Detailed response to the reviewer of the manuscript

"Mechanismfor potential strengthening of Atlantic overturning prior to collapse"

by D.Ehlert and A. Levermann

The responses are followed by the reviewed manuscript in which all changes are highlighted in bold font.

Referee #1 (A Cimatoribus)

The article presents results from a simple analytical model of the AMOC, building on the classical Gnanadesikan pycnocline model, and discussing in particular a mechanism for the strengthening of the overturning rate before a collapse of the AMOC. Overall, I think that the paper is clear, concise and presents interesting results. How- ever, before publication, I think that a few points should be considered in more detail. My main comments are in fact linked to our recent papers "Meridional overturning circulation: stability and ocean feedbacks in a box model" on Climate Dynamics (2014) and "Reconciling the north-south density difference scaling for the Meridional Overturning Circulation strength with geostrophy", under review in Ocean Science Discussions. It should be clear, however, that this does not imply that the authors should follow our approach.

Response: We would like to thank the reviewer for taking the time to review our paper and this very positive assessment. His comments have significantly improved our manuscript and we are confident that we were able to address all issues.

- The authors follow Marotzke (1997) considering the pole-tropics density differ- ence as the one controlling the overturning rate. In the papers mentioned above, we suggest that the definition of the density difference may be essential for re-producing some results of numerical models. How essential is the definition of for obtaining the results shown in the present paper?

  Response: This is a very important issue. The main point of the paper was to introduce an additional density difference which controls the Southern Ocean eddy transport. This is motivated by the "double-crossing" of the Atlantic by the AMOC (once in the South and once in the North). Since this open-ocean circulation in the upper layer has to be balanced geostrophically it needs to correspond to a density difference. These density steps can be seen in observations and models alike. While there is some freedom in where exactly to locate the density difference in the real ocean, it is important for our model to apply a meridional density difference both in the northern and southern Atlantic.

- From lines 7-10 on pag. 38 I understand that mN can have negative values, a result which is not discussed in much detail. This amounts to an enhanced upwelling in the high latitudes of the north Atlantic. Is this the case and, if so, how can this be physically justified?

  Response: We do not consider negative values of mN in itself, which would be an interesting addition, but would require the discussion of an entirely different circulation state which would in our opinion be out of the scope of this study. What we rather discuss is the possibility of the discriminant to become negative which defines our critical values and is discussed in the text.

- pag. 38, line 18: pycnocline depth increases with increasing freshwater forcing in fig. 4a. Furthermore, do the negative values of FN in the figure mean that freshwater is transported from the high to the low latitudes? Surface fluxes tend to transport freshwater from the low to the high latitudes in the real world, so how should be FN interpreted?
  Response: Positive F_N refers to freshwater flux by atmospheric dynamics from low to high latitudes. It dilutes salinity in the northern box and increases salinity in the low latitudes. Thus negative FN would mean decreased salinity in lower

latitudes and increased salinity in higher latitudes. This behavior might not be applicable with surface transports. Thank you for the comment. We have added an explanation to this end. However, the threshold behavior and the increase in overturning prior to the threshold occurs for positive freshwater fluxes, thus a freshwater flux from lower to higher latitudes. In any case the main result of the paper is not affected by this the behaviour in the case of negative freshwater flux.

- pag. 41, Freshwater-induced MOC strengthening: a strengthening of the AMOC under increasing freshwater forcing is shown in Fig. 5 of Cimatoribus et al. 2014, but the definition of the forcing freshwater is different therein. Is there a link between these two results, in particular concerning the mechanism causing it?

  Response: We are very sorry that we have not mentioned your work. This was merely due to the fact that we had submitted the paper a few years ago and in the mean time working contracts and places had changed. We simply missed your paper and are sorry about that. Now we have added a note on the fact that this has been seen previously. Thanks for pointing us in this direction.

- pag. 43, lines 1-4: Could this mechanism be interpreted as a change in the baroclinic modal structure (i.e. in the depth of the first baroclinic mode)? In the real ocean, gradients tend to decrease with depth; could this change the results?
  Response: As shown in an earlier study (Fürst & Levermann, Clim. Dyn.) we find that the vertical density structure as represented by the pycnocline is rather stiff in comparison to the meridional density differences. Compared to the Stommel model the Gnanadesikan models with meridional density differences tend to show a more stable meridional circulation. Changes in the circulation when they occur are then however associated with meridional density differences. We thus believe that this will not be easily measured in the vertical but rather in the meridional density differences of the real ocean (or even in the sea-level pattern.)

Minor comments:

- pag. 33, lines 15-17: "The four meridional tracer transport processes..." could this sentence be rephrased more clearly?
  Response: Rephrased sentence which hopefully clarifies its content that the tracer transport processes control on the one hand the horizontal and vertical density structure and they control the strength of the overturning on the other hand.

- pag. 35, lines 16-...: Since the use of this parameterisation for the eddy flow is one of the main new elements in the model, I would suggest that a more detailed motivation for the parametrisation is given, even if it has already been discussed in Levermann and Fu rst 2010.
  Response: Thank you for the suggestion. A more detailed description of the experiments and results in regard to the eddy return flow in Levermann and Fu rst 2010 has been included here. However, in our opinion going into further detail would unnecessarily expand the article, as the interested reader can find the an extended and very detailed description in the Levermann and Fürst 2010 article.

- pag. 36, eq. 6a: it seems to me that the equation should read ... SU(mN + mE)... instead of ... SU(mN +mW).... Even if at the steady state they are equal, I think that the equation would be more easily understandable this way.
  Response: Thank you for the detailed reading, the typo has been corrected.

- pag. 37, lines 8-11: as far as I understand all the results presented refer to steady states of the system. I think that the last sentence of this paragraph could be misunderstood as saying that time-dependent states are considered.
  Response: We agree with the suggestion. The sentence has been rephrased and includes now that we only refer to steady states.

- pag. 37, line 21: "provide" instead of "provided"

Response: Thank you for the detailed reading, the typo has been corrected

- pag. 41 line 7: Can this result be obtained more rigorously by taking the limit mE 0?
  Response: In this part of the paper the goal was to find a value of F_N,krit that is variable independent. Using m_E lead to including the pycnocline depth D into the formula, therefore we used the southern meridional density difference instead. However, if there is a more straight forward way to derive F_N,krit in the wind driven case we would be happy to include in into the manuscript.
- pag. 41, lines 21-22: please rewrite the sentence.
  Response: The goal of this sentence was to summarize the main finding, that the overturning creases prior to reaching a threshold under freshwater forcing. We rephrased that sentence, so that the content should be clearer now.

- pag. 42, line 21: “strong” should read “strongly”
  Response: Thank you for the detailed reading, we corrected the mistake.
- pag. 42, line 25: missing “depth” at the end of the line.
  Response: Thank you for the detailed reading, we corrected the mistake.
- pag. 44, line 23: the two papers cited mostly deal with numerical models. Comparisons with observations are found, to my knowledge, in the works of Talley and Bryden.

  Response: We included citation from Bryden 2011 (Journal of Marine Research) into the manuscript here. Thank you for the very helpful comment.

- Fig. 2: I could not find where Figure 2 is discussed in the text.
  Response: Thank you for pointing this out. We included the missing reference in the first part of section 3. There, possible solution of the pycnocline are discussed.

Referee #2

Based on the conceptual model by Fuerst and Levermann (2012), the authors develop a model with an additional degree of freedom by the inclusion of a parameterisation of SO eddies. This allows a new type of behaviour, where the steady state AMOC can increase with increasing FW flux into the North Atlantic, all the way to the bifurcation point. The main effect found is that fresh water-induced MOC strengthening in response to a fresh water flux from low latitudes to high northern latitudes could take place in the wind-driven case. By continuity, due to a resulting reduction in eddy return flow m_e arising from a denser low latitude box and subsequent reduction in density difference between the SO and the low latitude box, northern sinking must compensate in the conceptual model. (Unlike Gnanadesikan 1999, the authors parameterize m_e in terms of this density difference in Eq. 4.)
Response: We would like to thank the reviewer for taking the time to review this manuscript and the positive review. The insightful comments have improved the quality of this paper, especially in regard to bringing a more applied background to this theoretical approach. We are positive about being able to address all of the stated comments.

Main comments:

The following main comments are mostly limited to the wind-driven case.

I recommend that the authors look for this MOC enhancement effect in a numerical ocean or climate model as this is not very difficult to do (I realize that the paper points to future work to be done in this regard). The conceptual parameterization is based on GM, and many models use this parameterization: the effect should show up if it exists by programming the FW flux in the way envisioned in this study (I imagine using low vertical diffusivity). Based on many assumptions and parameter choices, the assertions made in the paper seem to be a stretch and in need of additional ways of illustration and validation.

Response: We agree with the reviewer and performed simulations with freshwater forcing using the University of Victoria Earth System Climate Model, version 2.9 (Uvic ESCM 2.9). A brief description of Uvic ESCM, the simulations and their results have been added to the manuscript (see section 5).

There are (too?) many parameters in the conceptual model. How physical are the parameters chosen for the eddy return flow term? How sensitive is the solution to these values? Towards the end of the paper, the authors appear to state the purpose of the paper as showing only that this type of effect (strengthening MOC with FW flux in wind-driven case) is possible in reality and/or a climate model, rather than showing that it actually exists. This might be an appropriate scope for the material covered so far in this paper, but that then also severely limits the weight carried by this study. More could be said about the likelihood of the main result to be real by conducting a parameter sensitivity study.

Response: We understand the reviewer's desire to bring our results closer to the real ocean. We have tried to motivate the choice of the parameters as good as we can in the paper. More importantly, we have provided an explicit condition on the parameters for which the main feature of the manuscript holds. Since the model presented here is however only conceptual in nature, we feel that we cannot push it any further than this. That is especially true because (like in the Stommel model) the meridional density differences require the definition of boxes in the ocean which are not very well constraint. We would be really grateful if the reviewer would allow us the mere qualitative statements that we are making in the text which we hope can make some contribution due to the analytic conditions that we provide.

A continuity argument is invoked to explain the MOC increase (p41, l25). Although strictly true (from eq. 5 and steady state), it would be helpful to state that in addition to satisfying continuity, M_N   Dˆ2 delta rho here (referring to Eq. 1) and that this must also be satisfied. As a result, the system must adjust in a very specific way to allow a steady solution. So only solutions where D increases (in the right way) with the FW flux are consistent with continuity and have a chance of being allowed. The paper could benefit from more physical explanations of the mathematics in general.

Response: We agree with the reviewer and have added to the manuscript that the qualitative result does depend on the assumption that M_N increases with D2 and Delta rho and that thereby only solutions where D increases with the freshwater flux are allowed. We have further added the physical interpretation that the meridional density differences are the main driver of the changes both in M_N and M_E while the vertical density changes (as represented by changes in the pycnocline depth D) play an important role in stabilizing the circulation (as mentioned for the model without Southern Ocean density difference in Fürst & Levermann, Clim. Dyn.).

The validity of the main result appears to depend on how valid the GM parameterisation is for the real ocean. There are of course more complex parameterisations, and there are also results from eddy resolving and permitting models. The paper should contain some discussion of how the results might depend on GM specifically, using existing literature.

Response: The original parameterization of Gent and McWilliams is very well motivated the baroclinic instability and provides a very clear and qualitative large-scale representation of the eddy flux across a sloping density front. The wealth of adjustments that have been suggested to this formula might be very valid in specific situation and it is not our desire to dismiss any of this work. However, there must be about 1000 studies published on different variations of the GM parameterizations. For the current paper it is the representation of the slope by both a horizontal (meridional) and vertical measure for the density structure, we find that a serious discussion of the various variations of the GM-parameterization would require a lot of additions to the manuscript and any reduced discussion would be unfair to those left out. We would again be grateful if we could omit this or would be grateful if the reviewer would suggest specific issues that he or she considers particularly important. We are happy to include them. It is just that we feel that a comprehensive coverage of the matter would expand the paper a lot.

The definition and significance of terms like “bistability” and “threshold” is not very clear in the paper. Bistability should refer to a situation where two steady states exist under the same forcing, whereas by “threshold” I think the authors mean a FW flux above which no northern sinking occurs. This should be spelled out more in the paper, and something should also be said about the circulation when northern sinking is absent. Any negative strength overturning states should also be interpreted.

Response: This is an important shortcoming of our description and we would like to thank the reviewer for pointing this out. There have been a number of studies and speculations about the circulation in the absence of northern sinking. Some models show an inverse circulation, which is sometimes associated with the Antarctic Bottom Water filling up the Atlantic (Rahmstorf et al. GRL, 2005) and some show a seemingly stagnant ocean (Stouffer et al. J Clim. 2005). Neither of these patterns would be properly captured by the physics that is incorporated in the conceptual model that we present. It is thus true that we cannot really describe a bistable situation but rather a threshold behaviour that describes that beyond a certain freshwater flux the circulation in the Atlantic cannot be captured by the conceptual model and is thereby not a classic overturning circulation as presently observed. We have now added a brief explanation to this end stating that the conceptual model does not capture the “off-state”. Furthermore we have exchanged the term bistability with threshold behaviour and cleaned the text in this respect.

The results depend on a very specific spatial pattern of moisture transport changes, namely an increase of FW transport from the low to the high northern latitude box. How do the authors expect the overturning to change under more general moisture flux changes? For instance, a freshening of the SO could negate this effect?

Response: The application of freshwater to the North Atlantic and infact a number of different places in the global ocean has been under intense investigation since the groundbreaking papers by Manabe and Stouffer starting in 1988. The behaviour of our conceptual model is rather straight forward when it comes to the “normal” addition of freshwater into the north Atlantic which is then compensated in the Southern Ocean. The really new issue in our model is the fact that the overturning **can** get stronger prior to collapse if the net-freshwater transport occurs between the high and low latitudes in the Atlantic. While we would be reluctant to discuss these “normal” freshwater forcings in the paper because we believe it would dilute the content of the study, we have studied the low-to-high-latitude case in detail: We have carried out experiments with freshwater forcing from the lower latitudes into the southern ocean and freshwater forcing from lower latitudes towards both higher southern and northern latitudes. All experiments lead to the same behavior in the overturning: there is an increased overturning prior to m_E becoming negative. In all cases, this can be explained with a decline in the meridional southern density difference, which in turn leads to a decline in m_E and due to continuity, m_N, the measure of the overturning strength, increases. We included a note about these results in the manuscript. However, including other freshwater fluxes into the model would make it necessary to recalculate the governing equations, i.e. a different version of this box-model. We feel this would expand the paper a lot and we would therefore prefer not to include this in the manuscript.

I don’t understand (immediately) why the threshold is reached when the eddy return flow becomes negative (Fig. 5), or whether this is physical. What does this change in sign mean? Why is the threshold reached there? Is this physical or an artifact? This should be explained clearly.

Response: We agree with the reviewer that this point might not have been clearly enough explained in the manuscript. This is also a matter of the validity of our model assumptions. The eddy return flow becomes negative because the southern meridional density difference becomes negative. Our parametrization for the eddy is only valid for a positive eddy return flow. That means our model describing the current state of the overturning circulation is not valid anymore which is interpreted as a point if instability of the circulation pattern the model describes. Physically this means that there is no outcropping of iso-pycnals in the Southern Ocean anymore in this case and thus the eddy return flow does not follow the physics that is described by the baroclinic instability and thereby the Gent-and-McWilliams equation. This also establishes a qualitatively different circulation pattern. We included a more detailed description in the manuscript.

The work is only valid for steady states. This makes the conclusions less relevant to any global warming scenarios (they generally are not in steady state), even though the paper alludes to “monitoring activities” in the abstract. This should be stated.

Response: We agree with the reviewer that this is a very important point and included it in the conclusion section.

More specific comments:

The density difference between low and high northern latitudes sets the overturning strength. However, studies have shown the importance of the density difference be- tween the Southern Ocean and the North Atlantic (Saenko and weaver, GRL 2003). How important is this choice in the context of this paper?

Response: While the introduction of the meridional density difference between low latitudes and the southern ocean in the parametrization of the eddy return flow is the main point of the paper and was motivated with the AMOC crossing the Atlantic in the south and in the north, the overall strength of the overturning as driven by the northern and low latitude box is not central. Picking the Southern box would make the analysis considerably more complicated. The main motivation for this choise is as follows: The circulation of the surface waters need to be in geostrophic balance, thus these circulations have to follow a density difference. These strong changes in density have been observed in the real ocean and numerical models and it is a key point in our model to include both a northern and southern meridional density difference in our model. However, the exact location of this density change can vary for the real ocean.

p32. l1. “”it was shown that this kind of model. . .” This sentence is not clear, it would be good to first explain that Gnadesikan 1999 does not appear to be concerned with the role of the density difference, that it is treated as a constant there.

Response: We agree with the reviewer. Thank you for the detailed reading. We rephrased this sentence accordingly.

p32. l21.”The threshold behaviour found here is consistent with the salt-advection feed- back in. . .”. These sentences are not clear. What is meant by “net salinity transport by the overturning”? Does it mean the overturning exports fresh water from the Atlantic basin?

Response: This sentence refers to the salt advection feedback and the transport of salt rich water by the AMOC into the northern Atlantic. We rephrased this sentence in the manuscript in order make this point more clear.

p32. l24. Reference to Huisman 2010. This paper only deals with a Rahmstorf type box model and a fully implicit global ocean model, not “a number of climate models” as stated in the text here. Also, unlike what is implied in the paper here, dynamics are not examined for observations in Huisman 2010.
Response: We included citations from Weaver 2012 (GeophysicalResearch Letters) and Bryden 2011 (Journal of Marine Research) into the manuscript here.

Manuscript prepared for Earth Syst. Dynam.
with version 2014/07/29 7.12 Copernicus papers of the LaTeX class copernicus.cls.
Date: 12 August 2014

Mechanism for potential strengthening of Atlantic overturning prior to collapse

D. Ehlert[1,2,*] and A. Levermann[1,2]

[1]Potsdam Institute for Climate Impact Research and Institute of Physics, Potsdam, Germany
[2]Potsdam University, Potsdam, Germany
[*]now at: Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence to: D. Ehlert (dehlert@sfu.ca)

Abstract. The Atlantic meridional overturning circulation (AMOC) carries large amounts of heat into the North Atlantic influencing climate regionally as well as globally. Paleorecords and simulations with comprehensive climate models suggest that the positive salt-advection feedback may yield a threshold behaviour of the system. That is to say that beyond a certain amount of freshwater flux into the North Atlantic, no meridional overturning circulation can be sustained. Concepts of monitoring the AMOC and identifying its vicinity to the threshold rely on the fact that the volume flux defining the AMOC will be reduced when approaching the threshold. Here we advance conceptual models that have been used in a paradigmatic way to understand the AMOC, by introducing a density-dependent parameterization for the Southern Ocean eddies. This additional degree of freedom uncovers a mechanism by which the AMOC can increase with additional freshwater flux into the North Atlantic, before it reaches the threshold and collapses: an AMOC that is mainly wind-driven will have a constant upwelling as long as the Southern Ocean winds do not change significantly. The downward transport of tracers occurs either in the northern sinking regions or through Southern Ocean eddies. If freshwater is transported, either atmospherically or via horizontal gyres, from the low- to high-latitudes, this would reduce the eddy transport and by continuity increase the northern sinking which defines the AMOC until a threshold is reached at which the AMOC cannot be sustained. If dominant in the real ocean this mechanism would have significant consequences for monitoring the AMOC.

1 Introduction

The Atlantic meridional overturning circulation (AMOC) is being considered as one of the tipping elements of the climate system (Lenton et al., 2008). While the definition by Lenton et al. (2008) is based on the idea that tipping elements respond strongly to a small perturbation, the AMOC might also be a tipping element in the dynamic sense of the word (Levermann et al., 2012). That is to say that a small external perturbation induces a self-amplification feedback by which the circulation enters a qualitatively different state. This self-amplification is due to the salt-advection feedback (Stommel, 1961; Rahmstorf, 1996) and has been found in a number of comprehensive ocean as well as coupled climate models (Manabe and Stouffer, 1993; Rahmstorf et al., 2005; Stouffer et al., 2006b; Hawkins et al., 2011). A cessation of the AMOC would have far-reaching implications for global climate (Vellinga and Wood, 2002) which include (1) a strong reduction of northern hemispheric air and ocean temperatures (Manabe and Stouffer, 1988; Mignot et al., 2007), (2) a reduction in European precipitation and (3) its wind pattern (Laurian et al., 2009), (4) a dynamic sea level increase in the North Atlantic (Levermann et al., 2005; Yin et al., 2009), (5) a perturbation of the Atlantic ecosystem (Schmittner, 2005; Kuhlbrodt et al., 2009), (6) a southward shift in the tropical rain belt and associated impacts on vegetation (Stouffer et al., 2006a) and (7) a perturbation of the Asian monsoon system (Goswami et al., 2006).

Conceptual models to capture the basic aspect of a meridional overturning circulation can be divided into models in which the overturning strength is determined by the meridional density difference in the Atlantic (Stommel, 1961; Rahmstorf, 1996) and those in which its strength is linked to the vertical density structure (Gnanadesikan, 1999). Stommel's model captures the salt-advection feedback in a pure form by resolving only the advection of the active tracers in two fixed-size boxes representing the northern downwelling and southern upwelling regions. The overturning strength is assumed to be proportional to the meridional density difference which was found to be valid in a number of ocean and climate models (e.g. Griesel and Morales-Maqueda, 2006; Rahmstorf, 1996; Schewe and Levermann, 2010). The Stommel-model is however missing a representation of the energy-providing processes for the overturning, such as the Drake-Passage effect and low-latitudinal mixing (Kuhlbrodt et al., 2007) as well as the influence of the Southern Ocean eddy circulation.

These processes are captured in a conceptual way by the model of Gnanadesikan (1999) which links the overturning to the vertical density profile as represented by the pycnocline depth **but treats meridional density differences as a constant**. It was shown that this kind of model is not consistent with the fact that the meridional density gradient indeed changes with changing overturning in a number of different climatic conditions (Levermann and Griesel, 2004; Griesel and Morales-Maqueda, 2006). By construction it does not capture the salt-advection feedback and can thereby not be used to study the possibility of a threshold behaviour of the overturning.

There have been a number of attempts to combine these two approaches and thereby to comprise the horizontal tracer-advection with the vertical one (Marzeion and Drange, 2006; Johnson et al., 2007; Fürst and Levermann, 2011).

Here we advance the simplest of the suggested models (Fürst and Levermann, 2011) by introducing an additional parametrisation for the Southern Ocean eddy flux. As found in a comprehensive coarse resolution ocean model (Levermann and Fürst, 2010) the horizontal scale of the Southern upwelling region can change and neglecting this change leads to a misrepresentation of the circulation within the Gnanadesikan (1999) framework. We attempt to complement the conceptual model in order to correct for this shortcoming. To this end we add a variable, meridional density difference in the southern Atlantic ocean in the scaling of the eddy-induced return flow. As will be shown, this allows for a qualitatively different response of the AMOC under freshwater forcing compared to earlier studies: a growth of the northern deep water formation with increasing freshwater flux from low- to high northern latitudes within the Atlantic before the threshold is reached and no AMOC in the modelled sense can be sustained. The threshold behaviour found here is consistent with the salt-advection feedback in the sense of a net-salinity transport **from lower latitudes into the northern Atlantic** by the overturning as suggested by Rahmstorf (1996). **This threshold behaviour has been shown in box models and complex climate models (Huisman et al., 2010; Weaver et al., 2012) but also in observations (Bryden et al., 2011)**.

This paper is structured as followed: firstly we describe the parametrisation of the transport processes, pycnocline dynamics and salinity dynamics, i.e. horizontal density distribution (Sect. 2). The transport processes include the two fundamental driving mechanism (Kuhlbrodt et al., 2007) which are low-latitudinal upwelling (Munk, 1966; Munk and Wunsch, 1998; Huang, 1999; Wunsch and Ferrari, 2004) and wind-driven upwelling in southern latitudes (Toggweiler and Samuels, 1995, 1998). In order to examine the behaviour of the model we derived governing equations for the two driving mechanisms separately as well as for the full case. The threshold behaviour, as described by Stommel (1961) is caused by the salinity advection. For simplicity we keep the temperatures fixed through-out the paper (Sect. 3). Section 4 discusses the change in the AMOC with increasing freshwater flux into the North Atlantic for the wind-driven case and the full case. **Also section 5 discusses the behavior of the AMOC under freshwater forcing, but for simulations using a complex climate model**. We conclude in Sect. 6.

2 Model description

We use a standard inter hemispheric model with four varying boxes (Fig. 1): (1) a northern box representing the northern North Atlantic with deep water formation, (2) an upper low-latitudinal box and (3) a deeper low-latitudinal box below the pycnocline, (4) a southern box with southern upwelling and eddy return flow (Gnanadesikan, 1999). The northern and southern boxes are fixed in volume while the low-latitudinal boxes vary in size according to the dynamically computed pycnocline depth. **The four meridional tracer transport processes between the boxes control the horizontal and vertical density structure on the one hand and they control the overturning on the other hand.** The density structure, in turn, determines the transport processes. Changes in the vertical density structure are described by variations in the pycnocline depth. The horizontal density structure is expressed by a southern and a northern meridional density difference. They depend on the dynamics of the active tracers, temperature, T, and salinity, S. For simplicity we assume a linear density function $\Delta\rho = \rho_0(\beta_S \Delta S - \alpha_T \Delta T)$ (Stommel, 1961). In order to capture the main feedback for a threshold behaviour while keeping the model legible, we include salinity advection and neglect changes in temperature. The simplification further is justified because the temperature in the upper layers is strongly coupled to atmospheric temperature which is to first order determined by the solar insulation. We thus assume, the ocean temperature in the upper layers to be constant. The high-latitudinal boxes represent strong out-cropping regions which homogenizes the water column and extends the argument to depth. In steady state, the fourth box, deeper low-latitudes ocean, is determined by the three other boxes. That means the approximation is valid for the whole model in equilibrium and temperature is used as an external parameter. The base of our work is the model by Fürst and Levermann (2011). We use the same parametrisations for the northern deep water formation and the upwelling processes. For the eddy return flow we introduce a different scaling by implementing southern meridional density difference which has strong influences on the behaviour of the model (Sects. 3 and 4).

2.1 Tracer transport processes

Different scaling for the deep water formation (as summarized in Fürst and Levermann, 2011) have been suggested. Here we use a parametrisation suggested by Marotzke (1997) and apply a β-plane-approximation to it. The resulting northern sinking scales linearly with the meridional density difference and quadratically with the pycnocline depth following geostrophic balance and vertical integration.

$$m_N = \frac{Cg}{\beta_N L_y^N} \frac{\Delta\rho}{\rho_0} D^2 = C_N \Delta\rho D^2. \tag{1}$$

Because all values are external parameters (Table 1) except the meridional density difference $\Delta\rho = \rho_N - \rho_U$ and the pycnocline depth D, the parameters are comprised into one constant C_N. In contrast to previous approaches (e.g. Rahmstorf, 1996) the meridional density difference does not span the whole Atlantic but instead is taken between low and high northern latitudes in accordance with the geostrophic balance between the meridional density difference and the North Atlantic Current.

The low-latitudinal upwelling follows a vertical advection-diffusion balance (Munk and Wunsch, 1998). That is to say, downward turbulent heat flux is balanced by upward advection. This balance with a constant diffusion coefficient for the full upwelling region yields an inverse proportionality between upward volume transport and pycnocline depth. Again all external parameters are expressed by one constant C_U to obtain

$$m_U = B\, L_U \frac{\kappa}{D} = \frac{C_U}{D}. \tag{2}$$

The southern upwelling term is considered to be independent of the pycnocline depth and results from the so-called Drake-Passage effect (Toggweiler and Samuels, 1995):

$$m_W = B\,\frac{\tau_{\mathrm{Dr}}}{|f_{\mathrm{Dr}}|\rho_0} = C_W. \tag{3}$$

The eddy return flow is parametrised following Gent and McWilliams (1990) which yields a tracer transport proportional to the slope of the outcropping isopycnals. In the formulation of Gnanadesikan (1999) this is represented by a linear dependence on the pycnocline depth divided by a horizontal scale for the outcropping region which is taken to be constant. The assumption of a constant horizontal scale for the outcropping region is not consistent with freshwater hosing experiments in a comprehensive though coarse resolution ocean model (Levermann and Fürst, 2010). **Levermann and Fürst (2010) show that the eddy return flow is proportional to pycnocline depth over a variable horizontal scale of the outcropping**. Here we attempt to capture variations in the meridional horizontal length scale of the outcropping region by the meridional density difference between the low-latitude ocean and the Southern Ocean, $\Delta\rho_{\mathrm{SO}} = \rho_S - \rho_U$. We thus use the parametrisation

$$m_E = B\,A_{\mathrm{GM}}\frac{\Delta\rho_{\mathrm{SO}}}{\rho_0}\frac{D}{H} = C_E\,\Delta\rho_{\mathrm{SO}}D. \tag{4}$$

As before, all quantities except D and $\Delta\rho_{\mathrm{SO}}$ are external parameters and compressed into one constant C_E.

2.2 Pycnocline and salinity dynamics

The temporal evolution of the pycnocline is determined by the tracer transport equation following Marzeion and Drange (2006).

$$BL_U\frac{\mathrm{d}D}{\mathrm{d}t} = m_U + m_W - m_E - m_N \tag{5}$$

Salinity equations for each box are derived from the advection in and out of the box, conserving salinity, as well as the surface fluxes, F_N and F_S which represent atmospheric freshwater transport as well as the horizontal gyre transport. The advection scheme follows the arrows shown in Fig. 1. In computing the temporal changes in total salinity the changes in volume due to the pycnocline dynamics needs to be accounted for.

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_U S_U) = m_U S_D + m_W S_S - S_U(m_N + m_W) + S_0(F_N + F_S) \tag{6a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_N S_N) = m_N(S_U - S_N) - S_0 F_N \tag{6b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_S S_S) = m_W(S_D - S_S) + m_E(S_U - S_S) - F_S S_0 \tag{6c}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_D S_D) = m_N S_N + m_E S_S - S_D(m_U + m_W). \tag{6d}$$

With finite difference method applied to Eqs. (1)–(6), we made numerical simulations which reached in equilibrium the values shown in Table 2 with the parameters given in Table 1.

3 Governing equation

Here we derive an equation for the steady-state solution of Eqs. (1)–(6) by comprising them into one equation of the oceanic pycnocline, D. We derive governing equations for the full case as well as for the analytically solvable cases of a purely mixing- and a purely wind-driven cases. The model is limited to positive and real solutions for the pycnocline (**see figure 2**) as well as for non-negative tracer transport values. A parameter combination that does not allow for a solution of this kind is thereby inconsistent with an overturning circulation as represented by this model. We denote a parameter region for which no such a physical solution exists as an “AMOC-off-state-region”. As in the earlier version of the model (Fürst and Levermann, 2011) we find a threshold behaviour with respect to an increase of the freshwater flux, F_N, for all three cases. The focus of this study is not to show the existence of such a threshold of all parameter values**. But, it is** to present a mechanism by which the overturning can increase **between steady states under different freshwater forcings** before the threshold is reached and no AMOC can be sustained.

3.1 Full case

In the full case the governing equation is a polynomial of 10th order in the pycnocline depth (Appendix A1, Eq. A7). Thus solutions can only be found numerically. Of the 10 mathematical roots, two are positive and real but of two adjacent solutions only one can be stable. Numerical solutions were obtained in two ways. First by finding the roots of the polynomial (Appendix A1, Eq. A7) and second by time forward integration of the original set of Eqs. (1)–(6) with different initial conditions. The time integration naturally selects the stable solutions. Though this is not a proof by any means, we feel confident to say that the solution with $D = 616\,\mathrm{m}$ is the stable of the two physical solutions (Fig. 3a). The corresponding tracer transport values are **provided** in Fig. 3b. The northern sinking decreases with increasing freshwater forcing for the parameter set of Table 1. The equation for the northern sinking as it results from the scaling (Eq. 1) and the salinity equations:

$$m_N = -\frac{1}{2}C_N D^2 \alpha \Delta T \pm \sqrt{\frac{1}{4}C_N^2 D^4 \alpha^2 \Delta T^2 - C_N D^2 \beta F_N S_0} \tag{7}$$

was also valid in the earlier version of the model (Fürst and Levermann, 2011). Rahmstorf (1996) provides a similar solution for the northern deep water formation with k as proportionality factor between the northern sinking and the north-south density difference:

$$m_N = -\frac{1}{2}k\alpha(T_S - T_N) \pm \sqrt{\frac{1}{4}k^2\alpha^2(T_S - T_N)^2 + k\beta F_S S_0}. \tag{8}$$

In these earlier models only positive roots of the solution yield stable equilibria. That differs from our model where for certain amounts of freshwater forcing the negative sign of the root in Eq. (7) (respectively Eq. 8) needs to be considered, as for example in the wind-driven case discussed below.

The threshold of the overturning is reached when the eddy return flow becomes negative (Fig. 3b, grey shaded area) **because the parametrization of the eddy return flow is only valid for positive**

values. That means the model presented here is not valid under negative eddy return flow. This is interpreted as a point of instability of the circulation pattern the model describes. Physically reaching the threshold means that there is no outcropping of iso-pycnals in the Southern Ocean anymore. Thus the eddy return flow does not follow the physics that is described by the baroclinic instability and thereby it does not follow the parametrisation by Gent and McWilliams (1990). This also establishes a qualitatively different circulation pattern.

It should also be noted that also negative freshwater forcing was applied, which might not be applicable with surface transport. However, the threshold freshwater forcing is in the positive range. This is also true for the mixing- and wind- driven case.

Besides freshwater forcing from lower latitudes into the northern box only, experiments were performed with freshwater forcing from the lower latitudes into the southern box and from lower latitudes into both southern and northern box. All experiments showed the same behaviour in the overturning. The reason cloud be that all experiments affect the meridional density differences in the same way.

3.2 Mixing-driven case

The purely mixing-driven case is defined by $C_E = C_W = 0$. In this case the pycnocline dynamics in steady state (Eq. 5) reduces to $m_N = m_U = C_U/D$. As the eddy return flow is eliminated from the equation, this case has not changed compared to the model of Fürst and Levermann (2011): the governing equation is a polynomial of fourth order in pycnocline depth and has one physical solution which decreases with increasing freshwater forcing (Fig. 4a). The overturning decreases until a threshold level (Fig. 4b) which is reached when the pycnocline and therefore the tracer transport processes become complex. The critical northern freshwater flux can be calculated by zero-crossing of the discriminant of the polynomial.

$$F_{N,\text{mixing}}^{\text{crit}} = \frac{3(2C_N)^{1/3} C_U^{2/3} \alpha^{4/3}}{8\beta S_0} |\Delta T|^{4/3} \tag{9}$$

3.3 Wind-driven case

The purely wind-driven circulation is defined by $C_U = 0$. Thus the tracer-transport balance in steady state (Eq. 5) reduces to $m_N = m_W - m_E$ into which the eddy return flow and the northern sinking are included as functions of the pycnocline depth and external parameters of Table 1 (see Appendix A for a detailed derivation). For the northern sinking the northern salinity difference is calculated via the salinity balance of North Atlantic (Eq. 6b) and inserted into the scaling of the northern sinking (Eq. 1), similarly for the eddy return flow by using the Southern Ocean salinity balance (Eq. 6c). The emerging governing equation is a third order polynomial of the pycnocline depth D which we solve analytically.

$$D^3 C_E C_N \alpha \Delta T \left[\frac{\beta S_0 (F_N + F_S)}{C_W} + \alpha \Delta T_{\mathrm{SO}} \right]$$

$$+ D^2 [C_N F_N S_0 \beta + C_N C_W \alpha \Delta T + \frac{C_E^2}{C_W^2} (S_0 \beta (F_N + F_S) + C_W \alpha \Delta T_{\mathrm{SO}})^2]$$

$$+ D 2 C_E [\beta S_0 (F_N + F_S) + C_W \alpha \Delta T_{\mathrm{SO}}] + C_W^2 = 0$$

The solutions depend on the sign of the discriminant of the polynomial $d = (q/2)^2 + (p/3)^3$ with p and q defined as:

$$\frac{q}{2} = \frac{1}{2} \left(\frac{C_N F_N S_0 \beta + C_N C_W \alpha \Delta T + C_E^2 A^2}{3 C_E \alpha \Delta T A} \right)^3 - \frac{C_W F_N S_0 \beta + C_W^2 \alpha \Delta T}{3 C_E C_N \alpha^2 \Delta T^2 A}$$

$$- \frac{C_E C_W A}{3 C_N^2 \alpha^2 \Delta T^2} + \frac{C_W^2}{2 C_E C_N \alpha \Delta T A}$$

$$\frac{p}{3} = \frac{6 C_W C_N \alpha \Delta T - 1}{9 C_N^2 \alpha^2 \Delta T^2} - \frac{F_N S_0 \beta + C_W \alpha \Delta T}{9 C_E^2 C_N^2 \alpha^2 \Delta T^2 A^2}$$

with $A = \left(\frac{S_0 \beta}{C_W} (F_N + F_S) + \alpha \Delta T_{\mathrm{SO}} \right)$.

A polynomial of third order has either one root (Appendix A2, Eq. A9) if the discriminant is positive, or three roots (Appendix A2, Eq. A10) if the discriminant is negative which is the case for the parameters of Table 1 near the threshold (Fig. 5). Only one of the three mathematical roots is a physical solution of equilibrium state of the model because one root is negative (Fig. 5, solution 1) and the other solution has a negative northern sinking and the pycnocline values are out of range of the ocean depth (Fig. 5, solution 0). No physical solution exists, when the eddy return flow becomes negative. At this threshold the discriminant of the governing equation has a negative pole which can be used to calculate the critical freshwater flux. In the following we describe a more straight forward way to give dependencies of the critical freshwater flux. Assuming steady state for the salinity balance of the upper low-latitudinal box (Eq. 6a equal to zero, with $m_U = 0$) and for the tracer transport balance ($m_E + m_N = m_W = C_W$), the salinity difference between the Southern Ocean and the upper low-latitudes emerges:

$$\Delta S_{\mathrm{SO}} = S_S - S_U = -\frac{S_0}{C_W} (F_N + F_S).$$

The salinity difference contains no variables. As the temperature dynamics are not considered in this model, the horizontal density difference between these two boxes is constant for a fixed set of parameters.

$$\Delta \rho_{\mathrm{SO}} = \beta \Delta S_{\mathrm{SO}} - \alpha \Delta T_{\mathrm{SO}} = -\beta \frac{S_0}{C_W} (F_N + F_S) - \alpha \Delta T_{\mathrm{SO}} \tag{10}$$

The critical eddy return flow is equal to zero. Using the definition of the flow (Eq. 4) and the fact that the critical pycnocline depth is far in the positive range, Eq. (10) can be set to zero at the threshold level. The critical freshwater flow becomes:

$$F_{N,\text{wind}}^{\text{crit}} = -\frac{\alpha \Delta T_{\text{SO}} C_W}{S_0 \beta} - F_S$$

The critical northern freshwater flow depends linearly on the southern temperature difference and on the southern wind stress (via C_W) and a higher southern freshwater flux would lower the critical northern freshwater flow. Please note that this is a significant difference to previous approachers (Fürst and Levermann, 2011; Rahmstorf, 1996), where the critical freshwater flow is in first or higher order (Eq. 9) sensitive to the northern temperature difference which has no influence onto the critical freshwater flux in this case.

4 Freshwater-induced AMOC strengthening

The introduction of the southern density difference as a variable changing the eddy return flow results in a mechanism that **has rarely been reported before**: **an increasing overturning under northern freshwater forcing prior to a threshold in freshwater beyond no AMOC, as described here, can be sustained. Cimatoribus et al. (2014) found a similar behaviour in a different box model, but under freshwater forcing from the southern into the northern Atlantic. The mechanism in the model described here** is simple: a freshwater flux from low-latitudes into the high northern latitudes reduces the eddy return flow. If this reduction is not compensated by a reduction in mixing-driven upwelling (as for example in a mainly wind-driven AMOC) then due to continuity northern sinking has to increase since Southern Ocean upwelling is constant. **Furthermore, it should be noted that this result depends on the assumption that the northern sinking, m_N, is a function of the square of the pycnocline depth and the meridional density difference (see equation 1). Consequently, only solutions of the pycnocline depth are allowed which increases in the right matter with increasing freshwater flux. In general, the changes in the meridional density differences are the main driver for changes in the northern sinking and the eddy return flow, i.e. the drivers for the freshwater induced strengthening of the AMOC. Changes in the vertical density differences, implemented here as changes in the pycnocline depth, stabilize the overturning circulation.** The mechanism **of an AMOC strengthening under freshwater forcing** is always dominant in the wind-driven case which we will proof at the end of this section. In the full case the mechanism takes not effect for the parameters of Table 1 but it emerges if the Southern Ocean temperature difference is changed in such a way as to make the mixing less relevant (Fig. 6).

4.1 Full case

In order to gain a better understanding of this behaviour, the tracer transport processes balance in steady state (Eq. 5 equal to zero) is differentiated with respect to the northern freshwater flux. That gives an equation for the derivative of northern sinking:

$$\frac{\mathrm{d}m_N}{\mathrm{d}F_N} = -\frac{\mathrm{d}m_E}{\mathrm{d}F_N} + \frac{\mathrm{d}m_U}{\mathrm{d}F_N} \tag{11}$$

Using the parametrisations of the eddy return flow (Eq. 4) and low-latitudinal upwelling (Eq. 2), Eq. (11) yields

$$\frac{\mathrm{d}m_N}{\mathrm{d}F_N} = -\left(\frac{C_U}{D^2} + C_E \Delta\rho_{\mathrm{SO}}\right)\frac{\partial D}{\partial F_N} - C_E D \frac{\partial \Delta\rho_{\mathrm{SO}}}{\partial F_N}.$$

The polynomial consists of two terms of opposing sign: the first term on the left depends on the change of pycnocline depth (representing the vertical density structure) with increasing freshwater flux. Since this derivative, $\frac{\partial D}{\partial F_N}$, is generally positive the full term is negative. The second term is positive since the horizontal density difference in the Southern Ocean declines when F_N is increased. The sign of the derivative of the northern sinking is determined by the ratio between the two terms. Thus **strongly** increasing pycnocline depth, i.e. strong positive changes in vertical density structure, shift the overturning to a deceasing threshold behaviour. If the southern meridional density difference decreases stronger (in absolute values), then the overturning rises under freshwater forcing. The crucial point is that the absolute value of pycnocline **depth** is present in the term with the derivative of southern meridional density difference. That means rising pycnocline depth also amplifies the term that depends on horizontal density structure and vice versa for the meridional density difference. A stronger statement can be derived for the purely wind-driven case.

4.2 Wind-driven case

Upwelling in the lower latitudes amplifies the decreasing of northern sinking with increasing freshwater flow. Therefore, the wind-driven case provides a better example and a clearer image. Without low-latitudinal upwelling the derivative of northern sinking (Eq. 11) equals the negative derivative of the eddy return flow ($\mathrm{d}m_N/\mathrm{d}F_N = -\mathrm{d}m_E/\mathrm{d}F_N$). From the scaling of the eddy return flow (Eq. 4) and the derivative of the southern horizontal density difference (Eq. 10) the derivative of the northern sinking emerges.

$$\begin{aligned}\frac{\mathrm{d}m_N}{\mathrm{d}F_N} &= -C_E \Delta\rho_{\mathrm{SO}} \frac{\partial D}{\partial F_N} - C_E D \frac{\partial \Delta\rho_{\mathrm{SO}}}{\partial F_N} \\ &= C_E \left(\frac{S_0 \beta}{C_W}(F_N + F_S) + \alpha \Delta T_{\mathrm{SO}}\right)\frac{\partial D}{\partial F_N} + C_E D \frac{S_0 \beta}{C_W}\end{aligned}$$

Now, solely the term depending on the negative southern density difference could diminish the derivative. For the values given in Table 1, $\frac{\partial D}{\partial F_N} \simeq \frac{100\,\mathrm{m}}{0.1\,\mathrm{Sv}}$, and $D \simeq 1000\,\mathrm{m}$, the derivative is far in the positive range ($\frac{\partial m_N}{\partial F_N} \simeq 5000$). In order to calculate the critical derivative, we use again the fact that the southern density difference equals zero at the threshold.

$$\left(\frac{\mathrm{d}m_N}{\mathrm{d}F_N}\right)_{\mathrm{crit}} = C_E D_{\mathrm{crit}} \frac{S_0 \beta}{C_W} > 0$$

The emerging critical derivative depends only on positive constants and the positive critical pycnocline depth, i.e. the overturning always increases close to the threshold. This result is not surprising in light of the heuristic explanation given above, but it is not trivial due to the still complex vertical and horizontal density dynamics.

5 Climate model experiments

In order to investigate the possibility of the occurrence of a freshwater induced AMOC strengthening in a more complex climate model experiments were performed with the University of Victoria Earth System Climate Model, version 2.9 (UVic ESCM). UVic ESCM 2.9 is a model of intermediate complexity, with a simple 1-dimensional atmosphere but a 3-dimensional dynamic ocean (Weaver et al., 2001; Eby et al., 2009). The model was forced with a constant freshwater flux and run to equilibrium over 4300 years. The constant freshwater flux ranged from 0.025Sv to 0.2Sv. Freshwater was transferred from the southern Atlantic (10°S to 30°S) into the northern Atlantic (10°N to 30°N) in all simulations. The maximum overturning, averaged over 1000 years, for these equilibrium simulations are shown in figure 7. The overturning increases for a freshwater forcing of 0.075Sv before it declines at 0.1Sv. The overturning creases under a freshwater forcing of 0.16Sv or higher. The AMOC strengthening is less pronounced compared to the box model behavior. However, due to the strong differences between the box models and UVic ESCM slightly different behaviors can be expected. Furthermore, the southern ocean is not well represented in complex climate models, especially eddy flows. These experiments show that an increase of the AMOC under freshwater forcing is a possible behavior of the overturning. However, further experiments would be needed to investigate the robustness of this behavior.

6 Conclusion and discussion

The conceptual model of the Atlantic overturning presented here builds on a previous model (Fürst and Levermann, 2011) and advances the model by the introduction of a dynamic southern ocean density difference for the eddy return flow as imposed by comparison with comprehensive ocean model results (Levermann and Fürst, 2010). As a first result the model reproduces the qualitative result that a threshold behaviour is a robust feature that is independent of the driving mechanism, i.e. it is present in a mixing-, a wind-driven as well as in a combined case. The regime of existence of a solution for the overturning for a specific parameter combination is defined by the simultaneous compliance of a number of conditions, e.g. positive volume fluxes and pycnocline depth. In the presented model the threshold is generally reached when the eddy return flow becomes negative. Similar to the predecessor of the model also here the threshold is associated with the salt-advection feedback. As suggested by Rahmstorf (1996), a threshold thus only exists when the salinity in the low-latitude box is higher than in the northern box. This is the case here (see Table 2). Whether the real ocean is in a bistable regime and thereby exhibits a threshold behaviour is of yet unclear. According to a diagnostic by Rahmstorf (1996), an overturning is bistable if the overturning carries a net salinity transport at 35 N. This diagnostic was confirmed to be valid in a comprehensive climate model (Dijkstra, 2007) and is discussed in depth by Hofmann and Rahmstorf (2009). **Following this diagnostic most climate models do not show a threshold behaviour in earlier studies (Drijfhout et al., 2010). However, in a more recent model in-comparison study the majority of climate models do show a threshold behaviour (Weaver et al., 2012). Also observational data indicates that the real ocean is in a bistable regime (Bryden et al., 2011), i.e. the current circulation pattern could change after reaching a threshold. It should be noted that the model presented here does not capture an "off-state" of the circulation, i.e. describing a circulation pattern after the threshold in freshwater forcing has been crossed. There are models showing an inverse circulation, which is sometimes associated with the Antarctic Bottom Water filling up the Atlantic (Rahmstorf et al., 2005) and other models show a seemingly stagnant ocean (Stouffer et al., 2006b). Neither of these patterns would be properly captured by the physics that is incorporated in the conceptual model presented here. Therefore, a bistable situation can not be described but rather a threshold behaviour. This threshold behaviour shows that beyond a certain freshwater flux the circulation in the Atlantic cannot be captured by the conceptual model and is thereby not a classic overturning circulation as presently observed.**

The main result is the observation that the overturning can increase prior to its collapse in response to a freshwater flux from low-latitudes to high northern latitudes. Previous models including the base models (Johnson et al., 2007; Marzeion and Drange, 2006; Fürst and Levermann, 2011) show the opposite behaviour, similar to the bifurcation in the initial model of Stommel (1961). The emergence of the effect depends on the inclusion of Southern Ocean winds as a driving-mechanism for the overturning and the inclusion of a dynamic southern ocean horizontal density difference. It thus does not include in the mixing-driven case. Thus our model has opposite behavior prior to reaching the threshold depending on whether the circulation is wind- or mixing-driven.

This has strong implications for potential monitoring systems that aim to detect the vicinity to the threshold. Methods that depend on the decline of the overturning prior to the threshold for example in order to detect an increase in variability might not be suitable in a situation (Lenton, 2011; Scheffer et al., 2009) in which the presented mechanism is relevant. **However, applicability of these findings for monitoring purposes are limited as the presented results refer to a system in equilibrium, and not a time dependent state as we see under current global warming.**

Whether the mechanism described here is dominant in the real ocean is beyond the scope of this paper. This study presents the physical processes which need to be investigated with comprehensive quantitative models and verified against observation in order to assess its relevance. Though a large number of so-called water hosing experiments have been carried out (e.g. Manabe and Stouffer, 1995; Rahmstorf et al., 2005; Stouffer et al., 2007), few studies have focussed on freshwater transport from low- to high-latitudes. **We were able to show a strengthening of the AMOC under freshwater forcing prior to a decline of the overturning by prescribing different amounts of constant freshwater transport from low latitudes in the southern Atlantic into the northern Atlantic. However, this behavior is not strongly pronounced. Thus further experiments** are needed in order to find whether the mechanism is indeed relevant for the real ocean.

Appendix A: Analytical calculations

A1 Full case

The salinities are exchanged by salinity differences between the boxes except the salinity of the northern box. The new salinity variables are defined as $\Delta S = S_N - S_U$, $\Delta S_D = S_N - S_D$, $\Delta S_{\mathrm{SO}} = S_S - S_U$, and S_N. The salinity balance of the northern box gives for the northern salinity difference:

$$\Delta S = -\frac{S_0 F_N}{mN}. \tag{A1}$$

The scaling of the northern sinking (Eq. 1) with the linearly scaling of the meridional density difference $\Delta\rho = \beta\Delta S - \alpha\Delta T$ and Eq. (A1) yields into a quadratic polynomial of m_N.

$$0 = m_N^2 + m_N C_N D^2 \alpha \Delta T + C_N D^2 \beta F_N S_0 \tag{A2}$$

It has the solution:

$$m_N = -\frac{1}{2} C_N D^2 \alpha \Delta T \pm \sqrt{\frac{1}{4} C_N^2 D^4 \alpha^2 \Delta T^2 - C_N D^2 \beta F_N S_0}. \tag{A3}$$

The salinity balance of the upper box can be used to calculate ΔS_D:

$$\Delta S_D = \frac{m_W}{m_U} \Delta S_{\mathrm{SO}} + \Delta S + \frac{S_0}{m_U}(F_N + F_S). \tag{A4}$$

The salinity balance of the southern box combined with Eq. (A4) results into an equation for the southern salinity difference.

$$\Delta S_{\mathrm{SO}} = -S_0 \frac{m_W(F_N + F_S) + m_U F_S}{m_W^2 + m_W m_U + m_E m_U} \tag{A5}$$

The scaling of the eddy return flow (Eq. 4), the linear density function for southern meridional density difference ($\Delta\rho_{\mathrm{SO}} = \beta\Delta S_{\mathrm{SO}} - \alpha\Delta T_{\mathrm{SO}}$), and Eq. (A5) can be collapsed into a quadratic equation for m_E.

$$0 = m_E + C_E D \beta S_0 \frac{m_W(F_N + F_S) + m_U F_S}{m_W^2 + m_W m_U + m_E m_U} + C_E D \alpha \Delta T_{\mathrm{SO}} \tag{A6}$$

It has the solution:

$$
\begin{aligned}
m_E = & -\frac{1}{2}\left(\frac{m_W^2}{m_U} + m_W + C_E D\alpha\Delta T_{\mathrm{SO}}\right) \\
& + \sqrt{\begin{aligned} & \frac{1}{4}\left(\frac{m_W^2}{m_U} + m_W + C_E D\alpha\Delta T_{\mathrm{SO}}\right)^2 - C_E D\beta S_0\left(\frac{m_W}{m_U}(F_N + F_S) + F_S\right) \\ & - C_E D\alpha\Delta T_{\mathrm{SO}}\left(\frac{m_W^2}{m_U} + m_W\right)\end{aligned}}.
\end{aligned}
$$

The governing equation of the pycnocline depth emerges by using Eq. (A6) and replacing the eddy return flow by $m_E = m_U + m_W - m_N$, m_N^2 by Eq. (A2), and the upwelling transport processes, m_U and m_W, by their scaling (Eqs. 2 and 3).

$$
\begin{aligned}
& D^{10} C_E C_U C_W C_N^2 \alpha^2 \Delta T^2 \left[S_0\beta(F_N + F_S) + C_W\alpha\Delta T_{\mathrm{SO}}\right] \\
& + D^9 C_N \alpha\Delta T \left[C_E C_W \left(C_W^2 + C_E C_U \alpha\Delta T_{\mathrm{SO}}\right)\left(F_N S_0\beta + F_S S_0\beta + C_W\alpha\Delta T_{\mathrm{SO}}\right)\right. \\
& \left. + C_U C_N \left(C_W^2 F_N S_0\beta + C_W^3\alpha\Delta T + 2C_E C_U C_W \alpha^2\Delta T\Delta T_{\mathrm{SO}} + C_E C_U S_0\alpha\beta(F_S\Delta T + F_N\Delta T_{\mathrm{SO}})\right)\right] \\
& + D^8 \left[C_E^2 C_W^2 (F_N S_0\beta + F_S S_0\beta + C_W\alpha\Delta T_{\mathrm{SO}})^2\right. \\
& + C_N^2 C_U^2 \left(F_N^2 S_0^2\beta^2 + 3C_W F_N S_0\alpha\beta\Delta T + \alpha^2\Delta T^2\left(3C_W^2 + C_E C_U\alpha\Delta T_{\mathrm{SO}}\right)\right) \\
& + C_N \left(C_W^4 F_N S_0\beta + C_W^5\alpha\Delta T + C_E C_U C_W^2 S_0\alpha\beta\Delta T(3F_N + 4F_S) + 6C_E C_U C_W^3\alpha^2\Delta T\Delta T_{\mathrm{SO}}\right. \\
& + C_E^2 C_U^2 S_0\alpha^2\beta\Delta T_{\mathrm{SO}}(F_S\Delta T + F_N\Delta T_{\mathrm{SO}}) \\
& \left.\left. + 2C_E C_U C_W\left(-F_N^2 S_0^2\beta^2 - F_N F_S S_0^2\beta^2 + C_E C_U\alpha^3\Delta T\Delta T_{\mathrm{SO}}^2\right)\right)\right] \\
& + D^7 \left[C_N^2 C_U^3 \alpha\Delta T(2F_N S_0\beta + 3C_W\alpha\Delta T)\right. \\
& + 2C_E C_W (F_N S_0\beta + F_S S_0\beta + C_W\alpha\Delta T_{\mathrm{SO}})\left(C_W^3 + C_E C_U F_S S_0\beta + 2C_E C_U C_W\alpha\Delta T_{\mathrm{SO}}\right) \\
& + C_N C_U \left(4C_W^3 F_N S_0\beta + 6C_W^4\alpha\Delta T + 12C_E C_U C_W^2\alpha^2\Delta T\Delta T_{\mathrm{SO}}\right. \\
& \left.\left. + C_E C_U\left(-2F_N F_S S_0^2\beta^2 + C_E C_U\alpha^3\Delta T\Delta T_{\mathrm{SO}}^2\right) + C_E C_U C_W S_0\alpha\beta(5F_S\Delta T + 2F_N(\Delta T + \Delta T_{\mathrm{SO}}))\right)\right] \\
& + D^6 \left[C_W^6 + 2C_U C_W^3 (C_E S_0\beta(3F_N + 4F_S) + 7C_N C_U\alpha\Delta T) + 10C_E C_U C_W^4\alpha\Delta T_{\mathrm{SO}}\right. \\
& + 2C_E C_U^2 C_W\alpha\Delta T_{\mathrm{SO}}(C_E S_0\beta(F_N + 3F_S) + 5C_N C_U\alpha\Delta T) + C_U^2 C_W^2\left(7C_N F_N S_0\beta + 6C_E^2\alpha^2\Delta T_{\mathrm{SO}}^2\right) \\
& \left. + C_U^2\left(C_E^2 F_S^2 S_0^2\beta^2 + C_N^2 C_U^2\alpha^2\Delta T^2 + 2C_E C_N C_U S_0\alpha\beta(F_S\Delta T + F_N\Delta T_{\mathrm{SO}})\right)\right] \\
& + D^5 C_U \left[6C_W^5 + 2C_U C_W^2 (3C_E S_0\beta(F_N + 2F_S) + 8C_N C_U\alpha\Delta T) + 20C_E C_U C_W^3\alpha\Delta T_{\mathrm{SO}}\right. \\
& \left. + C_E C_U^2\alpha\Delta T_{\mathrm{SO}}(2C_E F_S S_0\beta + 3C_N C_U\alpha\Delta T) + 2C_U^2 C_W\left(3C_N F_N S_0\beta + 2C_E^2\alpha^2\Delta T_{\mathrm{SO}}^2\right)\right] \\
& + D^4 C_U^2 \left[15C_W^4 + C_U C_W (2C_E S_0\beta(F_N + 4F_S) + 9C_N C_U\alpha\Delta T) + 20C_E C_U C_W^2\alpha\Delta T_{\mathrm{SO}}\right. \\
& \left. + C_U^2\left(2C_N F_N S_0\beta + C_E^2\alpha^2\Delta T_{\mathrm{SO}}^2\right)\right] \\
& + D^3 2C_U^3 \left[10C_W^3 + C_U C_E F_S S_0\beta + C_N C_U^2\alpha\Delta T + 5C_E C_U C_W\alpha\Delta T_{\mathrm{SO}}\right] \\
& + D^2 C_U^4\left[15C_W^2 + 2C_E C_U\alpha\Delta T_{\mathrm{SO}}\right] + D6C_U^5 C_W + C_U^6 = 0 \qquad \text{(A7)}
\end{aligned}
$$

A2 Wind-driven case

For a wind-driven overturning the upwelling in the lower latitudes is zero by setting $C_U = 0$. Thus the tracer transport balance in steady state (5 equal to zero) reduces to $m_W = m_N + m_E$. Differences in salinity are defined as in the full problem and the salinity balance in the northern box is the same as in the full problem. Therefore Eqs. (A1)–(A3) are valid. Using the salinity balance of the southern box, in this case the southern salinity difference reduces to:

$$\Delta S_{\mathrm{SO}} = -\frac{S_0(F_N+F_S)}{m_W}.$$

For the eddy return flow it follows:

$$m_E = -C_E D\beta\frac{S_0(F_N+F_S)}{C_W} - \alpha\Delta T_{\mathrm{SO}} C_E D. \tag{A8}$$

Replacing the northern sinking by Eq. (A3) and the eddy return flow by Eq. (A8) in the tracer transport balance the governing equation of the pycnocline depth emerges.

$$\begin{aligned}
&D^3 C_E C_N \alpha\Delta T\left[\frac{\beta S_0(F_N+F_S)}{C_W}+\alpha\Delta T_{\mathrm{SO}}\right]\\
&+D^2\left[C_N F_N S_0\beta + C_N C_W\alpha\Delta T + \frac{C_E^2}{C_W^2}(S_0\beta(F_N+F_S)+C_W\alpha\Delta T_{\mathrm{SO}})^2\right]\\
&+D2C_E[\beta S_0(F_N+F_S)+C_W\alpha\Delta T_{\mathrm{SO}}]+C_W^2=0
\end{aligned}$$

The solutions of the polynomial depend on the sign of the discriminant $d=(q/2)^2+(p/3)^3$ with p and q defined as:

$$\begin{aligned}
\frac{q}{2} &= \frac{1}{2}\left(\frac{C_N F_N S_0\beta + C_N C_W\alpha\Delta T + C_E^2 A^2}{3C_E\alpha\Delta T A}\right)^3 - \frac{C_W F_N S_0\beta + C_W^2\alpha\Delta T}{3C_E C_N\alpha^2\Delta T^2 A}\\
&\quad - \frac{C_E C_W A}{3C_N^2\alpha^2\Delta T^2} + \frac{C_W^2}{2C_E C_N\alpha\Delta T A}\\
\frac{p}{3} &= \frac{6C_W C_N\alpha\Delta T - 1}{9C_N^2\alpha^2\Delta T^2} - \frac{F_N S_0\beta + C_W\alpha\Delta T}{9C_E^2 C_N^2\alpha^2\Delta T^2 A^2}\\
&\text{with} A = \left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right).
\end{aligned}$$

If the disciminant is positive the governing equation has one real solution.

$$D = \sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}} \tag{A9}$$

For a negative discriminant there are three real solutions.

$$\begin{aligned}
D_1 &= 2\sqrt{-\frac{p}{3}}\cos\left(\frac{1}{3}\arccos\left(-\frac{3q}{2p\sqrt{-\frac{p}{3}}}\right)\right) - \frac{C_N F_N S_0\beta + C_N C_W\alpha\Delta T}{3C_E C_N\alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
&\quad + \frac{C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N\alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
&\quad - \frac{C_N F_N S_0\beta + C_N C_W\alpha\Delta T + C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N\alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
D_2 &= 2\sqrt{-\frac{p}{3}}\cos\left(\frac{1}{3}\arccos\left(-\frac{3q}{2p\sqrt{-\frac{p}{3}}}\right)+\frac{2}{3}\pi\right)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{C_N F_N S_0\beta + C_N C_W \alpha\Delta T}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)} + \frac{C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
&-\frac{C_N F_N S_0\beta + C_N C_W \alpha\Delta T + C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
D_3 = {} & 2\sqrt{-\frac{p}{3}}\cos\left(\frac{1}{3}\arccos\left(-\frac{3q}{2p\sqrt{-\frac{p}{3}}}\right)+\frac{4}{3}\pi\right)\\
&-\frac{C_N F_N S_0\beta + C_N C_W \alpha\Delta T}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)} + \frac{C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)}\\
&-\frac{C_N F_N S_0\beta + C_N C_W \alpha\Delta T + C_E^2\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)^2}{3C_E C_N \alpha\Delta T\left(\frac{S_0\beta}{C_W}(F_N+F_S)+\alpha\Delta T_{\mathrm{SO}}\right)} \qquad \text{(A10)}
\end{aligned}
$$

Acknowledgements. D. Ehlert would like to thank M. Eby for the support in setting up simulations with UVic ESCM.

References

Bryden, H. L., King, B. a., and McCarthy, G. D.: South Atlantic overturning circulation at 24S, Journal of Marine Research, 69, 38–55, 2011.

Cimatoribus, A. a., Drijfhout, S. S., and Dijkstra, H. a.: Meridional overturning circulation: stability and ocean feedbacks in a box model, Climate Dynamics, 42, 311–328, 2014.

Dijkstra, H. A.: Characterization of the multiple equilibria regime in a global ocean model, Tellus - Series A: Dynamic Meteorology and Oceanography, 59, 695–705, 2007.

Drijfhout, S. S., Weber, S. L., and van der Swaluw, E.: The stability of the {MOC} as diagnosed from model projections for pre-industrial, present and future climates, Climate Dynamics, pp. 1–12, 2010.

Eby, M., Zickfeld, K., Montenegro, a., Archer, D., Meissner, K. J., and Weaver, a. J.: Lifetime of Anthropogenic Climate Change: Millennial Time Scales of Potential CO 2 and Surface Temperature Perturbations, Journal of Climate, 22, 2501–2511, http://journals.ametsoc.org/doi/abs/10.1175/2008JCLI2554.1, 2009.

Fürst, J. and Levermann, A.: Minimal model of a wind- and mixing-driven overturning - threshold behaviour for both driving mechanisms, Climate Dynamics, pp. 1–22, 2011.

Gent, P. R. and McWilliams, J. C.: Isopycnal Mixing in Ocean Circulation Models, Journal of Physical Oceanography, 20, 150–155, 1990.

Gnanadesikan, A.: A simple predictive model for the structure of the oceanic pycnocline, Science, 283, 2077–2079, 1999.

Goswami, B. N., Madhusoodanan, M. S., Neema, C. P., and Sengupta, D.: A physical mechanism for North Atlantic SST influence on the Indian summer monsoon, Geophysical Research Letters, 33, L02 706, 2006.

Griesel, A. and Morales-Maqueda, M. A.: The relation of meridional pressure gradients to North Atlantic Deep Water volume transport in an OGCM, Climate Dynamics, 26, 781–799, 2006.

Hawkins, E., Smith, R. S., Allison, L. C., Gregory, J. M., Woollings, T. J., Pohlmann, H., and De Cuevas, B.: Bistability of the Atlantic overturning circulation in a global climate model and links to ocean freshwater transport, Geophysical Research Letters, 38, 1–6, 2011.

Hofmann, M. and Rahmstorf, S.: On the stability of the Atlantic meridional overturning circulation, Proceedings of the National Academy of Sciences, 106, 20 584–20 589, 2009.

Huang, R. X.: Mixing and Energetics of the Oceanic Thermohaline Circulation, Journal of Physical Oceanography, 29, 727–746, 1999.

Huisman, S. E., den Toom, M., Dijkstra, H. a., and Drijfhout, S.: An Indicator of the Multiple Equilibria Regime of the Atlantic Meridional Overturning Circulation, Journal of Physical Oceanography, 40, 551–567, 2010.

Johnson, H. L., Marshall, D. P., and Sproson, D. A. J.: Reconciling theories of a mechanically-driven meridional overturning circulation with thermohaline forcing and multiple equilibria, Climate Dynamics, 29, 821–836, 2007.

Kuhlbrodt, T., Griesel, A., Montoya, M., Levermann, A., Hofmann, M., and Rahmstorf, S.: On the driving processes of the Atlantic meridional overturning circulation, Reviews of Geophysics, 45, RG2001, 2007.

Kuhlbrodt, T., Rahmstorf, S., Zickfeld, K., Vikebo, F. B., Sundby, S., Hofmann, M., Link, P. M., Bondeau, A., Cramer, W., and Jaeger, C.: An Integrated Assessment of changes in the thermohaline circulation, Climatic Change, 96, 489–537, 2009.

Laurian, A., Drijfhout, S. S., Hazeleger, W., and Hurk, B.: Response of the Western European climate to a collapse of the thermohaline circulation, Climate Dynamics, 34, 689–697, 2009.

Lenton, T. M.: Early warning of climate tipping points, Nature Climate Change, 1, 201–209, 2011.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the National Academy of Sciences, 105, 1786–1793, 2008.

Levermann, A. and Fürst, J.: Atlantic pycnocline theory scrutinized using a coupled climate model, Geophysical Research Letters, 37, L14 602, 2010.

Levermann, A. and Griesel, A.: Solution of a Model for the Oceanic Pycnocline Depth: Scaling of Overturning Strength and Meridional Pressure Difference, Geophysical Research Letters, 31, L17 302, 2004.

Levermann, A., Griesel, A., Hofmann, M., Montoya, M., and Rahmstorf, S.: Dynamic sea level changes following changes in the thermohaline circulation, Climate Dynamics, 24, 347–354, 2005.

Levermann, A., Bamber, J., Drijfhout, S., Ganopolski, A., Haeberli, W., Harris, N., Huss, M., Krueger, K., Lenton, T., Lindsay, R., Notz, D., Wadhams, P., and Weber, S.: Potential climatic transitions with profound impact on Europe, Climatic Change, 110, 845–878, 2012.

Manabe, S. and Stouffer, R. J.: Two stable equilibria of a coupled ocean-atmosphere model, Journal of Climate, 1, 841–866, 1988.

Manabe, S. and Stouffer, R. J.: Century-scale effects of increased atmospheric CO_2 on the ocean-atmosphere system, Nature, 364, 215–218, 1993.

Manabe, S. and Stouffer, R. J.: Simulation of abrupt climate change induced by freshwater input to the North Atlantic Ocean, Nature, 378, 165–167, 1995.

Marotzke, J.: Boundary Mixing and the Dynamics of Three-Dimensional Thermohaline Circulations, Journal of Physical Oceanography, 27, 1713–28, 1997.

Marzeion, B. and Drange, H.: Diapycnal mixing in a conceptual model of the Atlantic Meridional Overturning Circulation, Deep-Sea Research II, 53, 226–238, 2006.

Mignot, J., Ganopolski, A., and Levermann, A.: Atlantic subsurface temperatures: response to a shut-down of the overturning circulation and consequences for its recovery, Journal of Climate, 20, 4884–4898, 2007.

Munk, W.: Abyssal recipes, Deep-Sea Research, 13, 707–730, 1966.

Munk, W. and Wunsch, C.: Abyssal recipes II, Deep-Sea Research I, 45, 1977–2010, 1998.

Rahmstorf, S.: On the freshwater forcing and transport of the Atlantic thermohaline circulation, Climate Dynamics, 12, 799–811, 1996.

Rahmstorf, S., Crucifix, M., Ganopolski, A., Goosse, H., Kamenkovich, I., Knutti, R., Lohmann, G., Marsh, B., Mysak, L. A., Wang, Z., and Weaver, A.: Thermohaline circulation hysteresis: A model intercomparison, Geophysical Research Letters, 32, L23 605, 2005.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., and Sugihara, G.: Early-warning signals for critical transitions, Nature, 461, 53–59, 2009.

Schewe, J. and Levermann, A.: The role of meridional density differences for a wind-driven overturning circulation, Climate Dynamics, 34, 547–556, 2010.

Schmittner, A.: Decline of the marine ecosystem caused by a reduction in the Atlantic overturning circulation, Nature, 434, 628–633, 2005.

Stommel, H.: Thermohaline convection with two stable regimes of flow, Tellus, 13, 224–230, 1961.

Stouffer, R. J., Russell, J., and Spelman, M. J.: Importance of oceanic heat uptake in transient climate change, Geophysical Research Letters, 33, 1–5, 2006a.

Stouffer, R. J., Yin, J., Gregory, J. M., Dixon, K. W., Spelman, M. J., Hurlin, W., Weaver, A. J., Eby, M., Flato, G. M., Hasumi, H., Hu, A., Jungclaus, J. H., Kamenkovich, I. V., Levermann, A., Montoya, M., Murakami, S., Nawrath, S., Oka, A., Peltier, W. R., Robitaille, D. Y., Sokolov, A. P., Vettoretti, G., and Weber, S. L.: Investigating the Causes of the Response of the Thermohaline Circulation to Past and Future Climate Changes, Journal of Climate, 19, 1365–1387, 2006b.

Stouffer, R. J., Seidov, B., and Haupt, B. J.: Climate Response to External Sources of Freshwater: North Atlantic versus the Southern Ocean, Journal of Climate, 20, 436–448, 2007.

Toggweiler, J. R. and Samuels, B.: Effect of Drake Passage on the global thermohaline circulation, Deep-Sea Research, 42, 477–500, 1995.

Toggweiler, J. R. and Samuels, B.: On the ocean's large scale circulation in the limit of no vertical mixing, Journal of Physical Oceanography, 28, 1832–1852, 1998.

Vellinga, M. and Wood, R. A.: Global climatic impacts of a collapse of the Atlantic thermohaline circulation, Climatic Change, 54, 251–267, 2002.

Weaver, A. J., Eby, M., Wiebe, E. C., Bitz, C. M., Duffy, P. B., Ewen, T. L., Fanning, A. F., Holland, M. M., MacFadyen, A., Matthews, H. D., Meissner, K. J., Saenko, O., Schmittner, A., Wang, H., and Yoshimori, M.: The UVic earth system climate model: Model description, climatology, and applications to past, present and future climates, Atmosphere-Ocean, 39, 361–428, http://www.tandfonline.com/doi/abs/10.1080/07055900.2001.9649686, 2001.

Weaver, A. J., Sedláček, J., Eby, M., Alexander, K., Crespin, E., Fichefet, T., Philippon-Berthier, G. l., Joos, F., Kawamiya, M., Matsumoto, K., Steinacher, M., Tachiiri, K., Tokos, K., Yoshimori, M., and Zickfeld, K.: Stability of the Atlantic meridional overturning circulation: A model intercomparison, Geophysical Research Letters, 39, n/a–n/a, 2012.

Wunsch, C. and Ferrari, R.: Vertical mixing, energy and the general circulation of the oceans, Annual Reviews of Fluid Mechanics, 36, 281–314, 2004.

Yin, J., Schlesinger, M. E., and Stouffer, R. J.: Model projections of rapid sea-level rise on the northeast coast of the United States, Nature Geoscience, 2, 262–266, 2009.

Table 1. Physical parameters for used for the model.

Parameter	Value	Unit	Description
			Geometry
H	4×10^3	m	Average depth of the Atlantic Ocean basin
B	1×10^7	m	Average width of the Atlantic Ocean
L_N	3.34×10^6	m	Meridional extend of the northern box
L_U	8.90×10^6	m	Meridional extend of the tropical box
L_S	3.34×10^6	m	Meridional extend of the southern box
			Stratification
ρ_0	1027	$\mathrm{kg\,m^{-3}}$	Average density of the Atlantic Ocean
S_0	35	psu	Average salinity of the Atlantic ocean
L_y^N	1.5×10^6	m	Meridional extent of the northern outcropping
A_{GM}	1×10^6	$\mathrm{m^2\,s^{-1}}$	Thickness diffusivity
κ	4×10^{-5}	$\mathrm{m^2\,s^{-1}}$	Background vertical diffusivity
α_{T}	2.1×10^{-4}	$1\,{}^\circ\mathrm{C}^{-1}$	Thermal coefficient for isobars
α		$\mathrm{kg\,(m^3\,{}^\circ C)^{-1}}$	Product of ρ_0 and α_{T}
β_S	8×10^{-4}	$1\,\mathrm{psu}^{-1}$	Haline coefficient for isobars
β		$\mathrm{kg\,(m^3\,psu)^{-1}}$	Product of ρ_0 and β_S
C	0.1	–	Constant accounting for geometry and stratification
			External rorcing
β_N	2×10^{-11}	$1\,\mathrm{ms^{-1}}$	Coefficient for β-plane approximation in the North Atlantic
f_{Dr}	-7.5×10^{-5}	$1\,\mathrm{s^{-1}}$	Coriolis parameter in the Drake Passage
τ_{Dr}	0.1	$\mathrm{N\,m^{-2}}$	Average zonal wind stress in the Drake Passage
F_N	0.1	Sv	Northern meridional atmospheric freshwater transport
F_S	0.1	Sv	Southern meridional atmospheric freshwater transport
T_N	5.0	$^\circ\mathrm{C}$	Temperature of the northern box
T_U	12.5	$^\circ\mathrm{C}$	Temperature of the tropical surface box
T_S	7.0	$^\circ\mathrm{C}$	Temperature of the southern box

Table 2. Numerical solution of the model by applying finite difference method on Eqs. (1)–(6). Equilibrium state is obtained after 2000 yr with a time step of 14 days and the starting conditions: Salinities set to 35 psu and the pycnocline depth set to 500 m.

Name		Value
Salinities	S_N	35.04 psu
	S_U	35.24 psu
	S_D	35.02 psu
	S_S	34.79 psu
Tracer transports	m_U	17.5 Sv
	m_U	5.8 Sv
	m_W	13.0 Sv
	m_E	1.2 Sv
Meridional density	$\Delta\rho$	1.45 $\mathrm{kg\,m^{-3}}$
differences	$\Delta\rho_{\mathrm{SO}}$	0.82 $\mathrm{kg\,m^{-3}}$
Pycnocline depth	D	615 m

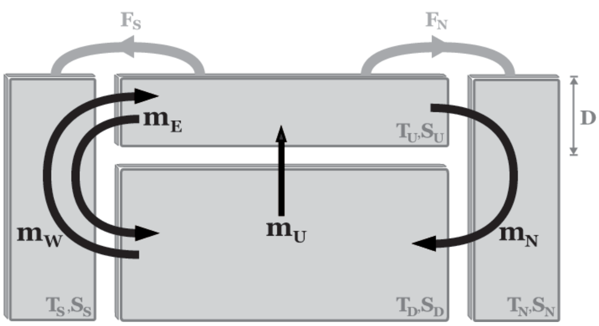

Figure 1. Schematic of the conceptual model as suggested in Fürst and Levermann (2011) and used here. The depth of the pycnocline D is determined by the balance between the northern deep water formation m_N, the upwelling in the low-latitudes m_U in response to downward mixing, the Ekman upwelling m_W and the eddy-induced return flow m_E. Salinity is advected along with these transport processes and determines together with a fixed temperature distribution the horizontal density differences. The differences are between low-latitudinal box and northern box, $\Delta\rho$, and low-latitudinal and southern box, $\Delta\rho_{\mathrm{SO}}$, respectively. The density differences, in turn, determines the northern sinking, $m_N \propto D^2\Delta\rho$, and the eddy-induced return flow, $m_E \propto D\Delta\rho_{\mathrm{SO}}$.

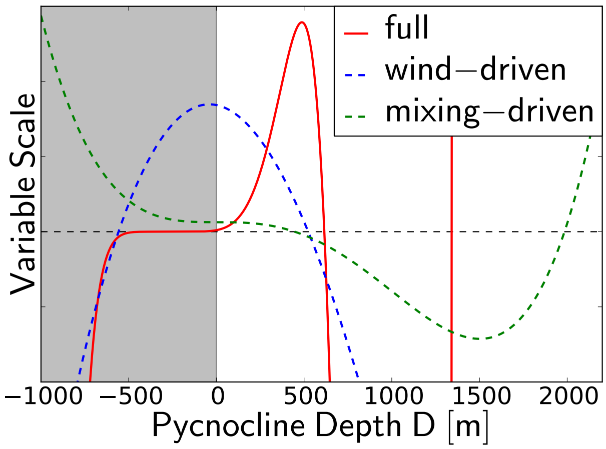

Figure 2. Trend of the governing equation for the full case (red line), the wind-driven case ($m_U = 0$, blue line) and the mixing-driven case ($m_W = m_E = 0$, green line). The intersections with zero (black dashed line) are solutions of the polynomial but those in the grey shadowed area correspond to a negative pycnocline depth. Therefore they are not physical. In all three cases there are two positive solutions, a lower stable, physical one D and a higher unstable or non-physical one $\hat{D}$. In the wind-driven case the non-physical solution is out of range of the pycnocline but is shown in Fig. 5. For the full case the solutions are $D = 616\,\mathrm{m}$ and $\hat{D} = 1342\,\mathrm{m}$, for the wind-driven case they are $D = 523\,\mathrm{m}$ and $\hat{D} = 6190\,\mathrm{m}$ and for the mixing-driven case they are $D = 446\,\mathrm{m}$ and $\hat{D} = 1985\,\mathrm{m}$.

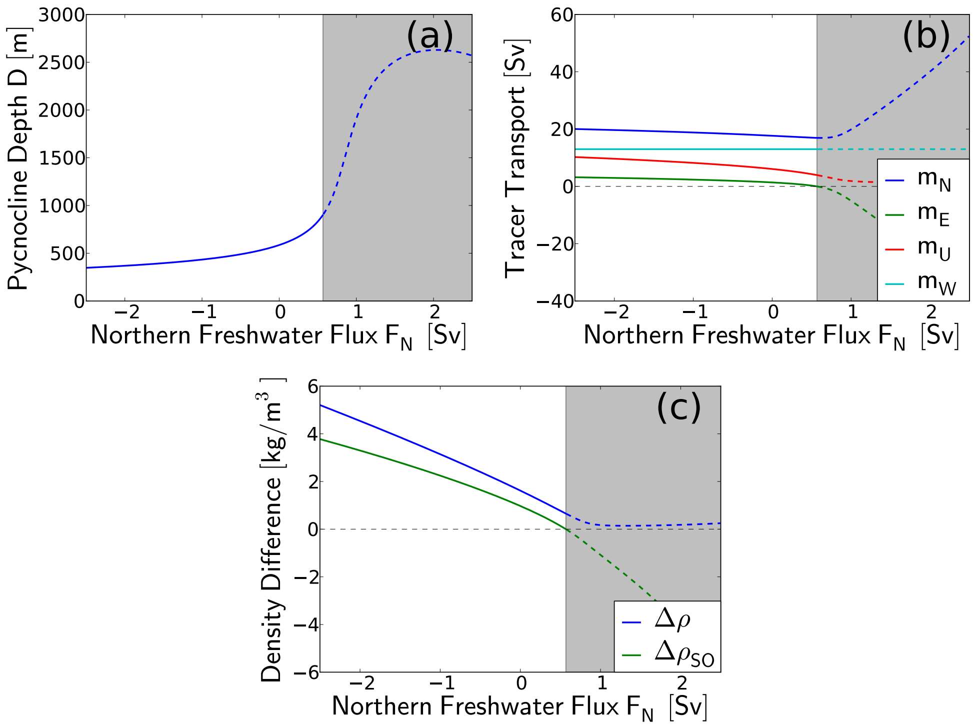

Figure 3. In steady state only one real stable solution of governing equation of the full exists which increases under freshwater forcing **(a)**. The tracer transport processes show different behaviours **(b)**. The eddy return flow m_E decreases (**b**, green line) until it becomes negative and the break down of circulation is reached (grey shaded area). Also the density difference between the southern box and the low-latitudinal box, $\Delta\rho_{\mathrm{SO}}$, crosses zero at the threshold level (**c**, green line).

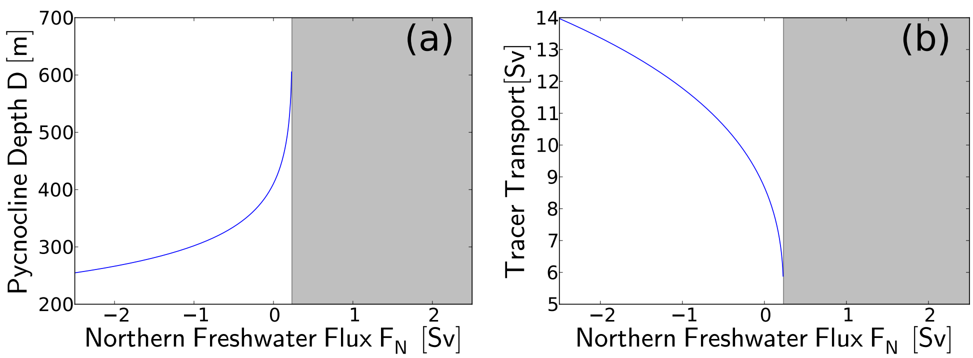

Figure 4. In steady state the governing equation for the mixing-driven case has one real, stable solution until a threshold level is reached **(a)**. Thereafter, no real solution exists. The tracer transports are upwelling in the mid latitudes and northern sinking which balance each other ($m_N = m_U$) and decrease under increasing freshwater forcing **(b)**.

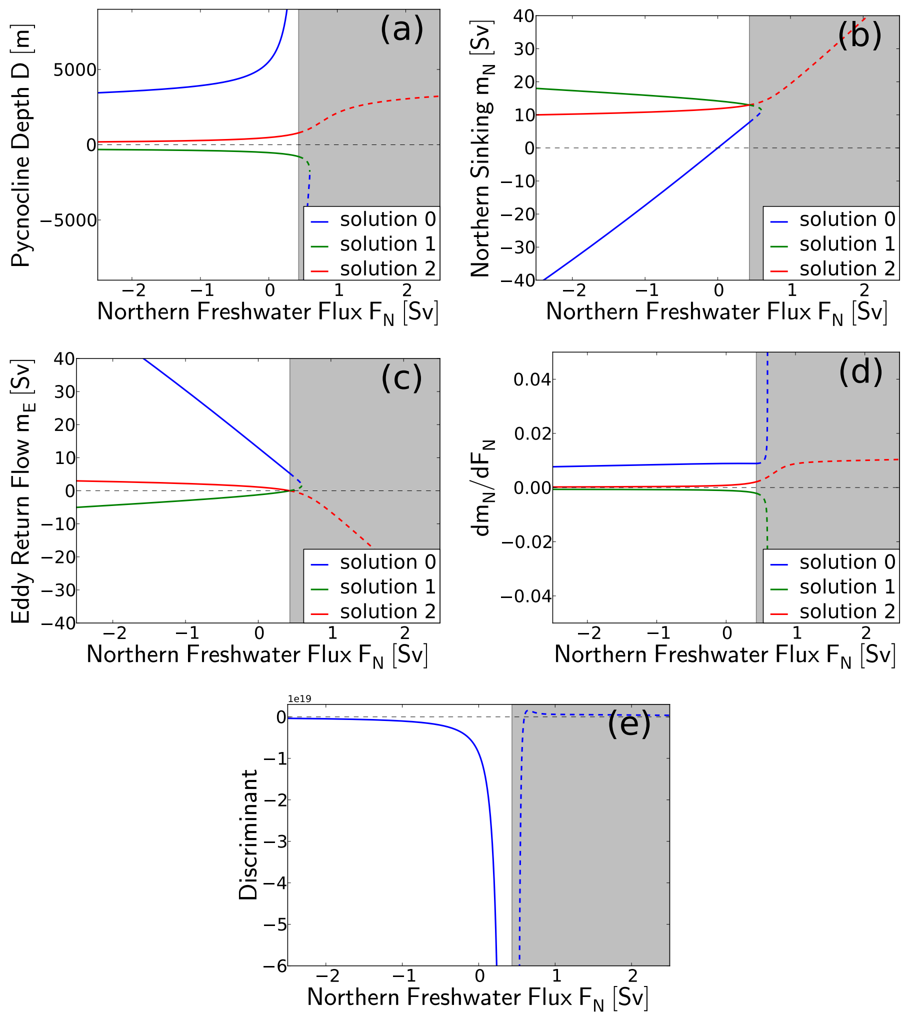

Figure 5. In steady state only one physical solution of governing equation for the wind-driven case exists. There are three real solutions before the circulation breaks down (**a**, white area) because the discriminant is negative (**e**). The physical branch is solution 2 (red line). The threshold (grey shaded area) is reached when the eddy return flow becomes negative (**c**, red line) and the discriminant of the governing equation has a negative pole (**e**). The zero crossing of the discriminant, which was in the parent model (Fürst and Levermann, 2011) the indicator for a cessation of the circulation, does not appear within the range applicability of our model. Within that range the northern sinking always increases (**b**, red line) and its derivative is positive (**d**, red line).

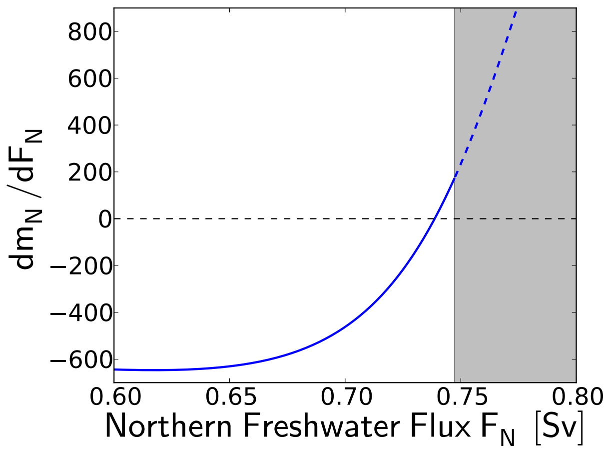

Figure 6. The derivative of the northern sinking with respect to freshwater forcing in full case. The derivative is positive before the circulation collapses (white area). This behaviour is caused by a change in the Southern Ocean temperature from $T_S = 7\,^\circ\mathrm{C}$ to $T_S = 5\,^\circ\mathrm{C}$.

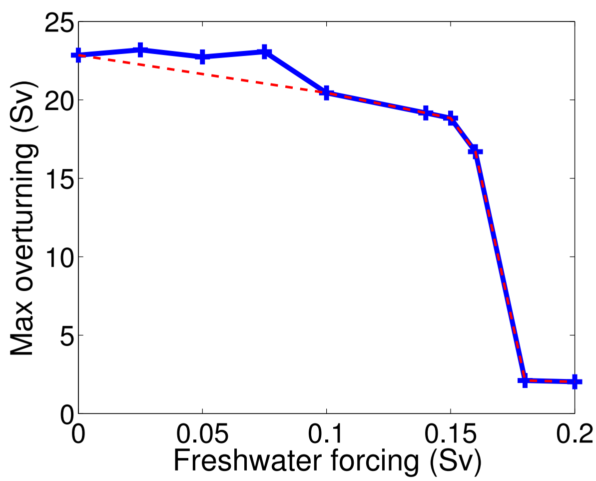

Figure 7. Maximum overturning for freshwater experiments with UVic ESCM 2.9. Freshwater was taken out at 10°S to 30°S and dumped into the Atlantic at 10°N to 30°N prescribing different constant amounts of freshwater. For each simulation the maximum overturning is averaged over the last 1000 years of the simulation (blue curve). Each blue cross corresponds to one equilibrium simulation. If the model would not show an AMOC strengthening, the modeled maximum overturning would be expected to follow the red curve.