“Final Response” to the Reviewers' Comments

We thank the two Reviewers for their helpful comments!

Response to Reviewer #1:

Explicit oceanic eddies are missing in this study and this could be biasing the result. In particular, the response of the Southern Ocean to wind changes has been shown to be markedly different in eddy vs. non-eddy models – the authors should cite studies relevant here: see for example Screen et al., 2009; Meredith and Hogg, 2006; Hogg et al., 2008; Sallée et al. 2008; Spence et al. 2010; Farneti and Delworth, 2010 (papers listed at end of review). Over short time-scales, an adjustment in the Southern Ocean winds generates a change in the horizontal Ekman transport, and in turn the vertical Ekman pumping and the location of fronts. However this adjustment is then pretty much entirely compensated on interannual time-scales by mesoscale eddy fluxes that transfer heat poleward and diffuse density gradients via baroclinic instability (see papers above). The results of the present paper would be entirely different in an eddy-resolving model, because eddy compensation would mitigate much of the response in the Ekman transport. As a result, the linear relationship could be not valid. Of course it would be impossible to run the suite of experiments examined here at eddying resolutions, but the authors need to discuss this shortcoming of their study in some more detail and point out it potential influence on the relationship between the sea level change and the AMOC strength. Similar discussion might be also applied for the freshwater forcing.

This is an important point that deserves more attention, and we have included it into the revised manuscript.

The influence of including eddies in simulations is still a matter of research, and of course one cannot say with certainty how much the results differ from what they would be if a higher resolution had been applied (or a different eddy-parametrization, which might capture parts of the potential differences, as also noted by Farneti and Delworth, 2010). But there may of course be relevant differences, especially regarding the impact of Southern Ocean wind stress changes in the context of the studies named by the reviewer.

We now emphasize in the manuscript that the price for the computational feasibility of the study was to run the experiments at a non-eddying resolution (which obviously has an impact on all experiments). We now also point out the mechanism of eddy-compensation as well as its relevance for our experiments and give citations of the studies in the literature which compare the models of different resolutions in this respect. We finally have to leave it to the reader as well as future studies with different models to judge the detailed implications for our results.

We would like to note that, even if one doubts that the strong AMOC change under these strong SO wind anomalies is realistic, the corresponding experiments could be still seen as an artificial remote forcing that produces an internal relationship between SL and AMOC strength surprisingly similar to the freshwater case and different from the CO2 forcing - which already is an interesting result and in the manuscript's line of argument.

We put this point into Section 2 in the context of the model description because it is an implication of the model's resolution and the potential impact is not limited to one of the sections presenting the results. However, we have again pointed out the fact that the study is done at a non-eddying resolution in the conclusions.

I think it would be necessary to mention in Section 2 (model simulations) in which exact latitude bands the Southern Ocean wind anomalies are applied.

We have now included the information (application of the wind stress anomalies south of 31.875°S) in Section 2 in the place where the location of freshwater forcing has already been provided.

In section 3, the authors conclude that both freshwater forcing and reduction of the Southern Ocean wind stress only result in a localized sea level reaction in the 100-year simulation. However, a total response of the AMOC to these forcing may take much longer time. In recent paper, Wei et al. 2012 found that the full AMOC response to the Southern Ocean wind perturbation takes several centuries (300 years) in their coupled GCM. What is the response time for both forcing in your specific model setup? Or some discussion is needed here based on the previous studies. In section 4.2, the authors provide the results from longer integrations. However, only figures from the CO2 experiment are shown. No further analysis or discussion on the longer timescale change of the sea level results from the other two forcing. It would be interesting to see the results from the freshwater and wind forcing experiments, or at least, the authors should have more discussion on it and either support or further explain the conclusion in Section 3.

<u>a) Response times of AMOC for freshwater and wind forcing; and their relevance for the top two patterns of Figure 2 showing only a localized reaction (Section 3):</u>
Of course, neither sea-level nor AMOC are in an equilibrium after the 100 years considered in Sections 3 and 4.1. We intentionally focused on the transient short-term behavior relevant for the 21st century, while we did of course try to find robust patterns on longer timescales during the preparation of the manuscript. We have continued the simulations on which Figure 2 is based, for 1000 years. Keeping in mind that our forcing strength is continuously increasing during the first 100 years (in contrast to the forcing in Wei et al., 2012), we observe the following regarding the response times:

- Southern Ocean wind forcing: The AMOC strength drops during the first 100 years (with a visible reaction already in the first decade); then it recovers and overshoots before going down again and oscillating around a rather stable mean value from the year 500 onwards.
- Freshwater forcing: Changes of the AMOC strength after the first 100 years are relatively minor, but equilibration also takes place around year 500.

We find the resulting sea-level anomalies to be indeed less localized, with a reaction in the Southern Ocean for the freshwater forcing and a reaction in the NA for the SO wind forcing. We have included an additional short paragraph in Section 3 right behind the paragraphs on freshwater and wind forcing (for 100 years).

<u>b) Longer timescales (Section 4.2):</u>
We are not sure if we understand the Reviewer correctly, but the discussion on longer timescales in Section 4.2 already focuses on all three forcing mechanisms, with the results after 1000 years being given for all of them in the three panels of Figure 6. The discussion on freshwater and SO wind stress forcing is between line 26 on page 334 and line 7 on page 335 (line numbers referring to the original manuscript). In Section 4.3, we indeed focus exclusively on CO2 forcing because of the different response that it causes as displayed in Figure 6.
If the Reviewer has in mind a discussion of the sea-level patterns as given in Fig. 7 for CO2 (in contrast

to the gradient given in Figure 6), then the comments added to Section 3 as discussed above should answer the question.

In Fig. 3a, Fig. 4a and Fig. 10b, one can see that the linear relation between sea level change and the AMOC is only valid in when the AMOC strength is larger than a certain value. Is there any threshold for this relation? One can also see that this value may depend on the freshwater amount. For each freshwater amount, the linear relation breaks down in different AMOC strength. In Fig. 13, the AMOC change is quite smooth under CO2 and SO wind forcing, but not the freshwater forcing, especially the last 50 years. I guess it is also related to the response time of different forcing as pointed in comment #3. Meanwhile, what causes the large fluctuation of the AMOC under the freshwater forcing in the last 50 years? Internal variability? Possible feedback? The authors should analyze more in section 5.2 to clarify this point.

These are several points, and we are going to comment on them step by step:

<u>a) Breakdown of the linear relations in Figs.3a, 4a, 10b:</u>
As already written by the Reviewer, the linear relation breaks down at different (small) AMOC strengths depending on the forcing strength. Also, it breaks down for a high AMOC maximum in case of significantly increased SO winds! So, it cannot be solely the AMOC strength, but it must be the forcing strength combined with the timescale that determine the breakdown. This is further supported by the fact that the linear relations break down when keeping the forcing fixed and waiting for another 100 years, as written in Section 4.2. This is also the case e.g. for the 0.09 Sv and the 0.1 Sv simulations, for which the AMOC does not drop much more in the second century. In this place (second paragraph of Section 4.2.), we have now written that this behavior supports the fact that it is not just the AMOC strength determining the breakdown of the linear relations found in Section 4.1.
The large SL-change in the last 30 years of the strongest forcing in Fig. 10b, could already be clearly attributed to a highly non-linear change in the subpolar gyre, as written at the end of Section 4.4 (cf. Levermann & Born, 2007).
In order to understand further what role the AMOC strength itself plays for the breakdown, one might try to extend the whole study by spinning up the model into a new equilibrium state with a much smaller AMOC strength (e.g. via a change in the vertical diffusivity) and then repeat the experiments. But it would be a problem to compare these results with the previous ones because the whole system and its reaction to the forcing also changes. That's why we think that the potential gain would not justify the difficulties of finding an appropriate equilibrium state and repeating the whole set of simulations on it, though it would in general be interesting to see in how far our results depend on the "base state" and model settings.

<u>b) Smoothness/fluctuations of the AMOC for freshwater-forcing in Fig. 13:</u>
While the largest fluctuations are indeed in the last 50 years, they already start in year 100, from when onwards the forcing strength is kept constant.
We have done the same simulation switching off the wind stress feedback. In this case, the small fluctuation between the years 100 and 150 still occurs; but the larger AMOC strengthening between year 150 and 200 does not.
As we have done this simulation with fixed wind stress for different forcing strengths and for a total of 1000 years, we have used them for comparison. In some of these runs, periodic variations of the same magnitude occur (roughly from the middle of the millennium onwards). And some show the AMOC behavior between year 100 and 200 as seen in Fig.13, but even without wind stress feedback.
Further analysis revealed that these AMOC variations correspond to variability of the subpolar gyre strength that seems to be triggered by our freshwater forcing. This explains why the wind stress

feedback plays a role without being a necessary prerequisite.
We have now summarized the above observations from the other simulations in Section 5.2.

Response to Reviewer #2

p. 327-329 Discussion setting up potential connection between height gradients and overturning is a bit confused. Suggest rewriting this discussion along the following lines. 1. Insofar as the overturning depends on the integrated geostrophic flow (as in Bryan, 1987 and Gnanadesikan, 1999) it can be represented as the surface height difference times a scale depth for integration, divided by a frictional resistance parameter. (Expand eqn 1 to show this). 2. If this scale depth and resistance parameter relatively constant (as assumed in Stommel, 1961 and found to first order by Hughes and Weaver and Thorpe et al.) then surface height alone may be a good proxy- on short enough time scales one would even expect this to be so as it takes a long time to alter the pycnocline depth. 3. Previous work suggests that SSH within subpolar gyre may be a good measure for overturning. 4. However, there are reasons to believe that surface height alone may not be a good proxy (wind forced runs of de Boer et al., 2010 show an increase in overturning but a decline in density contrast, implying that scale height must increase or "frictional resistance" must decrease). Griesel and Levermann also showed that the overturning in a coupled model where sinking can move around doesn't obey nice scaling laws. 5. If this is true, better data about stratification, i.e. from Argo floats, would be necessary to properly characterize the overturning. 6. So it makes sense to evaluate whether the "natural variability" we might see over decadal scales is a good proxy for different forcing mechanisms...

We thank the reviewer for the detailed suggestions on the structure of the Introduction! We have adopted the main aspect as we understand the Reviewer's comment. We have, however, not followed all the suggestions as explained below. But we would of course be open for discussion regarding these points if they are considered essential.

We have followed the suggestions for the beginning of the discussion: We are now starting the argument with geostrophy as well as the connection of the density gradient with SSH when assuming a fixed level of no motion. We cite Gnanadesikan (1999) already at this place, and additionally include the citation of Bryan (1987). Stommel (1961) is postponed.
However, we have not followed the recommendation to put the beginning into terms of SSH, scale depth and resistance parameter. We find it clearer to start the argument in terms of the meridional density difference and then, as also suggested by the Reviewer, to go on with the discussion on the depth scale which is the fixed integration limit in Thorpe et al. (2001)...

As suggested under 2., we now emphasize that the assumption of Stommel and the results of Thorpe et al. (scaling with a steric height difference which is based on a <u>fixed-depth integral</u>) are a simplification that indicates that SSH alone might be a good proxy (This would also demand that neither significant mass redistribution occurs, nor steric changes below the integration boundary.).

Instead of following the suggestion to continue directly with the previous work on SSH within the subpolar gyre, we keep this in the paragraph that focuses on previous attempts to use pure SSH(-patterns) for diagnosing the AMOC. We continue with the Reviewer's number 4 that closes our paragraph which discusses the history of theoretical considerations for AMOC scaling laws.

We leave the following paragraphs unchanged. That is, the first one on the actual current use of altimetry (combined with other measurements) for explicitly and locally calculating ocean velocities. The next one on attempts to use solely sea-level and not density data for diagnosing the AMOC. And the following one on the idea of using a non-complex north-south gradient in sea-level, which also puts the aim of our paper into relation with the work by Lorbacher et al.

p. 333: The linear relationship is presented as “confirming” earlier work, however, the point of the paper is that this is either accidental or true only for particular forcing cases. As such the discussion should talk about the “apparent robustness” of the relationship or its’ “consistency” with earlier work.

That's true. We have adopted the wording suggested by the Reviewer.

p . 334: Either put some discussion in of Figure 5 or eliminate it. (I suggest more discussion).

We prefer to keep the Figure, especially because the reader might find it valuable in case of future comparisons with different studies. We agree that some interpretation of it needs to be included in the text (last paragraph of Section 4.1, where the figure had so far been only mentioned). We have now highlighted the significant role of mass redistribution. And we have described the relation of thermosteric and halosteric contributions for the three forcings.

Insofar as you see different responses over different timescales it would be good to think about what’s changing. Vertical structure? Apparent efficiency between depthintegrated pressure gradient and overturning?

Regarding CO_2 forcing, Section 4.3 (together with Figure 9) already shows the relevant changes of vertical density structure over 1000 years, explaining the long time-scale changes in the lower panel of Figure 6 (compared to Figure 3). We had picked the CO_2-simulations for further analysis because for them, the difference between Figure 3 and 6 is by far most significant. But for the other two forcing types, for which there is initially a linear relation between $\Delta\eta$ and the AMOC, we have indeed not pointed out so far why this relation breaks down.

We unfortunately have full density data only with 25-year gaps for the freshwater and the wind forcing simulations. Checking some of theses years, significant changes in the depth-structure of density-anomalies after the first one or two hundred years are revealed (much stronger for freshwater forcing than for wind forcing). From this we conclude (and now put the information into Section 4.2) that the vertical structure of density is different during the first century compared to later which must influence any proportionality between depth-integrated pressure-differences and the sea-level difference, likely playing an important role for the breakdown of the original linear relations from Figure 3 (and the differences with respect to Figure 6).

It would also be interesting to see how well, how long and under which conditions a difference in depth-integrated pressures is linearly related to the AMOC strength for our forcings. However, it would be difficult to include this into the study, because we would have to rerun the experiments writing out the full density-fields and turn the focus away from the sea-level differences (which can be remotely diagnosed in contrast to the 3d density fields). So, this might be better placed and then properly analyzed in a different study.

Conclusions: Would be good to have some discussion about whether these results suggest different "fingerprints" for different forcing mechanisms (winds, freshwater, CO_2). For example, what's the cross-correlation between the patterns produced by the different RCP pathways?

Considering the coarse resolution of our model, we are hesitating to make a quantitative statement on detailed "fingerprints". That's why we had originally written in the conclusions that "it is desirable to focus on … more detailed patterns … for which more complex models … are likely to be better suited". Of course, Fig. 2 is given which allows for a qualitative check of patterns for each of the three forcing mechanisms (i.e. potential 'fingerprints') that can be compared to future analyses with higher resolution models, while the purpose of the figure for our analysis is to exclude a large-scale AMOC fingerprint common to all three forcing mechanisms and to motivate the analysis of the north-south gradient.

So, as the three patterns are indeed interesting on their own, we have now included a note on Figure 2 and fingerprint in the Conclusions.

Minor comments

We have adopted all suggestions given as minor comments. The question regarding the latitude band of wind forcing has already been answered above in response to the first reviewer.