Tipping points in overturning circulation mediated by ocean mixing and the configuration and magnitude of the hydrological cycle: A simple model

Anand Gnanadesikan,[a] Gianluca Fabiani,[b] Jingwen Liu,[a] Renske Gelderloos, [a,c] G. Jay Brett [d] ,Yannis Kevrekidis, [e] Thomas Haine, [a] Marie-Aude Pradal, [a] Constaninos Siettos,[f] and Jennifer Sleeman, [d]

[a] *Morton K. Blaustein Department of Earth and Planetary Science, Johns Hopkins University, Baltimore, MD, USA*

[b] *Modeling Engineering Risk and Complexity, Scuola Superiore Meridionale, Naples, Italy*

[c] *Department of Hydraulic Engineering, Delft University of Technology, Delft, The Netherlands*

[d] *Johns Hopkins Applied Physics Lab, Laurel, MD, USA*

[e] *Department of Chemical Engineering, Johns Hopkins University, Baltimore, MD, USA*

[f] *Department of Applied Mathematics, University of Naples Federico II, Naples, Italy*

Corresponding author: Anand Gnanadesikan,gnanades@jhu.edu

ABSTRACT: In the modern ocean, transformation of light surface waters to dense deep waters primarily occurs in the Atlantic basin rather than in the North Pacific or Southern Oceans. The reasons for this remain unclear, as both models and paleoclimatic observations suggest that sinking can sometimes occur in the Pacific. We present a six-box model of the overturning that combines insights from a number of previous studies. A key determinant of the overturning configuration in our model is whether the Antarctic Intermediate Waters are denser than northern subpolar waters, something that depends on the magnitude and configuration of atmospheric freshwater transport. For the modern ocean, we find that although the interbasin atmospheric freshwater flux suppresses Pacific sinking, the poleward atmospheric freshwater flux out of the subtropics enhances it. When atmospheric temperatures are held fixed, North Pacific overturning can strengthen with either increases or decreases in the hydrological cycle, as well as under reversal of the interbasin freshwater flux. Tipping-point behavior, where small changes in the hydrological cycle may cause the dominant location of densification of light waters to switch between basins and the magnitude of overturning within a basin to exhibit large jumps, is seen in both transient and equilibrium states. This behavior is modulated by parameters such as the poorly constrained lateral diffusive mixing coefficient. If hydrological cycle amplitude is varied consistently with global temperature, northern polar amplification is necessary for the Atlantic overturning to collapse. Certain qualitative insights incorporated in the model can be validated using a fully-coupled climate model.

SIGNIFICANCE STATEMENT: Currently, the global overturning circulation involves conversion of waters lighter than Antarctic Intermediate Water to deep waters denser than Antarctic Intermediate Water primarily~~light water to dense water~~ in the North Atlantic, rather than in the North Pacific or ~~Southern Ocean waters in the~~ Southern Oceans. Many different factors have been invoked to explain this configuration, with atmospheric freshwater transport, basin geometry, lateral mixing, and Southern Ocean winds playing major roles. This paper develops a simple theory that combines previous theories, presents the intriguing idea that alternate configurations might be possible, and identifies multiple possible tipping points between these states.

1. Introduction

The fact that the transformation of light surface waters to dense deep waters is dominated by processes in the North Atlantic basin has profound implications for the ocean's physical and biogeochemical structure (Gnanadesikan 1999; Marinov et al. 2006), as well as for global climate. While some of the cold, dense water that rises to the surface in the Southern Ocean cools further and sinks to form the Antarctic Bottom Water, some of it moves northward and is freshened and warmed as it is transformed into lighter Antarctic Intermediate (AAIW) and Subantarctic Mode Water (SAMW) (Lumpkin and Speer 2007). Additionally, downward diffusion of heat transforms some dense deep water into lighter surface waters. These processes are balanced by a sinking of North Atlantic Deep Water (NADW) in the North Atlantic. This meridionally asymmetric pattern is associated with a cross-equatorial heat transport (Trenberth et al. 2019), so that a given northern latitude is usually several degrees warmer than its southern counterpart. This asymmetry in temperatures also helps~~It also help~~ to keep the intertropical convergence zone and associated rainfall north of the equator (Zhang and Delworth 2005).

Inverse models constrained with transient tracers (DeVries and Primeau 2011) suggest that fully 65% of the water away from the surface mixed layer will first come into contact with the atmosphere within the surface layers of the Southern Ocean. A key driver of the dominance of this region is that the westerly winds within the unblocked latitudes of Drake Passage generate a net northward surface flow of water. As noted by a number of authors (Toggweiler and Samuels 1993, 1995; Gnanadesikan 1999), this water cannot be supplied via a western boundary current carrying light subtropical water poleward along a continental boundary. Instead, it must be upwelled from greater depths. Some fraction of this rising water is supplied by dense water flowing southward below the depth of ridges, while some is supplied by boluses of lighter low-latitude waters associated with mesoscale eddies (Johnson and Bryden 1989; Hallberg and Gnanadesikan 2001; Klinger and Haine 2019).

In ocean-only models (Fučkar and Vallis 2007; Johnson et al. 2007; Wolfe and Cessi 2011), the structure of the zonally averaged overturning circulation has been shown to be highly dependent on whether surface fluxes can make the Southern Ocean intermediate waters advected northward at the tip of Drake Passage lighter than NADW. If the nominal NADW becomes lighter than the nominal Antarctic Intermediate Water, then AAIW densities will not outcrop in the Northern Hemisphere. Given the low levels of diapycnal mixing observed away from the mixed layer, in such situations the net watermass transformation in the Southern Ocean must be small. If a steady-state is to be achieved throughout the ocean, the bulk of the northward flux of lighter waters in the Southern Ocean associated with northward Ekman transport must then be balanced by a southward flux of waters of similar density. This flux would be associated with some combination of mesoscale and stationary eddies (Johnson and Bryden 1989; Hallberg and Gnanadesikan 2001).

In addition to the north-south asymmetry in the overturning, there is also an interbasin asymmetry, whereby no counterpart to the NADW is formed in the Pacific. In a seminal paper Warren (1983) discussed reasons for this asymmetry, arguing that two factors play an important role in producing a relatively fresh surface. First, he claimed that the subpolar North Pacific receives a greater net air-sea flux of freshwater than the subpolar North Atlantic. Second, he noted that the relatively weak North Pacific overturning will be much less effective at removing this freshwater than the more vigorous Atlantic overturning. These two mechanisms, asymmetry in freshwater delivery and preferential flushing of the Atlantic, remain the two leading processes discussed in the literature today (Ferreira et al. 2018; Johnson et al. 2019).

In idealized coupled models, these two processes work together to localize overturning in the Atlantic. In models with a wide basin (representing the Pacific) and a narrow basin (representing the Atlantic), dense water formation tends to occur preferentially in the narrow basin. There are at least two reasons for this. In idealized energy moisture balance models without a dynamical atmosphere (De Boer et al. 2008; Jones and Cessi 2016, 2017; Youngs et al. 2020)~~Assuming~~ the atmospheric freshwater transport F_w^{basin} between the subtropical and subpolar gyres scales as the basin width L_x^{nbasin}. In more dynamically sophisticated models such as Ferreira et al. (2010) the mid-latitude storm track penetrates from the short to the long basin but not from the long to the short basin. In both cases the subpolar gyre of a wide basin will receive more freshwater than a narrow basin so that the overturning has to remove more freshwater. If the overturning circulation within a basin has magnitude M_{basin}, the salinity difference between high and low latitudes within that basin will scale as $\Delta S_{basin} \approx -(F_w^{basin}/M_{basin})S_0$ where S_0 is some average salinity. Thus, given two idealized basins with equivalent initial overturning, the wider basin with a larger freshwater flux $F_w^{wide} > F_w^{narrow}$ will exhibit a larger salinity difference ($\Delta S_{wide} > \Delta S_{narrow}$). But since the impact of this gradient on density is to make polar waters lighter with respect to the tropics, it will produce a weaker poleward density gradient. The weaker gradient will retard the overturning in the wider basin relative to the narrow basin. As less light water is transformed to dense water in the wide basin, the sea surface within the wider basin will stand higher and pump ~~tropical~~ water into the narrow basin, which will in general be saltier than water entering the narrow basin from the Southern Ocean. This interbasin transport as well as the additional salt that it carries then reinforces the overturning in the narrow basin particularly if the eastern boundary of this basin does not extend as far to the south (Nilsson et al. 2013; Jones and Cessi 2017). Note that the energy-balance models cited above do not resolve the high topography of the Rockies and Andes (which block the eastward mid-latitude transport of water vapor from the Pacific to the Atlantic) or the east African highlands (which block westward tropical transport of water vapor from the Indian to the Atlantic).

Insofar as the hydrological cycle is responsible for localizing of the overturning to the North Atlantic, we would expect Pacific overturning only when conditions are cooler, as the amplitude of the hydrological cycle tends to track global temperatures. However, there are a number of lines of evidence that suggest that this picture may be incomplete. First, both idealized models and realistic coupled models can generate some overturning in the Pacific given modern conditions (De Boer et al. 2010; Bahl et al. 2019). As noted in the latter paper, the overturning is highly sensitive to the lateral diffusion coefficient A_{Redi} associated with mesoscale eddies, whose spatial structure is poorly understood (Abernathey et al. 2022). Second, paleoceanographic evidence suggests that there have been times in the past when there was more deep water formation in the North Pacific (Rae et al. 2014; Burls et al. 2017; Ford et al. 2022), as there is evidence of lower levels of chemicals produced by the decomposition of organic material. This includes cold periods such as the Younger Dryas when the climate was colder and freshwater transport was presumably weaker than today, both between basins and from the subtropical to subpolar gyres. However it also includes the relatively warm Pliocene, during which the freshwater transports were likely even stronger. Third, although the modern atmosphere does appear to transport less freshwater from the low latitudes to the high latitudes between 40°S and 65°N in the Atlantic relative to the Pacific (as summarized in Ferreira et al. 2018), it also deposits significant freshwater flux in the Arctic. As we will argue below, this re-enters the global ocean in the Atlantic and *reduces* the extent to which the northern subpolar Atlantic is saltier and denser than the subpolar Pacific.

This paper seeks insight into the dynamics of how the configuration and magnitude of the hydrological cycle control the configuration and magnitude of the overturning using a relatively simple dynamical box model. Such approaches have a long history of giving insight into the dynamics and sensitivity of the overturning (Johnson et al. 2019). For example, the pioneering paper of Stommel (1961) showed that the opposing effects of high latitude cooling and poleward atmospheric freshwater transport could combine to produce a bistable overturning describeable by two fold bifurcations. Tziperman et al. (1994) showed that this mechanism could also operate in a fully coupled model, but that it only appeared for some initial conditions. Huang et al. (1992) showed the potential existence of multiple steady states when resolving deep and surface boxes in the Northern, low-latitude and Southern oceans. Gnanadesikan (1999) helped to explain how changes in Southern Ocean winds and eddies help determine the magnitude of the northern hemisphere overturning by controlling the transformation of dense water to light water. Johnson et al. (2007) extended the latter model to include prognostic equations for temperature and salinity and showed that, similar to the Stommel model, the balance between the hydrological and heat cycles could produce two stable states of overturning. In this case, the stable state is controlled by the density difference between the North Atlantic Deep Water and Southern Ocean surface waters. If the Southern Ocean surface is sufficiently light, the transformation of dense to light water in the Southern Ocean and low-latitude pycnocline is balanced by overturning in the North with a relatively shallow low-latitude pycnocline. If the Southern Ocean surface is denser than the subpolar North Atlantic, it is balanced by eddy fluxes of volume~~mass~~ to the Southern Ocean associated with a deep low-latitude pycnocline. Jones and Cessi (2016) extended the Gnanadesikan

(1999) model to include a Pacific basin- explicitly looking at how Southern Ocean winds control the flow of water between the Pacific-Indian and Atlantic basins, arguing that this produced a deeper pycnocline in the Pacific relative to the Atlantic. Gnanadesikan et al. (2018) extended the Johnson et al. (2007) model to include lateral tracer mixing. That paper examined what happened when hydrological fluxes were adjusted to make a model with ”incorrect” physics look like a model with ”correct” physics. While flux adjustment sometimes produced more realistic estimates of the stability of the overturning when the hydrological cycle amplitude was increased, as suggested by Liu et al. (2017), it did not always do so.

In this work we extend the model of Gnanadesikan et al. (2018) to include a separate low-latitude Pacific-Indian ocean box and a high-latitude Pacific box. We include a low-latitude exchange term between the Indo-Pacific and Atlantic similar to that formulated previously (Jones and Cessi 2016) thus potentially distinguishing the ”warm-water” and ”cold-water” pathways by which light water enters the Atlantic basin. The model can also be seen as extending Jones and Cessi (2016) by allowing for prognostic salinity and temperature and for a potential North Pacific overturning. The resulting model exhibits a phenomenologically rich interplay between freshwater fluxes and different regimes of overturning, predicting a number of new ”tipping points” between states.

2. Model description

a. Equations

A ”wiring diagram” of the model is shown in Figure 1. A full description of the model variables, parameters and equations is provided in the Supplemental Material. The model has two northern latitude boxes, each of which is capable of transforming warm, salty low-latitude water into colder, lower-salinity surface, intermediate, or deep water. The high latitude boxes also exchange water and tracers with low-latitude boxes in each basin. Freshwater fluxes are specified between the low-latitude boxes and the high latitude boxes ($F_w^{npac}, F_w^{natl}, F_w^{spac}, F_w^{satl}$) as well as between the two low-latitude boxes (F_w^{IB}). This allows us to examine the effect of changes in the magnitude of freshwater fluxes such as might be associated with climate change, as well as differences in the configuration of freshwater fluxes (i.e. the ratio of F_w^{npac} to F_w^{IB}), which may vary considerably from one climate model to another.

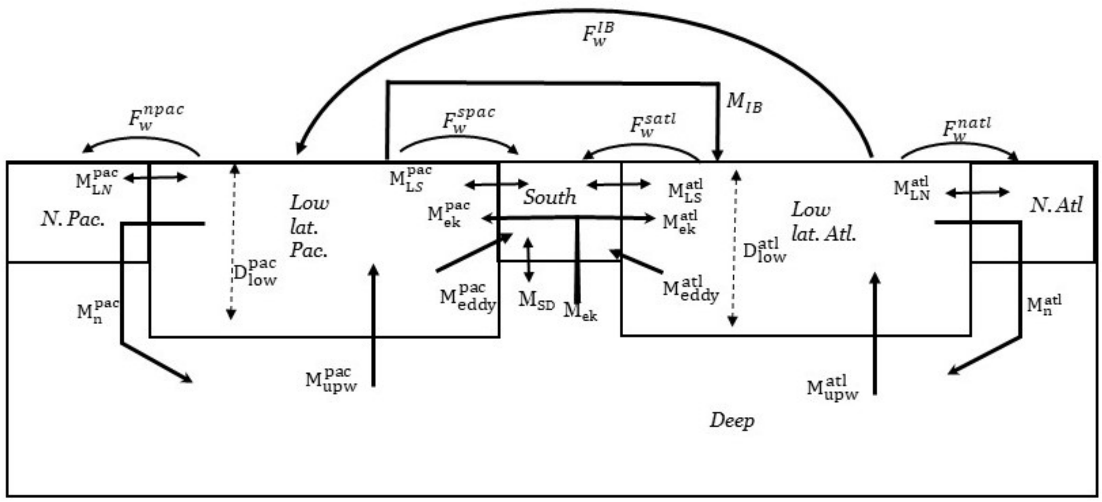


Fig. 1. Schematic of our six box model. Note that the circulation is centered on the Southern Ocean, as in Lumpkin and Speer (2007).

A key aspect of this model is that the pycnocline depth is not a fixed parameter. This contrasts with the classic box models (Stommel 1961; Rooth 1982; Huang et al. 1992; Tziperman et al. 1994) previously referred to as well as the more recent versions by Alkhayuon et al. (2019), where the overturning throughout the ocean is directly proportional to density differences between different regions. The depths of the pycnocline in the Atlantic (D_{low}^{atl}) and Indo-Pacific (D_{low}^{pac}) are state variables whose evolution is predicted by two volume~~mass~~ balance equations which depend in part on density difference between the northern and southern latitudes. We define $\rho_{natl,npac,latl,lpac,S,deep}$ as the densities of the North Atlantic, North Pacific, low latitude Atlantic and low latitude Pacific/Indian, Southern Ocean surface and deep ocean respectively. Then letting $Area_{latl,lpac}$ be the area of the low-latitude Atlantic and Pacific/Indian boxes, we can define

$$Area_{latl}\frac{\partial D_{low}^{atl}}{\partial t} = M_{ek}^{atl} + M_{upw}^{atl} - M_{eddy}^{atl} - M_n^{atl}(\rho_{natl} > \rho_{latl}, \rho_S) + M_{IB} - F_w^{satl} - F_w^{natl} - F_w^{IB}, \tag{1}$$

$$Area_{lpac}\frac{\partial D_{low}^{pac}}{\partial t} = M_{ek}^{pac} + M_{upw}^{pac} - M_{eddy}^{pac} - M_n^{pac}(\rho_{npac} > \rho_{lpac}, \rho_S) - M_{IB} - F_w^{spac} - F_w^{npac} + F_w^{IB}, \tag{2}$$

where the M_{ek} fluxes represent the upwelling of dense water into the Southern Ocean mixed layer and its subsequent export into the mid-latitude pycnocline; the M_{eddy} fluxes represent the supply of light water to the Southern Ocean driven by eddy thickness fluxes; the M_{upw} terms represent diffusive upwelling in the pycnocline; and M_{IB} represents the exchange of volume~~mass~~ between the basins. Terms of the form $\rho_A > \rho_B$ are set to 1 if true and zero if false. We note that the northern overturning only changes the volume of the low-latitude box if the high-latitude box is *denser than that of the corresponding low-latitude and Southern Hemisphere surface boxes*. If it is not, the assumption is that the water entering the high latitude box will return to the low-latitude box.

As in previous work (Gnanadesikan 1999; Gnanadesikan et al. 2018) the model describes the water fluxes between the boxes in terms of parameters and state variables. Following Gnanadesikan (1999) and Gnanadesikan et al. (2018) the overturning fluxes in each basin M_n^{atl} and M_n^{pac} are taken as proportional to the depth-integrated geopotential difference, which is tightly coupled to the available potential energy. Such a relationship has been found to hold in models (Bryan 1987; De Boer et al. 2010; Levermann and Fürst 2010) and in laboratory experiments (Park and Whitehead 1999). ~~The basic idea behind this is that the overturning velocity at a given depth is proportional to the meridional pressure gradient $v \propto -\frac{1}{\rho}\frac{\partial p}{\partial y}$~~ In Bryan (1987) this arises from assuming that the vertical shear associated with the overturning is proportional to the thermal wind shear~~overturning velocity is some fraction of the eastward geostrophic velocity~~

$$\frac{\partial v}{\partial z} = C_0 * \frac{\partial u}{\partial z} = C_0 * \frac{g\Delta\rho(z)}{\rho f L_y^n} \tag{3}$$

where f is the Coriolis parameter and L_y^n is the meridional scale over which the pycnocline shallows in the north) and $\Delta\rho(z)$ is an averaged meridional density difference at a given depth.~~and integrating this in the vertical from a level of no motion.~~ When we integrate over the width of the current L_y^n drops out and the horizontally-integrated velocity shear at a given depth becomes

$$\int \frac{\partial v}{\partial z} dy = C_0 \frac{g\Delta\rho(z)}{\rho f} \tag{4}$$

In Gnanadesikan (1999) the overturning velocity is related to the frictional balance in the western boundary current

$$v = -C_1 * \frac{L_B^2}{A_H}\frac{\Delta p}{\rho L_y^n} \tag{5}$$

which again implies that that

$$\frac{\partial v}{\partial z} = C_1 * \frac{L_B^2}{A_H}\frac{g * \Delta\rho(z)}{\rho L_y^n} \tag{6}$$

where $L_B = (A_H/\beta)^{1/3}$ is the thickness of the boundary layer, A_H is the lateral eddy viscosity, and $\beta = \partial f/\partial y$. Note that in this case there is an implicit assumption that $\Delta\rho(z)$ in the boundary current mirrors that in the basin as a whole.

When we double integrate (4) over depth we get that the overturning scales as.

$$M_n = \frac{g * \Delta\rho}{\rho\epsilon} * D^2 \tag{7}$$

where $\Delta\rho$ is an averaged density difference and D is the pycnocline depth. In this formulation, ϵ represents a resistance parameter

$$\epsilon = \left(\frac{C_0}{f}\int_{z=-z_0}^{0}\int_{z'=-z_0}^{z}\Delta\rho(z)/\Delta\rho * (dz'/D) * (dz/D)\right)^{-1} \tag{8}$$

Similarly, if we integrate (6)~~integrated~~ over the boundary layer the eddy viscosity drops out. If we then double-integrate with depth we get a functionally equivalent equation to (8), with C_0/f replaced with $C_1/\beta L_y^n$.~~transport at a given depth becomes proportional to~~ $\cancel{g\Delta\rho/\rho\beta L_y^n}$~~, though~~ $\cancel{C}$ ~~in (6) will not be the same as in (3)~~.

~~Both approaches assume that we can find the relevant pressure by integrating. Given that from the hydrostatic equation~~ $\Delta p(z) = -\int_{z'=z_0}^{z} g * \Delta\rho(z')dz'$ ~~where~~ $\cancel{z_0}$ ~~is a level of no motion, we can integrate in the vertical to get the transport~~ $M_{over} = C\int_{z=z_0}^{0}\left(\int_{z'=z_0}^{z}\frac{g*\Delta\rho(z')}{\rho}dz'\right)dz$ ~~If we nondimensionalize the density by~~ $\cancel{\Delta\rho}$ ~~and the depth by~~ $\cancel{D}$~~, this depth integrated velocity will be proportional to~~ $\cancel{M_n \propto g\Delta\rho D^2/\rho}$ ~~where the constant of proportionality, which as in previous work we define as~~ $\cancel{1/\epsilon}$~~,~~

Note that ϵ incorporates the relationship between the pycnocline depth D and the level of no motion z_0, the functional dependence of the meridional density gradient on depth and the small scale physics of the models (incorporated in the constants C_0, C_1) that allow for ageostrophic overturning. Using the definition of D discussed in Gnanadesikan (1999), if we assume that the pycnocline can be represented by a single jump in density at the level of no motion z_0, then the pycnocline depth $D = z_0/2$ and the integrated thermal wind shear will be $M_{thermal} = (2g\Delta\rho D^2)/\rho f$. If instead the density anomaly drops linearly to the level of no motion z_0, then $D = z_0/3$ and $M_{thermal} = (3g\Delta\rho D^2)/2\rho f$. Insofar as $M_{over} = C_0 * M_{thermal}$, in the first case $\epsilon = f/2C_0$, while in the second $\epsilon = 2f/3C_0$. Note also that our formulation here makes the overturning proportional to the available potential energy.

More specifically for our model we write

$$M_n^{atl,pac} = \frac{g(\rho_{natl,npac} - \rho_{latl,lpac}) D_{low}^{atl,pac^2}}{\rho_{natl,npac}\epsilon_{natl,npac}}. \tag{9}$$

where the density is computed using the full nonlinear equation of state using the temperature and salinity within the boxes with pressure referenced to the surface.

We note that this formulation then produces multiple possible configurations for the overturning circulation in each basin, outlined schematically in Fig. 2. When the northern basin is denser than either the low latitude or the Southern Ocean box it forms deep water. We refer to this as a ”DA” circulation when it is found in the Atlantic and a ”DP” circulation when it is found in the Pacific. When the northern basin has a density between that of the Southern Ocean and that of the low-latitudes (as is the case for the Pacific today), we consider it as forming intermediate water (and refer to it as ”IA” and ”IP” for the Atlantic and Pacific respectively). When the northern basin becomes sufficiently fresh, as in the ”off” state of the Stommel (1961) model, low-latitude water flows into the basin, gets lighter and is returned to the tropics. We refer to this situation as an ”SA” or ”SP” circulation for the Atlantic and Pacific respectively. Note that the designation refers to the *configuration in density space* of the pathways involved, and not necessarily to the *magnitude* of the pathways. Note also that these configurations will imply different relationships between z_0 and D_{low}, which as noted previously would be expected to affect ϵ.

We can then combine these designations to define a taxonomy of the Northern Hemisphere overturning. The modern ocean would then be described as ”DA-IP”, while an ocean in which the freshwater flux is strong enough to produce freshwater caps in both the North Atlantic and Pacific would be described as a ”SA-SP” ocean. The circulation during the Last Glacial Maximum was

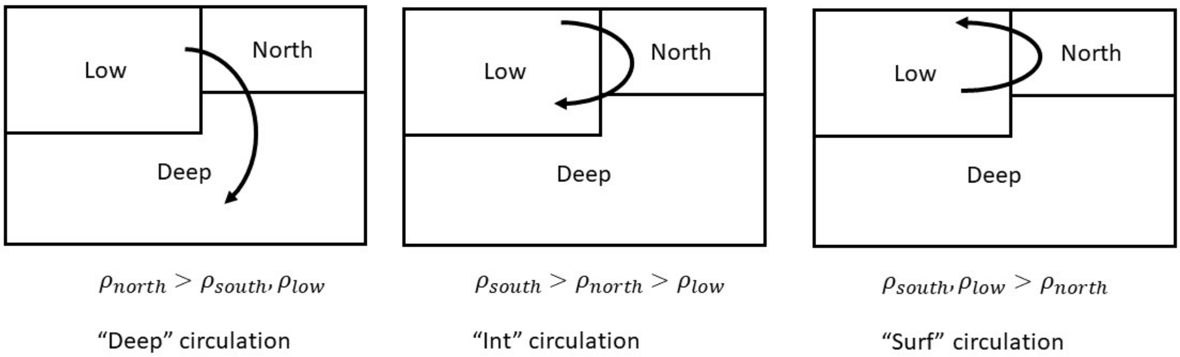


FIG. 2. Schematic of flows associated with different relationships between density in the southern, low-latitude and northern surface boxes. "Deep" circulation involves formation of water in the northern basin that is denser than both the Southern Ocean and the low-latitude surface boxes and thus is able to connect to the deep. Dynamically it is characterized by a positive overturning circulation as well as lower resistance ϵ to overturning. "Intermediate" circulation involves formation of water in the northern basin that is lighter than the Southern Ocean water but heavier than the tropics in that basin, similar to what we find in the Pacific today. Dynamically it is characterized by positive values of overturning but with higher resistance ϵ . "Surface" circulation involves low-latitude water entering the high latitudes and becoming lighter there- similar to what happens in the Arctic or Baltic today. For Northern Basins, this occurs when the density is lighter than both the low-latitude and Southern Ocean boxes. Dynamically it is characterized by negative values of overturning.

arguably an "IA-IP" circulation. (Rafter et al. 2022). This taxonomy will become relevant as we discuss solutions of the model in the Results section.

Following Gnanadesikan (1999) we allow for diffusive closures of the M_{eddy} and M_{upw} terms

$$M_{eddy}^{atl,pac} = \frac{A_{GM} D_{low}^{atl,pac} L_x^{satl,spac}}{L_y^s}, \tag{10}$$

$$M_{upw}^{atl,pac} = \frac{K_v Area_{latl,lpac}}{D_{low}^{atl,pac}}, \tag{11}$$

where A_{GM} is a thickness diffusion coefficient following (Gent and Mcwilliams 1990), K_v is a vertical diffusion coefficient, the L_x terms are the length of the Southern boundary of each basin, and L_y^s is the length scale over which the pycnocline shallows in the south. Note that Levermann and Fürst (2010) found that in realistic climate models the effective length scale L_y^s may depend on local density gradients, a process not included in the present study. Similarly we can define the Ekman flux as

$$M_{ek} = \int_s \vec{\tau} * d\vec{s} / \rho f \tag{12}$$

where s describes some closed pathway around the Southern Ocean. There has been some discussion in the literature about how this pathway should be chosen. Gnanadesikan (1999) defined it as the northernmost latitude circle unblocked by a continent, which for the ECMWF wind produce yields a flux of 28.5 Sv. However, Allison et al. (2010) argued that within idealized models averaging the transport over the ACC yielded a better prediction. Depending on which wind product is used and where one chooses the bounds of integration, one can find values ranging from 20-35 Sv. Our value of 24 Sv lies towards the lower edge of this range.

The exchange term is modified from Jones and Cessi (2016) as

$$M_{IB} = \frac{g\left[(\rho_{deep}/\rho_{lpac} - 1)D^{pac}_{low} - (\rho_{deep}/\rho_{latl} - 1)D^{atl}_{low}\right]\min(D^{pac}_{low}, D^{atl}_{low})}{\epsilon_{IB}} \tag{13}$$

so that it is proportional to the pressure difference between the two basins integrated over the minimum of the pycnocline depths. Note that this is slightly different than the original Jones and Cessi (2016) formulation, which turns out to be numerically unstable when the densities are allowed to vary separately in the two basins. Thus having an Indo-Pacific basin that is lighter than the Atlantic (as is the case in real life) allows for a circulation from the IndoPacific to the Atlantic. We do not consider in this paper the effects of wind stress curl in modulating the pressure gradient at the boundary between the Indian and Pacific Oceans as in Jones and Cessi (2016), but this represents a relatively straightforward future extension. It is worth noting, however, that in order to highlight the importance of these winds, Jones and Cessi (2016) assumed the density contrast between light and dense waters to be fixed (as was also the case in Gnanadesikan (1999)). For now the effect of the winds should be thought of as being implicitly included in ϵ_{IB}.

In this paper we also allow the temperatures and salinities to be determined by prognostic equations as in the single basin overturning models of Johnson et al. (2007) and Gnanadesikan et al. (2018). In addition to the volume~~mass~~ fluxes already described, the resulting balance equations also allow for terms due to lateral tracer stirring (Redi 1982), which produces mixing fluxes between low and high latitude boxes of the general form

$$M_{LN,LS}^{atl,pac} = \frac{A_{Redi} D_{low}^{atl,pac} L_x^{natl,npac,satl,spac}}{L_y^{natl,npac,satl,spac}}, \tag{14}$$

This mixing flux then produces transports of the form $M_{LN,LS}^{atl,pac,S} = (T_{latl,lpac} - T_{natl,npac,S})$ for temperature and $M_{LN,LS}^{atl,pac} = (S_{latl,lpac} - S_{natl,npac,S})$ for salinity.

We also allow for a mixing flux M_{SD} in the Southern Ocean that simulates the impact of bottom water formation/mixing of intermediate water into the deep. This flux is held constant in all the simulations described in this paper. It allows salt added to the Southern Ocean to escape to the deep ocean rather than being necessarily injected into the low-latitude pycnocline.

Finally, heat exchange between each surface box and the overlying atmosphere is handled using a restoring equation of the form

$$\frac{\partial}{\partial t}(D_X * T_X) = D_{mix}/\tau_{rest} * (T_X^{atm} - T_X) \tag{15}$$

Note that this means that the high latitude boxes (for which $D_X = D_{mix}$ are more tightly tied to the atmospheric temperature than are the low-latitude boxes. A full list of all the parameters is given in Table 1.

b. Calibrating the model

In Gnanadesikan et al. (2018) we calibrated the parameters governing the overturning in the northern hemisphere based on what is known about the rate of overturning and mean watermass properties. As in previous work, we calculate the observed pycnocline depth using data from the World Ocean Atlas 2013 (Locarnini et al. 2013; Zweng et al. 2013) as

$$D = \frac{\int_{z=-2000}^{0} z(\sigma_1(z) - \sigma_1(z=2000))dz}{\int_{z=-2000}^{0} (\sigma_1(z) - \sigma_1(z=2000))dz} \tag{16}$$

which for an exponential profile gives the e-folding depth. As shown in Fig. 3a, the exact value of the horizontal averaged $D_{low}^{atl,pac}$ that we use as a model variable will depend sensitively on the exact bounds of integration. Choosing the range 30°S to 45°N in the Atlantic and 30°S to 40°N in

TABLE 1. List of key parameters used in the model. Values that are varied from the control are shown in bold with alternatives shown in italics

Parameter	Description	Values
$Area_{ocean}$	Area of the ocean	3.6×10^{14} m^2
$Area_{latl}$	Area of the low latitude Atlantic	0.64×10^{14} m^2
$Area_{lpac}$	Area of the low latitude IndoPacific	1.98×10^{14} m^2
$Area_S$	Area of the Southern Ocean	0.62×10^{14} m^2
$Area_{natl}$	Area of the N.Atl+Arctic	0.22×10^{14} m^2
$Area_{npac}$	Area of the North Pacific	0.1×10^{14} m^2
L_x^{satl}	Length of Southern boundary of low-latitude Atlantic.	6.25×10^6 m
L_x^{spac}	Length of southern boundary of low-latitude IndoPacific.	18.75×10^6 m
L_x^{natl}	Length of northern boundary of low-latitude Atlantic	5×10^6 m
L_x^{npac}	Length of northern boundary of low-latitude Pacific (m)	10×10^6 m
$L_y^{satl,spac,natl,npac}$	Length over which pycnocline shallows	1×10^6 m
D_{oc}	Depth of ocean	3680 m
D_{mix}	Depth of High Latitude surface layers	100 m
K_v	Vertical diffusion coefficient	1×10^{-5} m^{2}s^{-1}
M_{ek}^{atl}	Ekman flux in the Atlantic	6 Sv
M_{ek}^{pac}	Ekman flux in the Pacific	18 Sv
A_{GM}	Thickness diffusion coefficient	1000 m^{2}s^{-1}
A_{Redi}	Tracer diffusion coefficient	**1000**,*400,2400* m^{2}s^{-1}
$\epsilon_{natl0,npac0}$	Baseline resistance to overturning	1.4×10^{-4}s^{-1}
$\Delta\rho_{trans}$	Width of transition over which resistance increases	0.1 kg m^{-3}
ϵ_{IB}	Resistance to interbasin flow	0.7×10^{-4} s^{-1}
F_w^{natl}	Baseline freshwater transport from low-latitude to N. Atl.	0.45 Sv
F_w^{npac}	Baseline freshwater transport from low-latitude to N. Pac.	**0.34 Sv**,*0.6 Sv*
F_w^{satl}	Baseline freshwater transport from low-latitude Atl. to SO	0.275 Sv
F_w^{spac}	Baseline freshwater transport from low-latitude Pac to SO	0.825 Sv
F_w^{IB}	Baseline interbasin flux	0.15 Sv
M_{SD}	Mixing between Southern surface and deep	15 Sv
$T_{natl,npac,latl,lpac,SO}^{atm}$	Baseline atmospheric restoring temperatures	0.3,1.3,16.8,18.2,2.8 $^\circ C$
τ_{rest}	Restoring time for temperatures	1 yr

the IndoPacific (reflecting the differences in the northward drift of the North Atlantic current vs. the Kuroshio) gives us pycnocline depths of 420m in the Atlantic and 380m in the Pacific.

Beginning with the Southern Ocean, we have a target mean salinity of around 34 PSU and temperature of around 4°C for the Antarctic Intermediate Waters. Given that modern deep waters are roughly equally fed from the south and the north, we set M_{SD} to 15 Sv. When we add up all the salt fluxes entering and leaving the Southern Ocean surface box, we find that we need a freshwater

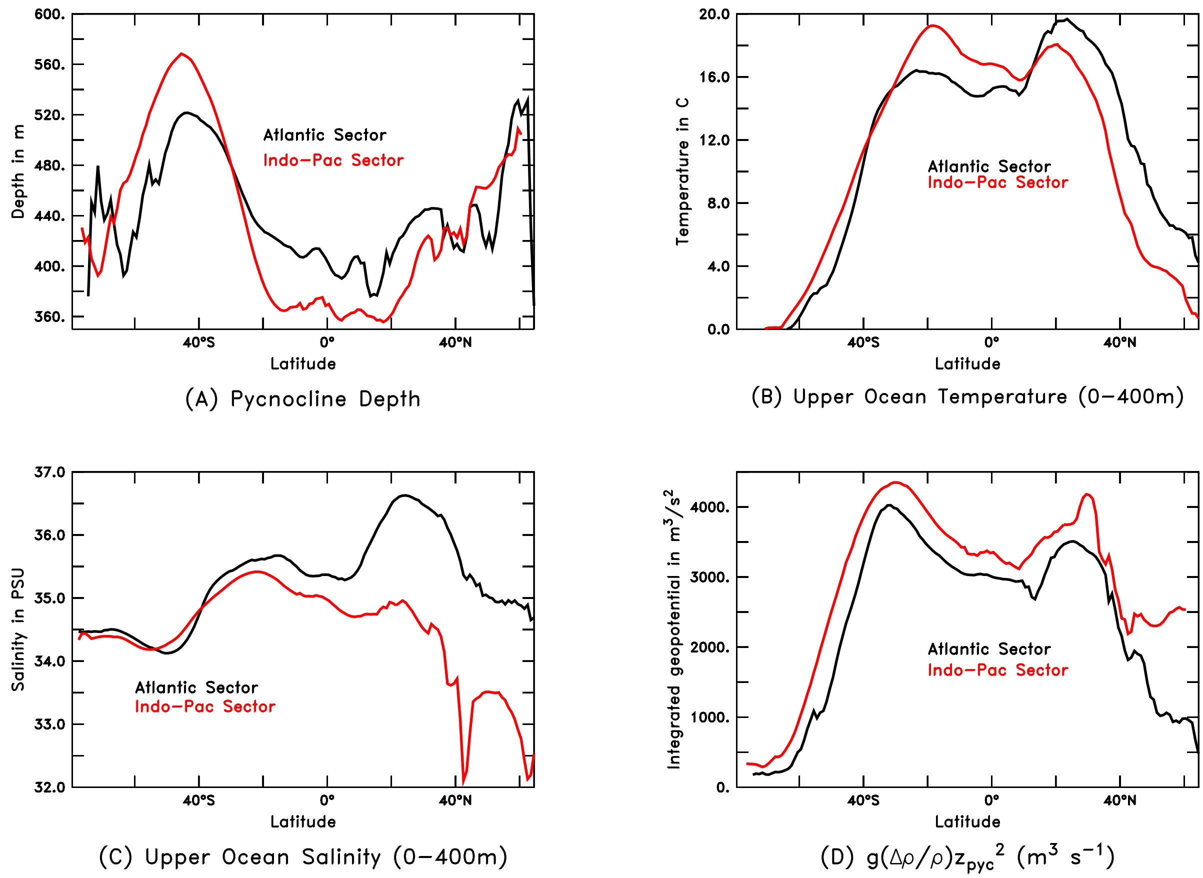


FIG. 3. Zonally averaged structure of upper ocean hydrography computed from World Ocean Atlas 2013 (Locarnini et al. 2013; Zweng et al. 2013). Black and red lines show values averaged over the Atlantic and IndoPacific respectively. (a) Pycnocline depth; (b) Upper ocean temperature; (c) Upper-ocean salinity; (d) Upper ocean $g'D_{pyc}^2$.

flux of about 1.1 Sv to balance them. While this value is close to the 1.04 Sv found by Tsukernik and Lynch (2013) using the ECMWF reanalysis, we note that such reanalyses are quite uncertain (Yu et al. 2017). We divide the Southern Ocean into an Atlantic sector whose southern boundary comprises 25% of the total length of the northern edge of the Southern Ocean ($L_x^{satl} = 6.25 \times 10^6 m$) and let the IndoPacific comprise the rest of the boundary ($L_x^{spac} = 1.875 \times 10^6 m$). We set the freshwater flux to the Southern Ocean in our baseline case to mirror this partitioning, so that $F_w^{satl} = 1.1Sv * 0.25 = 0.275Sv$ and $F_w^{spac} = 1.1Sv * 0.75 = 0.825Sv$.

We then turn to the Atlantic. The mean temperature and salinity above the pycnocline in the low-latitude Atlantic are 16.2°C and 35.8 PSU respectively in the 2013 World Ocean Atlas. We want our northern box to produce deep waters analogous to those found between 800 and 3000m at longitudes within the Atlantic from 45-50°N, giving us mean T≈4°C and mean S≈ 35 PSU. The integrated geopotential difference $g'D^2$ (roughly proportional to APE and shown in Fig. 3d) gives a difference of around 2500 m^3 s^{-2}. This is roughly consistent with what is seen when we calculate differences latitude by latitude. Given an overturning of 16-20 Sv, this would give us a value of $\epsilon_{natl} = 1.235 - 1.545 \times 10^{-4} s^{-1}$. For our baseline run we let $\epsilon_{natl} = 1.4 \times 10^{-4} s^{-1}$. Taking a baseline value of $A_{Redi} = 1000$ m^2 s^{-1}, the mixing flux computed from (14) is 2 Sv, so that the freshwater flux required to match the observed salinity difference is 0.45 Sv.

In the North Pacific, we assume that the there is no deep water formation and so the relevant volume~~mass~~ flux is the formation rate of North Pacific Intermediate Water, with a rough transport of about 6 Sv (Talley 1997; Lumpkin and Speer 2007). Given a low-latitude pycnocline depth of around 380m the low-latitude temperatures and salinities above the pycnocline are 17.2°C and 35 PSU respectively. The subpolar waters below the seasonal mixed layer (reflecting North Pacific Intermediate water) have an average temperature of 5.2°C and salinity of 33.8 PSU. This then produces a integrated geopotential difference of 1700 m^3 s^{-2}, resulting in a value of $\epsilon_{npac} = 2.8 \times 10^{-4}$s^{-1}, about twice that in the North Atlantic. Allowing for an additional mixing flux of 4 Sv due to the wider basin, this gives us a freshwater flux of 0.34 Sv. Note that this flux is actually smaller than in the North Atlantic despite the greater width of this basin. This undermines ~~, undermining~~ the assumption made in idealized models that it is a larger freshwater flux to the subpolar gyre in the Pacific relative to the subpolar gyre in the Atlantic~~the meridional atmospheric flux of freshwater~~ that localizes the overturning to the Atlantic.

The difference in the efficiency of overturning between the two basins reflects the difference in flow configuration induced by differences in density. A key difference between the North Atlantic Deep Water and North Pacific Intermediate Water is that the density of the former ($\sigma_\theta = 27.8$) is much heavier than the latter ($\sigma_\theta = 26.7$). The Antarctic Intermediate Water density ($\sigma_\theta = 27.0$) lies between the two. This means that the overturning generated by the near-surface APE difference in the Atlantic has a deeper level of ~~of~~ no motion, implying that it receives an extra ”kick” from Antarctic Intermediate water. This results in a smaller ϵ_{natl}. By contrast, in the Pacific, the Antarctic Intermediate Water slows the overturning, resulting in a larger value of ϵ_{npac}. Rather than focus on directly representing the dynamics of the intermediate water here, we instead represent this effect in our model as a transition between low resistance when the northern basin is denser than the south to a higher resistance when it is lighter.

$$\epsilon_{natl,npac} = 1.4 \times 10^{-4}\left[1.5 + 0.5\tanh\left(\frac{\rho_{natl,npac} - \rho_s}{\Delta\rho_{trans}}\right)\right]. \tag{17}$$

For now, we set $\Delta\rho_{trans} = 0.1$ kg m^{-3}, which allows for the configuration we want in the modern ocean. In the absence of observational evidence for an additional transition when the North Pacific becomes lighter than the tropical Pacific we do not include a further transition in resistance in this paper. However, we recognize that further investigation as in Fučkar and Vallis (2007) is warranted.

Finally, the interbasin exchange can be gotten by comparing the integrated geopotential difference relative to the deep water in the two basins. We find that this is around 950 m^3 s^{-2}. Given an interbasin transport of around 13 Sv (Lumpkin and Speer 2007), this implies a resistance $\epsilon_{IB} = 7 \times 10^{-4} s^{-1}$. This resistance is similar to the Coriolis parameter at the Southern tip of Africa, as would be expected from Jones and Cessi (2016). However, as previously noted it may also reflect the effect of wind stress curl in deepening the pycnocline as discussed in this paper. If this were the only process contributing to the contrast between the basins, the interbasin freshwater flux could then be backed out from this transport and the interbasin salinity difference of 0.8 PSU as 0.3 Sv. However, the fact that the Atlantic and IndoPacific receive different amounts of relatively fresh Southern Ocean surface and deep water also contributes to the interbasin difference, and we find a better fit in our model with a baseline value of 0.15 Sv. A full set of the relevant freshwater fluxes is shown in Table 1.

Using these baseline parameters we then vary our restoring temperatures in the surface layers to produce a solution that roughly agrees with the target observations. As shown in Table 2, the temperatures, salinities, densities, and transports in this solution do not diverge wildly from our target values. We also present results from a counterfactual simulation in which the baseline F_w^{npac}=0.6 Sv. This configuration results in a DA-SP circulation regime-with a negative overturning

TABLE 2. Target values (left-hand column) and final circulation for two versions of the box model. Control (center column) has a freshwater flux in the North Pacific F_w^{npac} of 0.34 Sv, smaller than the $F_w^{natl} = 0.45Sv$. Counterfactual case (right-hand column) sets F_w^{npac} to 0.6 Sv, higher than in the North Atlantic, and consistent with what is often found in idealized models

Parameter	Observed	Control	Counterfactual
N. Atl. T,S (°C,PSU)	4.0, 35.0	4.00,35.06	3.98,35.11
LL. Atl. T,S (°C,PSU)	16.2,35.8	16.21,35.81	16.21,35.86
D_{low}^{atl} (m)	420	429.4	427.9
M_n^{atl} (Sv)	16-20	19.0	18.9
N. Pac T,S (°C,PSU)	5.2,33.8	5.19,33.83	3.67,31.56
LL. Pac. T,S(°C,PSU)	17.2,35.0	17.20,34.95	17.28,35.02
D_{low}^{pac} (m)	380	381.4	378,1
M_n^{pac} (Sv)	2-8	6.4	-1.7
M_{IB} (Sv)	11-15	15.1	14.9
SO T,S (°C,PSU)	4.0,34	4.08,34.09	4.07,34.10
Deep T,S(°C,PSU)	4,34.5	4.03,34.5	4.02,34.50

in the North Pacific, indicative of warm salty water being converted to light surface waters, and a very fresh and cold surface North Pacific. However, changes in the other basins are relatively small.

c. *Numerical Continuation*

In the context of studying the proposed six-box overturning circulation model, continuation algorithms for numerical bifurcation analysis play a crucial role in identifying tipping points. Specifically, these methods help in analyzing the behavior of the overturning circulation as it undergoes a "hard" bifurcation, such as a saddle-node/~~, a~~ limit point, or a *subcritical* Hopf bifurcation. The present discussion focuses on continuation past limit points ~~(saddle-node bifurcations),~~ without aiming to provide a comprehensive guide to all bifurcation scenarios, for which one can refer to a number of published studies (Dhooge et al. 2008; Doedel 2007; Doedel and Tuckerman 2012; Fabiani et al. 2021).

Consider a parameter-dependent dynamical system, described by a system of *autonomous* ordinary differential equations (ODEs)

$$\frac{d\boldsymbol{y}}{dt} = \boldsymbol{f}(\boldsymbol{y};\lambda), \qquad f : \mathbb{R}^{n+1} \to \mathbb{R}^n \tag{18}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is the n-dimensional state variable vector, $\lambda \in \mathbb{R}$ is a scalar parameter and the function $\boldsymbol{f}$ is time-independent and sufficiently smooth. The goal is to construct a *solution curve* Γ for the system of nonlinear algebraic equations:

$$\Gamma := \{(\boldsymbol{y};\lambda) \in \mathbb{R}^{n+1} \text{ such that } \boldsymbol{f}(\boldsymbol{y},\lambda) = \boldsymbol{0}\}, \tag{19}$$

corresponding to the equilibria of the system (18) for various values of the parameter λ. The main concept underlying numerical continuation methods (Allgower and Georg 2012) is to generate a sequence of pairs $(\boldsymbol{y}_i, \lambda_i)$, $i = 1, 2, \ldots$ that approximate a specific branch of steady-states, satisfying a chosen tolerance criterion ($||\boldsymbol{f}(\boldsymbol{y}_i;\lambda_i)|| \leq tol$ for some small $tol > 0$) and involves a *predictor-corrector* process. We start from a known point on the curve, $(\boldsymbol{y}_i;\lambda_i) \in \Gamma$, and the tangent vector $\boldsymbol{v}_i$ to the curve there, computed through the implicit function theorem. To compute a new point $(\boldsymbol{y}_{i+1};\lambda_{i+1})$ we need two steps: (a) finding an initial guess for $(\boldsymbol{y}_{i+1},\lambda_{i+1})$ and (b) iteratively refining the guess to converge towards a point on the curve Γ (19). We denote the initial guess for $\boldsymbol{x}_{i+1} \equiv (\boldsymbol{y}_{i+1},\lambda_{i+1})$ as $X^{(0)}_{i+1}$, given by:

$$X^{(0)}_{i+1} = \boldsymbol{x}_i + h\boldsymbol{v}_i, \tag{20}$$

where h is a chosen step size. For a small enough h the prediction $X^{(0)}_{i+1}$ is close to the solution curve and can be corrected via e.g. a Newton-like scheme. Beyond critical points, where the Jacobian matrix becomes singular, solution branches can be traced with the aid of numerical bifurcation theory. For example, solution branches past saddle-node bifurcations (limit points) can be traced by applying the so called pseudo arc-length continuation method. This involves the parametrization of both $\boldsymbol{y}$ and λ by the arc-length s on the solution curve. The solution is sought in terms of both $\boldsymbol{y}(s)$ and $\lambda(s)$ in an iterative manner, by solving until convergence an augmented system, involving eq. (19) and the following pseudo arc-length condition:

$$N(X^{(k)}_{i+1}) = (X^{(k)}_{i+1}(s) - X^{(0)}_{i+1})^T \cdot \boldsymbol{v}_i = 0. \tag{21}$$

The tangent vector v_{i+1} to the curve at the new point is then computed. The direction along the curve must be preserved, i.e. $\boldsymbol{v}_i^T \boldsymbol{v}_{i+1} = 1$, and $\boldsymbol{v}_{i+1}$ must be normalized. Here, to construct the bifurcation diagrams of the 6-box model, we have employed Cl_Matcont version 5.4. Cl_MatCont (Dhooge et al. 2008) is a user-friendly Matlab package that relies on a collection of routines for numerical bifurcation analysis. Continuation codes such as the ones we are using here are capable detecting certain types of bifurcations (for cases where eigenvalues go from real and negative to real and positive). However, in the case of global bifurcations, where limit points intersect stable orbits, it may be more difficult to classify the bifurcation unambiguously. We therefore only label the cases where an unambiguous bifurcation is detected.

d. Coupled model

A full comparison between the box model and a coupled model is beyond the scope of this manuscript. However, we do use a subset of previously published simulations to show that certain key assumptions used in constructing our box model may also hold in more complex models. The model used is a coarse-resolution version of the GFDL CM2M model, denoted CM2Mc, which contains fully dynamic atmosphere, ocean, ocean biogeochemical, and sea ice components. The baseline simulation is described in Galbraith et al. (2011), to which the reader is referred for a more complete description. In a series of papers, the Gnanadesikan group (Pradal and Gnanadesikan 2014; Gnanadesikan et al. 2015; Bahl et al. 2019; Ragen et al. 2022) has explored the impact of changing the lateral mixing coefficient A_{Redi}, which diffuses tracers horizontally within the mixed layer and along isopycnals within the ocean interior. A baseline case with $A_{Redi} = 800$ m^2 s^{-1} was initialized with modern ocean temperatures and salinities and spun up for 1500 years using preindustrial concentrations of greenhouse gasses. At that point, as described in Pradal and Gnanadesikan (2014), a suite of simulations with a range comparable to those in those in the CMIP5 models was branched off the control and run for 1000 model years. At 360 years after this branch, as described in Bahl et al. (2019), additional simulations in which CO_2 was abruptly doubled were initialized from each branch and run for 140 years. In this paper we consider model solutions with $A_{Redi} = 400$ m^{2}s^{-1} (referred to as AREDI400) and $A_{Redi} = 2400$ m^{2}s^{-1} (referred to as ARED2400) for both control and 2xCO2 cases. 100-year averages from the end of the simulation are used in each case.

3. Results

a. Interbasin transport and the sensitivity of the overturning configuration to changes in hydrological amplitude

We begin by examining the interplay between the configuration of the hydrological cycle and the sensitivity of the overturning to instantaneous changes in the amplitude of hydrological cycling. Such changes might be found in an idealized climate model experiment where the greenhouse gas concentrations are suddenly raised or lowered. Starting with our target initial conditions, we define a set of freshwater flux patterns using the inferred $F_w^{natl,npac,satl,spac}$=0.45, 0.34, 0.275 and 0.825 Sv, respectively, but allow the interbasin transport to vary from -0.3 Sv to +0.3 Sv (horizontal axes, Fig. 4a-d). We then take the resulting patterns of freshwater fluxes and scale them up and down, varying from 0.1 to 2 times the "present-day" case (vertical axes, Fig. 4a-d). Note that if we assume that freshwater transport to the subpolar regions scales as atmospheric water vapor content, we would expect a doubling or halving of the flux to be associated with a 10°C change in temperature. Each model is run for 2000 years. We then repeat this set of experiments using a counterfactual case where the freshwater flux is higher in the Pacific than in the Atlantic ($F_w^{npac} = 0.6Sv$).

Both the Atlantic (Fig. 4a and c) and Pacific overturning (Fig. 4 b and d) show a strong sensitivity to the relative size of the interbasin transport (horizontal axis) and amplitude (vertical axis) of the hydrological cycle. Starting with a baseline present-day configuration in which F_w^{IB}=0.15 Sv with hydrological amplitude set equal to 1, we find that we can turn off the overturning in the Atlantic (Fig. 4a) either by increasing the amplitude of the hydrological flux, or by changing the direction

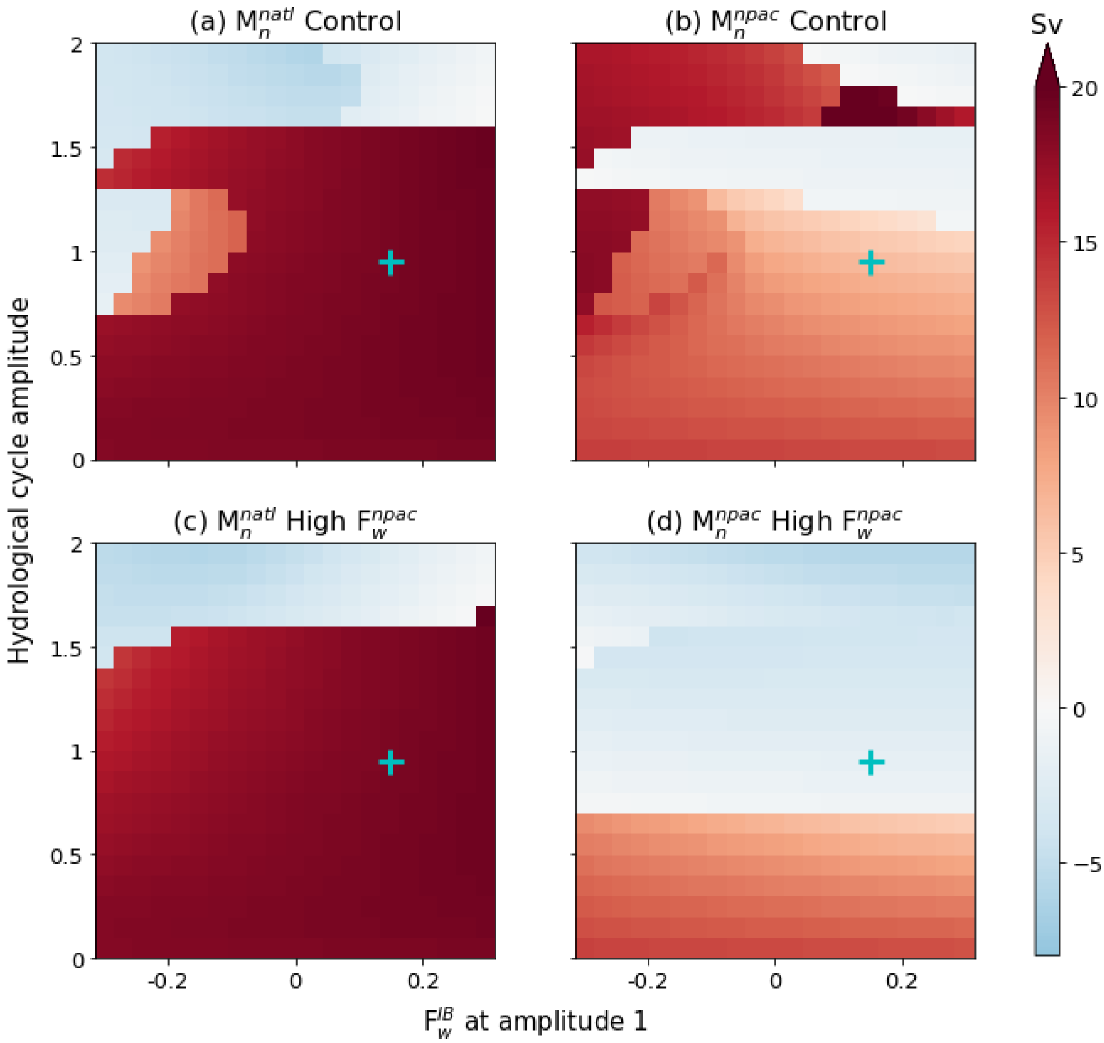


FIG. 4. Dependence of overturning circulation (Sv) in Atlantic (a and c) and Pacific (b and d) as a function of configuration and amplitude of the hydrological cycle for simulations started with observed initial conditions and run for 2000 years. In each subplot the horizontal axis shows a particular change in the configuration of the hydrological cycle (F_w^{IB} relative to all the other fluxes). The scale shows the value of F_w^{IB} in Sv when the hydrological cycle amplitude is 1. The amplitude of the hydrological cycle is shown along the vertical axis. Plus marks show our estimate of the present-day fluxes. The top row (a and b) shows an experiment where, at the hydrological amplitude of 1, the North Pacific freshwater flux F_w^{npac} is the observationally constrained value of 0.34 Sv (smaller than North Atlantic- so that the poleward freshwater transport acts to make the North Atlantic less salty with respect to the North Pacific). The bottom row (c and d) shows an experiment where F_w^{npac}=0.6 Sv at hydrological cycle amplitude of 1, so that the poleward freshwater transport makes the North Atlantic saltier with respect to the North Pacific.

of the interbasin atmospheric freshwater transport. The behavior of the Pacific overturning (Fig. 4b) is even more interesting. Starting from our baseline case, we can increase the overturning by decreasing the amplitude of the hydrological cycle (which reduces the contrast in salinity between the Pacific and the Atlantic), reversing the interbasin atmospheric transport so that it goes from the Pacific to the Atlantic (ditto) or *increasing* the amplitude of the hydrological cycle. *We are thus able reproduce the qualitative behavior whereby the overturning in the Pacific can strengthen in either warmer or colder climates.*

We can formalize these differences in overturning regimes by constructing a phase diagram of the overturning as a function of hydrological configuration and amplitude. As shown in Fig. 5 we color-code the different states, going from cooler to hotter colors as the overturning shallows in the Atlantic and darker to lighter colors as it shallows in the Pacific. In the control case the dominant regime at lower values of hydrological cycling is DA-IP, consistent with the fact that our target North Pacific Intermediate Water is lighter than both Antarctic Intermediate Water and North Atlantic Deep Water. A strong enough hydrological cycle with initial conditions similar to today can access either the SA-IP or the SA-SP states (orange colors in upper right of the plot). However, if we reverse the interbasin flux, we can enter a regime where both basins show deep water formation (dark blue region in Fig. 5a where the hydrological cycle amplitude is near 1 and there is a moderate flux from the Pacific to the Atlantic). A strong enough reverse flux can produce a SA-DP state (brown).

The counterfactual case shows a much simpler response: overturning shuts down as the amplitude of the hydrological cycle increases (move from bottom to top)- first in the Pacific (Fig. 4d), then in the Atlantic (Fig. 4c). Changing the interbasin atmospheric water transport F_w^{IB} has relatively little impact on the parameter dependence of overturning over the range shown. However, reversing it so that it dumps freshwater into the Atlantic does result in a Pacific overturning that is slightly more stable to increases in hydrological cycle amplitude, as well as an Atlantic overturning that is slightly less stable. The phase diagram for this case (Fig. 5b) shows only three of the six states seen in the Control simulation for the range of parameters covered here- DA-IP at low levels of hydrological cycling, DA-SP at levels comparable to the present day and SA-SP at high levels of cycling.

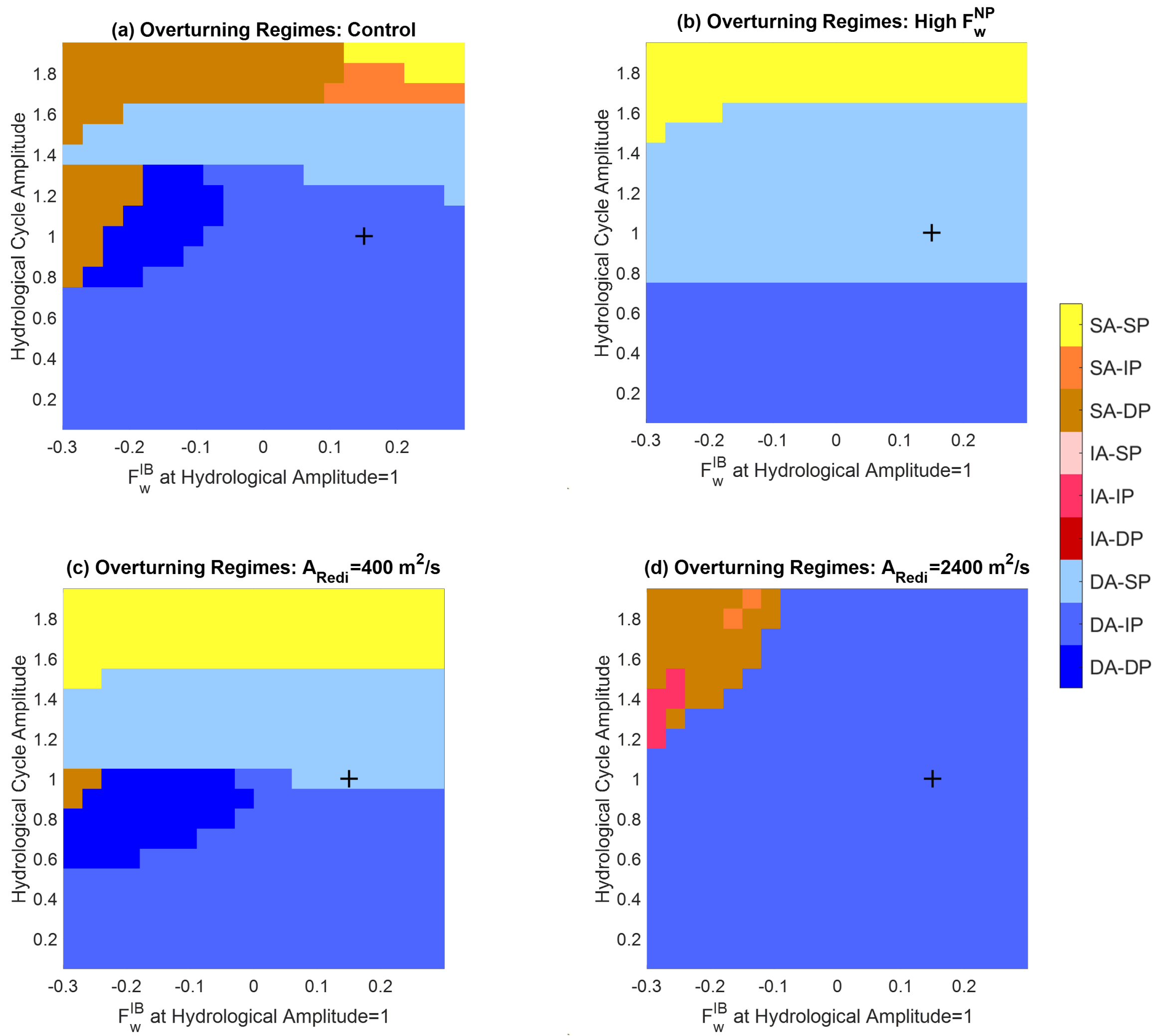


FIG. 5. Classifying global overturning configurations depending on whether Deep, Intermediate or Surface waters are primarily formed in the Atlantic and Pacific. As we move from cooler (blue) to warmer (brown/yellow) colors, Atlantic circulation shallows. As we move from darker shades to lighter ones, Pacific circulation shallows. Axes as in Fig. 4. Plus marks show "present-day" hydrological state. (a) Control simulation ($A_{Redi} = 1000m^2s^{-1}$,$F_w^{npac} = 0.34Sv$). (b) Higher Pacific freshwater flux. $F_w^{npac} = 0.6Sv$. (c) Lower lateral mixing ($A_{Redi} = 400m^2s^{-1}$). (d) Higher lateral mixing ($A_{Redi} = 2400m^2s^{-1}$).

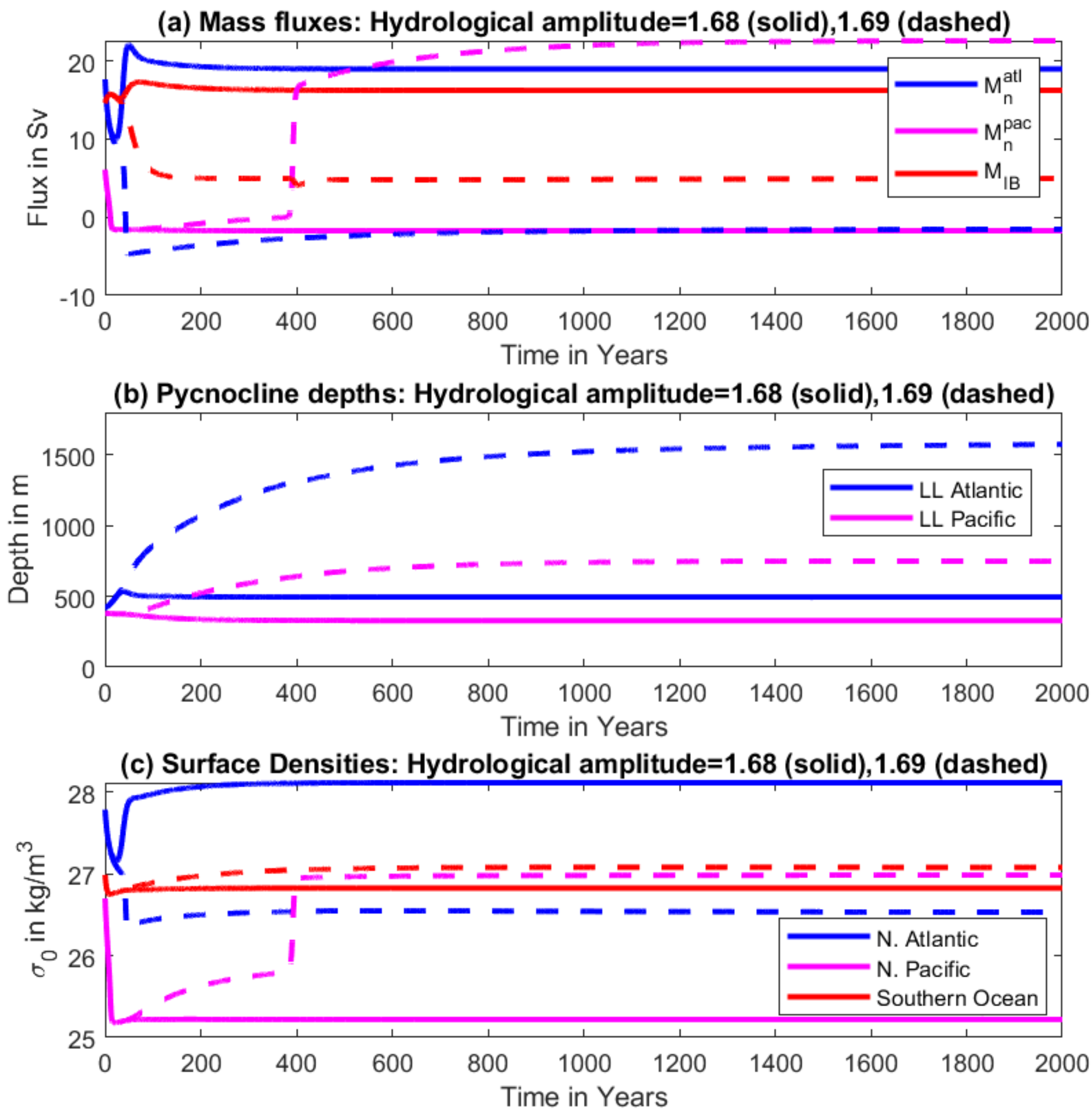


FIG. 6. Evolution of the (a) Large-scale circulation; (b) Pycnocline depths in Atlantic and Pacific; and (c) densities in two cases near a tipping point. Solid lines show case where a hydrological cycle with the baseline geometry has its amplitude instantaneously increased by a factor of 1.68. Dashed lines show a case where it is increased by a factor of 1.69.

b. Understanding how the overturning "tips" to the Pacific at high freshwater flux

The eventual tipping of the dominant overturning location to the Pacific within our model is sensitive to very small changes in freshwater flux. As shown in Fig. 6a, instantaneously increasing the hydrological cycle from our base case by a factor of 1.68 (solid lines) results in a collapse of the North Pacific overturning (solid magenta line). There is an initial drop, but then a recovery of the North Atlantic overturning (solid blue line), resulting in a final DA-SP regime. Increasing the scaling factor to 1.69 (dashed lines) produces an almost identical initial drop in both overturning circulations, but with the North Atlantic then proceeding all the way to collapse as well, giving us a temporary SA-SP configuration between years 50 and 400.After this we see another rapid in increase in the overturning in the North Pacific, which stabilizes at of 20 Sv. Note however, that the North Pacific subpolar box stays lighter than the Southern Ocean surface box (magenta dashed line stays below the red dashed line in Fig. 6c), so that the new state corresponds to a SA-IP ocean rather than a SA-DP ocean.

What accounts for the re-establishment of overturning in the Pacific in the case with a step change to 1.69 the hydrological amplitude? Note that within this case the establishment of an SA-SP circulation involves an initial reverse circulation of about -4.5 Sv in the Atlantic and -1.5 Sv in the Pacific, as the freshening of the North Pacific and Atlantic makes them lighter than the tropics in both basins. However, over time, the density in both subpolar basins rises and the magnitude of the overturning decreases (reflecting a smaller rise in pressure between the tropics and the subpolar gyre), pointing to a decrease in the density gradient between fresh, light subpolar water and saltier, heavier tropical water. This decerease in gradient is driven by both northern subpolar basins becoming saltier (not shown here) relative to their corresponding tropical basins. Because this increase in the salinity is occurring in the face of a weakening supply of salty water from the overturning, it must be driven by the only remaining term in the salt budget, namely the mixing flux. As can be seen in Fig. 6b, once the overturning circulation shuts off, the pycnocline in both the Atlantic and Pacific deepens. Over time this causes the mixing flux (which we have parameterized in equation 14 to scale as low-latitude pycnocline depth) to increase to about 6 Sv in both basins, even as the overturning drops. In the Pacific this increase in mixing allows the subpolar basin to become dense enough to start sinking, at which point the freshwater gets flushed out of the system.

c. Sensitivity to subgridscale eddy mixing

While there are a host of parameters that can affect the overturning circulation, we focus particularly on one, the lateral diffusion coefficient A_{Redi} that governs the horizontal diffusive exchange of heat and salt between the low-latitudes and the high latitudes. Within a recent generation of climate models A_{Redi} was found to vary from less than 400 m^2 s^{-1} to 2000 m^2 s^{-1} (Abernathey et al. 2022). As already noted, previous research from the Gnanadesikan group (Pradal and Gnanadesikan 2014; Bahl et al. 2019) has shown that this uncertainty is important, as changes over this range within a single model can affect the overturning in both the Atlantic and Pacific oceans.

Reproducing the sensitivity study of Fig. 4a and b with the control values for F_w^{npac} but with either lower (400 m^{2}s^{-1}, top row of Fig.7) or higher (2400 m^{2}s^{-1}, bottom row of Fig. 7) values for A_{Redi} reveals a strong sensitivity to this parameter in the box model. For lower values of A_{Redi}, given modern values of freshwater fluxes, the overturning in the Pacific reverses. The Atlantic overturning with observed fluxes is slightly stronger than with the control A_{Redi} = 1000 m^2 s^{-1} (with a value of 19.3 Sv). However, if we change the interbasin freshwater transport (moving to the left from the cross mark) the North Atlantic is more resistant to collapse. It is only a little less stable to increases in the hydrological cycle, with a collapse occurring near an amplitude of 1.58 rather than 1.68. This can also be seen by comparing the brown areas in Fig. 5a and c. A notable contrast with our control simulation is that turning off the North Atlantic overturning does not result in overturning switching to the Pacific for positive values of interbasin flux (Fig. 5c), because the lower mixing coefficient means that the lateral mixing is not sufficiently strong to degrade the freshwater cap in the Pacific. Only if the interbasin flux reverses do we see the overturning shift to the Pacific.

Increasing the mixing coefficient to 2400 m^2 s^{-1} (bottom row of Fig. 7), on the other hand, produces a somewhat larger (8 Sv vs 6 Sv in the control) Pacific overturning in the base case. The overturning configuration in this case is significantly more stable to changes in both interbasin flux and amplitude of the hydrological cycle. Interestingly, reversing the interbasin freshwater flux with higher mixing can produce an IA-IP state (dark pink, Fig. 5d), which does not appear in the other scenarios.

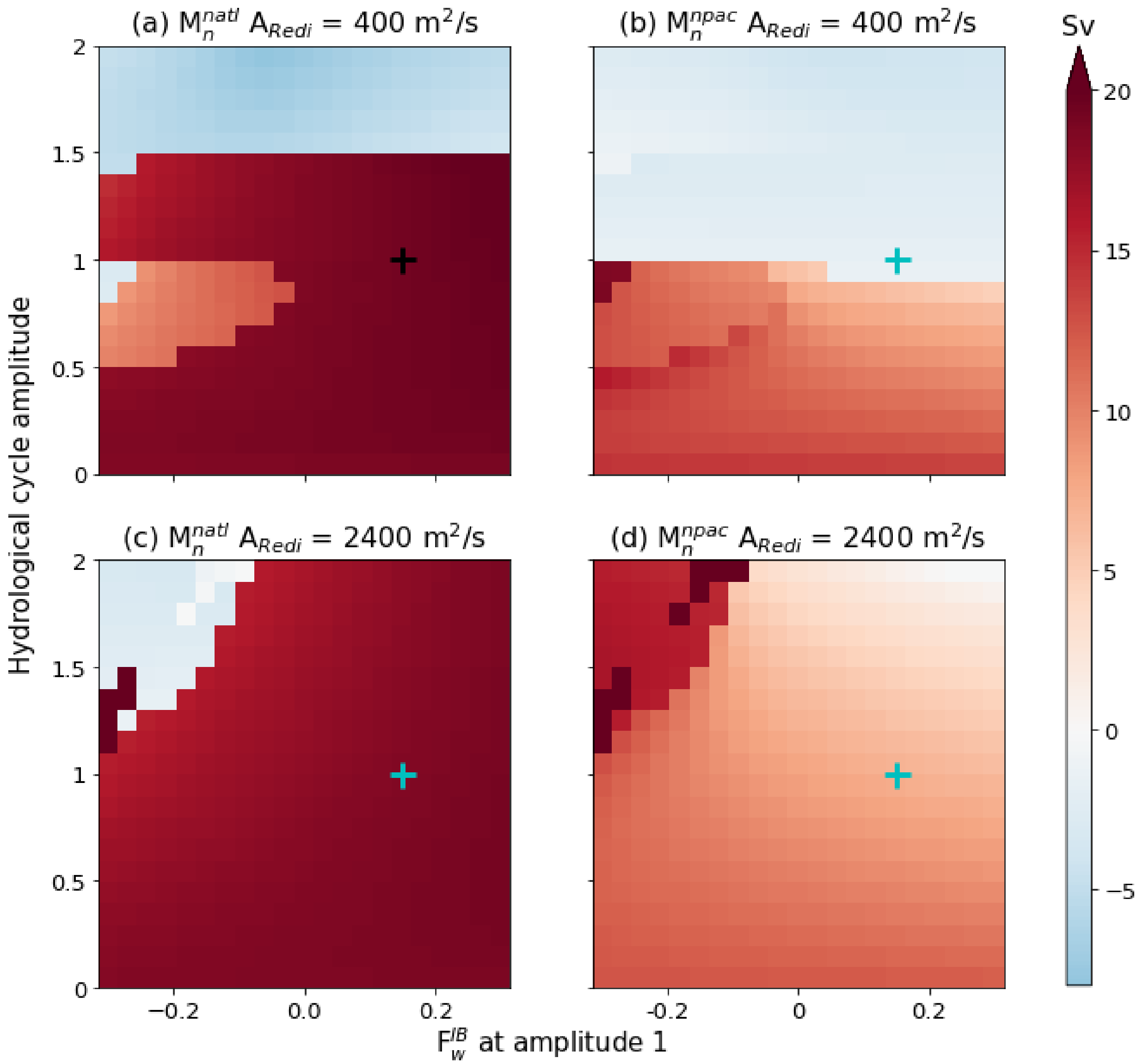
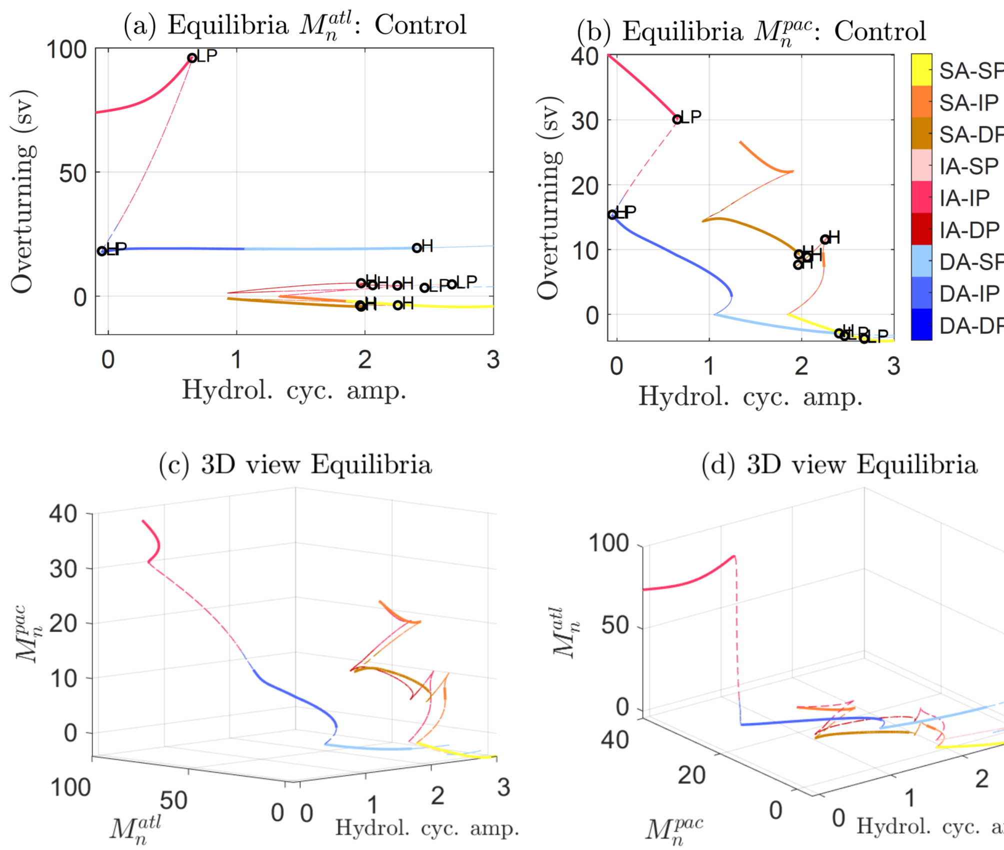


FIG. 7. Same as Fig. 4 except at varying values of subgridscale lateral eddy mixing parameter A_{Redi}. The top row (a and b) shows the Atlantic and Pacific overturning (Sv) respectively, with $A_{Redi} = 400$ m^2/s, near the lower end of the current value used in climate models. The bottom row (c and d) show the Atlantic and Pacific overturning (Sv) respectively, with $A_{Redi} = 2400$ m^2/s near the top end of the current value used in climate models.

d. Bifurcation analysis-Baseline case

The results presented up to this point assume a particular set of initial conditions (corresponding to the present-day) and an instantaneous change of the amplitude or configuration of the hydrological



599 FIG. 8. Numerical bifurcation diagrams with respect to the amplitude of the hydrological cycle showing steady
600 states of the overturning for the baseline parameter set. All states computed using the matCont code. Colors as
601 in Fig. 5. Limit points are marked with "LP", Hopf bifurcations with "H". Thick solid lines show stable steady
602 states, thinner dashed lines show unstable steady states. (a) North Atlantic overturning M_n^{atl} (b) North Pacific
603 Overturning M_n^{pac} (c,d). Two three-dimensional views of the solution space plotting $M_n atl$ and M_n^{pac} against
604 hydrological cycle amplitude.

cycle. However, it turns out that if we use numerical continuation methods to allow for quasi-static changes in parameters and to explore a wider range of initial conditions, we can find multiple equilibrium states while only changing the amplitude of the hydrological cycle. This is summarized

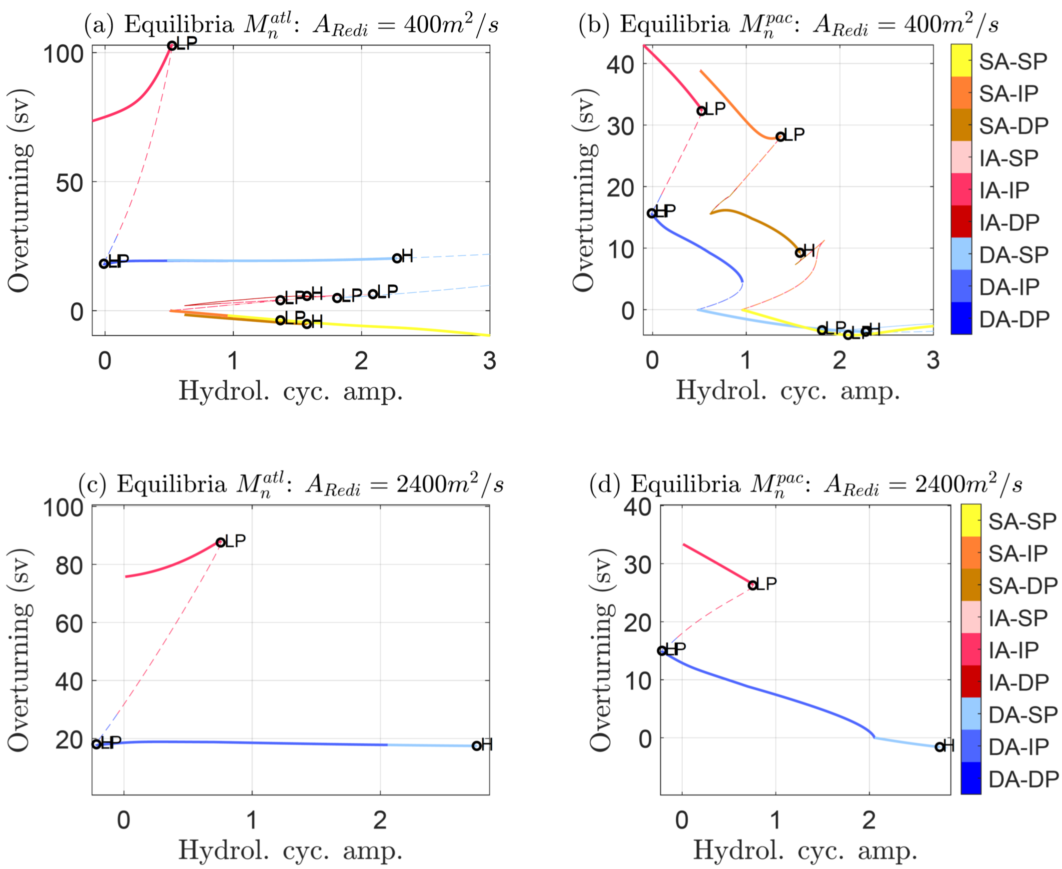


FIG. 9. Same as Fig. 8 a and b but now varying the lateral mixing coefficient A_{Redi}. (a) Atlantic overturning for low lateral mixing ($A_{Redi} = 400$ m^{2}s^{-1}). (b) Same as (a) but for Pacific. (c) Atlantic overturning for high lateral mixing. ($A_{Redi} = 2400$ m^{2}s^{-1}). (d) Same as (c) but for Pacific. Colors as in Figs. 5, 8.

for the baseline parameter set in Fig 8, which looks at steady states as functions of the amplitude of the hydrological cycle. In this and the following figure, geometric configurations of the overturning are denoted in the same colors as in Fig. 5 with stable branches denoted with thick solid lines and unstable branches denoted with thin dashed ones. The results show a number of surprises.

First, consider what happens if we start from our baseline case (which we would describe as DA-IP) at a value of 1 for the hydrological cycle amplitude and around 19 Sv for M_n^{atl}. As we increase the freshwater forcing (moving to the right along the blue curve in Fig. 8a) the overturning in the Atlantic is remarkably steady, while the Pacific overturning collapses. The collapse of the North Pacific overturning (blue lines in Fig. 8b) as we increase the freshwater flux looks very much like the classic Stommel fold bifurcation, with a transition to a DA-SP state. Analysis of the eigenvalues of the Jacobian at this point shows that this is, in fact, a limit point bifurcation. Though it is not visible on this plot, the collapse of the North Pacific overturning actually results in a slight *increase* in the North Atlantic overturning. As the freshwater flux continues to increase, there is eventually a Hopf bifurcation (indicated by the H on the light blue line in Fig. 8a) at a hydrological cycle amplitude of around 2.3 times the control value. At this point there is a transition to an SA-SP state (with the solid yellow lines being the only stable states at large hydrological cycle amplitudes), with deep pycnoclines in both basins and very fresh northern surface boxes.

The other configurations found in Fig. 5 when initializing our model from modern conditions also show up when tracing out stable and unstable manifolds using continuation methods. If we look the lower left of the Figs. 8a and b, we see a complicated network of curves. When plotted in 3 dimensions (Fig. 8c,d), we see that these curves correspond to several manifolds with weak surface or intermediate overturning in the Atlantic and intermediate or deep overturning in the Pacific. The solution illustrated in Fig. 6 is one of a manifold of states given by the thick orange curve between hydrological amplitudes of 1 and 2, M_n^{pac} in the 22-28 Sv range and M_n^{atl} between 0 and -5 Sv. This manifold is ”connected” to other states via an unstable IA-IP manifold (thin red curves). Additionally, there is a manifold corresponding to a SA-DP state (brown curve) with M_n^{pac} in the 10-15 Sv range and a M_n^{atl} between -5 and 0 Sv. In Fig. 5a, SA-DP states are only found by changing both the amplitude of the hydrological cycle and the relative size of the interbasin transport- here we are able to get them by changing initial conditions. Note that while the stable SA-DP manifold is connected to the SA-IP manifold by an unstable manifold (thin brown-orange line best seen in Fig. 8b,c) it is not obvious that a model reaching the end of one of these stable manifolds will transition between them.

There is also a branch with very high overturning in both the Atlantic and Pacific at low freshwater flux (red curves). This branch turns out to represent an IA-IP case in which there is a very strong flow of warm tropical water to high latitudes in both basins, such that the atmosphere is not able to cool the water enough to make it denser than the Southern Ocean. For our baseline parameter set, this branch is only stable at hydrological cycle amplitudes lower than the present day. Note that these states do not appear in Fig. 5a, so that they cannot be easily accessed from modern initial conditions via instantaneous changes in the hydrological cycle- again highlighting the utility of continuation methods in exploring the possible range of solutions.

The dependence of these regimes on hydrological cycle amplitude is sensitive to the parameterization of mixing. For A_{Redi} = 400 m^{2}s^{-1} (Fig. 9 a,b) the dependence is similar to that at A_{Redi}=1000 m^2 s^{-1} but all states shift to the left, with SA-SP (orange curves) states permitted at much lower hydrological cycle amplitudes and associated with larger values of M_n^{pac} (note difference between the range of the thick orange curves in Fig. 9b and 8b). Interestingly, while the IA-IP (red curves in upper left of plot) configuration collapses at a lower amplitude of hydrological cycle than is seen in Fig. 8a,b, lower A_{Redi} allows for higher overturning transports.

For the higher mixing case (A_{Redi} = 2400 m^{2}s^{-1}, Fig. 9c,d) we see the reverse effect: the IA-IP state is weaker at any given hydrological cycle amplitude, but persists to greater hydrological cycle amplitude. At high hydrological amplitudes we no longer see the SA states (no orange, yellow or brown curves near the bottom of Fig. 9c). Instead, when we run the model in this parameter range what appears is a limit cycle in which there are multicentennial bursts of DA-SP states which drain the pycnocline interspersed with multicentennial SA-IP states where the pycnocline slowly deepens. Detailed investigation of these states is deferred to a future manuscript.

e. Response to global temperature changes

Up to this point we have focused on changes in the geometry and amplitude of the hydrological cycle alone, without considering a primary driver of such changes, namely global temperatures. While a full discussion of the sensitivity to global warming is beyond the scope of this manuscript, we present some preliminary exploration of changing temperatures and hydrological cycle together. We consider two cases. In the first, we impose a globally uniform change in the atmospheric restoring temperatures with an associated increase of the hydrological cycle amplitude of 7% per degree. This scaling is what we would expect if transport scaled as water vapor content. In the second set of simulations, we allow for polar amplification in the northern hemisphere high latitudes but reduced warming in the Southern Hemisphere high latitudes. For this set of simulations we use a factor of 1.5 for the high latitude Northern Hemisphere and 0.6 for the high-latitude Southern

Hemisphere. These values are typical of the net warming of the atmosphere over the polar oceans in the GFDL ESM2Mc models, and are consistent with the CMIP6 suite of models (Hahn et al. 2021). In all cases we started our simulations from the same initial conditions and parameter sets as in the control simulation and integrated for 200 years in order to make our simulations more comparable with those carried out for the IPCC process.

As shown by the blue lines in Fig. 10a,b the overturning in both basins is surprisingly insensitive to a uniform change in temperature. This is understable if we consider the response of the surface densities to climate change (Fig. 10c,d). Under uniform temperature change (blue lines), all densities show a similar value of change. As temperature increases, the decrease in density gradient between the northern and low latitudes due to increased freshwater flux is largely compensated by the increase in the sensitivity of density to temperature. The idea that higher temperatures might compensate higher freshwater fluxes due to non-linearities of the equation of state was previously advanced by De Boer et al. (2007) and Schloesser (2020).

Under an asymmetric temperature change, however, the overturning shows more complex behavior (red lines, Fig. 10). At low temperatures, the North Pacific becomes denser than than the Southern Ocean due to a combination of a.) the Southern Ocean restoring temperature dropping less than the North Pacific and b.) the fact that in our model the restoring temperature is lower in the North Pacific than in the Southern Ocean to begin with. As a result, deep water can form there. The opening of a second deep water pathway "steals" some of the overturning from the Atlantic for cooling below about 2°C (note the dip in the red line at low temperatures in Fig 10a). As temperatures warm, on the other hand, the North Atlantic sees its density drop faster than the Southern Ocean. At a warming of around 4.7°C this results in a transition to an IA-IP circulation, with an associated deepening of the pycnocline (not shown) and for warming just short of 6°C we see a collapse in both basins and an SA-SP regime. We note that we expect these results to be strongly dependent on the degree of asymmetry in warming, the rate of warming and internal parameters–all of which will be explored in future work.

4. Comparison with a fully coupled model

While a full calibration of our six-box model against an ESM is beyond the scope of this paper (for reasons we will outline below), we can still use ESM simulations to support some of

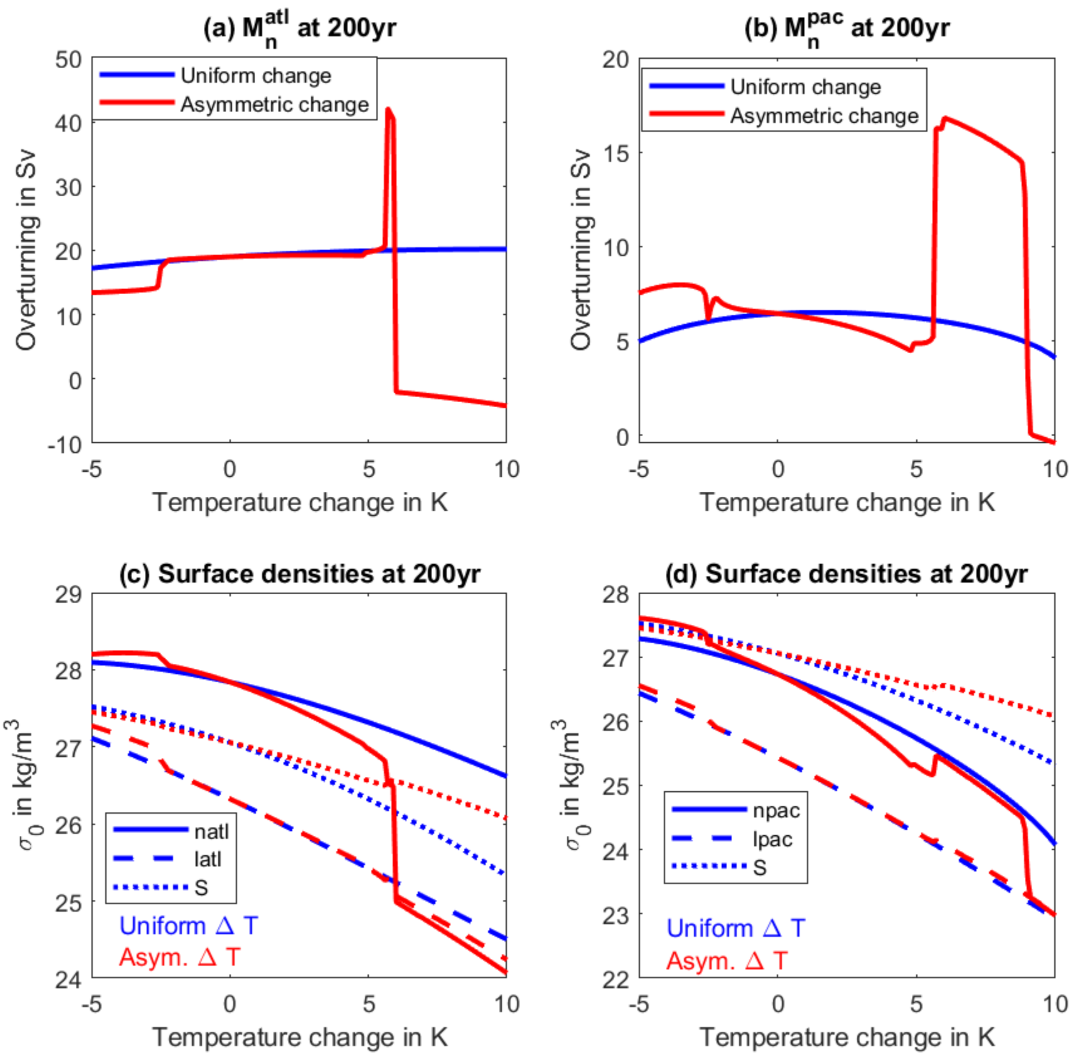


FIG. 10. Overturning and surface densities given the baseline physical and model initial conditions 200 years after an abrupt change in temperature and associated change in the amplitude of the hydrological cycle. Uniform temperature change is shown in blue, hemispherically asymmetric (1.5 tropical value in north, 0.6 in south) in red. (a) North Atlantic overturning M_n^{atl} in Sv. (b) North Pacific overturning M_n^{pac} in Sv. (c) Densities in North Atlantic (solid), Low Latitude Atlantic (long dashed) and Southern Ocean (dotted). (d) Same as (c) but for Pacific.

the assumptions made to construct our box model and to demonstrate some similar changes in overturning configuration. First, as CM2Mc is mass conserving in the ocean, we can use the difference between mass transport across various lines to compute how much freshwater is added to different regions. As shown in the top row of Table 3, the models all support the idea that net freshwater fluxes into the Arctic and subpolar North Atlantic are larger than net freshwater fluxes into the subpolar North Pacific.

Matching watermass properties is challenging in high latitudes, in part because the regions where deep water forms are often different in the model vs. in observations. For purposes of this paper, we take densities in the Atlantic from 55-65°N and 60-20°W at depths from 100-400m which is where the near-surface waters with densities corresponding to the model NADW are found in the main Atlantic. In the North Pacific, we look at waters between 100-400m between 55 and 65°N. In the Southern Ocean we look for waters between the surface and 400m between 55 and 50°S. The resulting watermass properties in the low mixing (AREDI400) control simulation are qualitatively similar to the our box model results, with the North Atlantic being the densest of the polar boxes, followed by the Southern Ocean and then the North Pacific. The density difference between the tropics and North Atlantic is 1.64 kg m^{-3} while that in the Pacific is about 1.49 kg m^{-3}, slightly larger than the 1.42 and 1.23 kg m^{-3} seen in our target values. With relatively realistic pycnocline depths (409 in the Atlantic and 388 in the Pacific) we can invert for the resistance parameter ϵ by plugging in the overturning in the two basins near its maximum latitude. This yields a resistance parameter of 1.21×10^{-4}s^{-1} in the Atlantic (again in line with our previous estimates) and 2.71×10^{-4}s^{-1} in the Pacific, qualitatively similar to the factor of 2 increase in equation (17).

When the A_{Redi} coefficient is increased, we see a salinification of the North Pacific (1.13 PSU) and Southern Ocean (0.44 PSU). The result is to increase the density in the North Pacific so that it is greater than in the Southern Ocean. The resulting overturning is actually larger in the Pacific than in the Atlantic, and the resulting ϵ is 0.94×10^{-4}s^{-1}. Meanwhile, M_n^{atl} drops slightly, from 21.7 to 18.0 Sv.

In our baseline box model a change in A_{Redi} from 400 to 2400 m^{2}s^{-1} results in North Pacific salinity increasing sharply from 31.5 PSU to 34.2 PSU, and the overturning switching from an DA-SP configuration with M_n^{pac}=-1.6 Sv to a DA-IP configuration with M_n^{pac}=7.9 Sv. However in the baseline box model the change in Southern Ocean salinity from 33.99 to 34.24 is smaller than in the coupled model and does not result in the Southern Ocean becoming denser than the

TABLE 3. Simulation results from four Earth System Models run with different values of lateral mixing coefficient A_{Redi} (400 and 2400 m^2s^{-1}) and two levels of CO_2 (control simulation is 286 ppmv, 2xCO2 is 572 ppmv). All results represent average over final century of simulation.

Output/Simulation	AREDI400 CTRL	AREDI2400 CTRL	AREDI400 2xCO2	AREDI2400 2xCO2
$F_w^{natl}, F_w^{npac}, F_w^{SO}$	0.34,0.28,0.65	0.39,0.23,0.65	0.41,0.31,0.76	0.45,0.28,0.77
$T_{natl}, S_{natl}, \rho_{natl}$	5.73, 35.10,27.65	5.54,35.20,27.76	4.53,34.65,27.43	5.88, 35.13,27.66
$T_{latl}, S_{latl}, \rho_{latl}$	16.08,35.50,26.01	16.05,35.74,26.19	17.35,35.80,25.95	17.53,36.03,26.07
$T_{npac}, S_{npac}, \rho_{npac}$	2.04,33.24,26.50	3.79,34.35,27.25	3.44,33.02,26.22	4.49,34.03,26.91
$T_{lpac}, S_{lpac}, \rho_{lpac}$	17.27,34.63,25.01	17.27,34.98,25.28	18.14,.34.52,24.71	18.38,34.89,24.93
$T_{SO}, S_{SO}, \rho_{SO}$	5.34,33.96, 26.79	5.59,34.4,27.09	5.78,33.86,26.65	6.58,34.31,26.90
$D_{low}^{natl,npac}$	409, 388	367,330	427,384	389,337
M_n^{atl}, M_n^{pac}	21.7, 8.0	18.0,21.6	16.6,7.1	16.7,13.3
ϵ_{natl}	1.21 x 10^{-4}	1.12x10^{-4}	1.55x10^{-4}	1.37x10^{-4}
ϵ_{npac}	2.71 x10^{-4}	0.94 x 10^{-4}	3.01x10^{-4}	1.63x10^{-4}

North Pacific. Increasing A_{Redi} also causes M_n^{atl} to decline (from 19.3 to 18.5 Sv), smaller than the decline in the coupled models. However, as this change in mixing is insufficient to form deep water in the North Pacific we do not see a reduction in the pycnocline depth, which would drive larger reductions.

Under doubled CO_2, the freshwater fluxes in the coupled simulations increase sharply. In the AREDI400 simulation the global mean surface air temperature warms by 1.4°C while in the AREDI2400 simulation it warms by 1.7 °C. This implies a 14%/°C sensitivity of F_w^{natl} in the the AREDI400 case but a 9%/°C sensitivity in the AREDI2400 case. F_w^{npac} has a slightly different sensitivity, 7.7%/°C in the AREDI400 case but 10%/°C sensitivity in the AREDI2400 case. This suggests that while the rough 7%/°C Clasius-Clapeyron-based scaling used in the box model is not wildly inaccurate, it likely underestimates some of the impacts due to changes in circulation. As expected, the increase in freshwater flux leads to increased salinity in the tropics and decreased salinity in the high latitudes. Note however, that this does not necessarily result in a decreased density contrast within a given basin. This highlights the utility of using a nonlinear equation of state in our box model, which leads to a greater focus on the role of changes in pycnocline depth and ϵ under climate change. It also supports the conclusion of De Boer et al. (2010) that a simple linear scaling between density contrast and overturning is insufficient to explain key aspects of overturning.

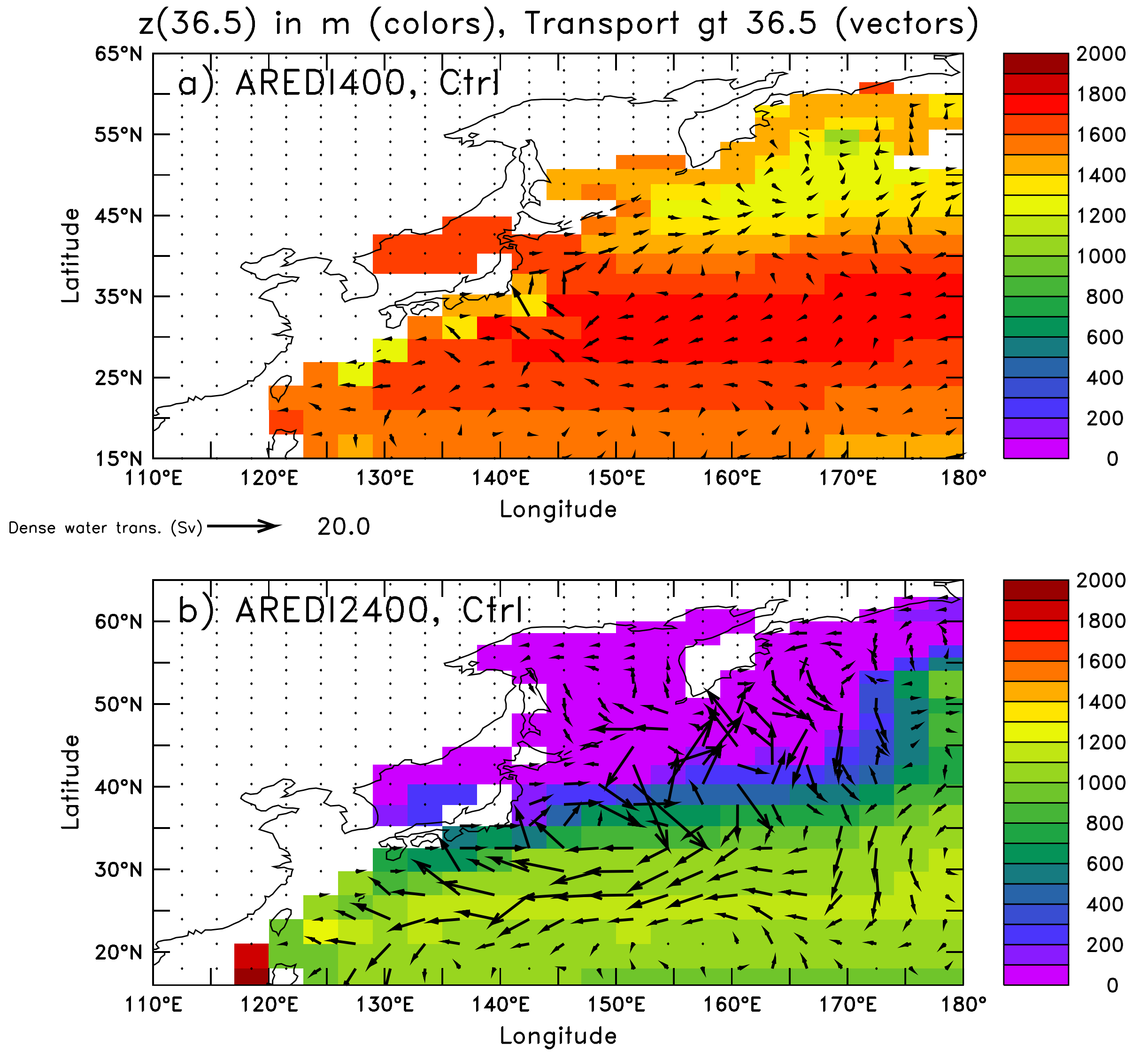


FIG. 11. Change in density structure and overturning associated with changing lateral mixing coefficient A_{Redi} in CM2Mc model of Galbraith et al. (2011). Colors show depth in m of the $\sigma_2 = 36.5$ isopycnal surface. σ_2 is used here because CM2Mc saves transports in density space in this coordinate system. Vectors show transport of water denser than this surface. (a) AREDI400 simulation showing that this surface is deep and circulation below it is relatively weak. (b) AREDI2400 simulation showing that the surface outcrops in the Northwest Pacific and that there is flow away from the outcrop in the eastern sector of the plot, but flow back towards the outcrop in western sector.

In the AREDI400 case, the overturning decreases by about 5 Sv (23%)in the North Atlantic and 0.9 Sv ($\sim$ 11%) in the North Pacific. In the AREDI2400 case, however, both the absolute and relative decline in the North Pacific (8.3 Sv, 38%) are much larger than the absolute and relative decline in the North Atlantic (1.3 Sv, 7%). Note that in the AREDI2400 case, the density of the Southern Ocean approaches that of the North Pacific, but remains much heavier than the North Atlantic and that this is associated with a sharp increase in ϵ. Again this qualitatively supports our approach of considering all three polar basins when thinking about changes in overturning configuration.

In summary, the coupled model simulations support a number of the choices made in configuring our box model. 1. Having a larger freshwater flux to the Arctic+Subpolar Atlantic than the North Pacific. 2. Including a nonlinear equation of state allowing for some compensation of changes in salinity and temperature. 3. Allowing for relatively larger responses in atmospheric freshwater transport to temperature change than is generally associated with global precipitation (which is dominated by the tropics). 4. Allowing for the resistance to overturning to depend on the contrast between the northern and Southern subpolar basins. While the dependence of ϵ on density structure is clearly more complicated than that in equation 17, we clearly see that having the density of the Southern Ocean approach that of the northern subpolar oceans is associated with a less efficient conversion of APE into overturning.

5. Discussion

We have developed a framework for understanding the competing roles of the geometric configuration of the atmospheric hydrological cycle, the amplitude of the atmospheric hydrological cycle, and oceanic eddy processes in setting the configuration and magnitude of the ocean circulation. Key lessons that emerge are: 1. Given a freshwater flux to the Arctic+Subpolar North Atlantic that is larger than the freshwater flux to the subpolar North Pacific, the North Pacific overturning can increase as a result of either increases or decreases in hydrological cycle amplitude. This is likely to be a very different result than would be found from using idealized models with strip continents-highlighting a potential deficiency in the configuration of such models. 2. We can qualitatively explain the sensitivity of the overturning circulation to the lateral eddy mixing A_{Redi}, a parameter that has been previously shown to have an important impact on overturning configuration in fully coupled models. 3. Including two basins and interhemsipheric control of overturning efficiency allows for a number of interesting transitions across a range of overturning configurations.

We have only begun to scratch the surface of the parametric dependence of this model. We note that, even for our simple model, we have over twenty initial conditions, physical parameters, and boundary conditions-making a comprehensive exploration of the search space challenging. We are currently exploring two approaches to this. One is to use the continuation methods outlined above to search for interesting phenomena. The second is using generalized adversarial networks to trace out the separatrices in state space between different dynamical states. Early results of the second approach are reported in Sleeman et al. (2023a,b).

In constructing our model we have tried to strike a balance between a parsimonious representation of the processes involved and a sufficiently comprehensive inclusion of key processes. That said, it should be recognized that there are a number of processes that could benefit from a more sophisticated treatment. One obvious shortcoming of our formulation is how we handle the transition as the Southern Ocean becomes lighter/denser than the high latitude boxes. Explicitly resolving an intermediate water box as in Alkhayuon et al. (2019) would allow for a better treatment of this, but would add additional parameters that are harder to constrain. This could be attacked by analyzing experiments with full general circulation models in which the densities of NPIW, AAIW, and tropical waters in the Pacific as well as Southern Ocean winds are changed separately, similar to the work of Fučkar and Vallis (2007), but with realistic geometry and fixed freshwater fluxes.

It is also worth noting that the full three-dimensionality of the system can complicate fitting box models to GCMs. When we examine transports in potential density space (computed in CM2Mc using σ_2) increasing A_{Redi} results in changes both to the density structure and the overturning. In the AREDI400 case, $\sigma_2 = 36.5$ never outcrops in the northwest Pacific (Fig. 11a), while in the AREDI2400 case (Fig. 11b) it does. Looking at the transport of water denser than this surface in AREDI2400 case reveals a recirculating flow. Dense waters are formed at relatively shallow depths and injected into the interior at a number of locations east of 150°E and flow southward below a sinking density surface. A significant fraction of this water, however, returns to the surface waters west of 150°E. When we look at the zonally-integrated average in either depth or density, we see significant cancellation of this flow, with only about 13 Sv of dense water moving southward at 15°N. While one can think of this circulation as being encompassed by the exchange terms

$M_{LN,LS}$, we cannot simply parameterize it as we have done in equation 14 using a global diffusion coefficient. Similar issues of how to divide recirculating flow from net transport arise in tracing the interbasin flow- one reason why we do not explore it in more detail here.

Additionally, at this point we have treated the deep ocean as a single box and do not resolve the different impacts of Antarctic Intermediate Water and Antarctic Bottom Water. Incorporating more structure into the deep water (for example following ideas of Nikurashin and Vallis 2011) would introduce additional time scales of variability and additional configurations of the circulation, allowing us to better resolve the deep ocean circulation. However, it would also introduce additional parameters, such as the magnitude and spatial distribution of the abyssal diapycnal and lateral mixing coefficients.

Finally, in this paper we have ignored the presence of noise in the climate system. Including noise in the model allows for a rich phenomenology of behavior including unsteady oscillations and more complex transition behavior between states. We plan to report on these phenomena in future publications.

Acknowledgments. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112290032. Approved for public release; distribution is unlimited.

Data availability statement. Matlab codes to generate all the results used in this manuscript are provided at 10.5281/zenodo.8126674.

References

Abernathey, R., A. Gnanadesikan, M.-A. Pradal, and M. A. Sundermeyer, 2022: Isopycnal mixing. *Ocean Mixing*, Elsevier, 215–256.

Alkhayuon, H., P. Ashwin, L. C. Jackson, C. Quinn, and R. A. Wood, 2019: Basin bifurcations, oscillatory instability and rate-induced thresholds for atlantic meridional overturning circulation in a global oceanic box model. *Proceedings of the Royal Society A*, **475 (2225)**, 20190 051.

Allgower, E. L., and K. Georg, 2012: *Numerical continuation methods: an introduction*, Vol. 13. Springer Science & Business Media.

Allison, L., H. Johnson, D. Marshall, and D. Munday, 2010: Where do winds drive the antarctic circumpolar current? *Geophysical Research Letters*, **37 (12)**.

Bahl, A., A. Gnanadesikan, and M.-A. Pradal, 2019: Variations in ocean deoxygenation across earth system models: isolating the role of parameterized lateral mixing. *Global Biogeochemical Cycles*, **33 (6)**, 703–724.

Bryan, F., 1987: Parameter sensitivity of primitive equation ocean general circulation models. *Journal of Physical Oceanography*, **17 (7)**, 970–985.

Burls, N. J., A. V. Fedorov, D. M. Sigman, S. L. Jaccard, R. Tiedemann, and G. H. Haug, 2017: Active Pacific meridional overturning circulation (pmoc) during the warm Pliocene. *Science Advances*, **3 (9)**, e1700 156.

De Boer, A., D. M. Sigman, J. Toggweiler, and J. Russell, 2007: Effect of global ocean temperature change on deep ocean ventilation. *Paleoceanography*, **22 (2)**, PA2010, https://doi.org/doi:10.1029/2005PA001242.

De Boer, A., J. Toggweiler, and D. M. Sigman, 2008: Atlantic dominance of the meridional overturning circulation. *Journal of Physical Oceanography*, **38 (2)**, 435–450.

De Boer, A. M., A. Gnanadesikan, N. R. Edwards, and A. J. Watson, 2010: Meridional density gradients do not control the Atlantic overturning circulation. *Journal of Physical Oceanography*, **40 (2)**, 368–380.

DeVries, T., and F. Primeau, 2011: Dynamically and observationally constrained estimates of water-mass distributions and ages in the global ocean. *Journal of Physical Oceanography*, **41 (12)**, 2381–2401.

Dhooge, A., W. Govaerts, Y. A. Kuznetsov, H. G. E. Meijer, and B. Sautois, 2008: New features of the software MatCont for bifurcation analysis of dynamical systems. *Mathematical and Computer Modelling of Dynamical Systems*, **14 (2)**, 147–175.

Doedel, E., and L. S. Tuckerman, 2012: *Numerical methods for bifurcation problems and large-scale dynamical systems*, Vol. 119. Springer Science & Business Media.

Doedel, E. J., 2007: Lecture notes on numerical analysis of nonlinear equations. *Numerical Continuation Methods for Dynamical Systems: Path following and boundary value problems*, 1–49.

Fabiani, G., F. Calabrò, L. Russo, and C. Siettos, 2021: Numerical solution and bifurcation analysis of nonlinear partial differential equations with extreme learning machines. *Journal of Scientific Computing*, **89**, 1–35.

Ferreira, D., J. Marshall, and J.-M. Campin, 2010: Localization of deep water formation: Role of atmospheric moisture transport and geometrical constraints on ocean circulation. *Journal of Climate*, **23 (6)**, 1456–1476.

Ferreira, D., and Coauthors, 2018: Atlantic-Pacific asymmetry in deep-water formation. *Annual Review of Earth and Planetary Sciences*, **46 (1)**, 1–26.

Ford, H. L., N. J. Burls, P. Jacobs, A. Jahn, R. P. Caballero-Gill, D. A. Hodell, and A. Fedorov, 2022: Sustained mid-Pliocene warmth led to deep water formation in the North Pacific. *Nature Geoscience*, 1–6.

Fučkar, N. S., and G. K. Vallis, 2007: Interhemispheric influence of surface buoyancy conditions on a circumpolar current. *Geophysical research letters*, **34 (14)**.

Galbraith, E. D., and Coauthors, 2011: Climate variability and radiocarbon in the CM2Mc earth system model. *Journal of Climate*, **24 (16)**, 4230–4254.

Gent, P. R., and J. C. Mcwilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, **20 (1)**, 150–155.

Gnanadesikan, A., 1999: A simple predictive model for the structure of the oceanic pycnocline. *Science*, **283 (5410)**, 2077–2079.

Gnanadesikan, A., R. Kelson, and M. Sten, 2018: Flux correction and overturning stability: Insights from a dynamical box model. *Journal of Climate*, **31 (22)**, 9335–9350.

Gnanadesikan, A., M.-A. Pradal, and R. Abernathey, 2015: Exploring the isopycnal mixing and helium–heat paradoxes in a suite of earth system models. *Ocean Science*, **11 (4)**, 591–605.

Hahn, L. C., K. C. Armour, M. D. Zelinka, C. M. Bitz, and A. Donohoe, 2021: Contributions to polar amplification in CMIP5 and CMIP6 models. *Frontiers in Earth Science*, **9**, 710 036.

Hallberg, R., and A. Gnanadesikan, 2001: An exploration of the role of transient eddies in determining the transport of a zonally reentrant current. *Journal of Physical Oceanography*, **31 (11)**, 3312–3330.

Huang, R. X., J. R. Luyten, and H. M. Stommel, 1992: Multiple equilibrium states in combined thermal and saline circulation. *Journal of physical oceanography*, **22 (3)**, 231–246.

Johnson, G. C., and H. L. Bryden, 1989: On the size of the antarctic circumpolar current. *Deep Sea Research Part A. Oceanographic Research Papers*, **36 (1)**, 39–53.

Johnson, H. L., P. Cessi, D. P. Marshall, F. Schloesser, and M. A. Spall, 2019: Recent contributions of theory to our understanding of the Atlantic meridional overturning circulation. *Journal of Geophysical Research: Oceans*, **124 (8)**, 5376–5399.

Johnson, H. L., D. P. Marshall, and D. A. Sproson, 2007: Reconciling theories of a mechanically driven meridional overturning circulation with thermohaline forcing and multiple equilibria. *Climate Dynamics*, **29 (7)**, 821–836.

Jones, C. S., and P. Cessi, 2016: Interbasin transport of the meridional overturning circulation. *Journal of Physical Oceanography*, **46 (4)**, 1157–1169.

Jones, C. S., and P. Cessi, 2017: Size matters: Another reason why the Atlantic is saltier than the Pacific. *Journal of Physical Oceanography*, **47 (11)**, 2843–2859.

Klinger, B. A., and T. W. Haine, 2019: *Ocean circulation in three dimensions*. Cambridge University Press Cambridge.

Levermann, A., and J. J. Fürst, 2010: Atlantic pycnocline theory scrutinized using a coupled climate model. *Geophysical Research Letters*, **37 (14)**, L14 602, https://doi.org/doi:10.1029/2010GL044180.

Liu, W., S.-P. Xie, Z. Liu, and J. Zhu, 2017: Overlooked possibility of a collapsed Atlantic Meridional Overturning Circulation in warming climate. *Science Advances*, **3 (1)**, e1601 666.

Locarnini, R. A., and Coauthors, 2013: *Volume 1: Temperature*. World Ocean Atlas 2013, NOAA Atlas NESDIS, S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 73.

Lumpkin, R., and K. Speer, 2007: Global ocean meridional overturning. *Journal of Physical Oceanography*, **37 (10)**, 2550–2562.

Marinov, I., A. Gnanadesikan, J. Toggweiler, and J. L. Sarmiento, 2006: The Southern Ocean biogeochemical divide. *Nature*, **441 (7096)**, 964–967.

Nikurashin, M., and G. Vallis, 2011: A theory of deep stratification and overturning circulation in the ocean. *Journal of Physical Oceanography*, **41 (3)**, 485–502.

Nilsson, J., P. L. Langen, D. Ferreira, and J. Marshall, 2013: Ocean basin geometry and the salinification of the atlantic ocean. *Journal of Climate*, **26 (16)**, 6163–6184.

Park, Y.-G., and J. Whitehead, 1999: Rotating convection driven by differential bottom heating. *Journal of physical oceanography*, **29 (6)**, 1208–1220.

Pradal, M.-A., and A. Gnanadesikan, 2014: How does the Redi parameter for mesoscale mixing impact global climate in an Earth System Model? *Journal of Advances in Modeling Earth Systems*, **6 (3)**, 586–601.

Rae, J. W., M. Sarnthein, G. L. Foster, A. Ridgwell, P. M. Grootes, and T. Elliott, 2014: Deep water formation in the North Pacific and deglacial CO2 rise. *Paleoceanography*, **29 (6)**, 645–667.

Rafter, P. A., and Coauthors, 2022: Global reorganization of deep-sea circulation and carbon storage after the last ice age. *Science Advances*, **8 (46)**, eabq5434.

Ragen, S., K. C. Armour, L. Thompson, A. Shao, and D. Darr, 2022: The role of Atlantic basin geometry in meridional overturning circulation. *Journal of Physical Oceanography*, **52 (3)**, 475–492.

Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, **12 (10)**, 1154–1158.

Rooth, C., 1982: Hydrology and ocean circulation. *Progress in Oceanography*, **11 (2)**, 131–149.

Schloesser, F., 2020: The Atlantic meridional overturning circulation and the cabbeling effect. *Journal of Physical Oceanography*, **50 (9)**, 2561–2572.

Sleeman, J., and Coauthors, 2023a: A generative adversarial network for climate tipping point discovery (tip-gan). *arXiv preprint arXiv:2302.10274*.

Sleeman, J., and Coauthors, 2023b: Using artificial intelligence to aid scientific discovery of climate tipping points. *arXiv preprint arXiv:2302.06852*.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13 (2)**, 224–230.

Talley, L. D., 1997: North Pacific Intermediate Water transports in the mixed water region. *Journal of physical oceanography*, **27 (8)**, 1795–1803.

Toggweiler, J., and B. Samuels, 1993: Is the magnitude of the deep outflow from the Atlantic Ocean actually governed by Southern Hemisphere winds? *The global carbon cycle*, Springer, 303–331.

Toggweiler, J., and B. Samuels, 1995: Effect of Drake Passage on the global thermohaline circulation. *Deep Sea Research Part I: Oceanographic Research Papers*, **42 (4)**, 477–500.

Trenberth, K. E., Y. Zhang, J. T. Fasullo, and L. Cheng, 2019: Observation-based estimates of global and basin ocean meridional heat transport time series. *Journal of Climate*, **32 (14)**, 4567–4583.

Tsukernik, M., and A. H. Lynch, 2013: Atmospheric meridional moisture flux over the southern ocean: A story of the amundsen sea. *Journal of Climate*, **26 (20)**, 8055–8064.

Tziperman, E., J. Toggweiler, K. Bryan, and Y. Feliks, 1994: Instability of the thermohaline circulation with respect to mixed boundary conditions: Is it really a problem for realistic models? *Journal of Physical Oceanography*, **24 (2)**, 217–232.

Warren, B. A., 1983: Why is no deep water formed in the North Pacific? *Journal of Marine Research*, **41 (2)**, 327–347.

Wolfe, C. L., and P. Cessi, 2011: The adiabatic pole-to-pole overturning circulation. *Journal of Physical Oceanography*, **41 (9)**, 1795–1810.

Youngs, M. K., R. Ferrari, and G. R. Flierl, 2020: Basin-width dependence of northern deep convection. *Geophysical Research Letters*, **47 (15)**, e2020GL089 135.

Yu, L., X. Jin, S. A. Josey, T. Lee, A. Kumar, C. Wen, and Y. Xue, 2017: The global ocean water cycle in atmospheric reanalysis, satellite, and ocean salinity. *Journal of Climate*, **30 (10)**, 3829–3852.

Zhang, R., and T. L. Delworth, 2005: Simulated tropical response to a substantial weakening of the Atlantic thermohaline circulation. *Journal of climate*, **18 (12)**, 1853–1860.

Zweng, M., and Coauthors, 2013: *Volume 2: Salinity*. World Ocean Atlas 2013, NOAA Atlas NESDIS, S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 73.