

Dear Editor,

Thank you for giving us the opportunity to revise and resubmit our manuscript entitled, Linking marine benthic biodiversity and ecosystem functions related to carbon cycling in a continental mud depocenter.

We sincerely appreciate the constructive comments provided by you and the reviewers. We have carefully revised the manuscript accordingly and prepared a point-by-point response addressing all comments and a track-changes file. We believe that these revisions have substantially improved the quality of the manuscript.

We hope that the revised manuscript is now suitable for publication in Biogeosciences, and we thank you for your time and consideration. We look forward to your response.

Sincerely,

Chueh-Chen Tung

On behalf of all co-authors

Reviewer 1

The authors have sampled macrofauna on an area of rather large spatial coverage to estimate the contribution of macrofauna biodiversity and ecosystem functions to carbon cycling on an environmental gradient from sand to mud. The questions presented are important but some clarifications are needed especially to the statistical tests in order to assess the accuracy of the methods and the estimations.

A lot of calculations were done based on one single sample of macrofauna from the same site. Good that the authors in the end do admit that direct measurements are still missing.

Response: We thank the reviewer for his thoughtful and critical comments. Please, see below how we dealt with the specific comments and suggestions made by Reviewer 1.

Comment 1.1: The abstract is well written and concise. However there was no methodological mention of how things were done (lab experiment? Observation? Calculation/model?).

Response 1.1: Thanks for the overall positive comment on the abstract. We have added the following methodological description to the abstract. The integrated methodological and results description in line 20 now read as follows: “Based on 171 macrofaunal taxa collected from 50 stations, hierarchical clustering based on abundance data categorized the macrofaunal communities into four distinct structural groups. Distance-based redundancy analysis (dbRDA) revealed that community functional composition was primarily structured by bottom hydrodynamics (salinity and bottom shear stress) and sediment characteristics (sorting, skewness, kurtosis, and D50).

Biomass-weighted functional trait vectors, mapped onto a non-metric multidimensional scaling (nMDS) space, showed a clear structural shift with communities in the deep clusters being dominated by mobile biodiffusors and subsurface filter feeders, whereas shallow clusters were characterized by less mobile, surface-modifying bivalves and polychaetes. These contrasting patterns led to pronounced differences in calculated potential ecosystem functioning: bioturbation and bioirrigation potentials were significantly higher in deeper communities, whereas the calculated community secondary production and respiration were higher in shallow communities. Across all stations, community secondary production and

respiration increased with taxonomic and functional richness, while bioturbation and bioirrigation potentials were negatively related to biodiversity and community evenness. Interaction regressions and simple slope analyses revealed that environmental gradients significantly modulated the biodiversity–ecosystem functioning (BEF) relationships. Our findings demonstrate that environmental gradients fundamentally shape both benthic community structure and the link between biodiversity and essential ecosystem functions related to carbon dynamics in fine-grained sediment and organic matter in shelf sea systems.”

Comment 1.2: Row 22: “...community *FUNCTIONAL* composition...”

Response 1.2: Thank you for this correction. Since we have reorganized the whole abstract, this sentence has been removed.

Comment 1.3: Row 25: “...led to pronounced differences in ecosystem functioning...”

You did actually not measure ecosystem functioning, you calculated the potential of the fauna to modify the sediment structure and ventilate their burrows and you calculated a theoretical secondary production and respiration rate; this should be clear also from the abstract

Response 1.3: We agree with the reviewer’s critical distinction. It is correct that we did not directly measure these ecosystem functions, but rather estimated the community's potential to modify the sediment and calculated theoretical secondary production and respiration rates using empirical models. Accordingly, we have replaced “ecosystem functioning” by “calculated potential ecosystem functioning” to reflect our methodological approach precisely.

The revised sentence now in line 26 reads: “These contrasting patterns led to pronounced differences in calculated potential ecosystem functioning: bioturbation and bioirrigation potentials were significantly higher in deeper communities, whereas the calculated community secondary production and respiration were higher in shallow communities.”

Introduction

Comment 1.4: Row 48: “...suggested that community evenness rather than species richness

controlS...”

Response 1.4: Thank you very much for catching this grammatical error. We have corrected this in the revised manuscript. The revised sentence now in line 51 reads: “Maureaud et al. (2019) suggested that community evenness rather than species richness controls ecosystem functions based on the observation that fish biomass in European seas tended to be higher in assemblages dominated by a few generalist species capable of exploiting both benthic and pelagic resources”

Comment 1.5: Starting row 65; the role of macrofauna in mediating biogeochemical processes. It is true that they play a pivotal role, but the proportional contribution is dependent on sediment characteristics and hydrodynamical forcing (see Bernard et al. 2019)

Response 1.5: We agree with this statement. We have integrated this point to provide a more balanced background. The revised section now reads in line 74: “Benthic invertebrates strongly affect sedimentary biogeochemical processes through sediment reworking (bioturbation) and ventilation (bioirrigation), thus affecting oxygen consumption, carbon remineralization, element cycling, and bacterial activity in sediments (Aller, 1994; Wenzhöfer and Glud, 2004; Glud, 2008; Kristensen et al., 2012; Gilbertson et al., 2012; Laverock et al., 2014). However, the proportional contribution of the macrofauna to these processes depends decisively also on local sediment characteristics and hydrodynamic forcing (Bernard et al., 2019).”

Comment 1.6: Row 78-79: not sure what the which is referring to, if it refers to the whole sentence “biological traits structure reflects local-scale environmental conditions”, the verbs thereafter should be singular. If it refers to the environmental conditions, then for what does it provide better ecological insight?

Response 1.6: We apologize for the ambiguity caused by the pronoun "which." Our intended meaning was that biological trait structure rather than taxonomic composition is more helpful in differentiating communities and provides deeper ecological insights.

To resolve this ambiguity, we have rephrased the section now in line 86 as follows:

“Moreover, biological trait structure reflects local-scale environmental conditions. This

approach helps differentiate communities and thus provides better ecological insight than methods considering the taxonomic composition only (Bremner et al., 2003).”

Materials and methods

Comment 1.7: Because only one replicate sample of fauna was taken, the comparison between sites includes a large amount of uncertainty. Abundance and biomass and the identity of taxa affect your calculation of bioturbation potential. Very well understanding the amount of work that goes into sieving and going through macrofauna samples I also know that the variation between replicates can be huge. It is estimated that taking five replicate samples gives you 90 % of the taxa present in a sampling area, three replicates gives 70 % so how do you incorporate this uncertainty into your results? Some areas can also represent a much more patchy distribution than others which further complicates the comparison between stations. Possibility to miss rarer, more patchily occurring taxa -> effect on diversity measures?

Response 1.7: We thank the reviewer for raising this critical point. We acknowledge that small-scale patchiness may lead to the omission of rare taxa and that the lack of replicate sampling introduces uncertainty particularly in our measures of biodiversity.

We have carefully addressed this issue in Materials and Methods to acknowledge the statistical uncertainty and patchiness, while contextualizing why our broad-scale gradient approach remains robust for capturing the macrofaunal baseline in this mud depocenter.

The revised section now reads in line 159: “We acknowledge that utilizing only a single sample per station may introduce uncertainty in our measures of biodiversity and increase the probability of missing rare or patchily distributed taxa. Instead of increasing the level of replication at the single stations, we decided for a broad coverage of the entire study area with a single sample at each station in order to capture the more pronounced variations in the infauna communities along environmental gradients. We expected the gradients in the infauna communities across the sampling area to allow for a more robust analysis of the variations in the relationship between biodiversity and ecosystem functioning than local variations among replicate samples at single stations.”

Comment 1.8: Row 130: "...permanently mixES..."

Comment 1.9: Row 132: "...and generally driveS..."

Response 1.8 & 1.9: Thank you very much for catching this grammatical error. We have corrected this in the revised manuscript. The revised sentence now reads in line 135: "Strong tidal and wind forcing permanently mixes waters in the areas shallower than 20 m (Otto et al., 1990; Huthnance, 1991), while mediating periodic and non-periodic stratification events in deeper zones (Burchard and Hetland, 2010; Chegini et al., 2020), and generally drives an anti-clockwise residual circulation along the coast (Kopte et al., 2022)."

Comment 1.10: Row 148-149: Was this wet weight, dry weight or ash-free dry weight?

Response 1.10: This sentence refers to the wet weight, which was measured. For the calculations, the wet weight was converted to ash-free dry mass using taxon-specific conversion factors provided by Brey et al. (2010) in lateral analyses. We specified this already in the original submission where it reads (line 149): "For each taxon, the total wet weight was measured with a precision of 0.001 g." Additionally, in line 241 it reads: "Macrofauna biomass (wet weight) was converted to ash-free dry mass (AFDM) to accurately depict metabolically active tissue, using taxon-specific conversion factors provided by Brey et al. (2010)."

Comment 1.11: Why did you choose to use summer values of salinity?

Response 1.11: We chose to use summer values of salinity because our sampling of the macrofauna was done in summer. We have revised the sentence in line 162 to describe our methodological approach precisely. The revised sentence now reads in line 171: "Seasonal bottom salinity was averaged over the multi-year period 2019–2022, and summer values were selected for subsequent analyses to match the seasonal salinity conditions with the timing of our macrofauna sampling, which also took place in summer."

Comment 1.12: What value of temperature was used for the production and respiration calculations? There was no mention of where these data came from, was it a single

measurement at the time of sampling or an annual mean/min/max? If a single measurement value is used, this should be mentioned and its effects on the calculated production and respiration discussed.

Response 1.12: We thank the reviewer for pointing out this critical methodological detail. The temperature used in the ANN models for calculating secondary production and respiration was measured directly in the grab sample. The recorded bottom temperatures ranged from 15°C to 18°C across the study area.

We agree that using a single seasonal measurement rather than an annual mean can cause a potential bias. We have updated the Materials and Methods to clarify the data source and add a detailed discussion regarding its potential effects in the revised Discussion.

The revised section in Materials and Methods reads in line 179: " Annual secondary production (P) of the macrofaunal community was estimated based on three continuous variables: body mass, water depth, bottom-water temperature. The bottom-water temperature was taken as the sediment temperature measured in each sample (ranging from 15 to 18 °C; with calculations for each species applied to the corresponding temperature of its sampling station), and 17 categorical parameters, including five taxonomic groups, seven lifestyles, four environmental types, and the state of exploitation, using the Artificial Neural Network (ANN) model developed by Brey (2012), carried out via the BenthicPro R package (Andresen and Brey, 2018)."

In the Discussion, the revised corresponding section reads in line 536: "Model-derived macrofaunal respiration was well in the range but slightly above sediment oxygen consumption measured empirically in adjacent areas of the southeastern North Sea (Provoost et al., 2013; Oehler et al., 2015). This slight overestimation likely occurred because the ANN models used summer bottom water temperatures instead of annual means. These higher temperature inputs could accelerate the calculated metabolic rates, artificially inflating the estimated annual production and respiration (Brey, 2010)."

Comment 1.13: Rows 184 and 185: Are there many taxa that occur throughout the environmental gradient and if there are, do they create similar burrow or affect the sediment

in a similar way in changing environmental conditions? E.g. burrow depth in muddy vs sandy sediment.

Response 1.13: We agree with the reviewer that many macrofauna species exhibit behavioral plasticity and may modify their burrow depth or reworking modes depending on sediment characteristics. However, as emphasized in the manuscript (lines 190–192), bioturbation and bioirrigation potentials are explicitly defined as relative, trait-based indices of functional potential rather than direct, real-time measurements of absolute bioturbation or bioirrigation rates. They assess the capacity of an assemblage based on its standard biological composition.

While individual behavioral adjustments (such as a species burrowing shallower in muddy vs. sandy sediments) do occur, macro-scale studies (e.g., Queirós et al., 2013; Gogina et al., 2017) have demonstrated that the shifts in species abundance and biomass across environmental gradients exert a far greater mathematical and ecological influence on the final index values than the localized behavioral variations of a single taxon. In our dataset, when a widespread taxon was present along the environmental gradient from mud to sand, its abundance and biomass changed accordingly, which directly scales its functional contribution in the equations. Finally, capturing fine-scale behavioral modifications for numerous taxa across multiple biotopes would require intensive, site-specific in situ or experimental imaging. Using established, standard scores ensures comparability with historical literature from the North Sea and other continental shelf systems.

To clarify this for the reader, we have modified the section of the manuscript starting in line 204 to acknowledge the potential for functional plasticity while validating the reliability of using standardized potential indices: “Both bioturbation and bioirrigation potentials represent relative (dimensionless), trait-based indices that integrate species abundance, biomass, and functional characteristics, rather than direct measurements of sediment reworking or chemical exchange measured from experiments. Although these indices inevitably overlook potential behavioral changes of the macrofauna across different sediments, macro-scale studies indicate that spatial fluctuations in abundance and biomass exert a greater influence on the final scores than local intra-specific variations (Queirós et al., 2013). Therefore, using standardized trait scores ensures comparable estimations of community functional potential and captures the major ecological signals across the study area.”

Comment 1.14: Row 193 to 194: Referring to the comment above and the uncertainty coming from the sampling design, isn't the same issue regarding comparison also affecting this comparison because of the underlying variables used in the calculation of these potentials?

Response 1.14: The reviewer is correct that since the calculation of bioturbation and bioirrigation potentials directly integrates species taxonomic identity, abundance, and biomass, any underlying uncertainty in the primary sampling data inherently propagates into these functional indices. However, from a mathematical perspective, the formulation of these indices (especially bioturbation potential) utilizes the square-root transformation of species biomass, which decreases the influence of extreme values caused by the random capture or neglect of large-bodied individuals in a single replicate. Moreover, this study does not focus on fine-scale, station-to-station comparisons, which would indeed be susceptible to localized patchiness. Instead, individual stations are treated as distributed spatial replicates to evaluate broad, macro-scale differences among the four distinct community clusters.

Consequently, while we acknowledge the inherent uncertainty from the sampling design, the aggregated cluster-level comparisons smooth out localized spatial noise. The broad ecological contrasts in functional potentials presented here remain robust and, therefore, we feel that no additional modifications to the manuscript text are required. Please, also see in this context our detailed response to comment 1.7 by Reviewer 1.

2.4. Statistical analyses

Comment 1.15: Row 227 > “A Bray-Curtis dissimilarity matrix was calculated based on the transformed data and analyzed using hierarchical clustering...”

Response 1.15: We have revised this sentence accordingly. The revised sentence now reads in line 243: “A Bray-Curtis dissimilarity matrix was calculated based on the transformed data and analyzed using hierarchical clustering with Ward’s minimum variance method.”

Comment 1.16: Row 229 > “The clusters were further tested by permutational multivariate analysis of variance (PERMANOVA)” To test what exactly was the PERMANOVA run?

Response 1.16: We thank the reviewer for raising this critical point, which led us to carefully re-examine our statistical workflow. Reviewer 2 also raised the same concern (see below

comments 2.7 and 2.10) that using PERMANOVA to compare groups already defined by clustering represents circular reasoning.

After reviewing our code and the underlying mathematical logic, we completely agree with the reviewers' points. Running a species-based PERMANOVA on cluster groups that were originally derived from the exact same Bray-Curtis distance matrix represents a circular analysis. Because the clustering algorithm (Ward's method) is mathematically designed to maximize between-group distances, a downstream significance test on those same groups will inevitably yield a significant p-value, making the PERMANOVA statistically redundant.

Therefore, we have decided to omit the PERMANOVA analysis and all related text descriptions from the revised Materials and Methods and Results sections. This proposed solution also applies to our responses to Comment 2.7 and Comment 2.10.

Comment 1.17: Row 240-241: compared among clusters or between clusters?

Response 1.17: Thank you for the comment. Since we analyzed a total of four distinct clusters, the global comparisons were performed among clusters using Kruskal-Wallis tests (line 266), and then the subsequent post-hoc pairwise tests were performed between specific pairs of clusters. We have rephrased this sentence to be precise.

The revision now reads in line 265: “Calculated ecosystem functions, including bioturbation and bioirrigation potentials, community secondary production, and respiration, were compared among clusters using Kruskal–Wallis tests, followed by post-hoc pairwise comparisons to identify specific differences.”

Comment 1.18: Row 291 > you cannot compare NMDS plots based on different species/functional traits saying that some stations were always on the right and some on the left and that the pattern is same across all the studied functions. If you are interested in the distribution of the functions across the study area, you should have them all in the same plot. Or do a PCO of the environmental characteristics and overlay the functional groups as vectors on that to see which functional groups actually correlate with what type of environment. Or am I missing something on the methods part, in that case please clarify the statistics behind.

Response 1.18: We thank the reviewer for this comment and appreciate the opportunity to clarify our workflow. We did not compare different, independent nMDS plots. Instead, all data were integrated into a single, unified community ordination plot as the reviewer suggested. We first computed the Community Weighted Mean (CWM) for each trait modality across stations using the functcomp function (FD package), where species biomass was utilized as the weighting variable. We then used the envfit function (vegan package) to fit these biomass-weighted CWM traits onto the single, underlying community-based nMDS space. This routine overlays the traits as passive vectors via linear regression, without changing the original coordinates of the sampling stations. Therefore, our plot contains both the community clusters, which are identical among all plots, and the functional vectors within the exact same multivariate space. We have revised the text in the Materials and Methods

The revision now reads in line 260: “The macrofaunal communities from each sampling station were mapped onto a single, unified non-metric multidimensional scaling (nMDS) ordination space based on their species composition. To evaluate functional trait distributions, we first calculated the Community Weighted Mean (CWM) for each functional trait modality across stations, using biomass as the weighting variable. These biomass-weighted functional traits were then fit onto the community-based nMDS space as passive vectors. The direction and length of each vector indicate its correlation strength with the ordination axes.”

Comment 1.19: Row 360 and 361: “Summary of the correlation coefficientS...” and “Bold STYLE indicates STATISTICALLY significant result / STATISTICAL SIGNIFICANCE”)

Response 1.19: We have revised the caption as: “Table 2. Summary of the correlation coefficients between diversity indices and derived ecosystem functions. P-values are indicated in parentheses. Bold style indicates statistically significant results ($p < 0.05$).”

Comment 1.20: Were the correlations between diversity indices and the bioturbation and bioirrigation potentials done for the whole dataset together? Given that there was such a clear clustering of the communities depending on environmental factors, wouldn't it had made sense to also run these cluster-wise to see the relative contribution of those withing respective cluster?

Response 1.20: The reviewer raises an important question regarding how to integrate the original pooled analysis with these new cluster-wise findings. We did actually explore cluster-wise correlation analyses within each of the four distinct community clusters. Interestingly, the results within individual clusters were mostly weak or non-significant except for a few scattered metrics, which led us to consciously choose the pooled, dataset-wide analysis.

We consider that the pooled approach is both mathematically more robust and ecologically more meaningful for the following reasons. First, subdividing the dataset into four separate clusters drastically reduces the degrees of freedom within each group. This loss of replicates decreases the statistical power, making the correlation tests highly susceptible to type II errors (false negatives) driven by intra-cluster noise. In addition, individual clusters were group stations with highly homogeneous environmental properties. Conducting correlations within a single cluster essentially eliminates the influence of environmental gradients. Therefore, pooling all stations gives us the macro-scale, region-wide baseline of BEF relationships in the study area. Then our interactions analysis enabled us to depict how the correlation slope shifts across different environmental gradients. Consequently, while the cluster-wise exploration was an insightful method, we retained the pooled analysis in the manuscript to emphasize a more comprehensive ecological narrative.

Discussion

4.1 Community structure and functional trait composition:

Comment 1.21: Row 413 and 414: is this the same reference than in the previous sentence (Vopel et al. 2003?). In that case, move the reference to the end of the second sentence.

Response 1.21: Yes, the data in the second sentence regarding the Skagerrak slopes is from the same study (Vopel et al., 2003). We apologize for the misplaced citation.

The modified the sentence now reads in line 443: “For example, *A. filiformis* contributes substantially to the total oxygen flux into soft sediments not only through its own respiration but also via diffusive processes across the additional sediment surface of its burrow walls. On slopes (65–90 m water depth) of the Skagerrak (western Sweden), *A. filiformis* accounted for at least 80% of the total oxygen flux (Vopel et al., 2003).”

Comment 1.22: Row 459: “supported sustained greater macrofaunal biomass” One verb too many here?

Response 1.22: We have revised this by removing "sustained". The sentence now reads in line 495: “In our study, higher-salinity parts of the muddy sediment supported greater macrofaunal biomass without a corresponding increase in species richness.”

Comment 1.23: You could be missing the rarer species, how would that affect the functional richness and redundancy?

Response 1.23: We thank the reviewer for raising this important question regarding the ecological implications of missing rare species on functional indices. Functional richness represents the total volume of functional space occupied by a community. It could be slightly underestimated if rare species with distinct trait combinations were missed due to single-replicate sampling. However, we argue that this potential omission has a negligible impact on both functional richness and redundancy. As demonstrated by Mouillot et al. (2013), not all rare species support distinct or unique functions; instead, a majority of rare species tend to possess traits that overlap heavily with more common species, thereby supporting redundant functions. Moreover, rarity and unique trait combinations are frequently associated with high habitat heterogeneity, specialized environmental tolerances, or restricted dispersal abilities (Gaston et al. 1997, <https://doi.org/10.2307/5951>; Ellingsen et al., 2007, <https://doi.org/10.1016/j.seares.2007.10.001>). Given that our study region is a quite homogeneous mud depocenter with uniform sediment profiles, the probability of hosting highly specialized, functionally unique rare species is low. The rare species present here are much more likely to be nested within the functional traits of the dominant fauna. In addition, functional redundancy is primarily sustained by dominant and common species that share similar, high-performing traits. Missing a few individuals of low-abundance, rare species will mathematically and ecologically leave the overall redundancy baseline of the macrofaunal community unaltered.

We have added a brief discussion in the revised manuscript (line 554) to address this aspect: “Therefore, species richness and functional richness are typically strongly linked with each other. Single-replicate sampling might theoretically lead to a slight underestimation of functional richness by potentially missing rare species. However, given the environmental

homogeneity of the study area, most rare taxa are likely to support common and redundant functions rather than unique ones (Mouillot et al., 2013).”

Comment 1.24: The introduction especially at the end focused quite a lot on carbon dynamics but this is a bit lacking in the discussion. There is a sentence in the conclusion stating that “...a direct measurement that would allow for linking macrofauna community to carbon sequestration is still missing” and that is good and it is true that you did not directly measure this. But there could be a concluding sentence of which compartment, according to your calculations, had more respiration and which one had more production etc.

Response 1.24: We thank the reviewer for this constructive feedback. We agree that while our Abstract highlighted the contrasting spatial patterns of ecosystem functions, the Conclusion section lacked an explicit summary.

The paragraph in the Conclusion section (line 609) has been revised as: “Together, these patterns suggest that changes in biodiversity may have important consequences for benthic ecosystem functioning and, ultimately, for carbon storage capacity (Ramalho et al., 2020). Specifically, our calculations show that carbon processing in our study region is split between distinct ecosystem functions: shallower communities display high secondary production and respiration, whereas deeper communities are characterized rather by elevated bioturbation and bioirrigation potentials.”

Reviewer 2

Comment 2.1: The authors present the results of an extensive study that aimed at (1) sampling benthic macrofauna with a wide range of sedimentary habitats of the southeastern North Sea, and (2) translate these macrofauna community data (at both the species and functional trait levels) into several proxies/indices of ecosystem functioning. If the study seems well conducted and the dataset valuable, the article generally suffers from a certain gap between the announced objectives from the title (carbon cycling in a depocenter) and the actual content of the paper that to my point of view more deals with the spatial variability of community structure and functional trait composition across a sedimentary/depth gradient. At this light, Introduction could be view simplified, and more strait to the point. The method

section needs to be extensively reviewed. This stands for the whole data acquisition process as well as the statistical analyses performed that clearly need to be explained in more details and/or justified. Please see below a few detailed comments:

Response 2.1: We sincerely thank the reviewer for the overall positive feedback regarding the value of our dataset and the conduction of this study.

Regarding the primary critique, i.e., the "gap" between the macro-scale implications (carbon cycling) and our empirical focus (spatial/functional trait variability), we would like to clarify the underlying concept for our framework. The study area, which is also referred to as the Helgoland Mud Area, is currently subject to intense investigations because of its role as major depocenter for carbon in the southeastern North Sea and its corresponding importance in the context of global climate change. Accordingly, our intention was to link our work to this important role of the study region without, however, providing direct measurements of biogeochemical fluxes of carbon sequestration, which has been done elsewhere (e.g. Müller et al., 2025, <https://doi.org/10.5194/bg-22-2541-2025>; Wei et al., 2025, <https://doi.org/10.1016/j.chemgeo.2025.122712>). Instead, our aim was to investigate the spatial dynamics of the specific ecosystem functions related to benthic carbon dynamics. The indices and proxies evaluated in this study, namely bioturbation, bioirrigation, secondary production, and respiration, are all ecosystem functions fundamentally related to the processing of carbon in marine sediments.

To address your concern and bridge this conceptual gap, we have modified the Introduction to better work out the actual focus of our study, which is the structuring of the benthic communities and the relationship between biodiversity and ecosystem functioning.

Nevertheless, we did not entirely give up the conceptual link to carbon sequestration, which we consider a solid justification and framing of our study given the importance of the study area as a major carbon sink in the southeastern North Sea. Instead, we have simplified and shorten the section in line 105-128 by 14% of the word counts in order to lower the emphasis on the topic of benthic carbon sequestration, which we did not measure explicitly. The condensed section of the Introduction now reads in lines 105: “We explored the relation between biodiversity and ecosystem functioning for the benthic macrofauna in an extensive area of muddy sediments in the southeastern North Sea. This area has been identified as a major depocenter for fine-grained material with considerable potential for benthic carbon sequestration (Müller et al. 2025; Wei et al. 2025). Recent geochemical investigations have

revealed the distinctive role of mud depocenters on sandy shelves as potential hotspots for carbon burial and their potential relevance in the context of global climate change (Wei et al., 2025). Fine-grained sediments accumulate primarily off the mouths of major rivers (Müller et al., 2025), driven by specific hydrodynamic conditions that promote sediment trapping and retention (Chen et al., 2025). As a result, these areas exhibit rapid organic matter deposition and exceptionally high carbon burial efficiency, highlighting the biogeochemical significance of local depocenters (Hebbeln et al., 2003; Wei et al., 2025; Müller et al., 2025). While the geochemical importance of the muddy sediments as a carbon sink is well established, a substantial portion of research historically focused on physical and sedimentological processes. Growing attention has been directed toward understanding how benthic macrofauna community composition varies spatially within depocenters (e.g., Kauppi et al., 2017; Bonifacio et al., 2018; Lamarque et al., 2022, 2025). Despite these valuable insights from various marine systems, characterizing the specific spatial variations and the relative contributions of benthic macrofauna in regulating carbon cycling within localized shelf-sea depocenters requires further resolution. Therefore, our study on the relationship between biodiversity and ecosystem functioning specifically focusses on functions of the benthic macrofauna related to carbon cycling. Benthic macrofauna contributes to carbon cycling in multiple, potentially antagonistic ways by incorporating organic carbon into secondary production, respiratory release of carbon as CO₂, and modifying sediment structure through bioturbation and bioirrigation. Through these activities, macrofauna redistribute organic matter in sediments and either promote or impede its remineralization, and thus affect the balance between carbon burial and release. Specifically, our objectives are: (1) to characterize the taxonomic and functional trait composition of the benthic macrofauna community across the muddy study region, (2) to identify environmental factors shaping the benthic macrofaunal community composition, and (3) to explore the relationships between biodiversity and ecosystem functions related to biogeochemical processes and their variations along environmental gradients.”

We have also thoroughly modified the Materials and Methods section according to the recommendations by the reviewer. Specifically, we have provided details regarding the data and analyses as required. Below, we provide point-by-point responses to each of your specific comments.

Comment 2.2: L 18-19: have you actually quantitatively measured proxy for ecosystem functions? It doesn't seem to be the case, the sentence has then to be modified

Response 2.2: We agree with the reviewer's comment. As Reviewer 1 also raised this question (please see our response to Comment 1.3), we would like to refer to our above response and the suggested revision of the text.

The revised sentence now in line 17 reads: "Here, we investigated how benthic community structure, functional traits, and the relationship between biodiversity and calculated potential ecosystem functioning related to carbon cycling vary along environmental gradients in muddy sediments of the southeastern North Sea."

Comment 2.3: L56-57: I would talk about diversity rather than just "species richness" since you already dealt with the species richness to productivity relationship in the previous paragraph. Moreover, it is important to notice here that environmental gradients can also shape single species trait expression, i.e. the way they could contribute to ecosystem functioning, see for example Needham et al. 2011. (<https://doi.org/10.1007/s10021-011-9468-0>) or more simply that sediment characteristics can modify the burrowing behavior and depth in several key species (see Wiesebron et al. 2021, <https://doi.org/10.3389/fmars.2021.707785>)

Response 2.3: We thank the reviewer for the suggestion, which supports the theoretical background of our Introduction. We have revised the section to incorporate this intra-specific/behavioral pathway and have integrated the recommended insights from Needham et al. (2011) and Wiesebron et al. (2021).

The revised section now reads in lines 60-71: "The relationship between biodiversity and ecosystem functions is not static but may vary in strength and direction along environmental gradients. In marine benthic systems, gradients in depth, sediment properties, and hydrodynamic forcing can alter community composition and functional trait distributions (van der Wal et al., 2017; van Son et al., 2013), thus regulating how biodiversity translates into ecosystem functioning. For example, the relationship between benthic macrofauna diversity and functions related to carbon fixation and mineralization depended on the levels of nutrients and turbidity in coastal ecosystems of New Zealand (Gammal et al., 2023). Such dependency can also result from environmental gradients shaping single-species trait expression and macrofaunal behavior. For instance, burrowing species behave differently in

response to variations in sediment characteristics (Needham et al., 2011) by exhibiting faster burrowing and more intense bioturbation in soft, low-bulk-density sediments (Wiesebron et al., 2021). Although numerous studies have examined patterns of biodiversity along environmental gradients, the extent to which environmental gradients shape BEF relationships remains unclear. Addressing this gap, we examine how BEF relationships in marine benthic macrofauna vary across sedimentary and hydrodynamic gradients.”

Comment 2.4: L110-113: This affirmation about a certain lack of studies focusing on biological aspects of depocenters and their links to ecosystem functioning needs to be downplayed. Although not from the North Sea a certain amount of studies have focused on these aspects, see for example Bonifacio et al. 2019 (<https://doi.org/10.1016/j.seares.2017.08.013>), Lamarque et al. 2022, 2025 (<https://doi.org/10.1016/j.csr.2022.104833>, <https://doi.org/10.1016/j.csr.2025.105637>), or Kauppi et al. 2017 (<https://doi.org/10.3354/meps12171>).

Response 2.4: We appreciate the reviewer's suggestion on this point and agree that our original statement was overstated. Following your recommendation, we have downplayed this claim in the revised manuscript and integrate the works of Bonifacio et al. (2019), Kauppi et al. (2017), and Lamarque et al. (2022, 2025) to illustrate the existing global foundations of benthic processing in organic-rich sediments.

The revised section in line 113 now reads: “While the geochemical importance of the muddy sediments as a carbon sink is well established, a substantial portion of research historically focused on physical and sedimentological processes. Growing attention has been directed toward understanding how benthic macrofauna community composition varies spatially within depocenters (e.g., Kauppi et al., 2017; Bonifacio et al., 2018; Lamarque et al., 2022, 2025). Despite these valuable insights from various marine systems, characterizing the specific spatial variations and the relative contributions of benthic macrofauna in regulating carbon cycling within localized shelf-sea depocenters requires further resolution. Therefore, our study on the relationship between biodiversity and ecosystem functioning specifically focuses on functions of the benthic macrofauna related to carbon cycling”

Comment 2.5: L117-118: “To address this knowledge gap, we investigated how structural and

functional aspects of benthic macrofaunal diversity regulate these processes linked to carbon sequestration” It’s see

Response 2.5: It appears that this comment may have been left incomplete. If the reviewer intended to inquire further about whether our study fully addresses how macrofaunal diversity regulates these carbon-process-related functions, we would like to reaffirm that our framework explicitly addresses this mechanism. Throughout the manuscript, we demonstrate that benthic macrofaunal diversity does not merely correlate with environmental variables, but actively modulates ecosystem functioning across the shelf gradient. Indeed, as also suggested by Reviewer 1 (please see our response to Comment 1.24), we have revised the Conclusion section with an explicit paragraph synthesizing the carbon dynamics.

Comment 2.6: L229: specify the Anova design used

Response 2.6: While the four clusters were defined by their taxonomic community structure (taxa composition and transformed abundance matrix), a one-way ANOVA was used to test whether these abundance-based structural differences led to variations in total community biomass. To examine whether mean AFDM differed significantly among the four identified faunal clusters, a one-way ANOVA was performed with 'Cluster group' as a four-level categorical factor. Prior to analysis, data were evaluated for normality (Shapiro-Wilk test) and homoscedasticity (Levene's test). All corresponding statistical outputs, including degrees of freedom, F-value and p-value are detailed in Table S2 of the Supplementary Material.

To clarify this for the reader, we have modified the sentence starting in line 245 of the revised manuscript to specify the ANOVA design: “To evaluate whether the structural differences captured by our abundance-based clustering led to variations in total community biomass, we conducted a one-way Analysis of Variance (ANOVA) with 'Cluster group' as a four-level categorical factor. The ANOVA tested for significant differences in total community biomass (mean AFDM) across the four identified infauna clusters. Prior to modelling, normality and homoscedasticity of the biomass data were verified using the Shapiro-Wilk test and Levene’s test, respectively.”

Comment 2.7: L229-230: What has been exactly tested using Permanova? If you have tested

if the groups defined by your clustering step were different based on the very same multivariate data, this would be somehow a typical circular reasoning.

Response 2.7: We understand the concern that using the same multivariate species dataset for both clustering and subsequent PERMANOVA significance testing can lead to circular reasoning and thus artificially inflate the significance. As Reviewer 1 also raised the same question, please refer to our comprehensive response to Comment 1.16. In brief, we have omitted the PERMANOVA to avoid circular reasoning in line with the recommendations by both reviewers.

Comment 2.8: L232-235: more details about the distm must be given here, type of distance/similarity used for both datasets? Model selection criterion used? Selection procedure? Preliminary collinearity assessment and potential transformations done on environmental variables?

Response 2.8: We thank the reviewer for this comment regarding the statistical approach of our distance-based redundancy analysis (dbRDA).

Regarding the distance metric, a fourth-root transformation was performed on the biomass data to downweigh the influence of highly dominant species, and a Bray-Curtis dissimilarity matrix was subsequently constructed using the vegdist function in R. For the environmental dataset, following standard dbRDA protocols, no distance matrix was calculated. Instead, the variables were directly utilized as linear constraints in the model.

Regarding the preliminary collinearity assessment, we appreciate the reviewer's sharp and insightful reminder. We admit that this was initially overlooked in our draft. In light of your comment, we completely reorganized and optimized our dbRDA approach. To eliminate severe multicollinearity, we exclude the overlapping grain size metrics (D10, D90) as they are inherently self-correlated. We also exclude the compositional sediment percentages (Sand, Silt, and Clay) because they are highly correlated to D50. Moreover, since Depth is positively correlated to Salinity and negatively correlated to Bottom shear stress, it is also excluded in dbRDA. We calculate the Variance Inflation Factor (VIF) values for all remaining environmental variables and ensure they drop to below 5. The remaining variables are Sorting, Skewness, Kurtosis, D50, OC, Salinity and Bottom shear stress.

For the model selection, we have retained all remaining environmental variables in this updated model without stepwise deletion, which allowed us to evaluate the combined effect of the entire environmental setting. Finally, we have verified the statistical significance of this overall model and each individual environmental term using permutation tests with 9999 randomizations. In summary, the revised methods support the conclusions drawn from the previous analysis. Accordingly, the revision of the methods does not require any corresponding changes in the Discussion or in summarizing sections like the Abstract.

We have updated the Materials and Methods section, along with the corresponding results and discussion in the revised manuscript to explicitly address this problem.

The revision in Materials and Methods, section 2.4 (Line 249) now reads: “A distance-based redundancy analysis (dbRDA) was performed to identify the environmental variables that best explained community dissimilarity. The initially selected environmental variables comprised water depth, modelled bottom salinity, bottom shear stress, and sediment-derived properties, including organic carbon content (OC), compositional percentages (sand, silt, and clay), grain size (D10, D50, D90), and statistical distribution measures (sorting, skewness, and kurtosis). To avoid severe multicollinearity within the dbRDA, a variable reduction protocol was implemented based on correlation analysis and Variance Inflation Factors (VIF). First, overlapping grain size metrics were excluded due to inherent self-correlation. Second, sediment compositional percentages (sand, silt, and clay) were omitted because they were highly collinear with median grain size (D50). Furthermore, depth was excluded because it was strongly correlated with both bottom salinity and bottom shear stress. Finally, VIF values were calculated for all remaining predictors to ensure that all selected variables dropped well below the conservative threshold of 5 (Zuur et al., 2010). Consequently, the optimized dbRDA model retained seven independent environmental drivers: sorting, skewness, kurtosis, D50, OC, salinity, and bottom shear stress.”

The revision in Results section (Line 296) reads: “In the dbRDA model, the selected environmental factors explained 44.1% of the overall variation in community structure, with the first three dbRDA axes accounting for 64.5%, 17.0%, and 6.4% of the explained variation, respectively (Fig. 2 and TableS3). A permutation test ($n = 9999$) indicated that the overall model was statistically significant (pseudo- $F = 4.73$, $p < 0.01$). Projection of the four clusters onto the first and third dbRDA axes differentiated Cluster 1 from Cluster 3 and 4. It reflected a gradient in the community composition primarily along the East–West axis.

Sorting, skewness, kurtosis, D50, salinity and bottom shear stress were identified as significant predictors of community composition ($p < 0.01$, Table S3). The first dbRDA axis indicated that communities in shallow regions were strongly associated with higher bottom shear stress, whereas communities in Clusters 1 and 2, located in deeper areas, were more strongly associated with higher salinity, suggesting that these two environmental gradients shaped the benthic community. Greater kurtosis was particularly associated with Cluster 2, while sediments in Cluster 4 were distinguished by greater sorting, indicating that sediment characteristics significantly promoted the distinction between these two clusters.”

Comment 2.9: L240: Wouldn't be clearer to talk about “derived ecosystem functions”, since no function have actually be measured?

Response 2.9: We completely agree with the reviewer's point. As also suggested by Reviewer 1 (please see our response to Comment 1.3 and Comment 2.2), we have applied the same modifications to the revised manuscript as described above.

Comment 2.10: L252-254: I'm not sure this makes sense, see my comment above.

Response 2.10: We understand the concern that using the same multivariate species dataset for both clustering and subsequent PERMANOVA significance testing can lead to circular reasoning and thus artificially inflate the significance. As Reviewer 1 also proposed the same question, please these our comprehensive response to Comment 1.16.

Comment 2.11: L271-282: This paragraph needs more precision (in relation with my comment above). It appears also important to state the order (and so the selection procedure) the variables are entered into the model. As well, it seems that there are some collinearities so it would be essential the authors state why they could have left some of them out. For example, salinity and depth appear correlated, but both included as potential explanatory variables. Therefore, as I read this, it seems that depth is not significant in the model because of the large amount of variance shared with salinity that may have been entered before in the model. But if salinity was not considered as explanatory variable, there might be a large chance that depth would be selected and tested as significant.

Response 2.11: We agree with the reviewer's concern that correlated variables can both be included as potential explanatory variables. As replied to Comment 2.8, we reorganized and redid the dbRDA. For the proposed revision of Materials and Methods and Results, please see our comprehensive response to Comment 2.8.

Comment 2.12: Figure 2: please indicate what the ellipses stand for

Response 2.12: We thank the reviewer for pointing to this missing information. In the original manuscript, these ellipses represented a 50% probability contour based on a multivariate normal distribution. Statistically, this means that any newly sampled station from that specific community would have a 50% probability of falling inside this ellipse. We chose a 50% probability level instead of the standard 95% to clearly contrast the core distributions of the four clusters while preventing the ellipses from overlapping and cluttering the graph.

However, following the complete reorganization and optimization of our dbRDA framework (including the removal of collinear environmental variables as detailed in Response 2.8), Figure 2 has been entirely redrawn. In the updated version of the figure, we have removed the ellipses altogether to maintain visual clarity. Therefore, this indication is no longer required in the figure caption.

Comment 2.13: L395: Highlighting carbon sequestration is a bit misleading since the link between the results presented and carbon sequestration is not that explicit. It needs to be downplayed/

Response 2.13: We agree with the reviewer that linking our calculated potentials directly to "carbon sequestration" was an overstatement and potentially misleading for the readers. We downplayed this aspect in the concluding paragraph by replacing "carbon sequestration" with "ecosystem functions linked to the cycling of organic matter and, thus, carbon in marine sediments", which more accurately reflects what secondary production and respiration estimates imply.

The revised sentence (Line 416) now reads: "Bottom shear stress and salinity likely modulated the observed biodiversity–ecosystem function relationships, disentangling the complex interplay among diversity, ecosystem functioning, and environmental variables, and

demonstrating how benthic macrofaunal diversity regulates ecosystem functions linked to the cycling of organic matter and, thus, carbon in marine sediments.”

Comment 2.14: L408: More context should be made to give these examples of experimental assessments, e.g. temperature, sediment type... As well, much more authors have measured sediment reworking in these species group, so a wider view of the literature would help in depicting generalities.

Response 2.14: We fully agree with the reviewer. We added context for the mentioned studies and expanded the literature review to improve the generality and robustness of this discussion.

This paragraph (line 430) now reads: “In contrast, the communities in the shallower parts of the region were dominated by bivalves (e.g., *Abra* spp.) and polychaetes (e.g., Nereididae). Experimentally derived mixing rates under controlled incubation at 10°C in surface sediments from the Gullmarsfjord (northeastern North Sea) indicate that *Abra nitida* exhibits a much lower baseline sediment-mixing intensity of approximately 0.5 cm² y⁻¹ (Lindqvist et al., 2016).

Nereidid polychaetes dominating these shallow communities, such as *Eunereis longissima* and *Nereis elitoralis*, have been characterized as surface deposit feeders that feed preferentially from the openings of their built tubes, resulting in discretely motile behavior (Fauchald and Jumars, 1979). Similarly, observations of *Alitta virens* in large glass aquaria kept at 12°C in both coarse sand and organic-rich mud from Conception Bay South, Newfoundland, Canada, demonstrated that these nereidids consistently perform shallow burrowing and near-surface deposit feeding across the entire 90-day observation period (Herringshaw et al., 2010).”

Comment 2.15: L437-439: see my comment above. It must be better precise in the mat met and results sections how this result arose

Response 2.15: As suggested by the Reviewer, we reorganized and rerun the dbRDA. For the proposed revision of Materials and Methods and Results, please see our comprehensive response to Comment 2.8. Here we suggest a revision in Discussion (section 4.2) based on the new results of dbRDA.

The new paragraph (line 469) now reads: “Sediment sorting, skewness, kurtosis, D50, salinity, and bottom shear stress clearly differentiated the four clusters in the dbRDA. Although water depth was excluded prior to modelling due to severe multi-collinearity with salinity and bottom shear stress, these remaining hydrodynamic and sediment characteristics captured the overall depth-related environmental gradient that shapes the benthic communities.”

Comment 2.16: L441-446: Also, it must be stated that flow velocity, as well as suspended particles, can induce a switch between deposit and suspension feeding in some species, including *A. filiformis* (see Loo et al 1996, MEPS), that can have consequences on bioturbation intensity

Response 2.16: The reviewer is correct that high flow velocities and increased suspended particle concentrations can induce a behavioral switch from deposit feeding to suspension feeding, which reflects on bioturbation intensity. We have revised the section in lines 475-480 to integrate this switch among feeding mechanisms alongside the physical disturbance effects.

However, as the reviewer mentioned to consider Loo et al. (1996), we also noted the authors' explicit conclusion that *predicting ingestion rates of passive suspension feeders is particularly difficult in the field*. This highlights the logistical necessity and robustness of using a trait-based potential index.

The revision (line 475) reads: “For example, Rowden et al. (1998) reported that the abundance of *A. filiformis* was negatively related to sediment surface current velocity, likely because of intense sediment erosion (Jago et al., 1993). Furthermore, elevated flow velocities and suspended particle concentrations can induce switches among feeding modes. For instance, *A. filiformis* can shift from deposit feeding to passive suspension feeding under active hydrodynamic conditions (Loo et al., 1996). This behavioral shift reduces their direct mechanical interaction with the seabed, which might further suppress community-level bioturbation intensity.”

Comment 2.17: L445-446: ok but bottom sheer stress also strongly influence grain size and

especially its variability as in figure 2 with the opposition of shear stress and kurtosis, so the structure of sediment might matter too?

Response 2.17: We agree with the reviewer that bottom shear stress does not act in isolation. It is the primary hydrodynamic driver shaping sediment characteristics, including grain size and kurtosis, as clearly reflected in the opposing vectors in our Figure 2.

We agree that the resulting sediment structure and stability are critical factors influencing macrofaunal colonization. High bottom shear stress creates a high-energy environment that alters sediment composition, which in turn creates unstable conditions that are unfavorable for large, deep-burrowing organisms. Therefore, both the direct physical disturbance of shear stress and the indirect modification of sediment structure suppress these functional traits. We revised the sentence to incorporate the role of shear-stress-induced sediment modification.

The revised sentence now reads in lines 480-483: “The combined analysis of functional traits and environmental factors indicated that bottom shear stress, along with its profound influence on shaping sediment structure (e.g., grain size and kurtosis), suppressed the abundance of strong bioturbators in the shallow cluster, which ultimately led to depressed bioturbation and bioirrigation potentials.”