We thank the reviewer for his/her positive comments. In the following response we will reply to both the comment he/she made in the Interactive discussion page (Online comments) and directly on the manuscript (Supplementary comments).

I-Scope

Anonymous Reviewer 1 (AR1 Online comments): Many thanks for giving me the opportunity to review the work of D. Langlet and coworkers concerning an experimental foram-anoxia study performed in the Adriatic Sea. The study concerns a 10 month long experiment where foraminifera and other organisms were subjected to artificial induced anoxia. At specific time points cores were collected and the live foraminifera were determined using celltracker green methodology. In a separate paper, submitted to the same special issue, the authors focus on the overall assemblages whereas in this paper they discuss the response from various species. However, there is quite a bit of overlap between the papers since it is difficult to discuss various species without discussing the overall concentrations and diversity.

The paper is overall a nice contribution, however, I lack a more clear aim why this study is performed (this is after all Biogeosciences which has a broader scope) […].

AR1 (Supplementary comments) P.12068 l.29 : Why are you doing this study? I lack a clear aim of why you are investigating the response of anoxia to foraminifera. I can think about a number of reasons, but I think it is worth justifying this in a broader context than simply we have a new method now to study live benthic foraminifera and we have done this experiment. Do you have any idea how the survive 10 months of anoxia, it is clearly not through denitrification.

Authors' response (Auth.): Following the Reviewer's comment we will re-phrase the end of the introduction to better explain and to widen the aims of the present study. The overall framework for this and the other studies in the special issue will also be outlined in the Preface.

The main goal of the study is to determine species-specific responses to long-term in situ generated anoxia. This approach will allow us to better understand the ecology of key-species of foraminiferal communities. It will also provide information which may help us to select species for future laboratory experiments in hypoxic/anoxic conditions, for instance to develop proxies of paleo-oxygenation.

The present study also aims to try to understand how benthic foraminifera can survive up to 10-months anoxia in the Adriatic Sea. In this context, we have tried to determine

whether denitrification is implied in the foraminiferal survival in anoxia. No previous contributions (Risgaard-Petersen et al., 2006, Piña-Ochoa et al., 2010) have tested nitrate-accumulation and denitrification capacity of the benthic foraminiferal species dominant in the northern Adriatic Sea. In the present experiment we identify whether some foraminiferal species from the Adriatic Sea can accumulate nitrates and/or use them via denitrification.

At a broader scale, the multi-taxa studies performed in the present special issue (macrofauna, soft-shelled and hard-shelled meiofauna) permits a better understanding of the effect of long-term anoxia on continental shelf benthic ecosystems.

AR1 (Supplementary comments) P.12068 l.29: Is this site particularly prone to anoxia? Doesn't sound like that when you describe the sediment.

Auth:

Yes, the Gulf of Trieste (Northern Adriatic Sea) is a recognized area for seasonal hypoxic/anoxic events and is also listed among the "spreading dead zones around the world" (Diaz & Rosenberg, 2008).

The oxygen concentration at this site (Piran) has been measured in the water column (at 22m depth: roughly 2 m above the sediment-water interface) since the 1980s. During this period several water-column hypoxic and anoxic events have been reported, along with repeated and extensive benthic mass mortalities (see Malej, A. and Malačič, V.: Factors affecting bottom layer oxygen depletion in the Gulf of Trieste (Adriatic Sea), Annales, 7, 33–42, 1995). Although the frequency of low-oxygen events has apparently declined after the 1990s in the Adriatic Sea (Giani et al., 2012), the Gulf of Trieste is still frequently affected by pelagic marine snow formation (Danovaro et al., 2009; Kamburska and Fonda-Umani, 2009). When such large amounts of organic matter sink to the sea floor its decomposition can locally lead to anoxia at the sediment-water interface.

More information on the history of the anoxia development at the study site is provided in the revised version of the manuscript, in page 12068.

II-Sites

AR1 (Online Comment): […] and in particular why at these two sites which from many aspects seem to be far from ideal.

AR1 (Supplementary comments) P. 12069 l. 9 : why did you choose this site? In my experience, if you choose a more muddy substrate you would have had more foraminifera and you wouldn't have to go through the quite destructive process of centrifuging the foraminifera in order to be able to count them. Would appreciate to see the rational behind the station selection.

Auth:

Because of the occurrence of "natural" hypoxic and anoxic events in the Adriatic Sea, this region is optimal for our study, which aims to understand the multiple responses of fauna to such low DO events.

The precise study site has mainly been chosen for practical reasons, i.e., the presence of the oceanographic buoy of the Marine Biological Station Piran. The presence of this buoy has numerous advantages:

1) The possibility to moor the boat to the buoy to facilitate the work of the scuba divers who installed the benthic chamber on the sea floor and sampled the sediment cores.

2) The project was originally designed to analyze the effect of anoxia on the macrofauna behavior (see Riedel et al., this issue). While the macroepifauna community consisting of interspecific aggregations (multi-species clumps or bioherms) in the region has visibly suffered from the combined effects of bottom fishing activities and repeatedly low DO events, the macroepifauna adjacent to the buoy is not directly affected by bottom fisheries and has not experienced hypoxia for at least 5 yr (V. Malacic, personal communication). Because of the presence of several well-developed macrofaunal multispecies clumps, this site was chosen for the macrofaunal studies. To compare the meiofaunal and the macrofaunal responses to anoxia we decided to install the chambers for meiofaunal analysis at the same site but at places with no visible macrofauna.

3) The presence of the oceanographic buoy permits to record the temporal variation of several physico-chemical parameters that can be used to compare the initial conditions of each set of experiments (9d, 1mo, 2mo, 10mo). For the present study we did not use such data because most of the benthic chambers in the macrofauna studies were only deployed for 4 days, essentially in stable conditions.

4) Finally, we have also chosen this site because of the long-term nature of our experiment. The benthic chambers had to remain on the sea floor for up to 10 months. There is important commercial fishing activity in the Gulf of Trieste and the sea floor is substantially impacted by such bottom-trawling and dredging activity. No fishing is

allowed a few hundred meters around the oceanographic buoy, which ensured safe retrieval of our scientific material after 10 months.

The centrifugation procedure was not chosen because of the sandy nature of the sediment. We decided to use this method to work on exactly the same cores for both foraminifera and soft-shelled meiofauna (see work of Grego et al.). The centrifugation procedure does not extract the foraminifera from the sediment (sand), as it does the soft-shelled meiofauna. The potential effect of centrifugation on foraminifera will be discussed below.

AR1 (Supplementary comments) P. 12070 l.4: (Mat and Meth. Section 2.1 Study Area) So the same specimens that were subjected to anoxia were not analysed? Why didn't you sample cores from the same site as your experiment was conducted?

Auth: Nitrate accumulation measurements were performed on foraminifera coming from two sites. As explained in the manuscript a large number of living individuals is needed to perform nitrate and denitrification measurements. Some individuals that were analyzed were sampled in Normoxic conditions in Piran in August 2011 (see Tables 6 and 7 of the manuscript). In such a sandy sample, rich in dead foraminifera, it was extremely time consuming to find a few living individuals. Because of this we then decided to complete the observations on a larger number of specimens. To do so, complementary nitrate and denitrification measurements were performed in 2012 on specimens collected in sediment from station D10a, which was the only "fresh" material available to us at that time.

The authors recognize that presenting results from these two different sampling sites can be confusing for the reader. Nevertheless, the intracellular nitrate analysis performed on samples coming from station D10a suggests that only the anoxia-resistant species can accumulate nitrates.

In the revised version of the manuscript we will better explain the origin of the foraminifera used in the nitrates and denitrification measurements and why we decided to work on these two sites.

III-results

AR1 (Online comment): At times the manuscript needs to be more structured in particular in the result section where there is a mix of size fractions and different time points which confuses the reader.

AR1 (Supp. Comments) P. 12075 l.23 : I find this paragraph somewhat unclear written there is a mix between fractions and time points. Maybe it would help to refer more to the figure and just point out the large differences.

Auth: This paragraph will be simplified as suggested by the reviewer.

IV-discussion

AR1 (Online comment): In the discussion each individual species is then discussed and I'm wondering if that long paragraph would be better off in a table, including all the references. It would make it easier to get an overview of which species belongs to which category of more or less sensitive species.

AR1 (Suppl. Comment): Section 4.3 this section were each species is not so much discuss but at least compared with the literature I wonder if this can't be shortened and presented in an illustrative table instead. It is quite repatvite to read about each species, even though in some case it is rather useful. I wonder if it wouldn't be more digestible if it was presented as a table. Still with the references included. The journal is Biogeosci not JFR.

Auth: as suggested by the reviewer we will try to summarize this part of the discussion partly as a Table. But we would rather not over-simplify the text. So we propose to present the discussion both as a (shortened) text and as a table.

V-data analysis

AR1: But my largest criticism is why the authors haven't used the large geochemical data set collected (and published in the same issue) by Metzger et al (also co-author on this paper). As it is now the authors compare concentration, diversity and several diversity measurements with time, time^2(?) and depth using linear modeling and then speculate about the importance of organic matter at a certain time point (1 month). They find out that time is the most important variable in most cases. This isn't particularly surprising. However, they have a lot of geochemistry data and it would be much more interesting to perform linear modeling using all available environmental data. Is there a significant relationship (or not) with Fe, Mn, redoxcline depth and so forth with certain species? This approach would be much more interesting and the data is there.

Auth: The geochemical data set was not used for the several reasons.

Geochemistry data are not available for all the sampling times (In Metzger et al. for logistical reasons we did not analyze the sediment from the "2 Months" experiment).

The main difficulty in designing a model to test the effect of sediment geochemistry on the foraminiferal fauna is the choice of the geochemical variables to be included in the model. In fact, we can choose between the following options:

1) **use the raw data of concentration of Mn, Fe, SO_4^{2-} and alkalinity with depth;**

2) **use Mn, Fe, SO_4^{2-} and alkalinity averaged per depth intervals (the same intervals as used in the faunal analysis: 0-0.5 cm, 0.5-1 cm to 4-5 cm) and per incubation time (average the two DETs replicates);**

3) **use modeled fluxes (production rates) of Mn, Fe and SO_4^{2-} per sample and averaged per incubation time. As it is not possible to model alkalinity fluxes we can use the average depth of alkalinity production.**

These 3 options are all somewhat problematic. Option 1 (raw concentration data) cannot be used as in the geochemistry dataset concentrations are available every 2 mm while the foraminiferal dataset is designed with depth intervals of 5 or 10 mm. To be able to use the concentration data we have to average them per depth interval and also, we have to average the two replicates (Option 2). In view of the large spatial variability of the pore water chemistry (see the difference in the two replicate profiles in Figure 1 for example) averaging the geochemical parameters does not really make sense. However, with option 2 we could eventually design a linear test of the effects of Time, depth and the concentration of Mn, Fe, SO_4^{2-} and alkalinity on the foraminiferal density. Unfortunately the latter 4 geochemical parameters are strongly correlated with each other (except Manganese and Iron and Manganese and Alkalinity; see Table 1) and with experimental duration. It would be unreasonable and statistically incorrect to use several "independent" variables which are strongly correlated in a linear model.

Consequently, we tried Option 3. Production rates of Mn and Fe, consumption rates of SO42- and the average alkalinity production depth were estimated and the results are presented in Table 2. Such an estimation of fluxes/production rates is problematic because of the experimental setup. The lack of stirring inside the benthic chambers renders the estimates of fluxes unreliable (See the interactive discussion on Metzger et al. paper for more details; available at: http://www.biogeosciences-discuss.net/10/12029/2013/bgd-10-12029-2013-discussion.html). Despite these strong

limitations (the lack of sense of averaging the fluxes values in the two heterogeneous replicates and the fact that estimated fluxes in un-stirred benthic chambers are not correct) we designed a complementary model testing the effect of Time, Mn, Fe, SO_4^{2-} and alkalinity on the total standing stocks of the 3 groups of species. After the backward variable selection procedure, only the time variable had a significant effect on the standing stocks of groups A and C, and none of the tested variables significantly impacted group B. This suggests that the tested geochemical variables have less impact on the foraminiferal standing stocks than incubation-time.

All of the 3 options have, in our opinion, numerous methodological flaws and we chose to not present its results. We would rather keep the approach that we used previously: describe the variation of the geochemical conditions with time, quantify the effect of time on the foraminiferal faunas and then try to understand how the geochemical conditions correlated to the time affect the foraminifera.

Nevertheless we agree with the referee that our manuscript lacks a direct analysis of the potential correlation between the foraminiferal faunas and the geochemistry data. We therefore propose to present an additional figure showing at each sampling time the vertical distribution of the geochemical elements presented in Metzger et al. next to the vertical distribution of either the total density of each foraminiferal group (A, B and C) or of each 9 of the major species (Figure 1).

In the statistical procedure we decided to add a Time^2 variable in order to model a potentially quadratic or squared relationship between the foraminiferal standing stock and the anoxia duration. This is important to be able to detect an increase of the Standing Stock between 0 and 30 days and then a decrease after 30 days such as observed for *Q. seminula*. The sentence explaining the interest of such models will be re-phrased in the methods section.

Table 1 – Spearman correlation matrix between the concentrations of Alkalinity, Sulphates, Manganese and Iron. Data are extracted from Metzger et al. (this issue). The concentrations of each compound were averaged per replicate DET and per the depth intervals used in the faunal analysis (0-0.5 cm; 0.5-1cm; … and 4-5 cm).

	Alkalinity		Sulphates		Manganese		Iron	
	r	p	r	p	r	p	r	p
Alk	1.00		-0.30	0.03	-0.19	0.17	0.37	0.00
Sulf			1.00		0.63	0.00	-0.37	0.01
Mn					1.00		-0.12	0.38
Fe							1.00	

Table 2 – Maximal production rates of dissolved Manganese, Iron and Sulphates (in nmol cm^{-3} s^{-1}) and minimum Alkalinity production depth (cm) for all the sampling times.

Incubation duration	Maximal production rate (nmol cm^{-3} s^{-1})			Minimum production depth (cm)
(days)	Mn^{2+}	Fe^{2+}	SO_4^{2-}	Alkalinity
0	2.46E-05	9.84E-05	-4.70E-05	10
9	3.44E-05	1.99E-04	-1.46E-04	4.5
29	5.25E-06	1.75E-04	-8.50E-04	0
315	1.20E-05	3.19E-05	-2.30E-03	2.5
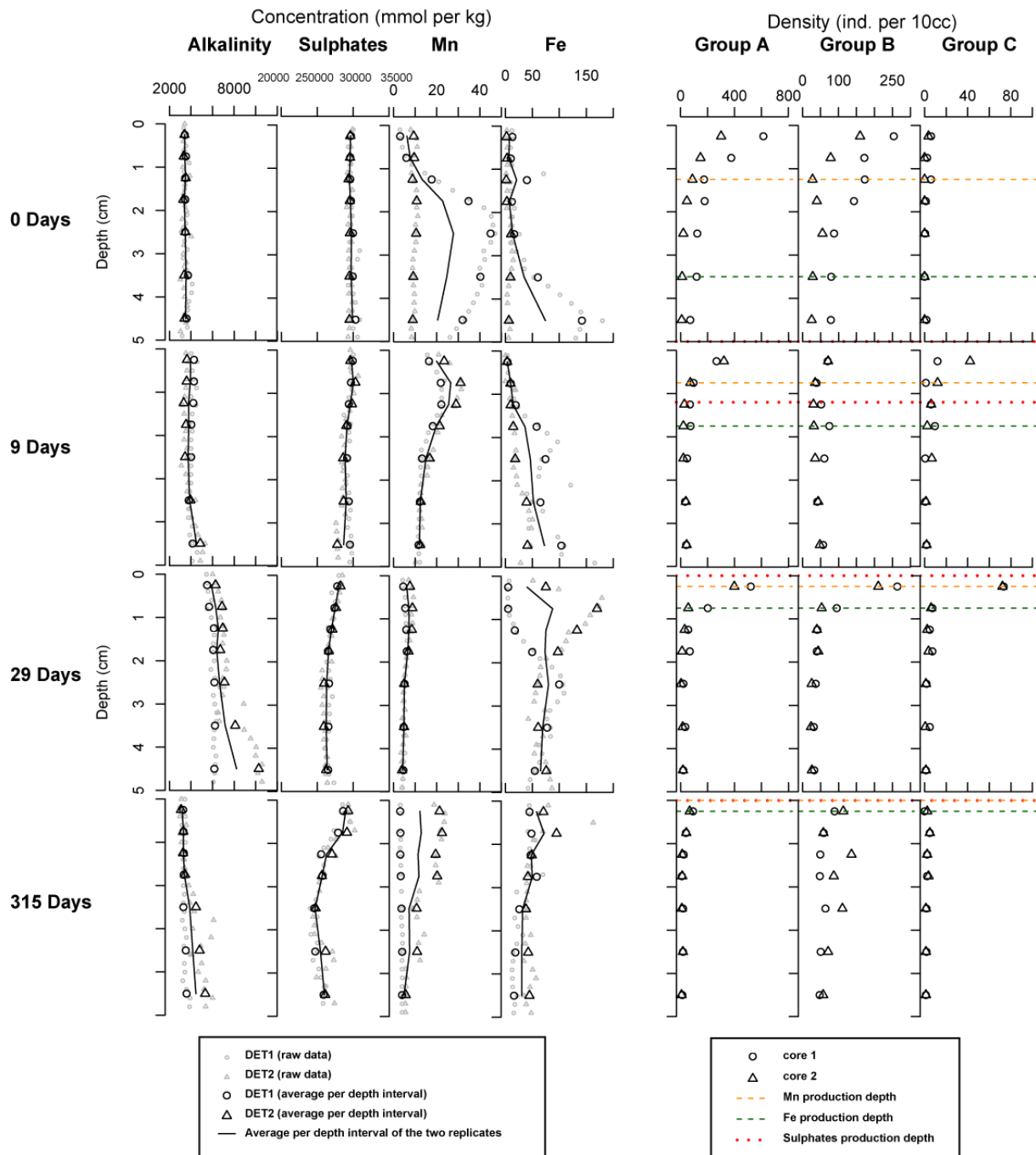
Figure 1 – Preliminary version of the Figure that would replace Figure 4 in the revised version of the manuscript. This figure presents the variation of each of the 4 geochemical parameters with depth and time (with raw data, averaged per depth intervals and with the two replicates averaged) on the left panel. The right panel presents the vertical variation in density of the three groups of species with time. Only Manganese, Iron and Sulphates production or consumption fronts are presented.

AR1: I have made plenty of comments in the attached pdf as well. The language is in general ok, but it can also be polished and I hope BG provide some text editing service, there is some unnecessary show, show, show repetition and capitalization of Time etc.

To sum it up, it is a study well worth publishing but they can lift this to another more interesting level.

Others-Main remarks on the Supplementary comments:

AR1: P. 12070 l.24: Wow, the poor foraminifera. How sure are you that the forams did become broken and fragmented after this treatment? Especially thin delicate species like L. scott, which is one of your key species.

P. 12081 l.6: the sampling treatment with centrifug most likely was quite devastating to L. scotti, did you do any comparison with samples you haven't centrifuged?

Authors: The centrifugation treatment was applied to all analyzed cores. If this treatment had an effect on fragile species, we expect that it would be of the same magnitude in all samples. The differences in densities between all samples would therefore still due to the experiment and not to the treatment procedure. Non-centrifuged samples were not analyzed but the observation of the centrifuged samples did reveal intact specimens (no broken chambers) of several very fragile species such as *Reophax nanus* and *Leptohalysis scottii*. In general, we did only rarely observe broken specimens in the sediment samples.

The major drawback of the centrifugation method is that some specimens of light-shelled species were occasionally found in the supernatant solution after centrifugation. However, the number of individuals in the supernatant was relatively insignificant compared to the large amounts observed in the sediment samples. The fact that the centrifugation method could underestimate the density of light and fragile species in a similar way in all samples will be discussed in the revised version of the manuscript.

AR1: P. 12077 l.26 : Why did you do that (Auth : introduction of brittle stars in the chamber)?

Auth: Due to technical limitations, a continuous measurement of the oxygen concentration in the benthic chamber was only possible in the "9 days chamber". We therefore needed a rough indicator of the development of anoxia that could be seen by the scuba divers without causing any perturbation to the chamber. The changes in the macrofaunal behavior with oxygen concentration have been documented by Riedel et al. (2013). The brittle stars were therefore used as a biological indicator of anoxia. This information will be added to the discussion and methods section of the revised manuscript.

AR1: P. 12078 .l6 and l.8: But you didn't measure the organic matter content at each time point. And still you don't include these data (H2S) in your linear modelling?

Auth: The organic matter content was indeed not measured, but the observations of the decaying dead macrofaunal organisms and the increase in the intensity of most geochemical reactions point to such considerable changes in the organic matter content. This part of the discussion will be re-phrased to better present which data support these interpretations.

H_2S production was not quantified in Metzger et al.. However, H_2S production is observed with non-quantitative DGT-like (Diffusive Gradient in Thin-films) probes. This non-quantitative information could not be included in the models.