Reply to reviewer Paul Renaud

We would like to thank Paul Renaud for the constructive comments on our manuscript 'Deep-sea benthic communities and oxygen fluxes in the Arctic Fram Strait controlled by sea-ice cover and water depth'. We will first address the reviewer's 'specific comments' and secondly reply to 'technical corrections/detail' with stating the planned improvements.

In the following, author responses starts with the term 'Reply' and changes, that will be included in the manuscript, are given in blue.

Specific comments:

1. Time scales of response. Direct (and only linear) correlations between environmental parameters and O2 flux may be misleading, or non-representative, depending on when the samples are taken (and when relative to the bloom/flux phenologies among different stations). Benthic biomass/density/structure likely respond to various factors (especially food-related parameters which are often covariates of depth) in a more seasonally (or up to decadally) integrated fashion, whereas O2 consumption/C remineralization are often more responsive to food inputs on a much shorter time scale (approx. weeks) (e.g. Renaud et al. 2008 DSR II). This must be considered in your interpretation.

Reply: We thank the reviewer for this important aspect. As spring bloom data in this region for the studied period (2014-2015) are not yet citable and the authors are no experts in satellite data acquisition and analyses, we will cite the results of Cherkasheva et al. 2014 (http://dx.doi.org/10.1016/j.jmarsys.2013.11.008) to provide information regarding the date of the spring bloom. Furthermore, we will acknowledge that correlations do not necessarily prove causal relationships and that oxygen flux measurements only represent a temporal snapshot. As we did not perform measurements during or right after the bloom deposition, we might have missed the immediate and short-term reaction of the benthos to the fresh organic matter. Nevertheless, we can expect a lower influence of macrofauna on the measured oxygen fluxes, compared to the findings in Renaud et al. 2008 (DSR II), owing to the high contribution of microbial benthic mineralization to the total benthic mineralization in the deep sea (Donis et al., 2016, Sauter et al. 2001, Wenzhöfer and Glud,2002), which is also expressed by the mean DOU/TOU ratio of 0.79 presented in our results.

2. As noted in the Methods, the difference between the eastern and western Fram Strait and potential consequences for benthic processes go beyond ice cover. Advected POC/PON/dissolved nutrients and warmer temperatures on the eastern side are far greater than on the western side. Of course that is linked to why the ice is there, but in this case, ice is more of a covariate and perhaps less likely a causative factors. In addition, different zooplankton and microbial communities can well lead to different 'food' deposition. This must be considered in detail if the two transects are to be comparable.

Reply: We agree with the reviewer and will discuss the complexity of advective and vertical pelagic food input influencing processes in Fram Strait in more detail. Furthermore, we will point out that we used the parameter 'sea-ice cover' as a proxy for primary production patterns. The sea-ice in the western Fram Strait represents a suppressed light availability and a reduced nutrient supply (owing to the main currents WSC and EGC). Both light availability and nutrient supply are the main drivers of primary production. This suits the findings of Pabi et al. (2008, doi:10.1029/2007JC004578), showing contrasting primary production quantities among the western and eastern Fram Strait.

3. Methods: it appears that most of the variables measured were only assessed from the top 1 cm of sediment. Can you provide a justification (data-based) for this? For meiofauna, it is often the top 2-3 cm that contains the majority of the fauna, and for macrofauna, at least the top 5 cm, even at deep-sea depths.

Reply: Indeed, microbial and meiofauna data were assessed from the top 1cm. Macrofauna data and the biogenic sediment compounds, however, were assessed from the top 5cm (MUC cores) and from even deeper sediments (benthic chamber sampled sediments). We will improve the method section to clarify this. Regarding the Meiofauna, we refer to Gorska et al. 2014 (http://dx.doi.org/10.1016/j.dsr.2014.05.010) and regarding microbial data, we refer to Quéric et al., 2004 (https://doi.org/10.1016/j.mimet.2004.02.005). Both studies show that most of the investigated organisms in the Hausgarten area occur in the top 1 cm.

4. Ice cover in the two 'regions' is essentially 70-80% vs 1-10% (heavy ice/no ice). Except for EG V (and N5 which is often excluded from analysis), there is nothing in between. How might this affect your results/interpretation? Many of the results from N5 are more similar to the LSC than the HSC stations (see comment 2 above).

Reply: We will acknowledge in our discussion, that comparing only two sites (heavy ice/no ice) does not allow us to estimate the actual relationship between ice cover and the response variables. We would like to point out that only station SV I station was often excluded from analyses, owing to its exceptional shallow water depth, compared to all other stations. The introduction of a third category (intermediate ice) would only be based on two stations (EG IV and N5). As these two stations are from the same water depth, they would not include the potential impact of water depth, which was identified as important and therefore would weaken the outcome of this approach.

5. Ice cover as the key factor. Related to comment 2 above, have you evaluated whether correlations/differences between benthic parameters and ice cover are the strongest relationships among your data? Primary productivity, vertical flux attenuation, and essentially food supply to the sediment surface may or may not be caused by sea ice in any way. Or it could be a feature of Arctic vs Atlantic water supply that causes a 'cascade of processes' and sea ice cover may just be a covariate with limited or even no direct causative effect (hence a logic problem on p 15 l 23-25). Your discussion implies that ice is the overriding factor but I do not see where you tested for this, or if it is even possible to disentangle all these variables to isolate depth as the key factor. If you ran similar analyses but grouped stations based on water mass characteristics instead of ice cover you would find the same result.

Reply: We verified whether correlations/differences between benthic parameters and ice cover are the strongest relationships among your data by running the PCA. The eigenvalues indicated that 'TOC', 'Chl a' and 'Macrofauna biomass' were responsible for the gradient along the x-axis and 'water depth', organic matter' and 'sea-ice concentration' for the gradient along the y-axis. However, 'sea-ice' is a proxy for light availability and nutrient supply in Fram Strait and therefore represents primary production, whereas water depth represents pelagic mineralization and therefore the fade of organic matter in the pelagic zone. Both process are responsible for the final 'TOC' and 'Chl a' concentrations at the seafloor. To make this clearer, we will include the proxy characterization in the method section, add the eigenvalues of the PCA as to the manuscript and integrate our argumentation in the discussion.

6. P 12 l 6: unclear what water column nutrients, presented as a snapshot without context of 'preformed' (winter) concentrations add here. Bloom phenology certainly is responsible for e.g. the lower nitrate in WS vs EG. Consider removing these data. The discussion on p 15-16 and then sec 4.2 is not really based

on the data collected, but more of a general pattern documented in the literature. I agree some of this should be included, but wouldn't a more extended and balanced discussion of benthic process rates and the other factors (proximal) responsible for variation in these rates be appropriate here?

Reply: We agree with the reviewer and we will remove the nutrient data from the manuscript. Further, we will discuss benthic mineralization and the other proximal factors responsible for variation in these rates in more detail.

7. Nutrient supply under the ice in EG is extremely low and not expected to increase with further melting of sea ice (e.g. Mauritzen et al. 2011 Prog Oceanogr). This casts serious doubt into any scenario where increased PP due to more light is invoked.

Reply: We will rewrite the potential future scenario, include spatial limitations and will point out that this scenario only holds true for areas, where sea-ice disappears and nutrient supply will increase. It will be changed to 'Our scenario is only suitable if sea-ice disappears and nutrient supply increase, which will result in enhanced primary production The development of future Arctic Ocean primary production patterns and changes is still under debate (Wassmann, 2011, Arrigo et al., 2012; Nicolaus et al., 2012, Boetius et al., 2013). However, it is likely that the described scenario becomes true in the Chukchi Sea and the Beaufort Sea, owing to the predicted strengthening of the nutrient rich Pacific inflow (Harada, 2015). Furthermore, owing to an increased atlantification, an increased nutrient supply is also likely for the continental margin at the Barents Sea (Neukermans et al., 2018). In addition, nutrient inflow by glacial and permafrost soil melt is also predicted to increase (Vonk et al., 2015). However, this riverine load might only enhance primary production at the shelf areas and therefore is not relevant for the deep sea. An enhanced primary production in the western Fram Strait is unlikely even if the light availability will increase, as the required nutrient supply increase is not expected for this region (Mauritzen et al., 2011).'

Technical corrections/details

1. P4 l6-9: unclear sentence. Perhaps just unnecessary (same for l 14-17 as it just repeats what you have just written)

Reply: The sentence will be removed.

2. P4, l 24: controlling the benthic ecosystem? Be more specific, including what you mean by 'labile organic matter' (different from benthic chlorophyll?)

Reply: The sentence will be changed to "However, the principal factor controlling microbial activity was most likely the supply of labile organic matter such as CPE, proteins and dissolved free amino acids (Boetius and Damm 1998)."

3. P6 l 10: if the algorithm can estimate ice cover at over 100% then couldn't values between 0 and 100 also be mismeasurements? Could there be some (automated) check to assure that adjacent pixels are 'similar', or some other way of testing for mismeasurement in this range?

Reply: Whenever there was a mismeasurement, the algorithm output was "128". So it is not the case that a sea-ice concentration of 101% or 105% or 112% and so on, could be measured. Therefore, the algorithm does not estimate ice cover over 100%. The sentence will be rewritten to make this point that clearer. Furthermore, we will add information regarding the quantity of these mismeasurements.

4. Also, is an annual average (vs some other ice cover parameter) the most relevant measure of ice cover?

Reply: We will provide some alternative sea-ice concentration periods (mean of 1-month before sampling, mean since first of May (assumed spring bloom onset) till sampling) for the reader. However, as pointed out by the reviewer, “benthic biomass/density/structure likely responds to various factors (especially food-related parameters which are often covariates of depth) in a more seasonally integrated fashion”. To acknowledge this, we used the annual sea-ice cover in the PCA.

5. P 6 l 25: frozen not frosted

Reply: We will follow the suggestion of the reviewer and change “frosted” to “frozen”.

6. P 8 l 25ff: Was non-local mixing (i.e. non-linear profiles) observed? How was this accounted for in the O2 flux calculation?

Reply: Non-local mixing was observed in some cases and therefore the reported DOUs for those cases are underestimations. However, only eight out of 81 ex situ obtained oxygen microprofiles at various stations and in one out of 34 in situ obtained oxygen microprofiles showed signs of non-local mixing. This information will be added to the method section.

7. P 9 l 1-5: How much of the sediment mass could be attributed to salt from the drying process?

Reply: It was 4.5% ± 1.9 over all samples. We will add this information to the method section.

8. P 10 l 5: consecutive not subsequent

Reply: We will follow the reviewers’ suggestions and change ‘subsequent’ to ‘consecutive’

9. P 10 l 14: ’x to zero mean and unit variance’ is unclear

Reply: In most applications of a PCA (e.g. as a factor analysis technique), variables are often measured in different units. For such data, the data must be standardized to zero mean and unit variance, a common standardization procedure. If this is not done, high values (e.g. macrofauna biomass with values of ten thousands of mg m^{-2}) will get a greater importance than low values (e.g. DOU with values of max. 2.1 mmol O2 $m^{-2}d^{-1}$). Similar terms used for this procedure are ‘data normalization’ or ‘z-scoring’. Though, as we followed the suggestions provided by Buttigieg and Ramette (2014), we decided to follow their term of ‘standardization’ (https://mb3is.megx.net/gustame/indirect-gradient-analysis/pca).

10. P 10 l 21: you must exclude EG II from the analysis. You cannot make the assumption and assign a value. It was fine to exclude the shallow station, and you should do the same with EG II

Reply: We agree with the reviewers’ comment that the assumption of a solute exchange value for EG II is not a valid approach to deal with data gaps. However, as the other parameters included in the PCA from EG II were actually measured, we rather prefer to perform the PCA without the parameters of ‘solute exchange’ from all stations. With this suggestion, EG II would still be part of the central analysis of the paper. Furthermore, as ‘solute exchange’ is well correlated with other parameters such as macrofauna biomass, it will still be represented in the PCA.

11. P 10 l 30: You need to indicate whether there was a different depth relationship between the two regions and then say what you did if this was (or was not) the case.

Reply: We will add the information, that the depth relationships in terms of the bottom slope were similar between the two regions to the method section ('study site'). Furthermore, we will add to the results section, that water content, phaeo, and CPE showed a similar water depth relationships within the HSC and LSC categories (Figure S4) compared to the water depth relationship of DOU (Figure 4). We will discuss that the microbial mineralization is the main driver of benthic deep-sea mineralization (see reply to reviewers' specific comment no°1). We will further add to the discussion that microbial density did not show differences between the HSC and LSC categories and therefore was not the biotic link which connected the food input pattern (Figure S4) with the mineralization pattern (Figure 4). There are no studies regarding benthic microbial biomasses or community structure across Fram Strait. Although we identified parameters well correlated with mineralization processes, our study analyses are not able to fully explain the contrasting mineralization pattern between the western and the eastern Fram Strait.

12. P 12 l 6: unclear what water column nutrients, presented as a snapshot without context of 'preformed' (winter) concentrations add here. Bloom phenology certainly is responsible for e.g. the lower nitrate in WS vs EG. Consider removing these data. The discussion on p 15-16 is not really based on the data collected, but more of a general pattern documented in the literature.

Reply: We will follow the reviewers' suggestion and remove the nutrient data from the manuscript.

13. P 12 l 10ff (and Fig 3). Please indicate any statistical results such that the figures correspond to what is written in the text regarding comparisons between the two regions. Only statistically significant results should be expressed as 'differences' (e.g. solute exchange is likely NS but significance is implied). Also, please clarify how many stations (and depth profile) each bar represents. This has some bearing on your comments about variability between the two locations.

Reply: We will improve the text and use the term 'differences' only in case of significant differences. Further, we will indicate significant differences between the stations in the figure and add the number of observations for each bar

As an example, the results of the sediment compounds will read than 'The mean DOU in the EG area ranged between 0.4 ± 0.1 mmol O_2 $m^{-2}d^{-1}$ (n=10) at EG V and 1.0 ± 0.1 mmol O_2 $m^{-2}d^{-1}$ (n=10) at EG II. In the WS area, DOUs at stations within the same water depth range as the EG stations ranged between 0.5 ± 0.2 mmol O_2 $m^{-2}d^{-1}$ (n=8) at HG IV and 2.1 ± 0.6 mmol O_2 $m^{-2}d^{-1}$ (n=8) at SV IV. At the shallow station SV I the DOU reached 3.0 ± 1.7 mmol O_2 $m^{-2}d^{-1}$ (n=6, Table 3). The mean TOU in the EG area ranged between 0.9 ± 0.3 mmol O_2 $m^{-2}d^{-1}$ (n=2) at EG I and 1.6 mmol O_2 $m^{-2}d^{-1}$ (n=1) at EG II. Similar mean TOU values were measured in the WS area, at stations within the same water depth range as the EG stations. TOU values ranged between 0.5 ± 0.2 mmol O_2 $m^{-2}d^{-1}$ (n=5) at HG IV Lander and 1.9 ± 0.6 mmol O_2 $m^{-2}d^{-1}$ (n=5) at HG I. At the shallow SV I station TOU reached 5.1 ± 0.3 mmol O_2 $m^{-2}d^{-1}$ (n=3, Table 3). DOU differed significantly between the WS and EG area, while TOU was similar among the areas (Fig. 3, Supplement Table S4).

14. P 12 l 18: 'pelagic food supply indicating parameter in the sediment' rephrase to clarify that these are sediment values and careful about how you define food quality. Not all organisms eat chlorophyll (in fact

maybe few actually do). Bacteria themselves are likely food for many organisms, and phaeopigments and other OC may also be quite high quality food for others.

Reply: Following the suggestion of the anonymous second reviewer, we add ranges of values regarding benthic food supply representing parameters. See our reply to comment no°13.

15. P13 l 21: 'which indicates bacterial activity and bacterial remineralisation as the major oxygen consumer' please indicate why you conclude this. Why would bacterial oxygen consumption not be reflected in DOU data. These are effectively two different techniques to measure the same thing, each with underlying assumptions. The conclusion you make regarding the ratio is not supported.

Reply: We will rephrase the sentence to 'The mean DOU/TOU ratio, which describes the fraction of the total community mediated oxygen flux (TOU) covered by the microbial mediated oxygen flux (DOU, Glud, 2008,) across the entire Fram Strait was 0.79 ± 0.30, with 0.63 ± 0.22 in the EG area and 0.92 ± 0.30 in the WS area, indicating that the total oxygen uptake is mainly microbially mediated.'

16. P 14 l 7-13: I would focus on the differences among EG and WS stations as revealed by PCA, and not individual variable correlations (which are NOT real correlations but instead are ordination-based relationships! If you want to look for correlation then run that analysis on the raw data).

Reply: We will follow the suggestion of the reviewer and rewrite the paragraph to emphasize the differences among EG and WS stations. It will be changed to 'The stations of the WS and EG area both followed the water depth gradient and shallower stations showed the higher oxygen fluxes. However, stations of the EG area were strongly influenced by the sea-ice cover, contained less organic matter and Chl a, and macrofauna biomass, compared to the WS stations.'

However, the reason to perform the PCA was to reveal the relationships between the multiple parameters. Therefore, we performed a PCA in the scaling II mode, which emphasize the relationships between parameters (Buttigieg and Ramette, 2014, https://doi.org/10.1111/1574-6941.12437). We will add this information to the method section. A performance on raw data as suggested by the reviewer is, however, not recommendable due to the reasons presented in the comment no°9. The correlation of single parameters with each other was already given in Table S2 and Figure S3. In addition, a PCA is a procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. Thus a PCA eliminates redundant information. As it also gives the contribution of the single parameter (=strength of influence) to each principle component, the parameters most likely control the investigated area can be identified (Boetius and Damm, 1998, doi: https://doi.org/10.1016/S0967-0637(97)00052-6).

17. P 16 l 17: but macrofauna biomass has a similar relationship with depth between the two ice-cover systems

Reply: We will rephrase the sentence to 'However, when taking both abiotic factors (sea ice and water depth) into account, the water depth-macrofauna density relationship differed between HSC and LSC (Supplement Fig. S4).'

18. P 16 l 18-20. Repetitive

Reply: We will follow the reviewers' suggestion and remove the repetitive content.

19. P 16 l 27: A CCA or RDA would find significant relationships.

Reply: We tested our data regarding the suitability of the usage of a CCA ('decorana'-command in R package vegan) with the outcome that our data showed linear correlations, so only a RDA would be a suitable approach. Usually, an RDA is used to correlate parameters of two information layers: environmental and biotic data, where the former influences the latter. However, from our point of view, we have three information layers: environmental, biotic, and flux data. It is common knowledge that environmental data influence biotic data and both are influencing the flux data. Therefore, we decided for an indirect ordination by a PCA as the direct approach of an RDA incorporates the measured parameters into the ordination, without considering the possible influence of other, unmeasured parameters.

20. Sec 4.3: first paragraph unnecessary.

Reply: We will follow the reviewers' suggestion and remove the first paragraph.

21. P 19 l 4: neither citation is in the references. Are you sure the Kortsch ref is appropriate?

Reply: The reference of Jones et al., 2014 was added to the references, whereas Kortsch et al., 2012 was removed as suggested by the reviewer and we added the reference of Harada (2015) instead.