and DBD and thus more reliable estimates of OC stock."

- *We have also changed many of the predictor abbreviations to be more intuitive and more easily remembered.*

2) Another point of concern which needs to be clarified is the choice of study area. It does not become clear what the advantages of choosing the Irish Sea are, although certainly there were some. The authors mention that only 3% of DBD data (the most important predictor variable!) are available for the study area. Either expanding the scope of the modelling to the entire NW European shelf, which is the source for this DBD data, or making a good case for limiting it to the Irish Sea are needed.
    - *Thank you for this observation. We have expanded our justification for choosing the Irish Sea in the regional setting section (lines 110–124), citing its ecological and economic importance, heavy bottom-trawling pressure, and known high-OC mud sediments.*
    - *Also we acknowledge that only 3% of the DBD training data come from within the Irish Sea. However, we argue that our adjusted DBD model is an improvement over the unadjusted method for several reasons:*
        i. *The unadjusted method, which has frequently been used in previous work (Diesing et al. 2017, Diesing et al. 2021, Smeaton et al. 2021), uses data solely from the Mississippi-Alabama-Florida shelf, assuming global applicability, which may not be valid for the Irish Sea.*
        ii. *Our adjusted DBD model includes data from the Irish Sea and surrounding NW European Shelf and underwent Area of Applicability (AOA) analysis, showing that 91.2% of the study area is within model applicability.*
        iii. *We now include these AOA figures (Supplementary S5, figures below) and corresponding clarifying text in the manuscript.*





Diesing, M., Kröger, S., Parker, R., Jenkins, C., Mason, C. and Weston, K., 2017. Predicting the standing stock of organic carbon in surface sediments of the North–West European continental shelf. Biogeochemistry, 135, pp.183-200.

Diesing, M., Paradis, S., Jensen, H., Thorsnes, T., Bjarnadóttir, L.R. and Knies, J., 2023. Organic Carbon Stocks and Accumulation Rates in Surface Sediments of the Norwegian Continental Margin. Authorea Preprints.

Smeaton, C., Hunt, C.A., Turrell, W.R. and Austin, W.E., 2021. Marine sedimentary carbon stocks of the United Kingdom's exclusive economic zone. Frontiers in Earth Science, 9, p.593324.

3) The authors should briefly state how the findings are transferable to other geographical regions, and what it would mean for existing OC stock assessments. Is this only applicable in areas with dense available data and therefore limited to well-studied zones, or can we improve global estimates? Is this only applicable in areas with dense available data and therefore limited to well-studied zones, or can we improve global estimates? How, e.g., would the results of Atwood et al. (2020) which are cited in the text change considering the findings of the MS? Adding a global relevance section to the discussion can help the reader better grasp the implications of the novel modelling framework.
    - *We agree and have included text in the discussion (lines 586–593) that our method relies on substantial data inputs to improve OC stock estimates. The improved methodology is applicable where large amounts of in situ data are available. We note that as more data becomes available in data-poor regions this method may be applied. Additionally, we have also included (lines 445–447) text to state that previous work using the porosity empirical relationship may have represented overestimates of DBD and consequently OC stock.*

Line comments

L15: the study addresses the short-term C cycle, not the geological one, consider rephrasing
We have removed the term 'over geological timescales'

L19: "Data gaps" may be misleading here. The study does not collect new data but rather improves the interpretation of existing data, rephrase
We have rephrased this sentence which now reads: "… however, regional predictions of OC often use global scale predictors which may have biases on smaller scales, reducing their utility for practical management decisions."

L22-23: State why Irish Sea was chosen; maybe due to sufficient data availability?
We disagree with the reviewer's intent here. We do not think that justification for the Irish Sea should be given in the abstract. We have removed "in the Irish Sea" from the sentence. Instead we have provided a more thorough justification for selecting the Irish Sea as the study is in section 2: Regional setting.

L28: "emphasize" instead of "emphasizes" as it refers to "findings"
This error has been changed. We have also changed the word to "highlight" instead of "emphasize".

L29: ensure consistent formatting of in situ (throughout the MS)
We have gone through the manuscript and standardized the formatting of "in situ"

L30: consider removing "for policy makers", as many more stakeholders are interested in improved OC stock assessment
"for policy makers" has been removed from the text

L49-50: harmonize phrasing "uppermost 10 cm", "surficial 0-10 cm" and "top 0.1 m" (in methods section) all refer to the same and should be consistent
We have standardized all mentions to the "top 10 cm", apart from one occurrence. We have kept "surficial (0 - 10 cm)" as it appears just after a sentence with "top 10 cm", we wanted to avoid repetition so have kept "surficial (0 - 10 cm)" for this one instance.

L56: possibly use "OC stock" instead of "OC content"; the sentence is hard to read
We have rewritten this sentence to be clearer for the reader. It now reads: "DBD is a scaling factor on OC content and adjusted the OC stock in a given volume based on the density of sediment."
L58: "sediment density" should be "soil density"
We have rephrased here to "soil density"

L59: consider removing "however" to improve flow
We have removed "however" from the sentence

L66-72: The repeated mention of "climate models" and references about them can be drastically shortened. Instead it could include a one sentence, plain-language summary of what bias adjustment entails

We have simplified the text in this paragraph. It now reads: "To address these discrepancies, bias adjustment techniques are commonly used in other scientific disciplines, for example in climate science, where large-scale models are adjusted to better align with local observational data (Laux et al., 2021; Luo et al., 2018). Bias adjustments reduce systematic errors in model outputs and ensures that projections match local conditions and are reliable for practical applications (Laux et al., 2021). Bias adjustments have been used to improve climate model utility in agricultural impact assessments, such as predicting planting dates and crop suitability in water-limited regions; to correct overestimations in soil moisture models and to improve predictions in sea ice thickness (Laux et al., 2021; Lee and Im, 2015; Mu et al., 2018). Despite their widespread use in climate science, bias adjustment methods are underutilised in other areas of spatial environmental modelling, including OC stock modelling." Additionally, we have provided a brief plain language of bias adjustment at the start of the methods section (lines 146 to 148)

L78: If OC and mud content measurements also use similar instrumentation maybe mention this (instead)
We have rephrased to sediment properties and OC content

L87: Be more clear how model improvement is assessed
We have added text at the start of the methods section (lines 148 to 152) that gives a plain language summary of how we assessed potential improvement in environmental predictors and random forest models. The text reads: "We evaluated the success of these improvements in two ways. First, we tested whether bias-adjusted predictors more closely matched local measurements, using an error metric (Root Mean Squared Error; RMSE) which measured how far predictions deviated from in situ observations. Second, we assessed whether these improved predictors led to more accurate predictions of OC content and DBD using machine learning models, using cross-validation and RMSE. The assumption underpinning this study is that predictors that better align with in situ data would produce more reliable predictions of OC content and DBD and thus more reliable estimates of OC stock."

L97: Does this mosaic of sediment types help modelling here?
Having variability in the sediment types, so long as that varies with parameters of interest (OC content and DBD) does help spatial modelling here. With little variation in sediment types but still substantial variation in OC content and DBD, it would be more difficult to use sediment type as a predictor of OC content and DBD. It is likely other parameters would be more useful. This makes intuitive sense, as in this hypothetical situation, sediment type would not be causing heterogeneity in OC content and DBD.

L104-105: Briefly mention how inshore area is defined in this data
We have added clarification on how the inshore area was defined on line 137 of the revised manuscript.

L147: Add a sentence as plain-language summary of QQ mapping
We have edited the sentence here to make it easier for the reader to follow on lines 213 to 215 on the revised manuscript: "This approach aligns the quantiles in observational and modelled data and preserves the spatial patterns of the original data while aligning their

statistical distribution with in situ data, and has QQ mapping bias adjusted models have been shown to outperform un-adjusted models (Ngai et al., 2017).”

L153-165: Consider moving to Supplementary, log ratio transformation is a standard procedure in compositional data and not crucial to the presented modelling approach
We agree with the reviewer and will move this text to the supplementary material for the fully revised manuscript.

L192: Is this the mud content from spatial averaging? Clarify
We have clarified here that we were referring to spatially averaged mud.

L206: refer to Fig 4 here (may be Fig 3 then)
We disagree with the reviewer on this point. As this is the methods section and still describing the concept of important predictor selection we do not think it is appropriate to refer to the plot that displays the important predictors partial dependence on OC content.

L214: could this k fold CV be replaced with the NNDM LOO CV, which is introduced later and said to perform better?
Here we are referring to the random k fold CV and that it does not perform as well as NNDM LOO CV. To make it clearer for the reader, we have rephrased the text to read (lines 299 to 303): “Random k fold cross-validation, can produce overly optimistic performance estimates by splitting spatially autocorrelated data across training and testing sets. By contrast, NNDM ensures spatial independence between folds, providing better estimates of model performance on spatially independent data (Milà et al., 2022).”

L236: “by a grid cell” instead of “by grid cell”
This typo has been corrected

L241: be more explicit than “all possible combinations”, there are 4 combinations; OC adj/unadj with DBD adj/unadj correct?
We have clarified in the text that we are talking about 4 possible combinations of adjusted vs unadjusted OC and DBD models.

L273: Sort plots by predictor importance in Figure 4
We agree with the reviewer and will sort the plots by predictor importance in the revised manuscript

L277: How would this approach perform in regions with even fewer DBD observations?
We have clarified in the discussion section (lines 582 to 589) the limitations of having fewer DBD points, which is why AOA analysis was performed. The text reads: “The refined estimates presented in this study rely on large amounts of in situ data and environmental predictors, making this approach most suitable for data-rich regions. Within our study area, the limited availability of DBD measurements required the use of an Area of Applicability (AOA) analysis to assess whether the adjusted DBD model could be reliably applied—highlighting potential limitations of this approach in data-poor settings. Nonetheless, our findings demonstrate that where sufficient observational data are available, OC stock estimates can be substantially improved. As more in situ datasets are generated in currently

under-sampled regions, this modelling framework can be replicated and further refined to support better-informed carbon assessments."

L318-320: Clarify how the presented findings would influence their estimates
We have included text between lines 441 and 443 that suggests how these findings would impact other estimates. "Additionally, these findings highlight the importance of using improved DBD models and suggests that previous estimates of OC stock that used the porosity empirical relationship may represent overestimates."

L326: move the definition of mud to its first mention in the MS
We have moved the definition to earlier in the manuscript. It is now on line 221 of the revised manuscript.

L331-335: The sentence is long and unclear; what is the "topography" of a mineral grain? Also the references need sorting
The sentence has been rephrased. It now reads: "The capacity for sediments to bind OC through clay-OC interactions can also vary with different mineral phases occurring in sediments, varying in particle-size as well as surface area, charge and distribution, and subsequent geochemical conditions constraining these characteristics (e.g. pH and ionic strength of pore water)."

Also the references have been corrected.

L346: Space missing between "(2024)" and "estimated"
We have corrected this typo

L356: "resuspension" instead of "suspension"
We have rewritten this sentence and it no longer needs correction

L362-63: "needs" instead of "need"
We have rewritten the sentence to be clearer.

L388: "carries" instead of "carry"
We do not agree with the reviewer here. The word "carry" here refers to the "predictions" mentioned at the start of the sentence. Since "predictions" is plural, "carry" should be used, not "carries"

L398: rephrase "increased in situ data" to "increased availability of in situ data"
We have edited this sentence. It now reads: "As more in situ datasets become available in currently under-sampled regions, this modelling framework can be replicated and further refined to support better-informed carbon assessments."

Figure 2: It seems the yellow shade is not reached/used, maybe adjust the colormap. Also the thick outline is not very visually appealing and might obscure data points. A dashed line should be tested.
We agree with the reviewer and this figure change will be made when the fully revised manuscript is submitted

Figure 4: The plot is not very visually appealing, it is not clear what the letters refer to. Maybe the distance between the a and b row can be slightly increases. Plots should be sorted in the order of relative parameter importance. Also the figure caption for 4 c is not correctly describing the presented plots.
We agree with the reviewer and this figure change will be made when the fully revised manuscript is submitted

Figure 5, 6 and 7 all show the same region, but different parameters. Maybe combining them all into one large, page filling figure would allow the reader to better appreciate all present trends at once. Another side note is, that the text mentions the “Isle of Man”, which is not labelled in a map and might be unfamiliar to many readers.
We agree with the reviewer and this figure change will be made when the fully revised manuscript is submitted