Dear Referees,

We have taken a close look at the referees' comments on our manuscript 2013-325 entitled "Meiofauna winners and losers of coastal hypoxia: case study harpacticoid copepods" by M. Grego et al. We addressed all points made, which helped to improve the manuscript in the sense suggested in every case. We thank both referees for their detailed remarks and contribution to our manuscript.

Comments of the referees are indicated in bold

[Anonymous referee #1's comment]:

The paper is generally clear and well written. However there are several typing mistakes and some of the sentences do not read well in English. I therefore suggest a full revision of the language before final submission.

We appreciate the reviewer's suggestion to revise the language of the manuscript before final publication. One of the authors (M. Stachowitsch) is a native speaker and professional scientific English copyeditor: he has re-read the manuscript, condensing and simplifying the text.

[Anonymous referee #1's comment]:

Besides that, I found the paper interesting and original. Nonetheless I have a major concern about methodology used. The first problem concern the lack of replication. Four chambers were used, one for each type of anoxia. therefore results cannot take into consideration the spatial variability of the response.

We thank the referee for the constructive remarks.

The appropriate spatial replication in benthic studies is presented in several books (Giere, 2009; Underwood and Chapman, 2005). The meiofauna is indeed known to have a patchy distribution. This patchiness, however, is typically found on a relatively small scale. Therefore, it was even suggested that a large number (36) of small cores (1 cm^2) should be taken on an area of only 15x15 cm to correctly evaluate the patchiness (Findlay, 1982; in Giere, 2009). While the standard meiofauna core typically encompasses 10 cm^2 of the sediment, the core used in the present study was slightly larger (i.e. 16.6 cm^2 of the sediment), to be able to analyse a greater amount of sediment for living fauna. Anoxia was generated in a volume of 50x50x50 cm (in each of the 4 plexiglas chambers), corresponding to a surface area of 2500 cm^2 each. Within this area, 3 replicate cores were taken for each procedure: anoxia and recovery, with a minimum distance between individual cores of approximately 10cm. Keeping the above-mentioned scale of meiofauna patchiness and their limited motility in mind, we consider these cores as independent cores. Chamber size, seafloor depth, and diving time constraints did not permit any more elaborate sampling protocol.

In the revised manuscript we have now stated that in each deployed chamber and for each treatment (9 days, 1 month, 2 months and 10 months), 3 independent replicate cores were taken. See the revised text in Methods 2.1.:

P 12389, L22: "... Four underwater chambers were deployed, *each for a different anoxic duration. Within each chamber, three independent cores, that were at least 10 cm apart from each other, were taken at the end of each experiment. These cores were treated as replicates. One chamber was deployed* on 2 August and was sampled on 11 August ..."

and

P 12389, L25: "... (on the intact site *triplicate cores were taken*) to monitor the recovery ...."

[Anonymous referee #1's comment]:

A second concern is about comparing different time periods. I appreciate that chamber deployment was done at same time (more or less) but then the evolution of the response should be compared vs. normal condition at the same time. Here, for instance, samples after 1 month are compared to the normal condition of 1 month before. This is important especially because of population dynamics of the species and other time-related variables which are not measured.

The focus in this experiment was the documentation of the different treatments (anoxia duration) on the initial harpacticoid copepod community, i.e., the community at the time of closure of the chamber. We were particularly interested in the potential of survival of different species to short- and longer-term oxygen depletion. The assumption was that the initial assemblage (in normoxic sediment outside the chamber) would maintain its structure and density over this time frame. This assumption is founded on previous research on seasonal variations of harpacticoids from this area (Vrišer, 1996). Moreover, the temperature and oxygen conditions were stable in this time period (see Table I), so that these variables did not impact the original community during our treatments.

Table I.: Bi-monthly monitoring under the oceanographic buoy (24 m depth) of the Marine Biology Station (where the chambers were placed).

Time	DO_ml	Temp	pH_Tc	sal
28.07.2010 10:46	7.369	17.3024	8.10297	37.0933
12.08.2010 09:56	4.467	17.4703	8.07428	37.1736
20.08.2010 12:34	6.96198	17.4782	7.99102	37.0445
26.08.2010 09:33	5.73146	17.1448	8.07909	37.3274
23.09.2010 08:55	6.9193	18.7754	8.12241	36.9316

Table I above was not included in the manuscript, but in order to clarify this issue, we have added the relevant information in the M&M:

P 12390, L1: "*Despite the difference in starting and termination points (3 August 2010 vs 23 September 2010), there was no substantial change in normoxic levels*

and temperature values measured in a monitoring program every 2 weeks at the same site.' The last chamber was deployed on...

[Anonymous referee #1's comment]:

A third point concern the use of ANOVA to analyse these data. What was the model of ANOVA used? How anova was used without replication? How the sex was introduced in the analyses as treatment? Individual inhabiting the same chamber were compared and it is very likely they affect eachothers therefore there are not independent. Therefore one of ANOVA assumptions is violated.

Three independent cores were taken from each chamber, which, for reasons outlined above, were used as replicate samples (see Fig 1). Maybe this misunderstanding originates from the fact that samples with no more live copepods after anoxia (i.e. one replicate of '2 months anoxia' and all replicates of '10 months anoxia') did not show up in the graphs. The factor tested on copepod density with ANOVA was the treatment (normoxia, 9 days anoxia, 7 days recovery, 1 month anoxia, 2 months anoxia).

The referee is correct that 'sex' or sex ratio of copepods should not be considered as a factor to test with ANOVA. It is not a factor to be tested, but it is a dependent variable that can vary as the treatment (induction of anoxia) changes. We sincerely thank the referee for this comment. We therefore re-calculated the analyses of sex with the G-test (Unplanned test of the homogeneity of replicates tested for goodness of fit) (Sokal, 1995), where our null hypothesis is that the sex ratio of copepods does not change with the drop of oxygen and duration of anoxia, or in the recovery after short anoxia.

We revised the following sentences:

P 12391, L 25 the "ratio of male/female" was deleted

P 12392, L 16 a sentence was added: *"To test whether the ratio of copepod males vs females changes from normoxia to different anoxic treatments and in the recovery phase, we analysed the sex ratio with the G-test (Unplanned test of the homogeneity of replicates tested for goodness of fit) (Sokal, 1995)."*

P 12394, L 10 the sentence "Regardless of the prevailing oxygen conditions, the male/female ratio remained relatively stable. The ratio is not balanced as the relative abundance of females was always significantly higher than that of males (Table 2) (2-way ANOVA with factors sex and treatment, $p<0.001$). *"*

was revised and reads as follows:'

"The male/female ratio of copepods did not change significantly from normoxia to anoxia of different duration and in the short recovery treatment (G-test, $p>0.05$). The ratio is not balanced: relative abundance of females was always significantly higher than that of males (Table 2)."

[Anonymous referee #1's comment]:

All in all, I think that the paper to be accepted should re-think the way of data analyses and as a consequence their interpretation. Samples exposed to hypoxia cannot be compared to natural conditions sampled weeks before.

We hope that we could dispel the reviewer's concerns regarding the data analysis in the present study with our detailed response (see above).

[Anonymous referee #1's comment]:

Probably an autocorrelation analyses would be more appropriated. In addition, animals within the same core cannot be compared in and analyses of variance. Probably in this case a correlation analyses should be more adequate.

The authors think that there is a misunderstanding here. We never compared copepod densities from the same core. The copepod densities from one core were compared to densities from the two replicated cores, and all the three together formed one treatment. A 1-way ANOVA was then used to compare the different treatments (normoxia, 9d anoxia, 7d recovery, 1m anoxia, 2m anoxia), as three replicates were available for each of them. Cores which did not show live copepods were excluded as replicates (one core of 2mA, and all cores from 10mA).

Indeed, an autocorrelation analysis would be the best option if the observations would represent repeated measurements on experimental units (Quinn and Keough, 2002). However, in the present case this is not valid: we never analysed an effect of treatment on the same sample or same set of animals/copepods.

We therefore consider the 1-way ANOVA as the optimal choice of analysis.

[Anonymous referee #2's comment]:

This manuscript describes the response of meiobenthic copepods to hypoxia and anoxia which are among the most common and harmful threats to marine benthic communities worldwide. This study is very interesting especially that concerns the effect of oxygen decline and anoxia on harpacticoid copepods - meiobenthic group considered as the most sensitive to oxygen depletion. The observation described in this manuscript shows that although the majority of copepods have not survived the first days of oxygen decline, some copepods belonging to one family were able to survive 2 months in the anoxic sediment. The clear advantage of this experiment is the experimental method used to induce hypoxia and anoxia that has been tested by the authors in their earlier experiments performed at the same study site and described in detail together with the results of these experiments in already published papers. This manuscript should be published but I would like to focus on some issues that require, in my opinion, more attention or explanation.

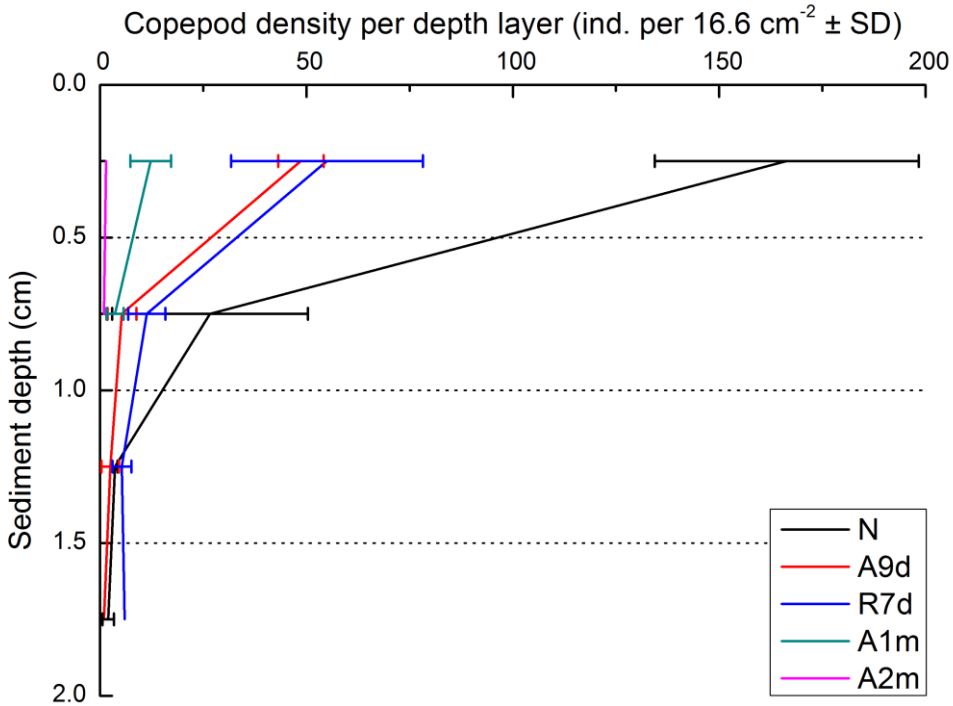
We sincerely thank the referee for his/her positive view of our manuscript and his/her interest in our research. The specific comments (see below) were very constructive and helped to improve the manuscript.

[Anonymous referee #2's comment]:

The authors sliced the sediment in very thin layers. It would be very interesting to see how the vertical distribution of copepods changed over time. The results

from separate slices are analysed based on the dendrogram but I would like to see a simple figure presenting vertical patterns of copepod concentrations after 9 days, 1 and 2 months. In normoxic conditions copepods were present to the depth of 2 cm and, I guess, the majority of them were concentrated at the sediment surface. It would be interesting to show how (if so) copepods migrated in response to oxygen decline and would probably help to better understand the processes in the upper two sediment cm.

We thank the referee for this suggestion and included a new Figure 5 (see below) with the copepod densities per depth (the original Figure 5 is now Figure 6).



The following text was added to the Material and Methods:

P 12391 L 28: *"*The graphs were drawn using Microsoft Office Excel *and OriginPro 8."*

P12391 L27: *"To test the distribution of copepods (absolute density) in the respective treatments (N, A9d, R7d, A1m, A2m, A10m) and different sediment depths (0-0.5, 0.5-1, 1-1.5 and 1.5-2 cm), a 2-way ANOVA was performed. In order to meet the assumption of normality (Kolmogorov-Smirnov test) and homoscedasticity (Levene's test), data were log transformed prior to the 2-way ANOVA.* All tests were performed using the R statistical software (Team, 2010)*."*

And in the results, section 3.3:

P 12395 L 5: *“Based on the copepod density in different sediment layers (0-0.5, 0.5-1, 1-1.5 and 1.5-2 cm) and different treatments (N, A9d, R7d, A1m, A2m, A10m), a 2-way ANOVA on logarithmically transformed data showed a significant effect of each factor on density ($p<0.001$ for factors treatment and depth, and $p=0.0512$ for the interaction of both factors) (Fig 5).“*

And in the discussion:

P12401 L3: *“Additionally, the densities in the 0.5 to 1 cm depth layer under normoxia resemble those of the uppermost sediment layer in the A9d treatment. Finally,...”*

[Anonymous referee #2’s comment]:

There is a lack of information on oxygen concentration changes in the sediment in this experiment and it is mentioned that since the chambers were not shaded the microphytobenthic production cannot be excluded, at least at the beginning of the experiment. Since there is no evidence on oxygen levels in the sediment, can be the first treatment (A9d) considered as really anoxic? From the method description we conclude that the A9d samples were taken after oxygen decline and very short term anoxia recorded in the overlying water (it is not entirely clear to me whether this anoxic period took 2 or 4 days: the A9d chamber was deployed on 2 August and was sampled on 11 August so after 9 days. But further in the text it is stated “that anoxia in this chamber was reached after 5 days, that is two days before the samples were taken”. It is unclear to me when exactly the samples were taken: after two or four days of real anoxia?). In this context, I think that the 9-days long treatment cannot be considered as a real A9d anoxia.

Indeed there is a mistake in the text, so we rephrased the sentence (P12390 L5) to: *“...and anoxia was reached on day 7 (Metzger et al., 2013), two days before the samples were taken.”*

To unambiguously clarify the sampling design and the name of the treatments, we added an additional sentence in the M&M section:

P12390 L3: *“Note that "9 day, 1 month, 2 months and 10 months" anoxia refer to the duration of the entire treatment (i.e. from the day the chamber was closed until samples were taken) and not to the actual duration of anoxia.”*

[Anonymous referee #2’s comment]:

Statistical analyses: ANOVA is performed to analyse the results but it is not mentioned whether the data meet the assumptions necessary to perform parametric analyses.

We double-checked the assumptions for ANOVA. The assumption of normality was tested with the Kolmogorov-Smirnov test and the assumption of homoscedasticity with Levene’s test, both by means of R statistical software. Data required a square root transformation in order to meet both ANOVA assumptions. Therefore the 1-way ANOVA to compare copepod densities in different treatments (Figure 1) was

recalculated on square-root transformed densities and the results were revised accordingly. This new approach did not change significantly the original results.

We added to the M&M section:

P12391 L25: *"To test the effect of factor treatment (N, A9d, R7d, A1m, A2m, A10m) on copepod density per core (16.6 cm^2, depths pooled), a 1-way ANOVA was used. The data were first tested for normality with the Kolmogorov-Smirnov test and for homoscedasticity with Levene's test (Sokal, 1995; Dytham, 2003; Zuur et al., 2010). Subsequently, the data on copepod densities were square root transformed, to meet both assumptions, prior to the 1-way ANOVA."*

Part of the Results was reformulated as follows:

P12392 L23: "*The 1-way ANOVA (factor oxygenation, on square root transformed data) revealed a significant difference ($p < 0.001$) among treatments. The Tukey HSD posthoc tests clarified the pairwise differences, i.e. all combinations were significantly different from each other ($p < 0.001$ or $p < 0.01$) except for A9d-R7d and A1m-A2m. Seven days after termination of the A9d-deployment, potential recolonisation from the surroundings was examined. The similar values of copepod densities in 7d recovery and the 9d anoxia treatment indicate that no substantial recolonisation took place."*

[Anonymous referee #2's comment]:

Multivariate analyses are performed with ANOSIM test (it is stated that ANOSIM test was performed "in addition" to MDS. In my opinion, it is more elegant to perform first the analysis of similarities and then use MDS plots to visualize the results), but given the number of replicates (3, while most of the Primer routines perform best with a minimum of 4 replicates) it seems that PERMANOVA would be an ideal statistical method to compare the differences among treatments. It would be then interesting to include the depth factor into PERMANOVA analysis.

We thank the referee for this remark; it was indeed a good suggestion to perform a 1-way PERMANOVA test (instead of the ANOSIM test) prior to MDS visualisation of the data. Moreover, we also performed a PERMANOVA test for the factors treatment and sediment depth in section 3.3 of the results ('Copepod assemblage in different depths and oxic conditions'). We accordingly changed/added the text.

in Materials & Methods:

P12392 L3: *"Based on the species densities (untransformed data) the copepod assemblages were analysed for similarity with the Bray-Curtis similarity index. Possible differences among treatments (N, A9d, R7d, A1m, A2m, A10m) were analysed with 1-way designed PERMANOVA and PERMDISP tests (Anderson et al., 2008). The similarity among samples was then visualised in the non-metric Multi-Dimensional Scaling (nMDS). A SIMPER analysis was used to investigate which species were responsible for dissimilarities among treatments. Moreover, potential differences among treatments (N, A9d, R7d, A1m, A2m, A10m) and depth layers (0-0.5, 0-5-1, 1-1.5 and 1.5-2 cm) were further analysed with a two-way crossed PERMANOVA with treatment and depth as fixed factors (Anderson et al., 2008)."*

in the Results:

section 3.1., P12393 L11: *“The PERMANOVA test and PERMDISP test (factor treatment; N, A9d, R7d, A1m, A2m, A10m) revealed that the treatments differ significantly from each other (P(perm)=0.001 and P(perm)=0.353, respectively).”*

section 3.3., P12395 L5: *“Moreover, based on species composition, the copepod assemblages in different treatments and at different sediment depths were analysed with a 2-way crossed PERMANOVA test and PERMDISP tests. From the tests we cannot formulate any strong conclusion for the individual factors (treatment or depth). Treatment did have an effect, but there is an interaction with depth. This can be explained by the fact that most of the fauna is concentrated in the top sediment layer.”*

[Anonymous referee #2’s comment]:

It is stated that ‘almost all chambers were deployed at the same time’, but further in the text (Page 12389, lines 21-28) we read that their deployment times varied.

The deployment times of the 9 days, 1 month and 2 month chamber varied only a few days from each other due to practical reasons such as limited available manpower/scuba divers, poor weather conditions, boat availability etc. . We also deployed the 10 month chamber in the same time period but recognized heavy burrowing activities by infaunal crustaceans after a few weeks. The chamber was moved and therefore the respective experiment started approximately 1 month later.

We inserted in the text, for clarification:

P12390 L4: *“The deployment times differed by a few days due to practical considerations: depth of the experiment (24 m), diving constraints (decompression), and boat availability (exception: 10m anoxia chamber, deployed several weeks later).”*

We believe we have addressed all the points made by the reviewers and look forward to seeing our manuscript published in Biogeosciences.

Sincerely,

Mateja Grego and co-authors

References:

Anderson MJ, Gorley RN, Clarke KR. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. Plymouth: PRIMER-E Ltd, 2008.

Findlay SEG. Influence of sampling scale on apparent distribution of meiofauna ona a sandflat. Estuaries 1982; 5: 322-324.

Giere O. Meiobenthology:the Microscopic Motile Fauna of Aquatic Sediments. Springer, 2009.

Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. New York: Cambridge University Press, 2002.

Sokal RR, Rohlf F. J. Biometry, the principles and practice of statistics in biological research. New York: W. H. Freeman and company, 1995.

Underwood AJ, Chapman MG. Design and Analysis in Benthic Surveys. Methods for the Study of Marine Benthos. Blackwell Science Ltd, 2005, pp. 1-42.

Vrišer B. Sezonska dinamika in variabilnost harpaktikoidov (Copepoda: Harpacticoida) v Tržaškem zalivu: triletna raziskava (Seasonal dynamics and variability of harpacticoids (Copepoda: Harpacticoida) in the Gulf of Trieste: a three-year study). Annales, Series Historia Naturalis 1996; 9: 53-60.