

A Saddle-Node Bifurcation may be Causing the AMOC Collapse in the Community Earth System Model

René M. van Westen¹, Elian Vanderborght¹, and Henk A. Dijkstra¹

¹Institute for Marine and Atmospheric research Utrecht, Department of Physics, Utrecht University, Utrecht, the Netherlands

Correspondence: René M. van Westen (r.m.vanwesten@uu.nl)

Abstract. Recently, a collapse of the Atlantic Meridional Overturning Circulation (AMOC) was found in the Community Earth System Model (CESM) under constant pre-industrial greenhouse gas forcing conditions. To determine the stability changes of the AMOC with changing (freshwater) parameters in models, it is important to determine the origin of the collapse behavior. In this paper, we argue that the classical picture of a saddle-node bifurcation holds for the AMOC collapse in the CESM. We provide specific supporting arguments by showing results of additional pre-industrial CESM simulations. The CESM results are compared with those of a five-box AMOC model, which is known to have saddle-node bifurcations, and with which many sensitivity experiments can be performed. Theoretical arguments are also provided showing that the essential dynamics of the CESM can be reduced to a low-dimensional model in which a saddle-node bifurcation causes the AMOC collapse. The underlying physical reason is that the AMOC behaviour in CESM is controlled by a small set of dominant feedback processes. This has important consequences for the value of conceptual AMOC models, for assessing the effect of model biases on the AMOC stability, and for the interpretation of AMOC behaviour under climate change scenarios.

1 Introduction

A hot issue in current climate research is the Atlantic Meridional Overturning Circulation (AMOC) response under future climate change. Climate models participating in the Coupled Model Inter-comparison Project Phase 6 (CMIP6, Eyring et al. (2016)) indicate a substantial AMOC weakening during the 21st century (Weijer et al., 2020). Beyond 2100 there is much more uncertainty as the AMOC may (partially) recover or fully collapse (Liu et al., 2017; Bonan et al., 2022; Drijfhout et al., 2025). Transient temperature responses are effective in causing the 21st century AMOC weakening but salinity responses are crucial in further destabilizing the AMOC (Gérard and Crucifix, 2024; van Westen et al., 2025). The dominant destabilizing AMOC tipping mechanism is the salt-advection feedback, where an AMOC weakening leads to a smaller northward salinity transport amplifying the initial AMOC weakening (e.g., Marotzke (2000)). The existence of the salt-advection feedback is why the AMOC is labelled as a tipping point in the climate system (Lenton et al., 2008; Armstrong McKay et al., 2022).

Stommel (1961) was the first to identify the salt-advection feedback in a simple two-box model and demonstrated that this feedback induces transitions between two stable AMOC steady states. The multi-stable AMOC regime is bounded by two saddle-node bifurcations in this model. Since then, studies using more detailed conceptual (box) models (Cessi, 1994; Cimadoribus et al., 2014; Wood et al., 2019) and numerically fully-implicit ocean-climate models (De Niet et al., 2007; Toom

et al., 2012; Mulder et al., 2021) have shown that saddle-node bifurcations bound the multi-stable regime of the AMOC in these models. Rahmstorf (1996) showed that the saddle-node bifurcation associated with the AMOC collapse is linked to a critical value of the freshwater transport carried by the AMOC at 34°S, represented by the quantity F_{ovS} . When including the stabilizing gyre responses (Sijp, 2012), a F_{ovS} minimum is found close to this saddle-node bifurcation (Dijkstra, 2007).

30 In numerically explicit ocean-climate models it is much harder (or not feasible) to determine the steady states versus (freshwater forcing) parameters and the boundaries of the AMOC multi-stable regime. An impression of the multi-stable regime can be obtained by performing quasi-equilibrium simulations, where a freshwater flux forcing is changed very slowly back-and-forth such that the model state stays close to the (slowly changing) statistical equilibrium. Such quasi-equilibrium simulations have been performed with many ocean-only models (Rahmstorf, 1995; Lohmann et al., 2024), Earth System Models of Intermediate Complexity (EMICs) (Rahmstorf et al., 2005; Cini et al., 2024), the FAMOUS model (Hawkins et al., 2011), the
35 Community Climate System Model (CCSM3) (Hu et al., 2012), and recently in the Community Earth System Model (CESM) (van Westen and Dijkstra, 2023; van Westen et al., 2024a).

When the salt-advection feedback is the dominant feedback, as is the case for the Stommel (1961) model, it can be shown that the stable ‘AMOC on’ state has a square-root (or quadratic) solution against varying freshwater flux forcing (see Appendix A) with the normal (most simple) form of $dx/dt = r - x^2$ with $r > 0$ (see Appendix B). This square-root relation in the
40 Stommel model can be understood from the fact that the AMOC strength is proportional to the salinity gradient, whereas the salinity gradient is also proportional to the AMOC strength. In more complex (climate) models that resolve more processes and climate feedbacks, a near square-root dependency is also found for the AMOC strength against forcing (Dijkstra, 2007; van Westen et al., 2024b; Vanderborgh et al., 2025). Finding indications of a square-root relation in quasi-equilibrium simulations
45 is challenging as it requires very slow rates to follow the steady states of the system (Rahmstorf, 1996). Even if the rate is sufficiently slow, this relation can be masked by relatively large (stochastic) noise (Berglund and Gentz, 2006). An alternative approach is by obtaining statistical equilibria for fixed forcing values, but this is computationally too costly for CESM. Nevertheless, as long as the salt-advection feedback remains dominant amid other AMOC-related feedbacks (Vanderborgh et al., 2025), a square-root dependency can be expected when the system is relatively close to its saddle-node bifurcation and hence
50 to tipping.

Here, we focus on the CESM results and address the issue whether its AMOC tipping behavior is also caused by the presence of a saddle-node bifurcation, similar to that in the fully-implicit ocean-climate models (Dijkstra, 2007). This is certainly a non-trivial issue as the CESM is an extremely high-dimensional dynamical system and the atmospheric fluxes create a high frequency forcing on the ocean component of the model. In addition, in the quasi-equilibrium CESM simulation (van Westen
55 et al., 2024a) the forcing rate is rather large compared to the equilibration time scale of the AMOC (van Westen et al., 2024b) and hence the (non-autonomous) dynamical system is not a fast-slow system (Kuehn, 2011). The existence of a saddle-node bifurcation in the CESM is important for assessing the role of model biases on the stability of the AMOC and for understanding the response of the model to transient climate change forcing (Ritchie et al., 2021).

The aim of this paper is to provide a convincing case that a saddle-node bifurcation is causing the AMOC collapse in the
60 CESM, as presented in van Westen et al. (2024a). Thereto, we have performed several additional CESM simulations which

were branched from the quasi-equilibrium CESM simulation, we will compare the CESM behavior with that of a five-box AMOC model for which a saddle-node bifurcation is known to exist (van Westen et al., 2024b). The advantage of this five-box model is that we can easily conduct multiple sensitivity experiments to better understand the CESM behaviour. Section 2 describes the model set-up and simulations for the CESM and five-box model. Next, in Section 3, the results on the (statistical) steady states and quasi-equilibrium results of both the CESM and five-box model are presented. Section 4 provides detailed arguments for the existence of a saddle-node bifurcation in the CESM, followed by Section 5, where the importance of this result for the behavior of the AMOC under climate change is shown. Finally, in Section 6, the results are summarized and discussed.

2 Models and Methods

2.1 CESM simulations

The CESM (version 1.0.5) is a fully-coupled climate model and the simulations here have a 1° horizontal resolution for the ocean/sea-ice components and a 2° horizontal resolution for the atmosphere/land components. For more details on the precise CESM set-up, we refer to van Westen and Dijkstra (2023) and van Westen et al. (2024a). In those studies, the pre-industrial forcing is used and in addition a freshwater flux forcing (F_H) is applied between 20°N and 50°N in the Atlantic Ocean and is compensated elsewhere (at the ocean surface) to conserve salinity. This is the same hosing region as in Hu et al. (2012) and Rahmstorf (1996), which has the advantage that the North Atlantic deep convection sites are not directly impacted under the hosing. The sensitivity of the hosing location will be thoroughly analysed below for the five-box AMOC model.

The quasi-equilibrium AMOC hysteresis simulation (van Westen and Dijkstra, 2023) is obtained by slowly increasing F_H from 0 Sv to 0.66 Sv and back to 0 Sv, at a rate of 3×10^{-4} Sv yr $^{-1}$, resulting in a 4400-year long simulation. This simulation remains close to the statistical equilibria, but the deviations become larger near the AMOC collapse and recovery (van Westen et al., 2024b). To determine statistical equilibria (i.e., steady states), two 500-year long CESM simulations were performed (van Westen et al., 2024b) at constant F_H , the steady states are indicated as $\overline{F_H}$. This was already done for $\overline{F_H} = 0.18$ Sv (starting at model year 600 of the quasi-equilibrium simulation) and at $\overline{F_H} = 0.45$ Sv forcing (starting at model year 1500). The last 100 years of these steady states show hardly any model drift, meaning that the AMOC and global climate are dominated by natural climate variability (van Westen and Baatsen, 2025). Below, we will show results of new CESM simulations performed under constant $\overline{F_H}$ forcing or with a slower rate of F_H , and closer to the values where the AMOC collapse occurs in the quasi-equilibrium simulation (around $F_H = 0.525$ Sv, van Westen et al. (2024a)).

We will (in Section 5) also use results from two climate change simulations that were initialized from the end of the steady state with $\overline{F_H} = 0.18$ Sv and $\overline{F_H} = 0.45$ Sv (van Westen et al., 2025). These climate change simulations were first forced under the historical forcing (1850 – 2005) and followed by either RCP4.5 or RCP8.5 scenario forcing (2006 – 2100, Representative Concentration Pathway). Subsequently, they were further integrated for 400 years under their 2100 radiative forcing conditions to study the equilibrium behaviour.

An overview of all the different CESM simulations are presented below in Table 1. In total, we present 11,670 model years of model output. Ideally, one would determine even more steady states or lower the varying F_H rate in the quasi-equilibrium simulation, but this is computationally not feasible. These additional simulations, however, can be done with the five-box AMOC model.

Table 1. Overview of the different simulations conducted with the CESM, which includes: simulation name, freshwater flux forcing (varying or fixed), radiative forcing, branched from simulation, duration, and the AMOC status at the end of simulation (on, transient or off). Note that the forward QE was branched from the 2800-year long pre-industrial control simulation from Baatsen et al. (2020). The simulations are sorted in order of appearance. Abbreviations: QE, quasi-equilibrium; PI, pre-industrial; RCP, Representative Concentration Pathway; E-RCP, Extended Representative Concentration Pathway.

Simulation name	Freshwater flux forcing Varying ($\partial_t F_H$) or fixed ($\overline{F_H}$)	Radiative forcing	Branched from simulation	Duration (years)	AMOC status at end of simulation
Forward QE	$\partial_t F_H = +3 \times 10^{-4} \text{ Sv yr}^{-1}$	PI at 1850 levels	2800 year PI control	2,200	Off
Backward QE	$\partial_t F_H = -3 \times 10^{-4} \text{ Sv yr}^{-1}$	PI at 1850 levels	Forward QE, $F_H = 0.66 \text{ Sv}$	2,200	On
Steady state #1	$\overline{F_H} = 0.18 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.18 \text{ Sv}$	500	On
Steady state #2	$\overline{F_H} = 0.45 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.45 \text{ Sv}$	500	On
Steady state #3	$\overline{F_H} = 0.465 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.465 \text{ Sv}$	500	On
Steady state #4	$\overline{F_H} = 0.48 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.48 \text{ Sv}$	500	Off
Steady state #5	$\overline{F_H} = 0.48 \text{ Sv}$	PI at 1850 levels	End of steady state #3	500	On
Steady state #6	$\overline{F_H} = 0.495 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.495 \text{ Sv}$	231	Transient
Steady state #7	$\overline{F_H} = 0.495 \text{ Sv}$	PI at 1850 levels	End of steady state #5	500	On
Steady state #8	$\overline{F_H} = 0.51 \text{ Sv}$	PI at 1850 levels	Forward QE, $F_H = 0.51 \text{ Sv}$	197	Transient
Steady state #9	$\overline{F_H} = 0.51 \text{ Sv}$	PI at 1850 levels	End of steady state #7	500	Off
Half QE	$\partial_t F_H = +1.5 \times 10^{-4} \text{ Sv yr}^{-1}$	PI at 1850 levels	End of steady state #2	1,050	Off
Historical #1	$\overline{F_H} = 0.18 \text{ Sv}$	Historical (1850 – 2005)	End of steady state #1	156	On
RCP4.5 #1	$\overline{F_H} = 0.18 \text{ Sv}$	RCP4.5 (2006 – 2100)	End of historical #1	95	Transient
E-RCP4.5 #1	$\overline{F_H} = 0.18 \text{ Sv}$	RCP4.5 at 2100 levels	End of RCP4.5 #1	400	On
RCP8.5 #1	$\overline{F_H} = 0.18 \text{ Sv}$	RCP8.5 (2006 – 2100)	End of historical #1	95	Transient
E-RCP8.5 #1	$\overline{F_H} = 0.18 \text{ Sv}$	RCP8.5 at 2100 levels	End of RCP8.5 #1	400	Off
Historical #2	$\overline{F_H} = 0.45 \text{ Sv}$	Historical (1850 – 2005)	End of steady state #2	156	On
RCP4.5 #2	$\overline{F_H} = 0.45 \text{ Sv}$	RCP4.5 (2006 – 2100)	End of historical #2	95	Transient
E-RCP4.5 #2	$\overline{F_H} = 0.45 \text{ Sv}$	RCP4.5 at 2100 levels	End of RCP4.5 #2	400	Off
RCP8.5 #2	$\overline{F_H} = 0.45 \text{ Sv}$	RCP8.5 (2006 – 2100)	End of historical #2	95	Transient
E-RCP8.5 #2	$\overline{F_H} = 0.45 \text{ Sv}$	RCP8.5 at 2100 levels	End of RCP8.5 #2	400	Off

2.2 The five-box AMOC model

The five-box AMOC model (Figure 1) was developed by Cimitoribus et al. (2014), extended by Castellana et al. (2019), and was recently further extended (hereafter the E-CCM, the Extended Cimitoribus-Castellana Model) by including oceanic temperatures (van Westen et al., 2024b). The E-CCM has four surface boxes, where the Atlantic Ocean is represented by boxes t and n, the Southern Ocean channel by box s, and the Southern Ocean Atlantic sector by box ts. There is one deep ocean box d, hence this model does not include the Indo-Pacific Ocean nor Arctic Ocean. The Atlantic Ocean pycnocline depth, indicated by the D , may vary in the E-CCM. The temperature and salinity are volume averaged over each box and heat and salinity are exchanged between the boxes, and also heat between the surface boxes and overhead atmosphere. Salinity is conserved in the E-CCM.

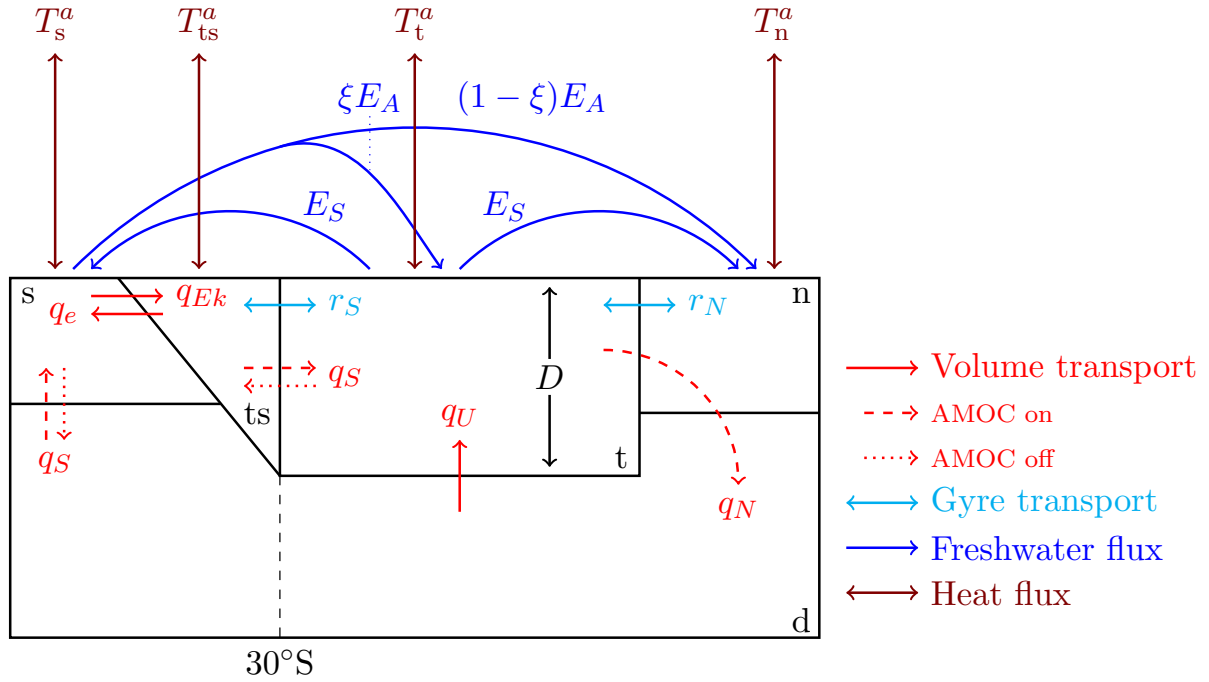


Figure 1. Schematic representation of the five-box AMOC model (the E-CCM), adapted from van Westen et al. (2024b). The red arrows represent volume transports, whereas the dashed and dotted arrows indicate the AMOC on and AMOC off states, respectively. The cyan and blue arrows represent the gyre transport and freshwater fluxes, respectively. The freshwater from box s is distributed linearly over box n and box t using a parameter ξ , where ξE_A is added to box t and $(1 - \xi)E_A$ to box n. The original E-CCM configuration van Westen et al. (2024b) is obtained when $\xi = 0$. The brown arrows are the heat fluxes with the overhead atmosphere for each surface box (i.e., box s, ts, t and n).

The AMOC strength in the northern box (q_N) in the E-CCM is given by:

$$q_N = \eta_h \frac{\rho_n - \rho_{ts}}{\rho_0} D^2, \quad (1)$$

where η_h is a hydraulic constant, $\rho_n - \rho_{ts}$ is the meridional density difference between box n and box ts, ρ_0 is a reference density, and D the pycnocline depth. The densities are determined from a linear equation of state. For full details and sensitivity experiments conducted with the E-CCM, we refer to van Westen et al. (2024b), where there is also a link to the publicly-available E-CCM code. We will show results for the version where sea-ice insulation effects are omitted and use the standard values of the parameters given in van Westen et al. (2024b), unless otherwise mentioned.

The E-CCM is forced through the asymmetric freshwater flux forcing (E_A) from box s to box n. Under varying E_A , the E-CCM has an ‘AMOC on’ state (clockwise circulation, red solid and dashed arrows) and an ‘AMOC off’ state (anti-clockwise circulation, red solid and dotted arrows). There is a multi-stable AMOC regime and this regime is bounded by two saddle-node bifurcations (van Westen et al., 2024b). To determine the sensitivity of the AMOC behavior to the hosing location (Rahmstorf, 1996; Ma et al., 2024), we make a modification to the E-CCM by distributing the freshwater flux forcing linearly over box n and box t using a parameter $\xi \in [0, 1]$. When $\xi = 0$, the freshwater flux forcing is only applied to box n and this is the original E-CCM configuration. The freshwater flux forcing is only over box t when $\xi = 1$.

The steady states of the E-CCM against varying parameters (i.e., bifurcation diagram), such as freshwater flux forcing, are determined using the continuation software AUTO-07p (Doedel et al., 2007, 2021). This code solves steady states using a pseudo-arclength continuation combined with a Newton-Raphson method (Wubs and Dijkstra, 2023). It is also able to detect Hopf bifurcations and saddle-node bifurcations. We used a value of 10^{-6} for the absolute and relative accuracy of each steady-state solution, and for the accuracy for locating special points, similar to van Westen et al. (2024b).

3 Results

3.1 Statistical equilibria in the CESM

The AMOC strength (at 1,000 m and 26°N) and the freshwater transport carried by the AMOC at 34°S (F_{ovs}) of the quasi-equilibrium CESM simulation (van Westen et al., 2024a) are shown in Figures 2a,b. The branched simulations from the quasi-equilibrium simulation at a constant forcing $\overline{F_H} = 0.18$ Sv (Figures 2c,i), $\overline{F_H} = 0.45$ Sv (Figures 2d,j) and $\overline{F_H} = 0.465$ Sv (Figures 2e,k) equilibrate after about 300 years. The branched simulation at $\overline{F_H} = 0.48$ Sv (Figures 2f,l) collapses and suggests that the upper bound of the multi-stable regime is around this $\overline{F_H}$ value. The branches initiated from $\overline{F_H} = 0.495$ Sv (Figures 2g,m) and $\overline{F_H} = 0.51$ Sv (Figures 2h,n) also collapse; these simulations were terminated before the 500-year mark because of computational costs. However, when the equilibrated $\overline{F_H} = 0.465$ Sv simulation is subjected to an instantaneous increase in freshwater flux to $\overline{F_H} = 0.48$ Sv ($\Delta\overline{F_H} = 0.015$ Sv), we still find a statistical equilibrium in the northward over-
turning regime (red curves in Figures 2f,l). We iteratively repeated the same procedure for $\overline{F_H} = 0.495$ Sv and $\overline{F_H} = 0.51$ Sv. The AMOC eventually collapses under a constant freshwater flux forcing of $\overline{F_H} = 0.51$ Sv. This means that the upper bound of the multi-stable regime is found for $0.495 \text{ Sv} \leq \overline{F_H} < 0.51 \text{ Sv}$. To obtain an even higher precision for this upper bound, we would need to increase $\overline{F_H}$ with even smaller increments, but is not done here because of computational limitations.

The AMOC in the quasi-equilibrium simulation starts to tip around $F_H = 0.525$ Sv (0.522 to 0.533 Sv, 10th and 90th percentiles, van Westen et al. (2024a)) and is at larger F_H values than the upper bound found from the statistical equilibria

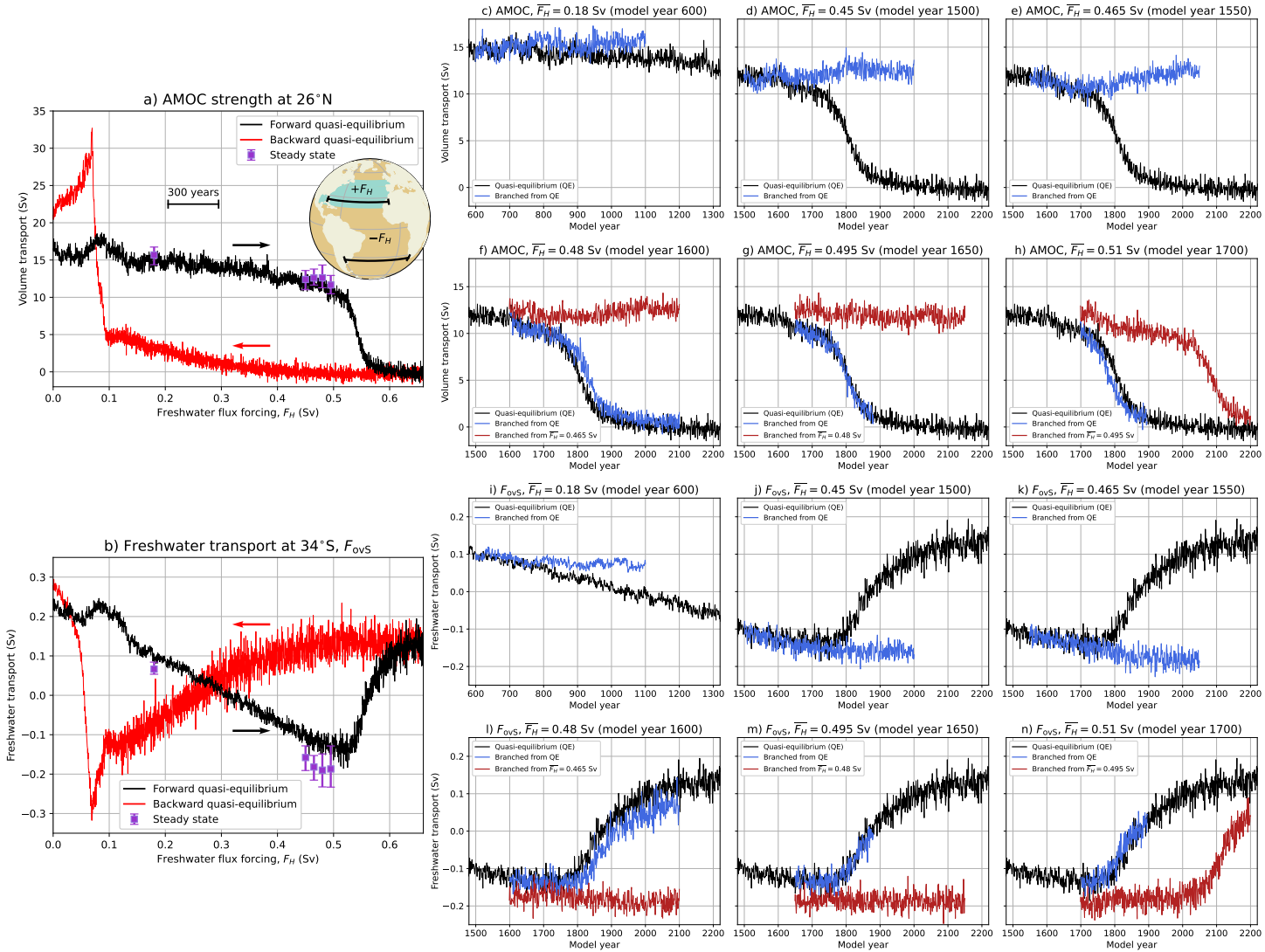


Figure 2. (a): The AMOC strength at 1,000 m and 26°N and (b): the freshwater transport by the AMOC at 34°S , F_{ovS} , for varying freshwater flux forcing F_H (i.e., the quasi-equilibrium simulation). Inset: The hosing experiment where fresh water is added to the ocean surface between 20°N – 50°N in the Atlantic Ocean ($+F_H$) and is compensated over the remaining ocean surface ($-F_H$). The statistical equilibria for various constant values of F_H (i.e., $\overline{F_H}$, steady states) in the northward overturning regime are also shown, where the marker indicates the mean and the error bars show the minimum and maximum over the last 50 years of the 500-year long branched simulations. The black sections indicate the 26°N and 34°S latitudes over which the AMOC strength and F_{ovS} are determined, respectively. The yellow shading in the two panels indicates observed ranges for the presented quantity (Smeed et al., 2018; Arumí-Planas et al., 2024). (c – n): Similar to panels a,b, but now the entire branched simulations for different $\overline{F_H}$ values. The branches are initiated from the quasi-equilibrium simulation (blue curves) or from the end of the previous statistical equilibria (red curves).

simulation ($0.495 \text{ Sv} \leq \overline{F_H} < 0.51 \text{ Sv}$). To determine the overshoot of the quasi-equilibrium simulation, we use a reference value of $\overline{F_H} = 0.5 \text{ Sv}$, but any other $\overline{F_H}$ value within the interval $\overline{F_H} \in [0.495, 0.51]$ can be used as a reference (giving slightly different numerical results). Using this reference, the quasi-equilibrium AMOC overshoots by $\Delta F_H = 0.025 \text{ Sv}$ (≈ 80 years). Do note that the AMOC collapses for the simulations branched from the quasi-equilibrium simulation for $\overline{F_H} \geq 0.48 \text{ Sv}$ (blue curves in Figure 2c – n). In other words, the branched simulations for $\overline{F_H} \geq 0.48 \text{ Sv}$ already surpassed a critical forcing value upon branching, which means that the standard quasi-equilibrium also surpassed its critical value and actually undershoots the upper bound of the multi-stable regime. This critical value for the quasi-equilibrium is located for $0.465 \text{ Sv} < \overline{F_H} \leq 0.48 \text{ Sv}$. The apparent overshoot with the reference value of $\overline{F_H} = 0.5 \text{ Sv}$ is then the result of inertia and the growth rate of AMOC feedbacks, in particular the destabilising salt-advection feedback. Indeed, these feedbacks develop on centennial timescales (Vanderborgh et al., 2025), which we will make more explicit below. The undershooting AMOC can already be seen when comparing the quasi-equilibrium with five different statistical equilibria (last 50 model years are used). The quasi-equilibrium simulation is about 1 Sv weaker than the statistical equilibria, but still reasonably agree. For F_{ovS} , on the other hand, the quasi-equilibrium is larger and (mostly) outside the ranges of the different statistical equilibria (Figure 2b).

When we lower the freshwater flux forcing rate, we expect that the system stays closer to the statistical equilibria (Hawkins et al., 2011). To test this, we branched off a quasi-equilibrium simulation with only half the hosing rate (i.e., $1.5 \times 10^{-4} \text{ Sv yr}^{-1}$) from the end of the statistical equilibrium at $\overline{F_H} = 0.45 \text{ Sv}$. This simulation was integrated for 1,050 model years, where F_H varied from 0.45 Sv to 0.608 Sv (red curves in Figures 3a,b). In the ideal case, the half-forcing quasi-equilibrium simulation should have been initiated from the same initial conditions as the standard quasi-equilibrium simulation for direct comparison, which would also allow to address the sensitivity in the overshoot/undershoot with the reference value of $\overline{F_H} = 0.5 \text{ Sv}$. Nevertheless, this half-forcing simulation can be used to check whether the AMOC collapse happens faster in F_H space. The faster transition (in F_H) is a characteristic of a saddle-node bifurcation (see Appendix B), but this is also the case for other bifurcation types (e.g., Hopf) (Berglund and Gentz, 2006).

The half-rate simulation remains (very) close to the different statistical equilibria for both AMOC strength and F_{ovS} . Following van Westen et al. (2024a), we used a break regression analysis (Mudelsee et al., 2014) to find the AMOC tipping event at $F_H = 0.534 \text{ Sv}$, with the 10th and 90th percentiles at $F_H = 0.533 \text{ Sv}$ and $F_H = 0.536 \text{ Sv}$, respectively. There is an overshoot of $\Delta F_H = 0.034 \text{ Sv}$ (227 years) compared to our reference value of $\overline{F_H} = 0.5 \text{ Sv}$, but keep in mind that AMOC feedbacks take a considerable time to develop. These feedbacks can be quantified by following the procedure outlined in Vanderborgh et al. (2025), see also Section 4 below. We decompose the different AMOC feedbacks for the $\overline{F_H} = 0.51 \text{ Sv}$ simulation (branched from the previous statistical equilibrium of $\overline{F_H} = 0.495 \text{ Sv}$) and the half-rate forcing simulation, where the most important feedbacks are shown in Figure 4; the standard quasi-equilibrium simulation decomposition is presented in Vanderborgh et al. (2025).

First the $\overline{F_H} = 0.51 \text{ Sv}$ simulation (Figure 4a), in which the AMOC weakens by about 1.5 Sv during the first 100 model years. This weakening is attributed to the slightly larger freshwater forcing (+0.015 Sv) compared to the starting equilibrium solution at $\overline{F_H} = 0.495 \text{ Sv}$. The destabilizing salt-advection feedback (linked to F_{ovS}) and surface (mainly sea-ice melt) feedback slowly grow over the following 250 years. Over the same period (model years 100 – 350), the gyres and overturning component at

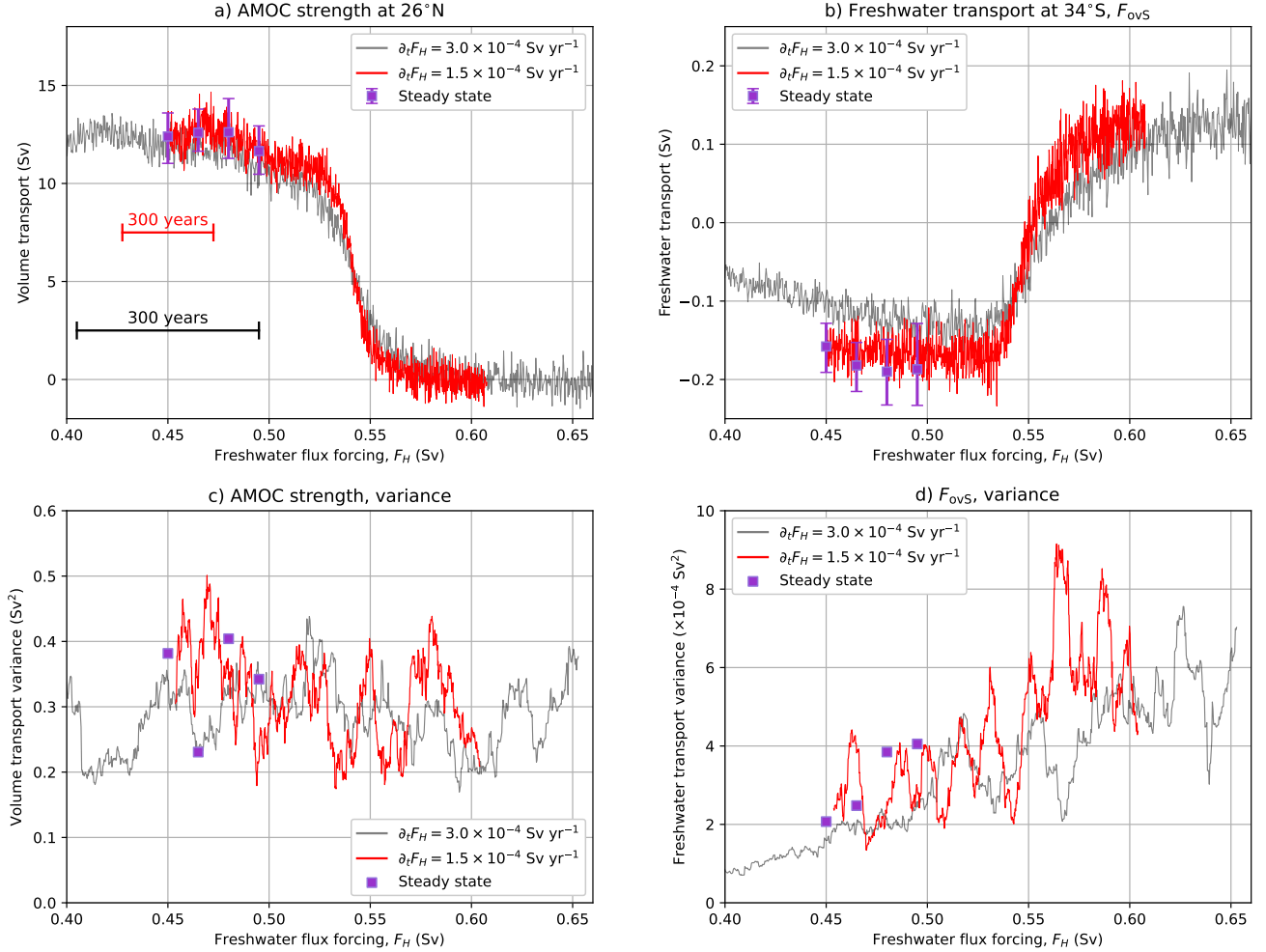


Figure 3. (a & b): The AMOC strength and F_{ovS} of the quasi-equilibrium simulations, one similar to Figures 2a,b, and including the simulation with varying $1.5 \times 10^{-4} \text{ Sv yr}^{-1}$ hosing rate (red curves). This quasi-equilibrium hosing with $1.5 \times 10^{-4} \text{ Sv yr}^{-1}$ was branched from the end of the statistical equilibria at $\overline{F_H} = 0.45 \text{ Sv}$. (c & d): The variance in AMOC strength and F_{ovS} , using a sliding window of 50 years. For each 50-year window, a linear trend was removed and then the variance was determined.

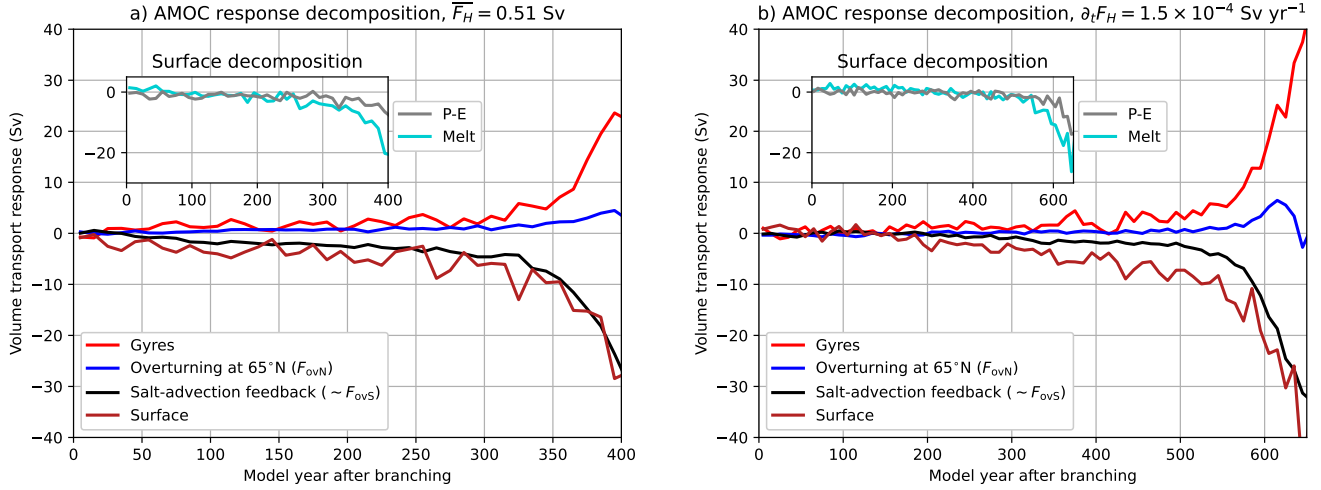


Figure 4. (a & b): Decomposition of the AMOC feedbacks (Vanderborgh et al., 2025) for the (a): $\overline{F_H} = 0.51$ Sv simulation (i.e., red curve in panel 2h) and the (a): half-rate forcing simulation (1.5×10^{-4} Sv yr $^{-1}$). The inset shows the two surface components of Arctic sea-ice melt and precipitation minus evaporation (P-E). The time series are presented as 10-year averages (to reduce the variance). Note the different horizontal ranges between the two panels.

65°N partly stabilize the AMOC. The combined effect results in an AMOC weakening of only 1.5 Sv over these 250 years and after model year 350 the AMOC fully collapses. The salt-advection feedback eventually becomes dominant and this destabilising feedback fully develops over centennial timescales (under constant freshwater flux forcing).

Next the half-rate forcing simulation (Figure 4b), where we find a similar centennial timescale for the destabilizing AMOC feedbacks. The AMOC feedbacks remain relatively small up to model year 350 ($F_H = 0.503$ Sv), then slowly increase in the following 200 years (model years 350 – 550) and thereafter the AMOC fully collapses. This gradual increase of the destabilizing feedbacks between model year 350 to 500, suggests that the AMOC will eventually tip and hence branching simulations with fixed F_H for $F_H \geq 0.503$ Sv will also result in an AMOC collapse, similarly as the standard quasi-equilibrium simulation. However, additional simulations are needed to find this critical value, which were not done here.

In other words, there is a certain critical value of forcing and, once crossed, the AMOC will eventually tip over centennial timescales (≈ 200 model years). This critical value is dependent on the initial condition and rate of forcing, which we will make more explicit with the E-CCM below. As argued above, for the half-forcing quasi-equilibrium simulation this critical value is likely around $F_H = 0.503$ Sv, which is well within the interval $0.495 \text{ Sv} \leq \overline{F_H} < 0.51 \text{ Sv}$. The AMOC collapse starts at $F_H = 0.525$ Sv in the standard quasi-equilibrium simulation, meaning that the destabilizing feedbacks were growing during the 200 model years ($\Delta F_H = 0.06$ Sv) prior to the collapse (Vanderborgh et al., 2025). This suggests that $\overline{F_H} = 0.525 - 0.06 = 0.465$ Sv is the latest statistical equilibrium which can be found when directly branching from the quasi-equilibrium simulation, which is indeed the case here (Figures 2e,k). This confirms again that the standard quasi-equilibrium simulation undershoots the upper bound of the multi-stable regime. The implication is that an overshooting (or undershooting) AMOC cannot be

assessed by only analysing the onset of the AMOC tipping event. In fact, the onset of the AMOC tipping event only indicates
 195 where the destabilising feedbacks become dominant and it is much more useful to analyse the changes in AMOC feedback strengths.

What is important here, is that the half-rate forcing's transition to the collapsed state is twice as fast (in F_H space), which is a typical characteristic of transitions near a saddle-node bifurcation (Berglund and Gentz, 2006) (see also Appendix B). The duration of AMOC transitions in both quasi-equilibria and in the statistical equilibrium simulations (Figure 2) is about
 200 100 years and the full equilibration to the collapsed AMOC state requires more than 500 years (van Westen et al., 2024a). Another characteristic of a saddle-node bifurcation is the loss of resilience (i.e., critical slow down) near the tipping point (van Westen et al., 2024b). This can be quantified by determining the variance and (lag-1) autocorrelation of specific observables. For the AMOC strength, we find no indications of critical slow down (not shown) which is consistent with the results in van Westen et al. (2024a). There is also no increase in the variance for the AMOC strength for both the quasi-equilibria and
 205 the statistical equilibria (Figure 3c). However, for the physics-based quantity F_{ovS} we find indications of critical slowdown (van Westen et al., 2024a; Smolders et al., 2025). Indeed, the F_{ovS} variance increases for larger F_H up to the tipping event (Figure 3d). This increase in variability indicates that the AMOC loses resilience and makes it more prone to transitions.

3.2 Equilibria in the E-CCM

The AMOC behaviour in the CESM can be reproduced with the E-CCM under varying freshwater flux forcing (now E_A).
 210 For the E-CCM, the steady states obtained using continuation techniques (cf. section 2b) are presented in Figures 5a,b for the AMOC strength and F_{ovS} , respectively. The continuation indicates two saddle-node bifurcations at $E_A^1 = 0.4861$ Sv (AMOC on) and at $E_A^2 = 0.1857$ Sv (AMOC off). The AMOC on and unstable steady states clearly show the square-root behaviour between AMOC strength and E_A , which arises from the dominant salt-advection feedback close to E_A^1 . The probabilities under (stochastic) noise for the transition from an AMOC on to an AMOC off state approach 1 when moving closer to E_A^1 (van
 215 Westen et al., 2024b), indicative of the loss of resilience. Here, we performed deterministic quasi-equilibrium and equilibrium simulations with the E-CCM, which are shown in Figure 5. Note that we used slightly different freshwater flux forcing (E_A) values in the E-CCM than in the CESM.

The quasi-equilibrium hysteresis simulation in the E-CCM is (qualitatively) comparable to that of the CESM (compare Figures 2 and 5); the large overshoot (> 35 Sv) in the E-CCM upon AMOC recovery is a model artefact (van Westen et al.,
 220 2024b). In the forward quasi-equilibrium simulation the AMOC strength is lower compared to the value at the steady states, while the F_{ovS} values are higher. The branches from the quasi-equilibrium eventually collapse for $\overline{E_A} = 0.477$ Sv and $\overline{E_A} = 0.486$ Sv, meaning that a critical $\overline{E_A}$ value was surpassed, which is then also the case for the quasi-equilibrium simulation.

In contrast to the CESM, it is computationally feasible to quantify this critical value in the E-CCM. Here, we define the critical branch as the branch from the quasi-equilibrium that collapses at the lowest possible $\overline{E_A}$ value. We use an accu-
 225 racy of $\Delta\overline{E_A} = 0.001$ Sv (but can be even higher when needed). Ultimately, the AMOC collapses when branching from the quasi-equilibrium simulation for $\overline{E_A} \geq 0.474$ Sv (Figure 6a). As was argued in the previous section, this critical value is also dependent on the initial condition and rate of forcing. For example, when we use the steady state at $\overline{E_A} = 0.45$ Sv as initial

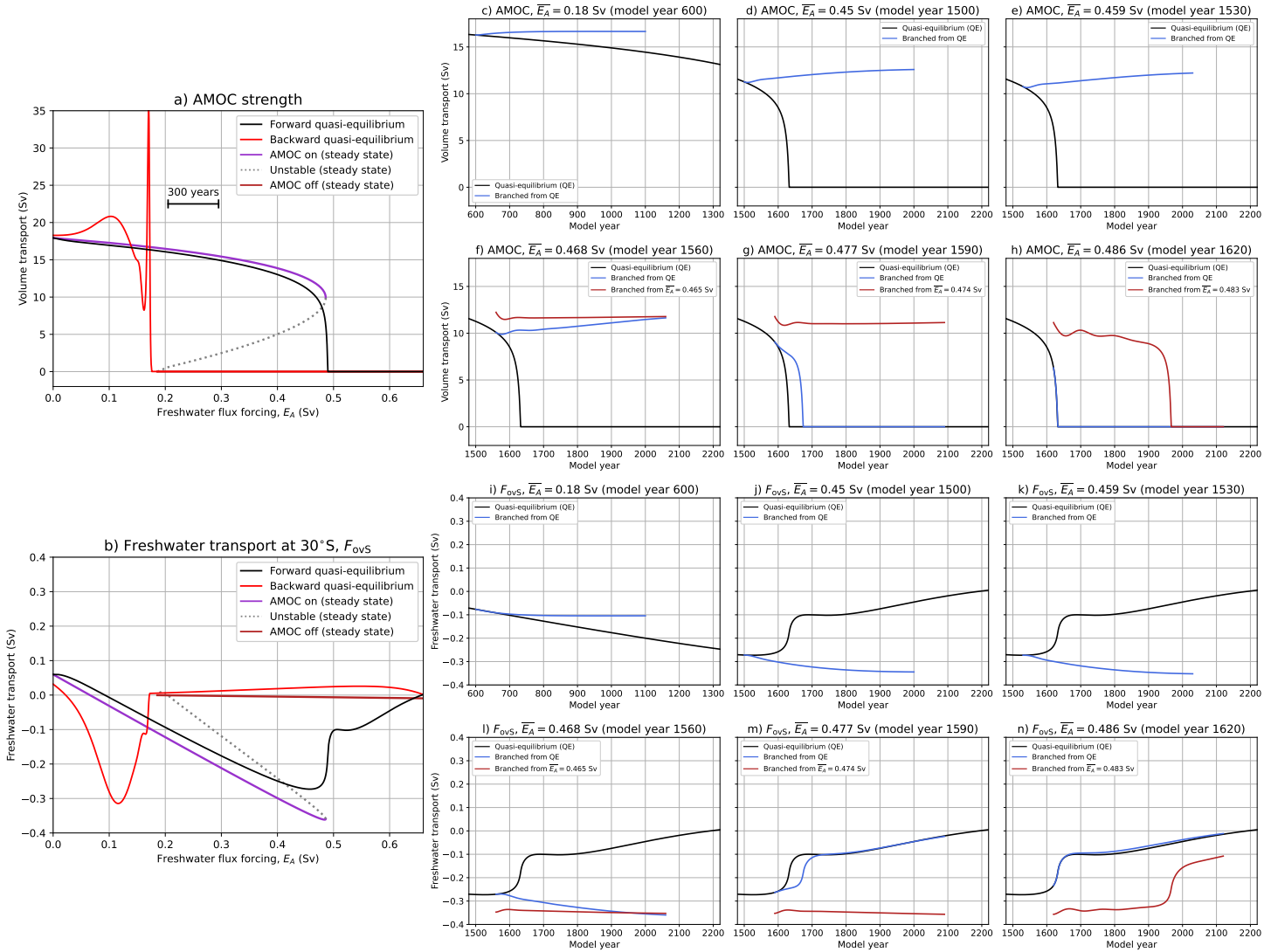


Figure 5. Similar to Figure 2, but now for the E-CCM. Note that in panels a and b the steady and unstable states (from the continuation) are also shown.

condition, we can increase E_A up to 0.479 Sv (with 3×10^{-4} Sv yr⁻¹) and then keeping the freshwater flux forcing constant. In this case, the AMOC eventually equilibrates to the AMOC on state (not shown), meaning that the critical value is found for $\overline{E_A} \geq 0.480$ Sv. When we initiate from $E_A = 0$ Sv while using a ten times smaller forcing rate (3×10^{-5} Sv yr⁻¹), the AMOC also equilibrates to the AMOC on state when increasing E_A up to 0.483 Sv and then keeping the freshwater flux forcing constant (not shown). The critical value for this other case is for $\overline{E_A} \geq 0.484$ Sv. Depending on the initialisation and forcing rate, the saddle-node bifurcation can only be reached with a limited accuracy.

Since the AMOC collapses at critical values lower than (i.e., undershooting) the saddle-node bifurcation (blue curve in Figure 6a), the system must cross the basin boundary of attraction between the AMOC on and AMOC off states. The continuation allows us to explore which variable (temperature, salinity, and pycnocline depth), or which specific combination of variables (e.g., AMOC strength, see (1)), crosses this boundary of attraction. Notably, the critical branch at $\overline{E_A} = 0.474$ Sv does not cross the basin boundary with respect to AMOC strength and one expects AMOC recovery to the AMOC on state, and yet the AMOC collapses (left inset in Figure 6a). This means that the AMOC strength is no good predictor for the future evolution of the system for the critical branch. When we analyse a different quantity, such as the salinity of box n (right inset in Figure 6a), it does cross the basin boundary. The salinity in box n is important here as it (partly) sets the AMOC strength (relation 1) and is influenced under the destabilizing salt-advection feedback, which gives rise to the quadratic relation between AMOC strength and freshwater flux forcing.

When we equally distribute the hosing over box n and box t ($\xi = 0.5$, Figure 6b), the saddle-node bifurcations shift to higher values of E_A . The quasi-equilibrium for this case has weaker AMOC strengths than the stable AMOC on state and close to the saddle-node bifurcation it has stronger strengths than the AMOC on state (left inset in Figure 6b). The critical branch (at $\overline{E_A} = 0.609$ Sv) has a stronger AMOC strength than the steady AMOC on state upon branching, but it still collapses. The salinity in box n does cross the basin boundary (right inset in Figure 6b), demonstrating again that AMOC strength is no good indicator for predicting the future AMOC trajectory. Only when the hosing is applied over box t ($\xi = 1.0$, Figure 6c), the AMOC collapses when increasing the freshwater flux forcing beyond the saddle-node bifurcation of $E_A^1 = 0.83495$ Sv. When we branch from the quasi-equilibrium for lower E_A than the saddle-node bifurcation (e.g., $\overline{E_A} = 0.8348$ Sv, not shown), the solution equilibrates to the stable AMOC on state.

The AMOC dynamics and the under- and overshooting behaviour can be understood from these three different cases. When a hosing perturbation is (partly) applied over box n, the AMOC strength directly reduces as the meridional salinity difference between box n and box ts increases. The largest part of the freshwater perturbation is carried away by the AMOC to box d, but a small part of the perturbation remains in box n (due to a weaker AMOC) and causes freshwater accumulation over box n. This freshwater accumulation results in a slightly weaker AMOC strengths compared to the steady states. Once the system has a sufficient amount of time to adjust to the imposed freshwater perturbation, the entire freshwater perturbation is redistributed over the boxes and the AMOC strength eventually increases (e.g., blue curves in Figures 5c,d,e,f). In other words, the advective (‘flushing’) timescale is slower than the hosing timescale, resulting in an enhanced AMOC strength decline. This makes the AMOC more prone to freshwater perturbations and explains why there is hardly any overshoot in the quasi-

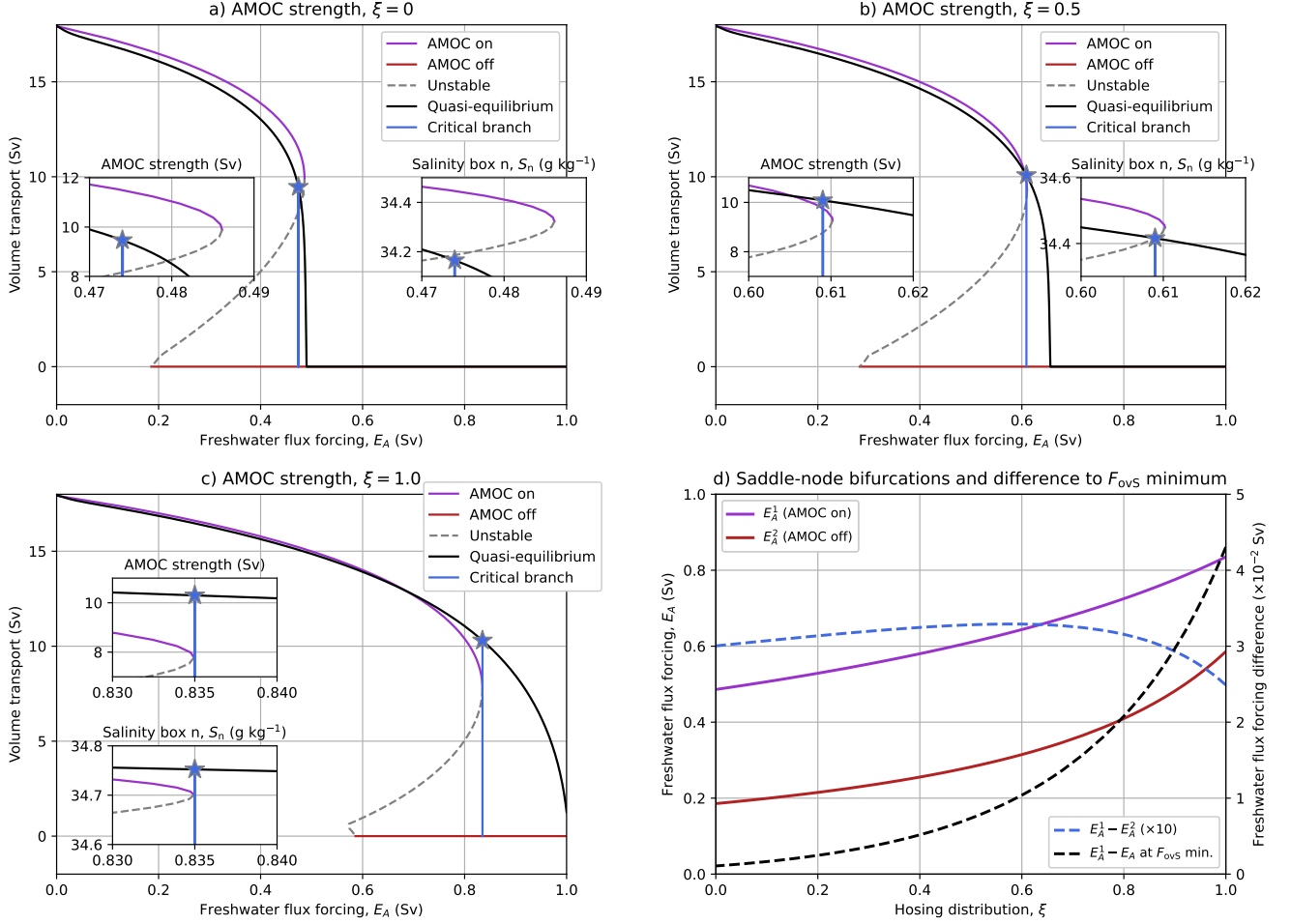


Figure 6. (a): The steady states for the AMOC strength and a quasi-equilibrium simulation (rate 3×10^{-4} Sv yr⁻¹) for the hosing over box n ($\xi = 0$). A simulation was branched from the quasi-equilibrium simulation for $\overline{E_A} = 0.474$ Sv (blue star), which was integrated into equilibrium. The two insets show zoomed-in versions of the AMOC strength and salinity of box n near the saddle-node bifurcation. (b & c): Similar to panel a, but now for b) $\xi = 0.5$ and c) $\xi = 1$, where the branched simulations were initiated at $\overline{E_A} = 0.609$ Sv and $\overline{E_A} = 0.835$ Sv, respectively. (d): The position of the saddle-node bifurcations of the AMOC on (E_A^1) and AMOC off (E_A^2) states (solid curves). The distance (expressed in ΔE_A) between E_A^1 and E_A^2 and between the E_A^1 and the F_{ovs} minimum.

equilibrium simulation with the saddle-node bifurcation (for $\xi = 0$). This is qualitatively different than the quasi-equilibrium CESM, meaning that $\xi = 0$ is not very likely for the CESM.

The direct AMOC weakening effect is smaller when adding (part of) the hosing over box t and there are two effects contributing to this different behaviour. First, the hosing is now distributed over the (much) larger box t than box n and making the salinity anomalies (averaged over box t) effectively smaller. Second, only a part of the salinity perturbations from box t is carried by the AMOC into box n and most of it is directly carried to box d (see also Figure 1). This implies that the role of the overturning contribution in redistributing salinity anomalies between box t and box n is getting smaller, while the (northern) gyre contribution is getting more important. These combined effects explain why the saddle-node bifurcations shift to larger E_A values for increasing ξ (Figure 6d). The larger gyre contribution is also reflected in a greater ΔE_A between the E_A^1 and F_{ovS} minimum, which also modifies the hysteresis width which is measured as the distance between the two saddle-node bifurcations (Figure 6d).

In the standard quasi-equilibrium CESM simulation (rate 3×10^{-4} Sv yr $^{-1}$), the AMOC strength is also smaller than that of the statistical equilibria. Thereafter, the AMOC appears to overshoot the upper bound of the multi-stable regime. The CESM trajectory shares similar characteristics as the E-CCM in the $\xi = 0.5$ configuration, which is consistent with the applied hosing region in the CESM (20°N – 50°N), though the CESM is much more complex than the E-CCM. Depending on the hosing region, one can change the relative contributions of important AMOC feedbacks and this results in differences in AMOC sensitivity, the onset of the AMOC tipping event and width of the multi-stable regime. It is therefore important to use a fixed hosing region, as was done for our CESM simulations or in the outlined procedure of the North Atlantic Hosing Model Intercomparison Project (NAHosMIP, Jackson et al. (2023)). Sensitivity experiments indicate that the northern portion of the North Atlantic (e.g., the Irminger basin) is most sensitive under hosing (Rahmstorf, 1996; Ma et al., 2024). Nevertheless, the destabilising salt-advection feedback becomes more dominant under increasing hosing strengths and causes the square-root dependency near the saddle-node bifurcation.

4 Feedback analysis in the CESM

The results from Section 3.2 demonstrate that as long as the salt-advection feedback dominates, one may expect a square root dependence in the AMOC on state under increasing freshwater flux forcing, similar to the Stommel model (see Appendix A). Although the AMOC is (highly) idealised in the E-CCM, it is qualitatively able to reproduce almost all AMOC characteristics of that in a much more complex and fully-coupled climate model (i.e., the CESM). This makes the existence of a saddle-node bifurcation in the CESM plausible, but this can not easily be demonstrated using only a limited number of equilibrium simulations. However, it turns out that from performing a feedback analysis as in Vanderborght et al. (2025), we can (under reasonable assumptions) derive a reduced model explicitly showing the dependence of AMOC strength on F_H .

4.1 Reduced model derivation

We start from the total Atlantic (34°S to 65°N) freshwater budget as governed by (Vanderborght et al., 2025):

$$\frac{dW}{dt} = F_{azS} - F_{azN} + F_{ovS} - F_{ovN} + F_{surf} + F_{res}, \quad (2)$$

295 where W is the total freshwater content. The Atlantic freshwater content can be modified through azonal (gyre) contributions (i.e., F_{azS} and F_{azN}), overturning contributions (i.e., F_{ovS} and F_{ovN}), surface contribution (i.e., F_{surf}) and residual contribution (i.e., F_{res}). The quantities F_{azS} and F_{ovS} are evaluated at 34°S, hence indicated with subscript ‘S’, and we follow a similar notation for the northern boundary (65°N) by using a subscript ‘N’.

Upon a freshwater perturbation, the evolution of the different contributions depends on the background state and the AMOC
300 strength (Vanderborght et al., 2025). The AMOC strength is fairly homogeneous over the Atlantic basin (van Westen et al., 2024a) and we assume a northward volume transport in the upper AMOC limb which we indicate here as Ψ ; the lower AMOC limb then carries Ψ southward. The velocity-weighted average salinity over the upper AMOC limb is indicated with $S_{\rightarrow}$, and similarly for the lower AMOC limb we use $S_{\leftarrow}$. The vertical salinity difference between the upper AMOC limb and lower AMOC limb is then indicated by $S_{\rightleftharpoons} = S_{\rightarrow} - S_{\leftarrow}$. Under this idealization it directly follows that:

$$305 \quad F_{ovS} = -\frac{S_{\rightleftharpoons}}{S_0} \Psi, \quad (3)$$

where $S_0 = 35 \text{ g kg}^{-1}$. Because the salinity transport in the lower AMOC limb is approximately adiabatic, the vertical salinity contrast at 34°S is closely related to a meridional salinity contrast between 34°S and the North Atlantic sinking region. This meridional salinity contrast is related to the AMOC strength via thermal wind balance (Butler et al., 2016). Therefore, the vertical salinity contrast scales with the AMOC strength as (Vanderborght et al., 2025):

$$310 \quad \Psi = \Psi_0 + c_2 (1 - c_1) (S_{\rightleftharpoons}(0) - S_{\rightleftharpoons}), \quad (4)$$

where c_1 represents the stabilizing thermal-advective feedback and c_2 is a scaling factor. Both c_1 and c_2 are positive constants and, for the CESM, their values are about 0.52 and 20 Sv kg g⁻¹ (Vanderborght et al., 2025). The terms Ψ_0 and $S_{\rightleftharpoons}(0)$ are the AMOC strength and vertical salinity difference for $F_H = 0$ Sv (no hosing), respectively.

Under the applied hosing (indicated by δF_H in the CESM) the value of F_{surf} increases and is primarily (i.e., to first order)
315 balanced by a declining F_{ovS} (van Westen et al., 2024a). On the other hand, the gyres flush freshwater anomalies out of the Atlantic Ocean and stabilize the AMOC (Vanderborght et al., 2025). Sijp (2012) argued that $S_{\rightleftharpoons}$ linearly scales with the integrated Atlantic freshwater content. This integrated freshwater content in turn scales with the anomalous freshwater transport by the gyres (Huisman et al., 2010), i.e.:

$$F_{gyre} = F_{azS} - F_{azN} = -g_1 S_{\rightleftharpoons} + g_2. \quad (5)$$

320 This linear relation is also applicable for the CESM, where $g_1 = 0.032 \text{ Sv kg g}^{-1}$ and $g_2 = 0.49 \text{ Sv}$ (Figure 7a). The last contribution which we consider is the overturning component at the northern boundary, F_{ovN} . The AMOC strength almost

vanishes at the northern boundary and the expression for F_{ovN} is different than that of the F_{ovS} (relation 3). The F_{ovN} scales linearly with S_{eq} and can be approximated by:

$$F_{\text{ovN}} = n_1 S_{\text{eq}} + n_2 \quad (6)$$

325 with $n_1 = 0.025 \text{ Sv kg g}^{-1}$ and $n_2 = -0.021 \text{ Sv}$ for the CESM as shown in Figure 7b. The contributions by the gyres and F_{ovN} scale linearly with increasing S_{eq} (or decreasing Ψ), whereas the F_{ovS} has a non-linear contribution. To be more precise, the F_{ovS} is determined by the product of the vertical salinity difference and the AMOC strength, where the latter scales linearly with the vertical salinity difference (i.e., relation 4). The F_{ovS} scales quadratically with AMOC strength, and conversely AMOC strength scales with the square root dependence on F_{ovS} . As the imposed freshwater flux forcing is primarily balance by F_{ovS}
 330 in the CESM (van Westen et al., 2024a), one expects a square root dependence in AMOC strength under increasing freshwater flux forcing.

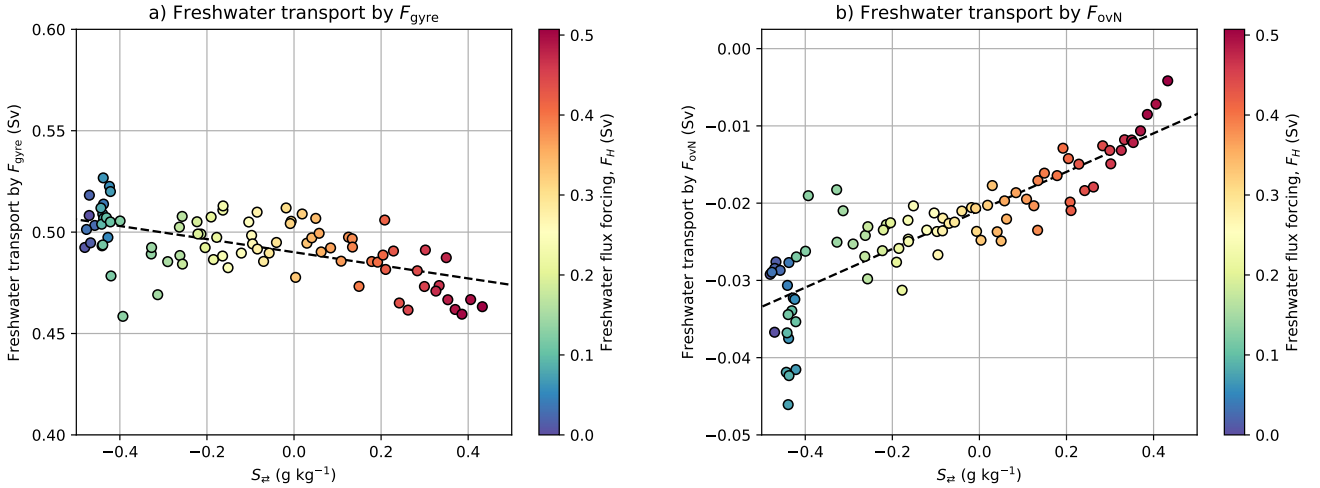


Figure 7. (a): The relation between F_{gyre} and S_{eq} , where the linear fit is determined over the 20-year averages up to model year 1,700 ($F_H = 0.51 \text{ Sv}$) of the standard quasi-equilibrium simulation. (b): Similar to panel a, but now for the F_{ovN} and S_{eq} .

A perturbation in the Atlantic freshwater content (cf. (2)) around an equilibrium state then gives:

$$-\delta F_{\text{ovS}} + \delta F_{\text{ovN}} - \delta F_{\text{gyre}} = \delta F_{\text{surf}}, \quad (7)$$

and using the expressions for F_{ovS} , F_{gyre} and F_{ovN} , this yields:

$$335 \quad \Psi \delta S_{\text{eq}} + S_{\text{eq}} \delta \Psi + n_1 S_0 \delta S_{\text{eq}} + g_1 S_0 \delta S_{\text{eq}} = S_0 \delta F_H \quad (8)$$

Using the relation between Ψ and S_{eq} (from 4) we find:

$$-\frac{\Psi}{c_2(1-c_1)} \delta \Psi + \left(-\frac{\Psi}{c_2(1-c_1)} + \frac{\Psi_0}{c_2(1-c_1)} + S_{\text{eq}}(0) - \frac{(n_1 + g_1)S_0}{c_2(1-c_1)} \right) \delta \Psi = S_0 \delta F_H, \quad (9)$$

which can be rewritten as:

$$(-2\Psi + \Psi_0 + c_2(1 - c_1)S_{\rightleftharpoons}(0) - (n_1 + g_1)S_0)\delta\Psi = c_2(1 - c_1)S_0\delta F_H, \quad (10)$$

340 and integrating both sides gives:

$$\Psi^2 - (\Psi_0 + c_2(1 - c_1)S_{\rightleftharpoons}(0) - (n_1 + g_1)S_0)\Psi + c_2(1 - c_1)S_0F_H + C = 0, \quad (11)$$

with integration constant C . The solution with $\Psi(F_H = 0) = \Psi_0$ is:

$$\Psi(F_H) = \frac{\Psi_0}{2} + \frac{c_2(1 - c_1)S_{\rightleftharpoons}(0)}{2} - \frac{(n_1 + g_1)S_0}{2} \pm \sqrt{\left(\frac{\Psi_0 - c_2(1 - c_1)S_{\rightleftharpoons}(0) + (n_1 + g_1)S_0}{2}\right)^2 - c_2(1 - c_1)S_0F_H} \quad (12)$$

Rather using $S_{\rightleftharpoons}(0)$, we express it as the initial F_{ovS} using (3), i.e., $S_{\rightleftharpoons}(0) = -\frac{S_0F_{\text{ovS}}(0)}{\Psi_0}$. The final expression becomes:

$$\Psi(F_H) = \frac{\Psi_0}{2} - \frac{c_2(1 - c_1)S_0F_{\text{ovS}}(0)}{2\Psi_0} - \frac{(n_1 + g_1)S_0}{2} \pm \sqrt{\left(\frac{\Psi_0^2 + c_2(1 - c_1)S_0F_{\text{ovS}}(0) + (n_1 + g_1)S_0\Psi_0}{2\Psi_0}\right)^2 - c_2(1 - c_1)S_0F_H} \quad (13)$$

345

Do note that several assumptions are required to arrive at this final expression. For example, various residual (F_{res}) and climate feedbacks were not considered, such as ocean-sea ice interactions (destabilizing), ocean-atmosphere fluxes (destabilizing), pycnocline deepening (stabilising), open Bering strait (stabilizing) and the effect of ocean eddies (stabilizing) (Vanderborght et al., 2025). The linear relation in F_{gyre} and F_{ovN} with $S_{\rightleftharpoons}$ is less accurate and c_1 is less constant close to the tipping point. Freshwater anomalies may be stored in the Atlantic Ocean and hence we assumed that changes in the freshwater content are much smaller than changes in the freshwater balance terms (i.e., $\frac{dW}{dt} \ll \Delta(F_{\text{azS}} - F_{\text{azN}} + F_{\text{ovS}} - F_{\text{ovN}} + F_{\text{surf}} + F_{\text{res}})$). These additional feedbacks and processes modify the idealized AMOC response and make it more difficult to derive an analytical solution for the northward overturning regime, as these processes (ideally) need to be expressed as a function of $S_{\rightleftharpoons}$ (if it exists). We stress that this idealized AMOC response under hosing should be interpreted with care and one needs to consider the appropriate feedback contributions for each (climate) model set-up. The key point is that the AMOC strength exhibits a square-root dependence on the freshwater flux forcing, leading to a saddle-node bifurcation when the dominant balance is between the applied freshwater flux forcing and the overturning component. As long as other contributions remain sufficiently small, their effect will not change the structure (and therefore the type) of the bifurcation diagram. Indeed, the F_{gyre} and F_{ovN} remain fairly linear up to $F_H = 0.51$ Sv (Figure 7) and this is beyond the critical forcing ($0.465 \text{ Sv} < \overline{F_H} \leq 0.48 \text{ Sv}$, Figure 2) for which the salt-advection feedback becomes dominant. Once the AMOC starts to collapse, the different AMOC contributions become much larger (e.g., Figure 4) and their responses are attributed to large-scale adjustments under a collapsing AMOC.

360

For the Stommel 2-box model, we can demonstrate that a similar AMOC response holds (see Appendix A). Under no freshwater flux forcing ($\eta = 0$) in this model, the salinity difference between the two boxes is zero. This constraint gives the initial AMOC strength of $\Psi_0 = k\alpha\Delta T^a$ and $F_{\text{ovS}}(\eta = 0) = 0$, where k is a hydraulic pumping coefficient, α the (dimensionless) thermal expansion coefficient, and ΔT^a the (dimensionless) atmospheric temperature difference. The northern boundary is

365

closed ($n_1 = 0$) and gyres are not represented ($g_1 = 0$) in the Stommel model. The oceanic temperatures in the Stommel model are fixed (under steady state assumption), and in this case $c_1 = 0$. Relation (13) for the Stommel model reduces to:

$$\Psi(F_H) = \frac{k\alpha\Delta T^a}{2} \pm \sqrt{\left(\frac{k\alpha\Delta T^a}{2}\right)^2 - c_2 S_0 F_H} \quad (14)$$

and is similar to relation A9, apart from some scaling coefficients.

370 4.2 Application of the reduced model

Using the reduced model, the critical value of F_H for an AMOC collapse in the CESM can be estimated by assuming that the freshwater flux forcing is (in its first order) balanced by the overturning and azonal (gyre) components, which is the case for the CESM (van Westen et al., 2024a). The critical freshwater flux forcing is obtained by setting the terms under the square root in equation (13) equal to zero. Solving this yields:

$$375 \quad F_H^c = \frac{1}{c_2(1-c_1)S_0} \left(\frac{\Psi_0^2 + c_2(1-c_1)S_0 F_{\text{ovS}}(0) + (n_1 + g_1)S_0\Psi_0}{2\Psi_0} \right)^2. \quad (15)$$

The F_H^c is dependent on the initial AMOC strength and initial F_{ovS} value. In the CESM, the Atlantic Ocean surface area outside $20^\circ\text{N} - 50^\circ\text{N}$ receives a negative freshwater flux as part of the global compensation (see inset Figure 2a). This makes the applied hosing 86% effective when considering the total Atlantic Ocean surface area ($34^\circ\text{S} - 65^\circ\text{N}$) and F_H^c needs to be adjusted by a factor $\frac{1}{0.86}$. The time-means (first 50 model years) in the CESM quasi-equilibrium simulation are $\Psi_0 = 16$ Sv and $F_{\text{ovS}}(0) = 0.22$ Sv, which give: $F_H^c = \frac{1}{0.86} 0.38 = 0.44$ Sv (Figures 8a,b). When using the maximum and minimum values (over the first 50 model years) for AMOC strength and F_{ovS} , we find $F_H^c = \frac{1}{0.86} 0.44 = 0.52$ Sv and $F_H^c = \frac{1}{0.86} 0.33 = 0.38$ Sv, respectively (Figures 8a,b).

The F_H^c determined from the reduced model is somewhat smaller (0.06 Sv for the mean) than our reference of $\overline{F_H} = 0.5$ Sv. By increasing the gyre (or northern overturning) responses, we can reduce this difference (Figure 8d). The gyre contributions also control the distance between F_H^c and value of F_H at the F_{ovS} minimum (Dijkstra, 2007; Huisman et al., 2010; Dijkstra and van Westen, 2024). For the reduced model and with standard values of the parameters n_1 and g_1 , this difference is about $\Delta F_H = 0.34 \times 10^{-2}$ Sv (Figure 8d), and decreasing with smaller g_1 (or n_1).

The actual F_{ovS} minimum in CESM is found for the statistical equilibrium of $\overline{F_H} = 0.48$ Sv (Figure 9a), whereas the F_{ovS} minimum in the quasi-equilibrium was found around $F_H = 0.52$ Sv (van Westen et al., 2024a). There is substantial overlap in the statistical properties of the four statistical equilibria closest to the tipping point, which complicates the F_{ovS} minimum assessment. Alternatively, van Westen et al. (2024a) used cubic splines to determine the F_{ovS} minimum, in which cubic polynomials are interpolated between so-called knots. For these knots, the F_{ovS} values from the four statistical equilibria can be used, but this results in spurious fits (thin curves in Figure 9a) due to the limited number of knots. To obtain an unbiased estimate of the F_{ovS} minimum, all F_{ovS} combinations of the four statistical equilibria (i.e., 6,250,000 combinations) are considered, from which the frequency of the F_{ovS} minimum per statistical equilibrium is determined. These frequencies are: 1.1% ($\overline{F_H} = 0.45$ Sv), 21.7% ($\overline{F_H} = 0.465$ Sv), 43.2% ($\overline{F_H} = 0.48$ Sv) and 34.0% ($\overline{F_H} = 0.495$ Sv), with the weighted

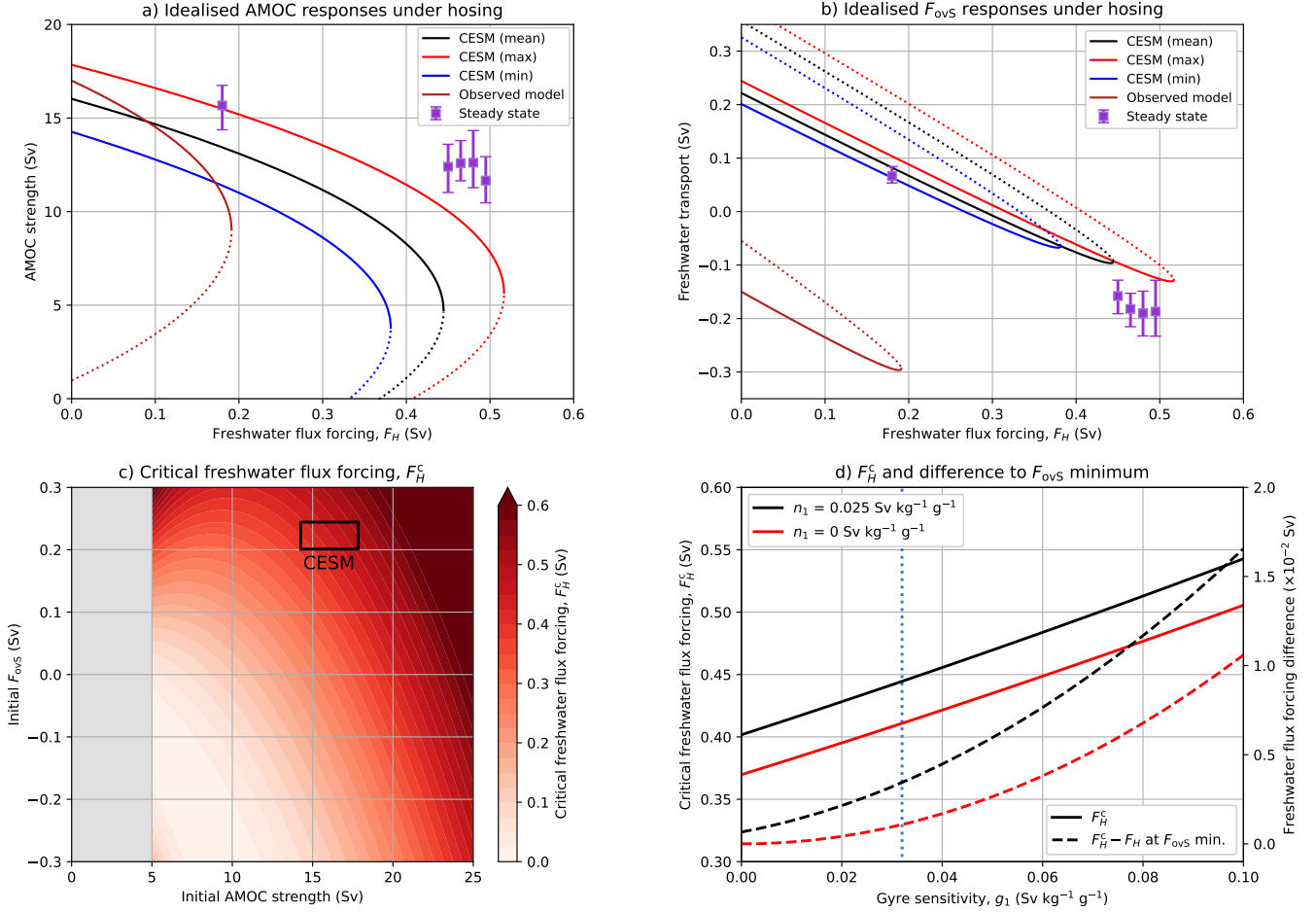


Figure 8. (a&b): The AMOC and F_{ovS} responses of the reduced model under the freshwater flux forcing (cf. equations (13) and (3), respectively), where the solid curves indicate the steady AMOC on state and dotted curves the unstable branch. The initial values for both the AMOC strength and F_{ovS} were obtained from the first 50 model years of the quasi-equilibrium. The AMOC strength values are 16.0 Sv (mean), 17.8 Sv (maximum) and 14.3 Sv (minimum), and F_{ovS} values are 0.22 Sv (mean), 0.24 Sv (maximum) and 0.20 Sv (minimum). For the ‘Observed model’, we use the reduced model in combination with observed values of 17 Sv (Smeed et al., 2018) and -0.15 Sv (Arumí-Planas et al., 2024) for the AMOC strength and F_{ovS} , respectively. (c): The critical freshwater flux forcing (F_H^c) for varying initial AMOC strength and initial F_{ovS} . The ranges for the CESM (first 50 model years of quasi-equilibrium) are indicated. The critical freshwater flux forcing was not determined for relatively weak AMOC strengths (< 5 Sv). (d): Values of F_H^c (solid curves) and difference to F_{ovS} minimum (dashed curves) for varying gyre sensitivity (g_1) and two cases for the northern overturning sensitivity (n_1), using the time-mean (first 50 model years) AMOC strength and F_{ovS} . The standard CESM values are $g_1 = 0.032$ Sv kg⁻¹ g⁻¹ (blue dotted line) and $n_1 = 0.025$ Sv kg⁻¹ g⁻¹ (black curves). For all CESM results, we consider the hosing over 20°N – 50°N (with global surface compensation), making the applied hosing 86% effective (see main text).

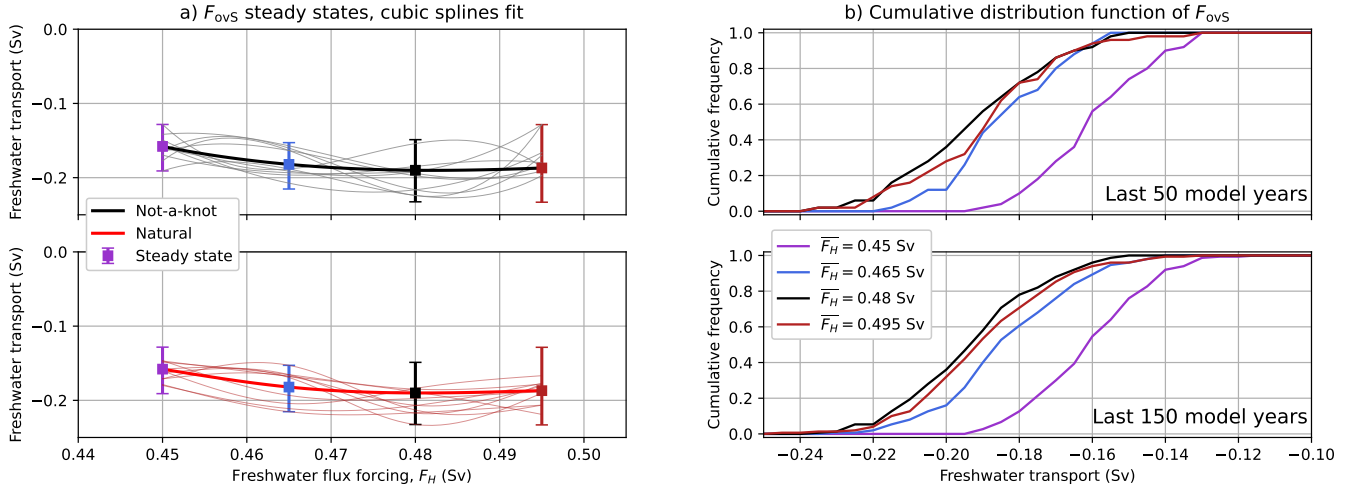


Figure 9. (a): Cubic splines fits (thin curves) using random F_{ovs} values from the four statistical equilibria. The mean over 100,000 random cubic splines are shown by the thick curves. We use the not-a-knot boundary condition (upper panel) and the natural boundary condition (lower panel). (b): The cumulative distribution function of the F_{ovs} for the four statistical equilibria, showing the last 50 model years (upper panel) and last 150 model years (lower panel).

F_{ovs} minimum at $F_H = 0.482$ Sv. This indeed confirms that the F_{ovs} minimum is most likely found for $\overline{F_H} = 0.48$ Sv, where 66% of the combinations has the minimum for $\overline{F_H} \leq 0.48$ Sv. The former is also reflected in the cumulative distribution function of F_{ovs} for the four statistical equilibria (upper panel in Figure 9b), where $\overline{F_H} = 0.48$ Sv (black curve) has the largest cumulative frequency for most F_{ovs} values. This result is robust when using a different 50-year window or the last 150 years of the equilibrium simulations (lower panel in Figure 9b). For the latter case, the F_{ovs} minimum frequencies are: 1.2% ($\overline{F_H} = 0.45$ Sv), 21.4% ($\overline{F_H} = 0.465$ Sv), 42.6% ($\overline{F_H} = 0.48$ Sv) and 34.8% ($\overline{F_H} = 0.495$ Sv) over all the combinations (i.e., 506,250,000), with the weighted F_{ovs} minimum also at $F_H = 0.482$ Sv. What is important here, is that the F_{ovs} minimum is found $\Delta F_H = 0.013$ to 0.028 Sv before the upper bound of the multi-stable regime. A similar freshwater flux forcing difference is found in a fully-implicit global ocean model (Dijkstra and van Westen, 2024), where it was shown that the F_{ovs} minimum is connected to a saddle-node bifurcation.

The overlap in the statistical properties of the four statistical equilibria closest to the tipping point also complicates the shape (i.e., square-root) estimate between AMOC strength and F_H . These four equilibria are clearly insufficient and one needs more equilibria to obtain a better estimate of the shape. This is computationally expensive for the CESM, but can easily be done for the E-CCM and also under stochastic noise. Even if more equilibria were available for the CESM, there is a possibility that the structure of multiple equilibria is much more complicated (Lohmann et al., 2024). The latter may explain the relatively strong AMOC strength for $\overline{F_H} = 0.48$ Sv, but this can not be verified from the results presented here. It is therefore more relevant to analyse the different AMOC feedback strengths over large F_H intervals, which clearly indicate a square root dependence between AMOC strength and F_H (Vanderborght et al., 2025) and this is also supported by the reduced model here.

415 Using the reduced model (with the c_1 , c_2 , g_1 and n_1 from the CESM), one can make a rough estimate of the critical freshwater flux forcing needed to collapse the present-day AMOC. For observed values, we used 17 Sv (Smeed et al., 2018) and -0.15 Sv (Arumí-Planas et al., 2024) for AMOC strength and F_{ovS} , respectively. We assume that all the Greenland Ice Sheet melt is added to the Atlantic Ocean surface, making the hosing 100% effective, and we find $F_H^c = 0.19$ Sv (Figure 8). Although this critical freshwater flux forcing is substantially smaller than the CESM, it still boils down to 25 times the present-day melt rate of the Greenland Ice Sheet (Sasgen et al., 2020). Nevertheless, what is most relevant here is that the present-day AMOC is more sensitive (i.e., relatively large $\frac{\partial \text{AMOC}}{\partial F_H}$) compared to CESM and typical CMIP6 models, as most climate models are positively biased in their F_{ovS} (Van Westen and Dijkstra, 2024; van Westen et al., 2025). In other words, the AMOC is overly stable when having positive F_{ovS} biases and underestimate the risk of AMOC tipping (Liu et al., 2017). As was argued in Vanderborght et al. (2025), the reduced model only holds under (quasi-)equilibrium conditions, making this analysis less
425 useful under transient climate change (van Westen et al., 2025).

5 Transient AMOC behavior under climate change

The existence of a saddle-node bifurcation in the E-CCM helps to understand how AMOC stability in CESM is influenced under climate change. Changes in the background climate conditions can be interpreted as a shift in the position of the saddle-node bifurcation. This can already be demonstrated in the Stommel model where the saddle-node bifurcation shifts to lower
430 freshwater flux forcing values under a smaller atmospheric temperature gradient (Figure A2).

We first analyse the CESM simulations under the Hist/RCP4.5 and Hist/RCP8.5 scenarios. The AMOC collapses in three out of the four CESM simulation under climate change (Figures 10a,b). The simulation under the higher freshwater flux forcing of $\overline{F_H} = 0.45$ Sv are closer to the tipping point (under PI conditions) and hence are more prone to undergo transitions, which is indeed the case. For $\overline{F_H} = 0.18$ Sv, only the Hist/RCP8.5 scenario shows an AMOC collapse while in the Hist/RCP4.5 scenario
435 the AMOC eventually recovers. In the latter scenario, the AMOC shows distinct centennial variability and this is associated with the typical overturning time scale (Winton and Sarachik, 1993).

The imposed transient climate change forcing induces above average surface temperature trends (compared to the global mean) at the higher latitudes (i.e., polar amplification, Figures 10c,d). This temperature response reduces the meridional (equator-to-pole) temperature gradient and may influence the multi-stable AMOC regime, as is the case for the Stommel
440 model (Figure A2). We can test this in the E-CCM by reducing the atmospheric meridional temperature gradient by imposing a (positive) atmospheric temperature anomaly (ΔT_n^a) over box n (and also over atmospheric box s as they are coupled (van Westen et al., 2024b)). We keep the atmospheric temperatures the same for boxes t and ts to limit the degrees of freedom.

The steady states (with $\xi = 0$) for the reference case ($\Delta T_n^a = 0^\circ\text{C}$) and climate change case ($\Delta T_n^a = 5^\circ\text{C}$) are shown in Figure 11a. Both saddle-node bifurcations shift to lower E_A values and the hysteresis width decreases from 0.30 Sv (reference)
445 to 0.22 Sv (climate change). This shift can be understood from the smaller meridional density difference between box n and box ts (equation (1)) due to higher temperatures and this requires a smaller freshwater flux forcing to reach the critical AMOC strength corresponding to the tipping point. The reduced meridional temperature gradient also weakens the AMOC on strength

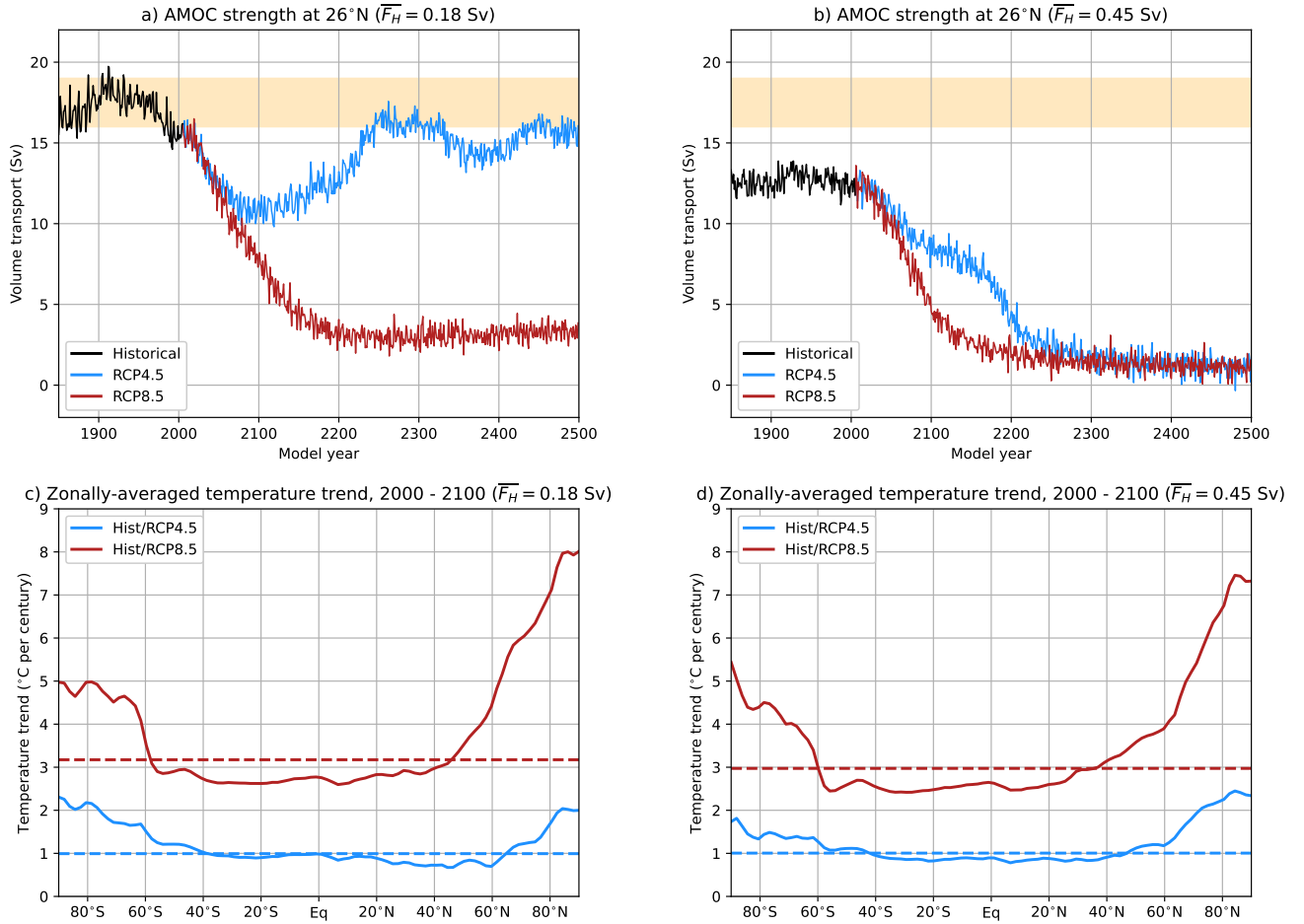


Figure 10. (a&b): The AMOC strength at 1,000 m and 26°N under the different climate change scenarios, the yellow shading indicates observed ranges (Smeed et al., 2018). (c&d): The zonally-averaged (2-meter) surface temperature trend (model year 2000 – 2100) under the different climate change scenarios. The globally-averaged temperature trend is indicated by the dashed lines.

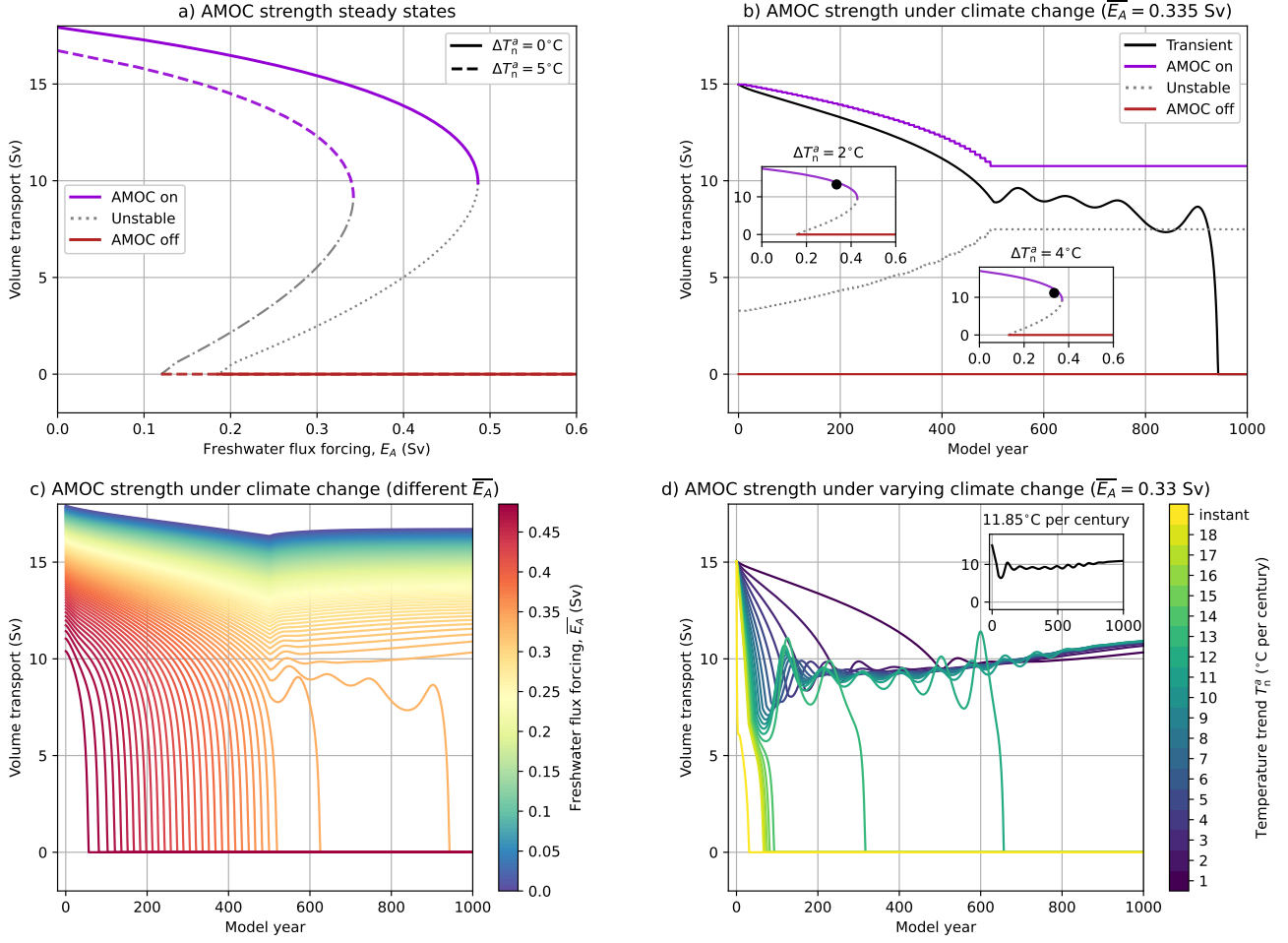


Figure 11. (a): The steady states for the AMOC strength for the standard set-up (solid curves) and under climate change (dashed curves). (b): The AMOC strength under transient climate change and $\overline{E_A} = 0.335$ Sv, where ΔT_n^a linearly increases up to 5°C up to model year 500 (trend of 1°C per century) and then remains constant. The steady states at $\overline{E_A} = 0.335$ Sv for each climate change anomaly (with an accuracy of 0.1°C) are also displayed. The insets show the steady states and the transient AMOC state (black dot) at $\Delta T_n^a = 2^\circ\text{C}$ (model year 200) and $\Delta T_n^a = 4^\circ\text{C}$ (model year 400). (c): Similar to panel b, but now for different values of $\overline{E_A}$ with $\Delta \overline{E_A} = 0.005$ Sv. (d): The transient AMOC strength under climate change and $\overline{E_A} = 0.33$ Sv, but now for varying temperature trends in ΔT_n^a . The inset shows the transient AMOC strength for a temperature trend of 11.85°C per century.

by a few Sv when comparing the two cases. The shift of the upper saddle-node bifurcation to lower E_A values indicates that the AMOC on state loses stability under climate change.

450 To study the transient climate change forcing in the E-CCM, we linearly increase T_n^a by 1°C per century up to model year 500 and then keep the temperature anomaly constant at $\Delta T_n^a = 5^\circ\text{C}$. The AMOC strength (black curve in Figure 11b) under climate change is shown for constant $\overline{E_A} = 0.335$ Sv, a similar set-up as in the CESM. For each temperature anomaly

ΔT_n^a we determined the steady states (with an accuracy of 0.1°C) and the values for the AMOC on, unstable branch and AMOC off states for $\overline{E_A} = 0.335$ Sv are also shown in Figure 11b. These steady states represent the ‘frozen’ bifurcation diagrams for a given temperature anomaly (insets in Figure 11b). The transient AMOC is clearly deviating from the AMOC on state. Up to model year 500, the AMOC gradually weakens and after a few oscillations eventually collapses in model year 900. These oscillations are related to a (sub-critical) Hopf bifurcation close to the saddle-node bifurcation. When lowering the T_n^a trend to 0.726°C per century and then keeping $\Delta T_n^a = 5^\circ\text{C}$ fixed, the AMOC strength also displays substantial oscillatory behaviour but does recover (not shown). This means that rate-induced effects are present and the AMOC collapses for T_n^a trends larger than 0.726°C per century for $\overline{E_A} = 0.335$ Sv.

When using a trend of 1°C per century for T_n^a (up to $\Delta T_n^a = 5^\circ\text{C}$) and varying $\overline{E_A}$ (Figure 11c), we always find an AMOC collapse for $\overline{E_A} > 0.342$ Sv as there are no stable AMOC on states at larger E_A values (Figure 11a). The AMOC always recovers for $\overline{E_A} \leq 0.33$ Sv, again demonstrating that rate-induced effects are present for $\overline{E_A} = 0.335$ Sv and $\overline{E_A} = 0.34$ Sv. Rate-induced effects are also present for $\overline{E_A} \leq 0.33$ Sv, however, the AMOC is much more stable compared to the previous presented case of $\overline{E_A} = 0.335$ Sv. This is also demonstrated in Figure 11d, where we vary the T_n^a temperature trend and then keeping $\Delta T_n^a = 5^\circ\text{C}$ fixed for $\overline{E_A} = 0.33$ Sv. Oscillatory behaviour becomes more pronounced when increasing the T_n^a temperature trend and the greatest AMOC weakening is found for relatively large temperature trends. For a temperature trend of 11.85°C per century (inset in Figure 11d), the AMOC strength (and other quantities) crosses the basin boundary between model years 43 and 87 and the AMOC displays oscillatory behavior. These oscillations decrease in amplitude after model year 800 and then the AMOC recovers. For larger temperature trends than 11.85°C per century the AMOC eventually collapses, which is a factor of 16 larger than the critical temperature trend of 0.726°C per century for $\overline{E_A} = 0.335$ Sv. This demonstrates that slightly lower E_A values can make the AMOC substantially more stable. It is possible to collapse the AMOC for $\overline{E_A} < 0.33$ Sv and this requires even larger climate change anomalies ($\Delta T_n^a > 5^\circ\text{C}$).

6 Summary and Discussion

The Community Earth System Model (CESM) as used here (version 1.0.5) is an extremely high-dimensional dynamical system, representing the interaction of the ocean, atmosphere, land and sea-ice processes. In a pre-industrial configuration, the AMOC collapses under a quasi-equilibrium input of freshwater in the $20^\circ\text{N} - 50^\circ\text{N}$ region, with surface freshwater compensation over the rest of the global domain (van Westen et al., 2024a).

In this paper, we have provided arguments for the case that, as in ocean-climate models lower in the model hierarchy (box models (Cessi, 1994) and fully-implicit ocean models (Dijkstra, 2007)), the AMOC collapse behavior in CESM is caused by the presence of a saddle-node bifurcation in the high-dimensional dynamical system. While one indeed would expect such a bifurcation in a deterministic dynamical system when varying a single parameter (where the saddle-node and the Hopf bifurcation are the only two generic codimension-1 bifurcations), this is far from trivial in the CESM. The ocean component of the CESM is much more complicated with several interacting positive and negative feedbacks (Vanderborcht et al., 2025) and which is forced by a rapidly varying atmosphere. So attractors of the CESM are expected to have a quite complicated

geometrical structure and transitions between those (such as between the AMOC on state and AMOC off state) could in principle be much more complicated than the traditional saddle-node bifurcation picture as suggested by conceptual models (Dijkstra, 2024).

For a saddle-node bifurcation, one would have to demonstrate a square root dependence of the AMOC strength on the freshwater forcing near the collapse point, which arises from the destabilising salt-advection feedback (Vanderborght et al., 2025). This is not feasible for the CESM due to its strong internal variability and hence our case is built using three more indirect arguments. The first argument is that in the CESM, there is a strict critical boundary of existence of the statistical steady ‘AMOC on’ state. We showed this by subsequent near-equilibrium computations near the collapse point in the quasi-equilibrium simulation, similar to the approach in Hawkins et al. (2011). Such a strict boundary is characteristic of a saddle-node bifurcation as shown for the E-CCM. The full AMOC hysteresis experiment (van Westen and Dijkstra, 2023) shows that the AMOC recovers at a much lower freshwater flux forcing ($F_H \approx 0.09$ Sv) compared to the collapse point ($0.495 \leq \overline{F_H} < 0.51$ Sv), demonstrating non-linear behaviour that is also essential to saddle-node bifurcations. The second argument is based on the CESM results with a slower freshwater forcing rate. Here, we show that the AMOC collapse precisely follows the behaviour (Ritchie et al., 2021) one would expect near a saddle-node bifurcation, i.e., with a steeper transition (in F_H space) than for the standard forcing rate. Do note that this characteristics is also found for other bifurcation types (Berglund and Gentz, 2006). The third, and probably strongest, argument relies on the assumption that overturning freshwater transport predominately compensates any freshwater flux forcing, which holds approximately for the CESM (van Westen et al., 2024a). In this case, one can show that the AMOC strength has a square-root dependence with the freshwater forcing using a reduced model (cf. section 4).

To these arguments, we can add the support from early warning indicators as found for the CESM (van Westen et al., 2024a). A characteristic property of saddle-node bifurcations is the loss of resilience (i.e., critical slowdown) near the tipping point, measured by the increase in variance and autocorrelation (van Westen et al., 2024b). Although these early warning indicators based on the AMOC strength were not giving any critical slowdown, optimal regions for early warning signal detection were found near 34°S (Smolders et al., 2025). The results presented here (cf. Figure 3) show an increase in the F_{ovS} variance close to the tipping point. This increase in variability indicates that the AMOC loses resilience, making it more prone to transitions, characteristic of approaching a saddle-node bifurcation (van Westen et al., 2024b).

The implications of this result are substantial. First of all, it shows that, for the AMOC tipping problem, conceptual models that capture only the dominant feedbacks are useful (Dijkstra, 2024). For example, in the E-CCM only the salt-advection feedback and gyre feedback are captured which are also dominant in CESM and hence it is relatively easy to tune the behavior of the E-CCM to the CESM. Similarly, Wood et al. (2019) tuned a box model (only representing the salt-advection feedback) to the FAMOUS (Hawkins et al., 2011) where likely due to its low resolution the gyre feedback is relatively weak. Sensitivity studies in the conceptual model can then be used to design useful simulations in the complex model and also physical explanations can be sought in the reduced model. Second, if the multi-stable regime of the AMOC is bounded by saddle-node bifurcations, then the effect of model biases can be studied in terms of shifts of the saddle-node bifurcations. In fully-implicit ocean models, it was recently shown that a bias in Indian Ocean precipitation leads to a right shift (i.e., to higher Atlantic freshwater flux

forcing strengths) of the bifurcation diagram (Dijkstra and van Westen, 2024; Boot and Dijkstra, 2025). Our reduced model (cf. Section 4.2) also shows that positive freshwater transport biases at 34°S make the AMOC more stable under hosing. If indeed a saddle-node bifurcation is present in all global climate models (GCMs), this would indicate that GCMs having such a bias would be too stable (Van Westen and Dijkstra, 2024; van Westen et al., 2025).

525 So far, the saddle-node bifurcation was discussed only in the case of an AMOC collapse when changing the freshwater flux forcing. However, under climate change mainly the heat flux forcing will change and not in a quasi-equilibrium way. Also in this case, we have shown that the existence of the saddle-node bifurcation is an important aspect to explain the transient behavior of the CESM. Climate change modifies the atmospheric meridional temperature gradient and shifts the saddle-node bifurcation to lower freshwater flux forcings, making the ‘AMOC on’ state less resilient. This was shown in greater detail
530 by the idealized results of the E-CCM, the collapse behavior can be viewed as crossing a moving saddle-node bifurcation in time (Ritchie et al., 2021). Rate-induced effects are also highly relevant under climate change (Hankel, 2025), with the strongest evidence for rate-induced tipping when comparing the RCP4.5 (AMOC recovery) and RCP8.5 (AMOC collapse) and $\overline{F_H} = 0.18$ Sv. Although the AMOC collapses for both the RCP4.5 and RCP8.5 under $\overline{F_H} = 0.45$ Sv, which suggests a moving saddle-node bifurcation under climate change, rate-induced effects cannot be dismissed and to test this we need to
535 conduct more climate change forcing experiments, this is out of the scope of this paper. Note that the E-CCM is limited in representing other (non-linear) climate change feedbacks, such as enhanced evaporation (due to higher temperatures) which could partly stabilize the AMOC (van Westen et al., 2025).

Finally, as the phase space of the CESM is so high-dimensional, why would a saddle-node bifurcation appear in such a model (as there are many instabilities)? This result can be possibly explained by looking at the Lorenz84-Stommel1961 model
540 or the PlaSim sea-ice model (Tantet et al., 2018), which both display chaotic behavior, but also show a large-scale transition under variation of one parameter. Here, the chaotic behavior is only in the atmosphere component and the large-scale transition dynamics is governed only by the slow component, which is then noise-forced. While in the total phase space, this may be a crisis bifurcation, in the reduced phase space of the slow component, this would appear then as a saddle-node bifurcation. However, more work is needed to make this more precise.

545 *Code and data availability.* All processed model output and Python scripts to generate the results are available at:
<https://doi.org/10.5281/zenodo.17123475>

Appendix A: The Analytical Solutions of the Stommel Box Model

The Stommel 2-box model (Stommel, 1961) consists of two well-mixed boxes (equal volume) and the boxes exchange water mass properties over time (Figure A1). The circulation strength, ψ , is set by the density difference between the high-latitude
550 (T_1, S_1) and equatorial box (T_2, S_2) :

$$\psi = k(\rho_1 - \rho_2) \tag{A1}$$

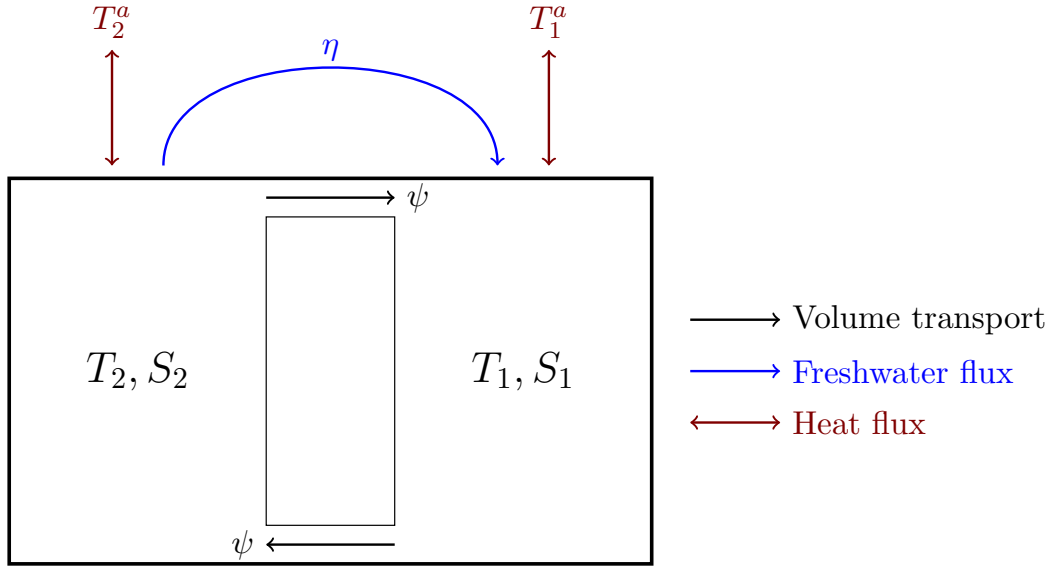


Figure A1. Schematic representation of the Stommel 2-box model in its northward overturning state with AMOC strength ψ . The blue and brown arrows are freshwater and heat fluxes, respectively. The hosing is directed from the equatorial box (with T_2, S_2) to the high-latitude box (with T_1, S_1).

where k is a hydraulic pumping constant. A linear equation of state ($\rho = \rho_0 - \alpha(T - T_0) + \beta(S - S_0)$) yields:

$$\psi = k(\alpha\Delta T - \beta\Delta S) \quad (\text{A2})$$

where $\Delta T = T_2 - T_1$ and $\Delta S = S_2 - S_1$. The governing (dimensionless) differential equation for the Stommel model are then

555 given by:

$$\frac{dT_1}{dt} = |\psi|\Delta T + \lambda_T(T_1^a - T_1) \quad (\text{A3})$$

$$\frac{dT_2}{dt} = -|\psi|\Delta T + \lambda_T(T_2^a - T_2) \quad (\text{A4})$$

$$\frac{dS_1}{dt} = |\psi|\Delta S - \eta \quad (\text{A5})$$

$$\frac{dS_2}{dt} = -|\psi|\Delta S + \eta \quad (\text{A6})$$

560 In these relations λ_T is the thermal exchange coefficient with the overhead atmosphere, the atmospheric temperatures are fixed.

Under the assumption that the thermal exchange with the atmosphere is much faster than the thermal exchange between the boxes ($\psi\Delta T \ll \lambda_T(T_i^a - T_i)$, with $i = 1, 2$), the steady state for the temperatures has $T_1 = T_1^a$ and $T_2 = T_2^a$. Using this steady state assumption, the time-evolution equation of the circulation strength (from A2 and A3 – A6) reduces to:

$$565 \quad \frac{d\psi}{dt} = -k\beta\frac{d\Delta S}{dt} = -k\beta\left(\frac{dS_2}{dt} - \frac{dS_1}{dt}\right) = 2k\beta(|\psi|\Delta S - \eta) \quad (\text{A7})$$

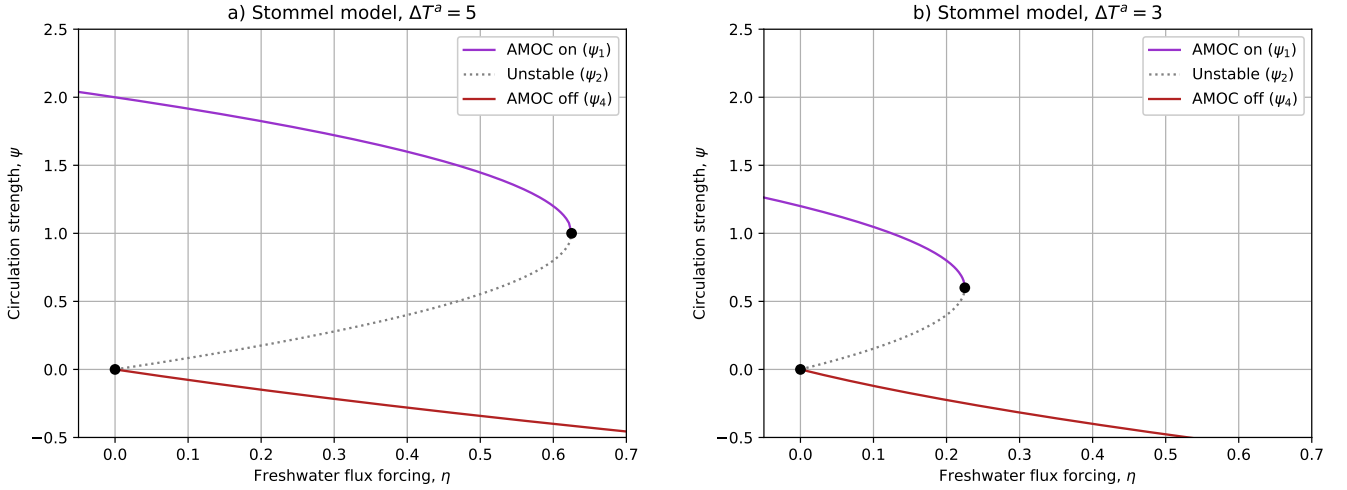


Figure A2. Bifurcation diagram for the Stommel 2-box model, where the black dots indicate saddle-node bifurcations. The atmospheric temperature differences are (a): $\Delta T^a = 5$ and (b): $\Delta T^a = 3$. For the other dimensionless coefficients, we used: $\alpha = 2 \times 10^{-4}$, $\beta = 8 \times 10^{-4}$ and $k = 2 \times 10^3$.

where the temperature contribution vanishes as the atmospheric temperatures are constant ($\frac{d\Delta T}{dt} = \frac{d\Delta T^a}{dt} = 0$). The final step is to substitute $\Delta S = \frac{k\alpha\Delta T^a - \psi}{k\beta}$ from (A2) to obtain:

$$\frac{d\psi}{dt} = -2|\psi|\psi + 2k\alpha\Delta T^a|\psi| - 2k\beta\eta \quad (\text{A8})$$

The steady states ($\frac{d\psi}{dt} = 0$) with northward overturning ($\psi > 0$) are given by:

$$\psi_{1,2} = \frac{k\alpha\Delta T^a}{2} \pm \sqrt{\left(\frac{k\alpha\Delta T^a}{2}\right)^2 - k\beta\eta} \quad (\text{A9})$$

For the reversed circulation ($\psi < 0$), these are:

$$\psi_{3,4} = \frac{k\alpha\Delta T^a}{2} \pm \sqrt{\left(\frac{k\alpha\Delta T^a}{2}\right)^2 + k\beta\eta} \quad (\text{A10})$$

but note that ψ_3 has to be rejected since $\psi_3 \not\leq 0$. The stable AMOC on state is given by ψ_1 , the stable AMOC off state by ψ_4 , and the unstable state by ψ_2 . The (dimensionless) solutions for two different atmospheric temperature differences are shown in

Figure A2.

Appendix B: The Normal Form of the Saddle Node Bifurcation

For the Stommel model, the dynamics of the AMOC strength in the AMOC on state is given by:

$$\frac{d\psi}{dt} = -2\psi^2 + 2k\alpha\Delta T^a\psi - 2k\beta\eta, \quad (\text{B1})$$

which can be generalised for the saddle-node bifurcation to:

$$580 \quad \frac{d\psi}{dt} = A\psi^2 + B\psi + C - Dt. \quad (\text{B2})$$

where A, B, C and D are constants, and the freshwater flux forcing is now varied linearly with time (i.e., $\eta(t) = Dt$). This generalised form also holds for the reduced model (Section 4.1).

Relation (B2) is rewritten as:

$$\frac{d\psi}{dt} = A \left(\psi + \frac{B}{2A} \right)^2 + \left(C - \frac{B^2}{4A} \right) - Dt. \quad (\text{B3})$$

585 and we follow the procedure outlined in Faure Ragani and Dijkstra (2025), where time t is considered as a parameter and the saddle-node bifurcation can be found by setting the last two terms on the right hand side of (B3) to zero. Solving for t yields:

$$t^{\text{SN}} = \frac{C}{D} - \frac{B^2}{4AD}. \quad (\text{B4})$$

To obtain the normal form, we apply a rescaling of the variables:

$$x = -At^{\text{SN}} \left(\psi + \frac{B}{2A} \right) \quad \text{and} \quad \tau = \frac{t}{t^{\text{SN}}} \quad (\text{B5})$$

590 and the dynamics of the AMOC in the rescaled variables are:

$$\frac{dx}{d\tau} = \frac{dx}{dt} \frac{dt}{d\tau} = -At^{\text{SN}} \frac{d\psi}{dt} t^{\text{SN}} = -A(t^{\text{SN}})^2 \left(A \left(\psi + \frac{B}{2A} \right)^2 + \left(C - \frac{B^2}{4A} \right) - Dt \right). \quad (\text{B6})$$

Now using (B4) and (B5) to find the normal form of:

$$\frac{dx}{d\tau} = -A(t^{\text{SN}})^2 \left(A \frac{x^2}{(-At^{\text{SN}})^2} + Dt^{\text{SN}} - Dt^{\text{SN}}\tau \right) = r - x^2 \quad (\text{B7})$$

where $r = -AD(t^{\text{SN}})^3(1 - \tau)$. Note that $r > 0$ for $\tau < 1$ as $A < 0$ and $D > 0$.

595 The non-autonomous system (B7) can be solved analytically (Li et al., 2019) and it was shown that the collapse time $t^* = 1 + 2.333\alpha^{-1/3}$, where $\alpha = -AD(t^{\text{SN}})^3 > 0$. If the forcing value at which the collapse occurs for a rate D is indicated by $\gamma_f = Dt^*$, then for the collapse forcing (γ_s) at half rate $D/2$, we find that $\alpha_s = 4\alpha_f$ and hence $\gamma_s = D(1 + 2.333\alpha_s^{-1/3}) = D(1 + 0.177\alpha_f^{-1/3}) < \gamma_f$. Hence, the transition occurs at lower forcing strength (and faster) when the rate is lower (see also Figure 3b and Figure 4 in Li et al. (2019)).

600 *Author contributions.* R.M.v.W., E.V. and H.A.D. conceived the idea for this study. R.M.v.W. conducted the analysis and prepared all figures, E.V. contributed to the AMOC feedback decomposition. All authors were actively involved in the interpretation of the analysis results and the writing process.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. We thank Michael Kliphuis (IMAU, UU) for performing the additional CESM simulations. The model simulations and
605 the analysis of all the model output was conducted on the Dutch National Supercomputer Snellius within NWO-SURF project 2024.013.
R.M.v.W., E.V. and H.A.D. are funded by the European Research Council through the ERC-AdG project TAOC (PI: Dijkstra, project 101055096).

References

- Armstrong McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton, T. M.: Exceeding 1.5 C global warming could trigger multiple climate tipping points, *Science*, 377, eabn7950, 2022.
- 610 Arumí-Planas, C., Dong, S., Perez, R., Harrison, M. J., Farneti, R., and Hernández-Guerra, A.: A multi-data set analysis of the freshwater transport by the Atlantic meridional overturning circulation at nominally 34.5 S, *Journal of Geophysical Research: Oceans*, 129, e2023JC020 558, 2024.
- Baatsen, M., Von Der Heydt, A. S., Huber, M., Kliphuis, M. A., Bijl, P. K., Sluijs, A., and Dijkstra, H. A.: The middle to late Eocene greenhouse climate modelled using the CESM 1.0.5, *Climate of the Past*, 16, 2573–2597, 2020.
- 615 Berglund, N. and Gentz, B.: Noise-induced phenomena in slow-fast dynamical systems: a sample-paths approach, Springer, 2006.
- Bonan, D. B., Thompson, A. F., Newsom, E. R., Sun, S., and Rugenstein, M.: Transient and equilibrium responses of the Atlantic overturning circulation to warming in coupled climate models: The role of temperature and salinity, *Journal of Climate*, 35, 5173–5193, 2022.
- Boot, A. A. and Dijkstra, H. A.: Physics of AMOC multistable regime shifts due to freshwater biases in an EMIC, *Earth System Dynamics*, 16, 1221–1235, 2025.
- 620 Butler, E., Oliver, K., Hirschi, J. J.-M., and Mecking, J.: Reconstructing global overturning from meridional density gradients, *Climate Dynamics*, 46, 2593–2610, 2016.
- Castellana, D., Baars, S., Wubs, F. W., and Dijkstra, H. A.: Transition probabilities of noise-induced transitions of the Atlantic Ocean circulation, *Scientific Reports*, 9, 20 284, 2019.
- Cessi, P.: A simple box model of stochastically forced thermohaline flow, *Journal of Physical Oceanography*, 24, 1911–1920, 1994.
- 625 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.
- Cini, M., Zappa, G., Ragone, F., and Corti, S.: Simulating AMOC tipping driven by internal climate variability with a rare event algorithm, *npj Climate and Atmospheric Science*, 7, 31, 2024.
- De Niet, A., Wubs, F., van Scheltinga, A. T., and Dijkstra, H. A.: A tailored solver for bifurcation analysis of ocean-climate models, *Journal*
- 630 *of Computational Physics*, 227, 654–679, 2007.
- Dijkstra, H. A.: Characterization of the multiple equilibria regime in a global ocean model, *Tellus A: Dynamic Meteorology and Oceanography*, 59, 695–705, 2007.
- Dijkstra, H. A.: The role of conceptual models in climate research, *Physica D: Nonlinear Phenomena*, 457, 133 984, 2024.
- Dijkstra, H. A. and van Westen, R. M.: The Effect of Indian Ocean Surface Freshwater Flux Biases On the Multi-Stable Regime of the
- 635 AMOC, *Tellus A: Dynamic Meteorology and Oceanography*, 76, 90–100, <https://doi.org/https://doi.org/10.16993/tellusa.3246>, 2024.
- Doedel, E. J., Paffenroth, R. C., Champneys, A. C., Fairgrieve, T. F., Kuznetsov, Y. A., Oldeman, B. E., Sandstede, B., and Wang, X. J.: AUTO-07p: Continuation and Bifurcation Software for Ordinary Differential Equations, 2007.
- Doedel, E. J., Paffenroth, R. C., Champneys, A. C., Fairgrieve, T. F., Kuznetsov, Y. A., Oldeman, B. E., Sandstede, B., and Wang, X. J.: auto-07p, <https://github.com/auto-07p/auto-07p>, 2021.
- 640 Drijfhout, S., Angevaere, J., Mecking, J., van Westen, R., and Rahmstorf, S.: Shutdown of northern Atlantic overturning after 2100 following deep mixing collapse in CMIP6 projections, *Environmental Research Letters*, 2025.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, 2016.

Faure Ragani, A. and Dijkstra, H. A.: Physical characterization of the boundary separating safe and unsafe AMOC overshoot behavior, *Earth System Dynamics*, 16, 1287–1301, 2025.

Gérard, J. and Crucifix, M.: Diagnosing the causes of AMOC slowdown in a coupled model: a cautionary tale, *Earth System Dynamics*, 15, 293–306, 2024.

Hankel, C.: The effect of CO₂ ramping rate on the transient weakening of the Atlantic Meridional Overturning Circulation, *Proceedings of the National Academy of Sciences*, 122, e2411357 121, 2025.

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, L10 605, 2011.

Hu, A., Meehl, G. A., Han, W., Timmermann, A., Otto-Bliesner, B., Liu, Z., Washington, W. M., Large, W., Abe-Ouchi, A., Kimoto, M., Lambeck, K., and Wu, B.: Role of the Bering Strait on the hysteresis of the ocean conveyor belt circulation and glacial climate stability, *Proceedings of the National Academy of Sciences*, 109, 6417–6422, <https://doi.org/10.1073/pnas.1116014109>, 2012.

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.

Jackson, L. C., Alastrué de Asenjo, E., Bellomo, K., Danabasoglu, G., Haak, H., Hu, A., Jungclaus, J., Lee, W., Meccia, V. L., Saenko, O., et al.: Understanding AMOC stability: the North Atlantic hosing model intercomparison project, *Geoscientific Model Development*, 16, 1–32, 2023.

Kuehn, C.: A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics, *Physica D-Nonlinear Phenomena*, 240, 1020–1035, 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 of the United States of America*, 105, 1786–93, <https://doi.org/10.1073/pnas.0705414105>, 2008.

Li, J. H., Ye, F. X.-F., Qian, H., and Huang, S.: Time-dependent saddle–node bifurcation: Breaking time and the point of no return in a non-autonomous model of critical transitions, *Physica D: Nonlinear Phenomena*, 395, 7–14, <https://doi.org/10.1016/j.physd.2019.02.005>, 2019.

Liu, W., Xie, S.-P., Liu, Z., and Zhu, J.: Overlooked possibility of a collapsed Atlantic Meridional Overturning Circulation in warming climate, *Science Advances*, 3, e1601 666, 2017.

Lohmann, J., Dijkstra, H. A., Jochum, M., Lucarini, V., and Ditlevsen, P. D.: Multistability and intermediate tipping of the Atlantic Ocean circulation, *Science advances*, 10, eadi4253, 2024.

Ma, Q., Shi, X., Scholz, P., Sidorenko, D., Lohmann, G., and Ionita, M.: Revisiting climate impacts of an AMOC slowdown: dependence on freshwater locations in the North Atlantic, *Science Advances*, 10, eadr3243, 2024.

Marotzke, J.: Abrupt climate change and thermohaline circulation: Mechanisms and Predictability, *Proc. Natl. Acad. Sci.*, 97, 1347–1350, 2000.

Mudelsee, M., T, B., Lear, C. H., and Lohmann, G.: Cenozoic climate changes: A review based on time series analysis of marine benthic ¹⁸O records, *Reviews of Geophysics*, 52, 333–374, 2014.

Mulder, T. E., Goelzer, H., Wubs, F. W., and Dijkstra, H. A.: Snowball Earth Bifurcations in a Fully-Implicit Earth System Model, *International Journal of Bifurcation and Chaos*, 31, 2130 017, 2021.

- Rahmstorf, S.: Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle, *Nature*, 378, 145–149, 1995.
- 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., March, R., Mysak, L., Wang, Z., and
685 Weaver, A. J.: Thermohaline circulation hysteresis: a model intercomparison, *Geophysical Research Letters*, L23605, 1–5, 2005.
- Ritchie, P. D. L., Clarke, J. J., Cox, P. M., and Huntingford, C.: Overshooting tipping point thresholds in a changing climate, *Nature*, 592, 517–523, <https://doi.org/10.1038/s41586-021-03263-2>, 2021.
- Sasgen, I., Wouters, B., Gardner, A. S., King, M. D., Tedesco, M., Landerer, F. W., Dahle, C., Save, H., and Fettweis, X.: Return to rapid ice loss in Greenland and record loss in 2019 detected by the GRACE-FO satellites, *Communications Earth & Environment*, 1, 1–8, 2020.
- 690 Sijp, W. P.: Characterising meridional overturning bistability using a minimal set of state variables, *Climate dynamics*, 39, 2127–2142, 2012.
- Smeed, D. A., Josey, S., Beaulieu, C., Johns, W., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L., et al.: The North Atlantic Ocean is in a state of reduced overturning, *Geophysical Research Letters*, 45, 1527–1533, 2018.
- Smolders, E. J. V., van Westen, R. M., and Dijkstra, H. A.: Optimal Observation Locations for Early Warning of the Onset of an AMOC Collapse, Under review in *Geophysical Research Letters*, <https://arxiv.org/abs/2406.11738>, 2025.
- 695 Stommel, H.: Thermohaline convection with two stable regimes of flow, *Tellus*, 13, 224–230, 1961.
- Tantet, A., Lucarini, V., Lunkeit, F., and Dijkstra, H. A.: Crisis of the chaotic attractor of a climate model: a transfer operator approach, *Nonlinearity*, 31, 2221–2251, <https://doi.org/10.1088/1361-6544/aaaf42>, 2018.
- Toom, M. D., Dijkstra, H. a., Cimadoribus, A. a., and Drijfhout, S. S.: Effect of Atmospheric Feedbacks on the Stability of the Atlantic Meridional Overturning Circulation, *Journal of Climate*, 25, 4081–4096, <https://doi.org/10.1175/JCLI-D-11-00467.1>, 2012.
- 700 van Westen, R. M. and Baatsen, M. L.: European temperature extremes under different AMOC scenarios in the community Earth system model, *Geophysical Research Letters*, 52, e2025GL114 611, 2025.
- van Westen, R. M. and Dijkstra, H. A.: Asymmetry of AMOC Hysteresis in a State-of-the-Art Global Climate Model, *Geophysical Research Letters*, 50, e2023GL106 088, 2023.
- Van Westen, R. M. and Dijkstra, H. A.: Persistent climate model biases in the Atlantic Ocean’s freshwater transport, *Ocean Science*, 20, 549–567, 2024.
- 705 van Westen, R. M., Kliphuis, M., and Dijkstra, H. A.: Physics-based early warning signal shows that AMOC is on tipping course, *Science advances*, 10, eadk1189, 2024a.
- van Westen, R. M., Jacques-Dumas, V., Boot, A. A., and Dijkstra, H. A.: The Role of Sea-ice Insulation Effects on the Probability of AMOC Transitions, *Journal of Climate*, 2024b.
- 710 van Westen, R. M., Vanderborght, E., Kliphuis, M., and Dijkstra, H. A.: Physics-based Indicators for the Onset of an AMOC Collapse under Climate Change, in press for *Journal of Geophysical Research: Oceans*, 2025.
- Vanderborght, E., van Westen, R. M., and Dijkstra, H. A.: Feedback Processes causing an AMOC Collapse in the Community Earth System Model, *Journal of Climate*, 2025.
- Weijer, W., Cheng, W., Garuba, O. A., Hu, A., and Nadiga, B. T.: CMIP6 Models Predict Significant 21st Century Decline of the Atlantic Meridional Overturning Circulation, *Geophysical Research Letters*, 47, e2019GL08 607, <https://doi.org/10.1029/2019gl086075>, 2020.
- 715 Winton, M. and Sarachik, E. S.: Thermohaline oscillations induced by strong steady salinity forcing of ocean general circulation models, *Journal Of Physical Oceanography*, 23, 1389–1410, 1993.

- Wood, R. A., Rodríguez, J. M., Smith, R. S., Jackson, L. C., and Hawkins, E.: Observable, low-order dynamical controls on thresholds of the Atlantic Meridional Overturning Circulation, *Climate Dynamics*, 53, 6815–6834, 2019.
- 720 Wubs, F. W. and Dijkstra, H. A.: *Bifurcation Analysis of Fluid Flows*, Cambridge University Press, 2023.