A Theory for Self-sustained Multicentennial Oscillation of the Atlantic Meridional Overturning Circulation. Part II: Role of Temperature

Kunpeng Yang[1,2], Haijun Yang*[1,2] and Yang Li[3]

[1]*Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Science and CMA-FDU Joint Laboratory of Marine Meteorology, Fudan University, Shanghai, 200438, China.*

[2]*Shanghai Scientific Frontier Base for Ocean-Atmosphere Interaction Studies, Fudan University, Shanghai 200438, China.*

[3]*Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China.*

Journal of Climate

Submitted

January 30, 2023

**Corresponding author address*: Haijun Yang, Department of Atmospheric and Oceanic Sciences, Fudan University, 2005 Songhu Road, Shanghai China, 200438.

Email: yanghj@fudan.edu.cn.

Abstract

In the first part of our research on self-sustained multicentennial oscillation of the Atlantic meridional overturning circulation (AMOC), we used a hemispheric box model considering only the salinity equations. In this follow-up paper, we consider both thermal and saline processes in the box model, so as to investigate the role of temperature in multicentennial AMOC oscillation. The thermal processes exert mainly three effects: shortening the oscillation period, stabilizing subpolar stratification and thus the oscillation system. These three effects are caused by the fast surface temperature restoring process, the stabilizing subpolar temperature stratification, and the negative temperature advection feedback, respectively. Nonlinear restraining effect from enhanced subpolar mixing, or a nonlinear relation between AMOC anomaly and meridional difference of density anomaly, is still needed to realize a self-sustained oscillation, whose mechanism can be generalized as follows: a combination of a linearly growing oscillation dominated by linear advection and a nonlinear restraining process. This study advances the theory reported in the first part of this research. Linear stability analyses reveal that the eigenmode of the system is sensitive to model geometry, flow properties, and meridional differences of sea-surface temperature (SST) and sea-surface salinity (SSS). Our theoretical results suggest that, a smaller (larger) meridional SST (SSS) difference weakens (strengthens) the negative temperature (positive salinity) advection feedback which may lead to a less stable AMOC. Such heuristic findings may be expected in the future due to more intense warming and freshwater hosing at the high latitudes of the Northern Hemisphere.

Keywords: AMOC, Box model, Nonlinearity, Self-sustained oscillation, Temperature feedback, Multicentennial timescale

1. Introduction

Compared to the glacial-interglacial climate variation, the mid-Holocene to pre-industrial climate has been considered to be relatively stable (Grootes et al. 1993), thanks to the lack of drastic fluctuation of natural forcing (Otto-Bliesner et al. 2017) and of anthropogenic forcing. Although the natural forced variability is important, it is reasonable to deduce that internal variability was crucial for climate variability during that period. On the multicentennial timescale, it has been suggested that climate variability is strongly linked to the Atlantic meridional overturning circulation (AMOC) (Oppo et al. 2003; Hall et al. 2004; Miettinen et al. 2012; Chabaud et al. 2014; Ayache et al. 2018). However, the fundamental mechanism governing the multicentennial AMOC internal variability remains unclear.

Multicentennial AMOC variability has been reported in pre-industrial control simulations of several coupled climate models. Mechanisms provided in these studies can be categorized into at least three groups. The first group relates the multicentennial AMOC variability to the Arctic Ocean. Jiang et al. (2021) identified a 200-year period AMOC oscillation in the IPSL-CM6-LR model, and suggested it was due to sea-ice-induced salinity anomaly exchange between the Arctic Ocean and North Atlantic convection region. Utilizing the EC-Earth3 model with the same ocean component (NEMO 3.6) to that of the IPSL-CM6-LR, Meccia et al. (2022) found a 150-year AMOC oscillation bearing a similar mechanism. Oscillation periods of the first group of studies are hardly distinguishable from the centennial timescale, probably because their mechanisms are merely related to surface ocean processes instead of deeper ocean processes. The second group emphasizes the roles of the Southern Ocean and processes at depth. Park and Latif (2008) reported a multicentennial peak in the AMOC strength spectrum of their Kiel Climate Model. Their following studies attributed such variability to a bipolar ocean seesaw teleconnection mechanism (Martin et al. 2013, 2015); that is, an increase in the Antarctic Bottom Water (AABW) formation caused by a sudden strengthening in Southern Ocean deep convection weakens the North Atlantic Deep Water (NADW) formation and hence the AMOC strength. Drastic Southern Ocean deep convection is triggered when deep Southern Ocean heat accumulation becomes too extreme, whose timescale is set by the advection of warm water from the North Atlantic. A slightly different example can be seen in Delworth and Zeng (2012), where an AMOC oscillation with a period of 200-500 years was discerned in their GFDL-CM2.1 simulation. The AMOC variation was directly controlled by salinity anomaly advected from the Southern Ocean all the way to the NADW formation sites, different from the aforementioned bipolar seesaw mechanism. The third group proposes that no inter-hemispheric or Arctic Ocean related process is needed to sustain the multicentennial AMOC oscillation. Li and Yang (2022) (hereafter LY22) found a 300-400-year AMOC oscillation in a CESM1 control simulation, where the salinity advection feedback between the subtropical and subpolar North Atlantic operates as the essential mechanism. As a side note, the AMOC oscillation in Vellinga and Wu (2004) is driven by atmosphere-ocean feedback instead of internal oceanic processes, which leads to its centennial timescale, excluding itself from the multicentennial paradigm. Inconsistency between those mechanisms calls for theoretical model studies.

Theoretical studies of the AMOC tend to focus on its thermohaline portion, which is buoyancy-driven thus determined by two foremost elements, temperature and salinity variations. Adopting different restoring coefficients for temperature and salinity in a simple two-box model, Stommel (1961) found multiple equilibria in the system, which stems from competing effects of thermal and saline processes. The follow-up theoretical studies inspired by Stommel's idea primarily focused on multi-equilibria phenomenon due to the combined effect of temperature and salinity (Welander 1982; Joyce 1991; Huang et al. 1992; Cessi 1994; Scott et al. 1999; Zhang et al. 2002; Lucarini and Stone 2005; Colin de Verdière 2007), revealing that both thermal and saline processes are indispensable. Given the relatively stable Holocene climate, we focus on small amplitude and sustainable climate variation around a single equilibrium, instead of abrupt climate shift suggested by the multi-equilibria phenomenon. Griffies and Tziperman (1995) (hereafter GT95) realized a stochastically sustained AMOC oscillation in their 4-box model. Using a 3-box model without separation between upper and deeper subpolar ocean, Rivin and Tziperman (1997) (hereafter RT97) found that the AMOC oscillation in their model could be sustained by either stochastic forcing or intrinsic nonlinearity. More recently, Wei and Zhang (2022) realized a self-sustained AMOC oscillation in their revised Stommel's 2-box model that consists of an Arctic Ocean box and a North Atlantic box. However, oscillations in those studies are on the multidecadal timescale. Roebber (1995) realized a 683-year stochastically sustained AMOC oscillation in a 3-box ocean model coupled with a Lorenz atmospheric model. Yet, this oscillation timescale is dominated by diffusion instead of advection process. Scott et al. (1999) and Lucarini and Stone (2005) demonstrated that there exist multicentennial oscillations around one of the several equilibria in their inter-hemispheric models, though the oscillations were unsustainable. In a loop model, Sévellec et al. (2006) found a 170-year self-sustained AMOC oscillation, while the regrets are the indistinguishability between meridional and vertical processes due to model simplicity, and the scarcity of discussion on the role of equilibrium salinity difference advected by anomalous flow. It is therefore necessary to come up with a theoretical model directed at revealing the underlying mechanism for sustainable multicentennial AMOC internal variability.

Driven by this desire, our first publication of LY22 focused on the multicentennial AMOC oscillation in a single-hemispheric 4-box model, while we further looked into its self-sustained oscillation behavior. Since salinity variation dominates over temperature variation in regulating NADW formation (Delworth et al. 1993; Dong and Sutton 2005; Jiang et al. 2021), and for an easier access toward analytical solutions, the model in LY22 only retains salinity variation. Such approach trades temperature variation for model simplicity, and has been adopted in several theoretical studies (Winton and Sarachik 1993; Huang and Dewar 1996; Rahmstorf 1996; Cimatoribus et al. 2014; Sévellec and Fedorov 2014). However, excluding all the thermal processes is unphysical, since it has been shown that thermal processes can potentially affect the stability (Zhang et al. 1993; Nakamura et al. 1994; Rahmstorf and Willebrand 1995; Tziperman and Gildor 2002) and period (Schmidt and Mysak 1996) of an oscillatory AMOC system. Moreover, on the multicentennial timescale, the role of temperature variation in AMOC variability is not well studied. We thus add temperature variation in the salinity-only model of LY22.

In this study, we extend the 4-box salinity-only model (hereafter 4S) in LY22 to a temperature-salinity one by adding temperature equations. For studying AMOC oscillation mechanism within the framework of internal oceanic processes and maintaining an explicit picture for our theory, we employ mixed boundary conditions for both temperature and salinity equations. We aim to reveal the effects induced by thermal processes and their underlying causes. We then work on realization of self-sustained oscillation with restraining terms affiliated, in order to further explore the essential mechanism for the self-sustained AMOC oscillation. Finally, we examine the sensitivity of eigenmode to model parameters controlling model geometry, flow properties, and meridional differences of equilibrium sea-surface temperature (SST) and salinity (SSS).

This paper is structured as follows. In section 2, a 4-box temperature-salinity model (hereafter 4TS) is introduced, followed by illustration of temperature and salinity feedbacks involved. In section 3, the role of temperature equations is analyzed. In section 4, we test two ways for realizing self-sustained oscillation and come up with a more profound self-sustained AMOC oscillation mechanism. In section 5, we examine sensitivities of eigenmode's period and stability to model parameters. Summary and discussion are presented in section 6.

2. Box model

a. Model description

The model used here is a hemispheric 4-box model with identical geometry to that in LY22 (Fig. 1a). The model domain is 60° in longitude, with the tropical and subpolar boxes spanning over 0°-45°N and 45°-70°N, respectively. The AMOC moves through the boxes clockwise. We do not discuss the possibility of a reversed AMOC cell, by excluding multi-equilibria. Analogous box models have been widely used (Joyce 1991; Huang et al. 1992; Griffies and Tziperman 1995). In the 4-box salinity-only model (hereafter 4S model) of LY22, only salinity equations were used. In the 4TS model, we employ both temperature and salinity equations:

$$V_1\dot{T}_1 = q(T_4 - T_1) + V_1\gamma(T_1^* - T_1) \quad (1a)$$
$$V_2\dot{T}_2 = q(T_1 - T_2) + V_2\gamma(T_2^* - T_2) \quad (1b)$$
$$V_3\dot{T}_3 = q(T_2 - T_3) \quad (1c)$$
$$V_4\dot{T}_4 = q(T_3 - T_4) \quad (1d)$$
$$V_1\dot{S}_1 = q(S_4 - S_1) + F_w \quad (1e)$$
$$V_2\dot{S}_2 = q(S_1 - S_2) - F_w \quad (1f)$$
$$V_3\dot{S}_3 = q(S_2 - S_3) \quad (1g)$$
$$V_4\dot{S}_4 = q(S_3 - S_4) \quad (1h)$$

It is an advection-dominated box model, with mixed boundary conditions where Haney-style restoring for SST (Haney 1971) and surface virtual salt flux (VSF) for SSS are adopted, leading to more relaxed temperature variation and freer salinity variation. Eqs. (1e-h) are used in the original 4S model in LY22. V_i, T_i, and S_i are volume, temperature, and salinity, respectively, in each box. Dot over variable denotes its temporal derivative. q stands for AMOC strength. F_w is surface VSF, representing surface freshwater flux in reality. T_1^* and T_2^* correspond to the restoring temperatures for boxes 1 and 2, respectively. The Newtonian restoring coefficient γ is the reciprocal of the restoring timescale for T_1 and T_2.

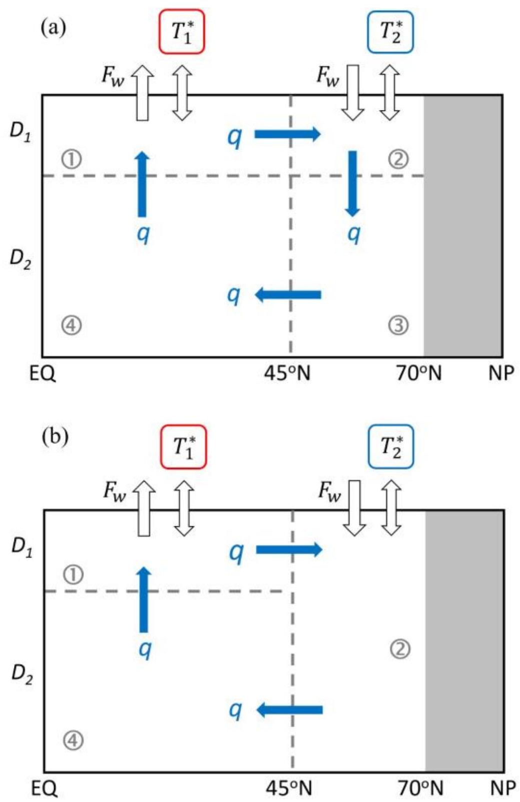

FIG. 1. Schematics of temperature-salinity box models. (a) The 4-box model; (b) the 3-box model reduced from the 4-box one. Numbers ①, ②, ③, and ④ denote the ocean boxes. Boxes 1 and 4 stand for the upper and deeper tropical ocean boxes, respectively, while boxes 2 and 3 stand for the upper and deeper subpolar ocean boxes, respectively. D_1 and D_2 represent the upper and deeper ocean box depths, respectively. Net VSF into (out of) the surface tropical (subpolar) ocean box is represented by F_w. T_1^* (T_2^*) is the restoring temperature of the tropical (subpolar) ocean box. q represents the AMOC.

By setting the terms on the left-hand side of Eq. (1) to 0, the equilibrium solutions at steady state can be written as follows:

$$\overline{T_1} = T_1^* - \frac{\overline{q}V_2(T_1^* - T_2^*)}{\overline{q}(V_1+V_2)+V_1V_2\gamma}, \qquad \overline{T_2} = \frac{V_1T_1^*+V_2T_2^*-V_1\overline{T_1}}{V_2} = \overline{T_3} = \overline{T_4} \tag{2a}$$

$$\overline{S_2} = \overline{S_3} = \overline{S_4}, \qquad F_w = \overline{q}\left(\overline{S_1} - \overline{S_2}\right) \tag{2b}$$

Variables with overbar denote their equilibrium values. Following LY22, the upper ocean boxes depth D_1, deeper ocean boxes depth D_2, and total depth D are still 500, 3500, and 4000 m, respectively. Note that our hemispheric model only incorporates the AMOC recirculating in the Northern Hemisphere. Consequently, $\overline{q}$ is set to 10 Sv, which is lower than the measured value of 20 Sv (McCarthy et al. 2015). Straightway, T_1^* and T_2^* are considered to be close to the averaged realistic SSTs in the tropical and subpolar regions, set to 25℃ and 7℃, respectively. Values for $\overline{S_1}$, $\overline{S_2}$, $\overline{S_3}$, and $\overline{S_4}$ are identical to that in LY22, set to 36.0, 33.5, 33.5, and 33.5 psu, respectively. Therefore, F_w is $25.0\,psu \cdot Sv$ following Eq. (2b). The choices of restoring temperatures and equilibrium salinities are based on the CESM1 control simulation analyzed in LY22 (Yang et al. 2015).

Typically, restoring timescale $1/\gamma$ is set to 1-2 months for a surface layer of a few tens of meters in depth (Marotzke and Willebrand 1991; Weaver and Sarachik 1991; Pierce 1996). Here, we permit the surface temperature restoring to happen over the full depth range of boxes 1 and 2; therefore, a much longer restoring timescale has to be used. A reasonable range of $1/\gamma$ should be around one year to several years. We set $\gamma = 3.171 \times 10^{-8}\ s^{-1}$, corresponding to 1-year restoring timescale.

The total AMOC strength q can be separated into a given equilibrium part $\overline{q}$ and a calculated anomalous part q'. We consider a linear relation between q' and thickness-weighted meridional difference of density anomaly $\Delta\rho'$. Both q' and $\Delta\rho'$ can be decomposed into temperature-driven part (q'_T, $\Delta\rho'_T$) and salinity-driven part (q'_S, $\Delta\rho'_S$); therefore, we have:

$$q = \overline{q} + q' \tag{3a}$$

$$q' = q'_T + q'_S = \lambda\Delta\rho'_T + \lambda\Delta\rho'_S = \lambda\Delta\rho' \tag{3b}$$

where

$$\Delta\rho'_T = -\rho_0\alpha[\delta(T'_2 - T'_1) + (1-\delta)(T'_3 - T'_4)] \tag{4a}$$

$$\Delta\rho'_S = \rho_0\beta[\delta(S'_2 - S'_1) + (1-\delta)(S'_3 - S'_4)] \tag{4b}$$

$$\delta = \frac{V_1}{V_1+V_4} = \frac{V_2}{V_2+V_3} = \frac{D_1}{D} \tag{4c}$$

The sensitivity of q' to $\Delta\rho'$ is represented by a linear closure coefficient λ. ρ_0, α, and β are the reference density, thermal expansion, and haline contraction coefficients for seawater, respectively. T'_i and S'_i are the temperature and salinity anomalies of box i, respectively. A summary of the standard parameter values is provided in Table 1. These basic model parameters are independent of each other, and are the same as those in LY22, except for the temperature related parameters.

TABLE 1. Standard values of the parameters used.

Symbol	Physical Significance	Value
V_2	Volume of box 2	$2.8 \times 10^{15}\ m^3$

V_1, V_3, V_4	Volumes of boxes 1, 3 and 4, respectively	$5V_2$, $7V_2$, $35V_2$
D_1, D_2, D	Thicknesses of the upper, deeper ocean boxes, and the entire ocean boxes	$500, 3500, 4000\ m$
T_1^*, T_2^*	Restoring temperatures of boxes 1 and 2	25℃, 7℃
γ	Restoring coefficient of boxes 1 and 2	$3.171 \times 10^{-8}\ s^{-1}$
$\overline{S_1}, \overline{S_2}, \overline{S_3}, \overline{S_4}$	Equilibrium salinities of boxes 1, 2, 3, and 4	$36, 33.5, 33.5, 33.5\ psu$
$\overline{q}$	Equilibrium strength of AMOC	$10\ Sv\ (10^6\ m^3\ s^{-1})$
λ	Linear closure coefficient	$12\ Sv \cdot kg^{-1}\ m^3$
ρ_0	Reference seawater density	$1.00\ \times 10^3\ kg\ m^{-3}$
α	Thermal expansion coefficient	1.468×10^{-4} ℃$^{-1}$
β	Haline contraction coefficient	$7.61\ \times 10^{-4}\ psu^{-1}$

We linearize Eq. (1) as follows:

$$V_1\dot{T}_1' = q'\left(\overline{T_4} - \overline{T_1}\right) + \overline{q}(T_4' - T_1') - V_1\gamma T_1' \quad (5a)$$

$$V_2\dot{T}_2' = q'\left(\overline{T_1} - \overline{T_2}\right) + \overline{q}(T_1' - T_2') - V_2\gamma T_2' \quad (5b)$$

$$V_3\dot{T}_3' = \overline{q}(T_2' - T_3') \quad (5c)$$

$$V_4\dot{T}_4' = \overline{q}(T_3' - T_4') \quad (5d)$$

$$V_1\dot{S}_1' = q'(\overline{S}_4 - \overline{S}_1) + \overline{q}(S_4' - S_1') \quad (5e)$$

$$V_2\dot{S}_2' = q'(\overline{S}_1 - \overline{S}_2) + \overline{q}(S_1' - S_2') \quad (5f)$$

$$V_3\dot{S}_3' = \overline{q}(S_2' - S_3') \quad (5g)$$

$$V_4\dot{S}_4' = \overline{q}(S_3' - S_4') \quad (5h)$$

Eqs. (5e-h) are the linearized 4S model of LY22. In LY22, it was assumed that under an extremely strong vertical mixing between subpolar boxes 2 and 3, the 4S model can be reduced to a 3-box salinity-only model (hereafter 3S). Applying the same treatment to the 4TS model, a 3-box temperature-salinity model (hereafter 3TS; Fig. 1b) can be obtained. Now, Eqs. (4a-c) become,

$$\Delta\rho_T' = -\rho_0\alpha[T_2' - \delta T_1' - (1-\delta)T_4'] \quad (6a)$$

$$\Delta\rho_S' = \rho_0\beta[S_2' - \delta S_1' - (1-\delta)S_4'] \tag{6b}$$

$$\delta = \frac{V_1}{V_1+V_4} = \frac{D_1}{D} \tag{6c}$$

and Eqs. (5a-h) are reduced to:

$$V_1\dot{T}_1' = q'\left(\overline{T_4} - \overline{T_1}\right) + \overline{q}(T_4' - T_1') - V_1\gamma T_1' \tag{7a}$$

$$V_2\dot{T}_2' = q'\left(\overline{T_1} - \overline{T_2}\right) + \overline{q}(T_1' - T_2') - V_2\gamma T_2' \tag{7b}$$

$$V_4\dot{T}_4' = \overline{q}(T_2' - T_4') \tag{7c}$$

$$V_1\dot{S}_1' = q'(\overline{S}_4 - \overline{S}_1) + \overline{q}(S_4' - S_1') \tag{7d}$$

$$V_2\dot{S}_2' = q'(\overline{S}_1 - \overline{S}_2) + \overline{q}(S_1' - S_2') \tag{7e}$$

$$V_4\dot{S}_4' = \overline{q}(S_2' - S_4') \tag{7f}$$

Here, boxes 2 and 3 in the 4-box models are well mixed to become a new box 2 in the 3-box models, and V_2 in the 3-box models equals to the sum of V_2 and V_3 in the 4-box models. For consistency, other background state parameters in Table 1 are identical in the 3-box and 4-box models.

Unless otherwise mentioned, results shown in this paper are obtained from forward numerical integration of Eq. (5) (the 4TS model) with the standard parameters listed in Table 1. The fourth-order Runge-Kutta method is used to solve Eq. (5), with $S_1'(t=0) = -0.02\ psu$ at the first time step, and $S' = T' = 0$ thereafter. The integration time step is 7.2 days, and the total integration length exceeds 10000 years, but only the initial parts are shown. Annual mean data is used for analysis.

b. Stability analysis

Let us first examine the eigenvalues of the 4TS model. Table 2 lists the eight eigenvalues of Eq. (5) under the parameters in Table 1. The eigenvalues in the 4S model of LY22 under the same parameters are also listed in Table 2 for comparison.

TABLE 2. Eigenvalues ($10^{-10}\ s^{-1}$) for the 4TS and 4S models under the parameters of Table 1.

4TS	4S	Physical Significance
$-0.55 \pm 6.59i$	$0.31 \pm 5.83i$	Oscillatory mode

0	0	Zero mode
-366	—	Damped mode
-324	—	Damped mode
-37.4	-37.4	Damped mode
-5.28	—	Damped mode
-0.78	—	Damped mode

There is still a pair of conjugate eigenvalues ($-0.55 \pm 6.59i$) in the 4TS model. The weakly unstable oscillatory mode ($0.31 \pm 5.83i$) in the 4S model becomes a weakly damped oscillatory mode in the 4TS model; that is, the e-folding time changes from 1025 years for growing oscillation in the 4S model to 576 years for decaying oscillation in the 4TS model, and the period is shortened from 340 years to 300 years. This seems to suggest that the thermal processes have a stabilizing effect on the system, and shorten the oscillation period slightly. The zero mode (eigenvalue 0) represents the climatological mean state. The other five eigenvalues in the 4TS model represent five purely damped modes, which are not of our concern here.

The stability of the box model system is strongly dependent on the linear closure parameter λ, i.e., the sensitivity of AMOC anomaly to the meridional difference of density anomaly as formulated in Eq. (3b). The critical role of λ and its physical explanation can be found in LY22. In this paper, we simply solve Eqs. (5) and (7) numerically to investigate how λ affects the stabilities and periods of the 4TS and 3TS models.

Figure 2 shows dependences of real and imaginary parts of eigenvalue ω on λ. The results from the 4S and 3S models in LY22 are also plotted in Fig. 2 for comparison. The intersections between line $Re(\omega) = 0/Im(\omega) = 0$ and the stability diagram of each model, (λ_C, 0) corresponds to the stability threshold (Fig. 2b), (λ_1, 0) and (λ_2, 0), to the lower and upper limits for the existence of the imaginary part (Fig. 2a), respectively. Their values are listed in Table 3. When $\lambda \geq \lambda_2$ or $\lambda \leq \lambda_1$, only purely growing or damped modes without oscillatory potential exist, suggested by the corresponding positive or negative real part (Fig. 2b). When $\lambda_1 < \lambda < \lambda_2$, the system exhibits oscillatory behavior because of the presence of the imaginary part (Fig. 2a). With the increase of λ, the models have the tendency to change from a damped to a growing oscillation. In comparison with the 4S and 3S models of LY22, permitting the temperature variation in the system has at least three consequences:

(a) An increase of oscillation frequency, evidenced by the larger imaginary parts in the 4TS and 3TS models (Fig. 2a, orange curves) than in the 4S and 3S models (Fig. 2a, black curves).

(b) An overall stabilization of the system, evidenced by the larger λ_c in the 4TS and 3TS models than in the 4S and 3S models listed in Table 3, and the smaller real parts in the 4TS and 3TS models (Fig. 2c, orange lines) than in the 4S and 3S models (Fig. 2c, black lines).

(c) Stabilization of subpolar stratification, evidenced by the larger λ_c in the 4TS model than in the 3TS model, but smaller λ_c in the 4S model than in the 3S model.

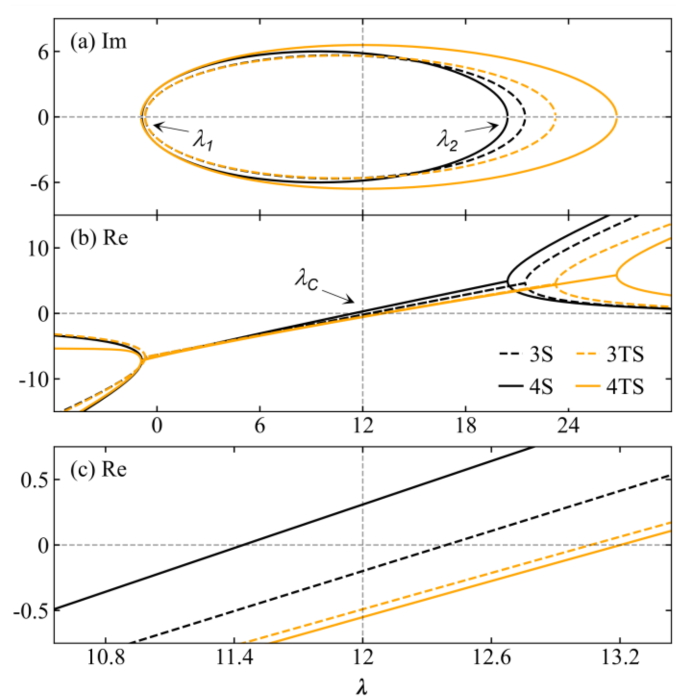

FIG. 2. Dependences of (a) imaginary parts and (b) real parts of eigenvalue ω on λ in the 4TS (solid orange curve), 3TS (dashed orange curve), 4S (solid black curve), and 3S (dashed black curve) models. (c) is the zoomed-in version of (b) near line $Re(\omega) = 0$. Results of the 4S and 3S models are from LY22. The units of the ordinate are 10^{-10} s^{-1}. The values of the other parameters are the same as those listed in Table 1. The vertical dashed line denotes the situation under the standard value $\lambda = 12\ Sv \cdot kg^{-1}m^3$.

TABLE 3. Values for λ_c, λ_1, and λ_2 (units: $Sv \cdot kg^{-1}\ m^3$) in different box models.

	4TS	3TS	4S	3S
λ_C	13.20	13.06	11.44	12.39
λ_1, λ_2	-0.92, 26.80	-0.70, 23.24	-0.89, 20.44	-0.69, 21.46

The stability analyses provide us mathematical fundamentals, showing how the behaviors of the system change when the thermal processes are included. Physical insight into why the system behaviors change will be deliberated next.

3. Effects of temperature equations

a. Temperature feedbacks

By permitting temperature variation, two feedbacks between thermal processes and AMOC are introduced in the 4-box model: the negative temperature advection feedback and the positive restoring advection feedback. Numerical results reveal that the former overcomes the latter, evidenced by the transition from a growing oscillation (Fig. 3a, green curve) into a damped oscillation (Fig. 3a, black and orange curves). Let us illustrate these two feedbacks using box 2 (Fig. 3b). Starting with a positive perturbation q', the anomalous advection $q'(\overline{T_1} - \overline{T_2})$ transports more warm water northward; T_2' is increased and thus $\Delta\rho_T'$ is lowered, causing a final decrease in q_T'. This is the negative temperature advection feedback, which can be physically and mathematically derived from Eqs. (5b) and (6a), and can be further illustrated by the lead/lag correlation between $q'(\overline{T_1} - \overline{T_2})$ and q_T': they have a negative correlation coefficient at lag 0 (Fig. 3e, orange curve). The correlation coefficient at lag 0 cannot be treated as a verification of causality between AMOC and terms on the right-hand side of Eq. (5). However, it provides an intuitional representation of the positive/negative feedback after the physical interpretation of model equations is given. The increased T_2' caused by the initial positive perturbation in q' also triggers relaxation via anomalous restoring $-V_2\gamma T_2'$, whose strength is proportional to the restoring coefficient γ. This limits the growth of positive T_2' itself, limiting the decreases of $\Delta\rho_T'$ and q_T'. Consequently, the net effect of the restoring advection feedback itself is to increase q', demonstrating it is a positive feedback. The positive restoring advection feedback is further illustrated in Fig. 3e (green curve): $-V_2\gamma T_2'$ has a positive correlation coefficient with q_T' at lag 0. Note that the restoring advection feedback is driven by the temperature advection feedback and only partly hampers the latter; thus, the net effect of temperature feedbacks is a negative feedback. There is another feedback coming from $\overline{q}(T_1' - T_2')$ of Eq. (5b), which is positive (Fig. 3e, dotted black curve), but rather weak (Fig. 3b, dotted black curve). The negative temperature advection feedback and the positive restoring advection feedback are both local feedbacks, since they are free of anomalies transported from one box to another.

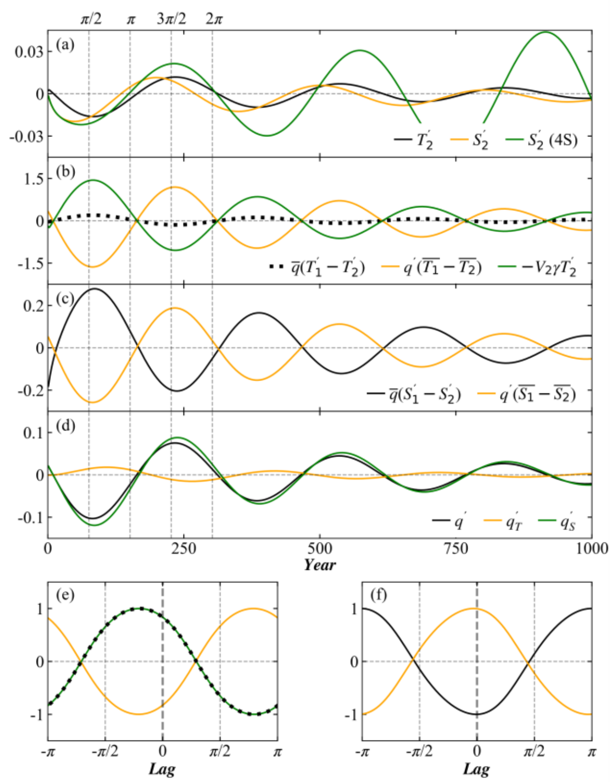

FIG. 3. (a) Damped and growing oscillations in the 4TS and 4S models under the standard parameters in Table 1. Black, orange, and green curves are the time series of T_2' (units: °C), S_2' (units: psu) in the 4TS model, and S_2' in the 4S model, respectively. (b) Time series for temperature terms (units: $Sv \cdot$ °C) on the right-hand side of Eq. (5b); (c) time series for salinity terms (units: $Sv \cdot psu$) on the right-hand side of Eq. (5f); and (d) time series for q', q_T', and q_S' (units: Sv) in the 4TS model. The vertical dashed lines in (a)-(d) mark the locations of π/2, π, 3π/2, and 2π of the period (300 years) in the 4TS model. (e) Lead/lag correlation coefficients between q_T' and $\overline{q}(T_1' - T_2')$ (dotted black curve), $q'\left(\overline{T_1} - \overline{T_2}\right)$ (solid orange curve), and $-V_2\gamma T_2'$ (solid green curve) in the 4TS model. (f) Lead/lag correlation coefficients between q_S' and $\overline{q}(S_1' - S_2')$ (solid black curve), and between q_S' and $q'\left(\overline{S_1} - \overline{S_2}\right)$ (solid orange curve) in the 4TS model. In (e)-(f), negative lag represents q' lags the other term.

The salinity advection feedback in the 4TS model is nearly identical to that in the 4S model of LY22. The positive and negative feedbacks come from terms $q'\left(\overline{S_1} - \overline{S_2}\right)$ and $\overline{q}(S_1' - S_2')$ (Figs. 3c, f), respectively. Note that q' is the sum of salinity-induced q_S' and temperature-induced q_T'. These two components are roughly out of phase, with the former being much bigger than the latter (Fig. 3d); it suggests that the salinity advection feedback has more remarkable effect on the AMOC than the temperature advection feedback does. However, the phase of q' is no longer identical to that of salinity-induced q'_S, and the variation of q' is obviously offset by temperature-induced q'_T (Fig. 3d), indicating that the behavior of the system is affected by the thermal processes.

b. Role of restoring advection feedback

The restoring advection feedback in the temperature equations can significantly affect the system's behavior. To understand this better, let us first examine how the restoring timescale affects the temperature advection feedback. Based on Eq. (2a), we have,

$$\overline{T_1} - \overline{T_2} = (T_1^* - T_2^*) / \left[\frac{\overline{q}(V_1+V_2)}{V_1 V_2 \gamma} + 1\right] \tag{8}$$

This depicts a larger γ (or a shorter restoring timescale) causes a larger $\overline{T_1} - \overline{T_2}$, thus stronger anomalous advection of meridional difference of mean temperature $q'\left(\overline{T_1} - \overline{T_2}\right)$. However, the negative temperature advection feedback is realized through an increase in T'_2, which is in turn limited by the stronger $-V_2\gamma T'_2$; so the restoring advection feedback tends to always offset the temperature advection feedback to a certain degree (Fig. 3b), regardless of the restoring strength.

TABLE 4. Conjugate eigenmodes in the 4TS model under different γ.

γ	Eigenvalue ($10^{-10}\ s^{-1}$)	E-folding Time (Years)	Period (Years)
$\gamma_0 = 0$	$0.31 \pm 5.83i$	1025	341
$\gamma_1 = (5\ year)^{-1}$	$-1.48 \pm 7.30i$	-215	273
$\gamma_2 = (1\ year)^{-1}$	$-0.55 \pm 6.59i$	-576	302
$\gamma_3 = (0.25\ year)^{-1}$	$0.036 \pm 6.09i$	8830	327
$\gamma_4 = (1\ month)^{-1}$	$0.21 \pm 5.92i$	1492	336
$\gamma_5 = (1\ minute)^{-1}$	$0.31 \pm 5.83i$	1025	341

There are two extreme situations. When $\gamma \to 0$ or $\gamma \to \infty$, that is, when the restoring timescale for SST goes to either infinity or zero, the oscillatory eigenmode ($0.31 \pm 5.83i$) in the 4TS model (Table 4) is identical to that in the 4S model (Table 1). Under these two situations, the thermal processes have no effect on AMOC variation, and the 4TS model is practically reduced to the 4S model. In the situation with $\gamma \to 0$, Eqs. (2a) and (8) give us $\overline{T_1} = \overline{T_2} = \overline{T_3} = \overline{T_4}$. There is no restoring advection feedback, and the temperature advection feedback becomes null ($q'(\overline{T_1} - \overline{T_2}) = 0$), prohibiting any temperature variation in the system. Therefore, the linearized temperature equations (Eqs. 5a-d) are identical to the linearized salinity equations (Eqs. 5e-h), and the combined temperature and salinity equations are equivalent to the salinity equations in the 4S model. Case $\gamma \to 0$ suggests that the system can only be affected by temperature equations when the restoring advection feedback is active. In the situation with $\gamma \to \infty$, we have $\overline{T_1} = T_1^*$ and $\overline{T_2} = T_2^*$ based on Eq. (2a). The extremely strong restoring kills any temperature perturbation immediately, which makes $T_1' = T_2' = 0$ (thus $T_3' = T_4' = 0$), and the 4TS system is equivalent to a system without active thermal process; so only the saline processes matter to the system behavior. In summary, under these two extreme situations, the results from linear stability analysis suggest the variations of salinity and AMOC are identical to those of the 4S model in LY22; as a result, the corresponding figures (similar to Figs. 2-3 but with $\gamma \to 0$ or $\gamma \to \infty$) are not shown.

Figure 4 shows dependences of imaginary and real parts of the oscillatory mode on γ in the 3TS and 4TS models (orange curves). The 3TS and 4TS models exhibit a damped oscillation only when γ is neither too small nor large (Fig. 4b, orange curves), namely, when the temperature effects are strong. Compared with the 4S model (Fig. 4b, solid black line), the 4TS model is generally more stable, manifested by the negative or longer positive e-folding time. Even for a very short SST restoring timescale (one to several months) (Table 4), the positive e-folding time of the oscillatory mode in the 4TS model is still much longer than that in the 4S model. This is because for any given γ and λ, temperature-induced q_T' is always opposite to salinity-induced q_S', so that the total q' is always smaller; in other words, the AMOC sensitivity to buoyancy perturbation is always weaker in the 4TS model than in the 4S model. In addition, including the fast thermal restoring process leads to a shorter oscillation period in the 4TS model than in the 4S model (Fig. 4a), because the superimposition of a quick timescale and a slow timescale leads to a timescale in between. Since a reasonable restoring timescale is always much shorter than the multicentennial timescale, we can practically neglect the effect of restoring timescale on the oscillation period of the system.

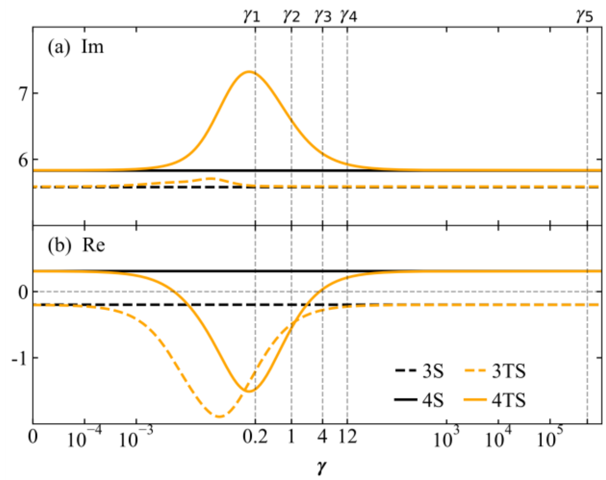

FIG. 4. Dependences of (a) positive imaginary parts and (b) real parts of the oscillatory mode on γ (units: year^{-1}) in the 4TS model (solid orange curve) and the 3TS model (dashed orange curve). The units of the ordinate are 10^{-10} s^{-1}. The vertical dashed lines from left to right denote the situations under γ_1, γ_2, γ_3, γ_4, and γ_5, respectively. The reference oscillatory modes in the 4S and 3S models are plotted as the solid and dashed black lines, respectively, which are independent of γ. Here, λ is set to 12 $Sv \cdot kg^{-1}\, m^3$. The values of the other parameters are the same as those listed in Table 1.

The restoring timescale also affects the relative stability of the 4TS and 3TS models. As shown in Fig. 1, under extremely strong vertical mixing in the subpolar ocean, the 4-box model (Fig. 1a) can be reduced to a 3-box model (Fig. 1b). LY22 showed that the 3S model is always more stable than the 4S model. Here, we find that including the thermal processes causes no obvious change of stability from the 4TS to 3TS model (Fig. 4b). To better understand the stability change, we should first recognize that whether the temperature and salinity anomalies stay in the subpolar upper or deeper ocean does not influence the meridional difference of density anomaly due to the vertically weighted volume-averaged treatment. However, the time consumed in transporting temperature and salinity anomalies from the upper to deeper ocean boxes is omitted; consequently, they are removed faster from the subpolar region in the 3-box model, which reduces their restraining and amplification effects on q'_T and q'_S, respectively. Therefore, the removals of temperature and salinity related subpolar stratification in the 3TS model have destabilizing and stabilizing effects, respectively, on the oscillation of the system.

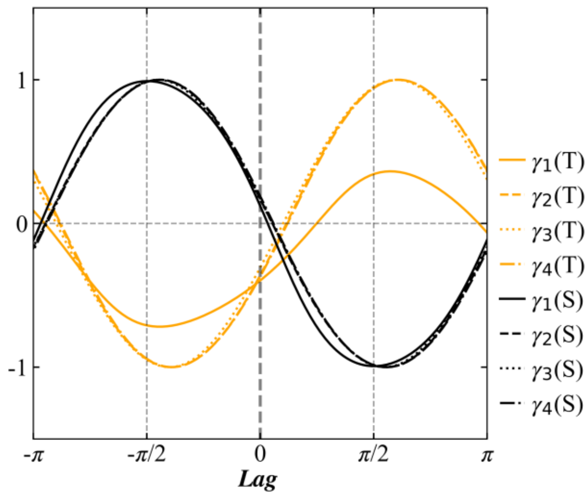

FIG. 5. Lead/lag correlation coefficients between $T_2' - T_3'$ and q_T' (orange curves) and between $S_2' - S_3'$ and q_S' (black curves) in the 4TS model under different γ. Negative lag represents q' lags the other terms. Here, λ is set to 12 $Sv \cdot kg^{-1}\ m^3$. The values of the other parameters are the same as those listed in Table 1.

In the 4TS model, the subpolar temperature (salinity) stratification $T_2' - T_3'$ ($S_2' - S_3'$) has negative (positive) correlation with q_T' (q_S') at lag 0 (Fig. 5). These correlations do not rely on the temperature restoring coefficient γ. This further confirms that the existences of subpolar temperature- and salinity-related stratification have stabilizing and destabilizing effects, respectively, on the system. Whether the total subpolar buoyancy stratification plays a stabilizing or destabilizing role, however, depends on γ. When γ lies in the range of about several years (from γ_1 to γ_2; Fig. 4b), the total subpolar buoyancy stratification plays a stabilizing role since the stabilizing effect of temperature stratification overcomes the destabilizing effect of salinity stratification. When γ is too small or too large, the temperature effect becomes weaker, while the salinity effect is not influenced. Hence, the stabilization from temperature stratification no longer overcomes the destabilization from salinity stratification, making the 3TS model more stable than the 4TS model. The 4S model can be qualitatively interpreted as the case with weak temperature effects in this study, that is, including extreme mixing in the subpolar ocean can stabilize the system. We conclude that under realistic range of γ, the 4TS model can be more stable than the 3TS model, due to stronger stabilization effect of subpolar temperature stratification than the destabilization effect of subpolar salinity stratification.

4. Realization of self-sustained oscillation

Self-sustained oscillation is still absent in the 4TS model. Under the same parameters, the 4TS model is more stable than the 4S model in LY22 (Fig. 2c), as discussed in section 3. However, this does not lead to a self-sustained oscillation in the 4TS model. Clearly, additional processes are needed. In LY22, an enhanced vertical mixing process is added explicitly in the subpolar ocean, to realize a self-sustained oscillation. To ensure that the system becomes unstable, we choose $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$ for the 4TS model from now on, and the subpolar vertical mixing is still destabilizing as a whole (Fig. 2b), in contrast to the purely stabilizing vertical salinity mixing in LY22. Will a self-sustained oscillation be realized with the destabilizing subpolar vertical mixing now? If so, what is the exact role of such vertical mixing in establishing the self-sustained oscillation?

a. Self-sustained oscillation with enhanced subpolar mixing

Similar to LY22, we introduce an enhanced mixing term between boxes 2 and 3 in the 4TS model. Eqs. (5b-c) and (5f-g) become,

$$V_2\dot{T}_2' = q'\left(\overline{T_1} - \overline{T_2}\right) + \overline{q}(T_1' - T_2') - k_m(T_2' - T_3') - V_2\gamma T_2' \qquad (9a)$$

$$V_3\dot{T}_3' = \overline{q}(T_2' - T_3') + k_m(T_2' - T_3') \qquad (9b)$$

$$V_2\dot{S}_2' = q'\left(\overline{S_1} - \overline{S_2}\right) + \overline{q}(S_1' - S_2') - k_m(S_2' - S_3') \qquad (9c)$$

$$V_3\dot{S}_3' = \overline{q}(S_2' - S_3') + k_m(S_2' - S_3') \qquad (9d)$$

The mixing coefficient k_m (units: m^3/s) is represented by:

$$k_m = \kappa q'^2 \qquad (9e)$$

where κ (units: $m^{-3}s$) is a positive constant. We set κ to $1 \times 10^{-4}\ m^{-3}s$ in this paper. No matter the sign of q', k_m is always positive and helps remove subpolar upper-ocean anomalies. Detailed physics of the enhanced mixing process was discussed in LY22. If this mixing is strong enough, the 4TS model is virtually equivalent to the 3TS one.

A growing oscillation (Fig. 6a, solid black curve) is turned into a self-sustained oscillation (Fig. 6a, solid orange curve) when enhanced subpolar mixing is included, which can be seen clearly in the phase diagram of T_2' vs S_2' (Fig. 6e, orange curve); that is, a limit cycle is formed eventually. With $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$, the intrinsic mode of the 4TS model is unstable. As q' grows (decreases), more warm and saline (cold and fresh) water is removed from the subpolar upper ocean into the deeper ocean via anomalous mixing of $-k_m(T_2' - T_3')$ and $-k_m(S_2' - S_3')$ (Figs. 6b, c, solid orange curves); thus, the anomalies exit the subpolar region more quickly. This vertical mixing has two effects. The first one is, from a nonlinear restraining view, to prevent the runaway tendency of subpolar temperature and salinity anomalies by pulling them back to a relatively confined range (Fig. 6e, orange curve); therefore, the drastic variation of q' is inhibited. The second one is, from a linear stability view, the mixing of temperature (salinity) destabilizes (stabilizes) the system, which is illustrated by the correlation coefficients at lag 0 between the mixing terms and AMOC (Fig. 6d). Permitting only salinity variation in LY22, it is natural to have an impression that it is the linearly stabilizing effect of the subpolar salinity mixing that turns the growing oscillation into a self-sustained one. With the presence of temperature variation and thus overall destabilizing subpolar vertical mixing, we propose that the realization of self-sustained oscillation is not controlled by subpolar mixing, whether it is stabilizing or not, but depends on the nonlinear restraining effect of such mixing. Note that the 3-box models are fully linear systems; thus, no self-sustained oscillation can be realized, no matter how stabilizing the subpolar vertical mixing is. The κ chosen here is one order smaller than that used in LY22, suggesting that even weak subpolar vertical mixing can turn a growing oscillation into a self-sustained one. This triggers another underlying problem: is vertical mixing the only media to transform a growing oscillation into a self-sustained one?

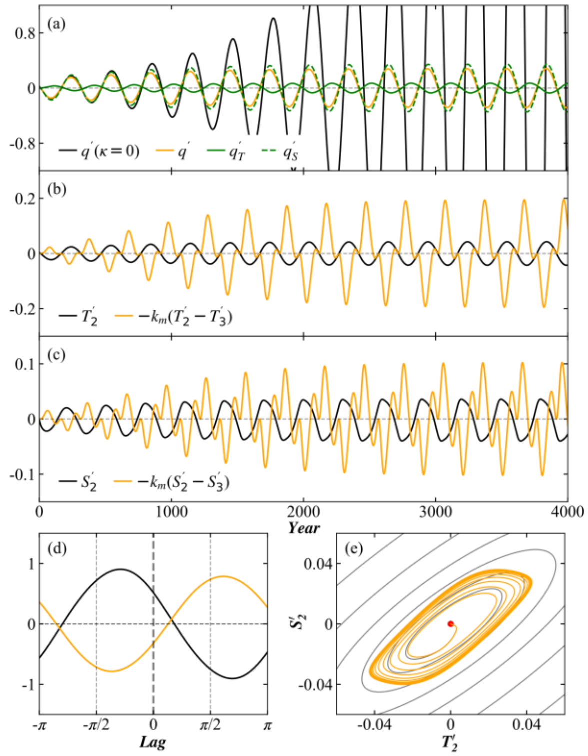

FIG. 6. Oscillations under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$. (a) Time series for q' (solid black curve) under $\kappa = 0$, q' (solid orange curve), ${q_T}'$ (solid green curve), and q'_S (dashed green curve) under $\kappa = 1 \times 10^{-4}\ m^{-3}s$ (units: Sv). (b) Time series for T'_2 (solid black curve; units: ℃) and $-k_m(T'_2 - T'_3)$ (solid orange curve; units: $Sv \cdot$ ℃). (c) Time series for S'_2 (solid black curve; units: psu) and $-k_m(S'_2 - S'_3)$ (solid orange curve; units: $Sv \cdot psu$). (d) Lead/lag correlation coefficients for $-k_m(T'_2 - T'_3)$ and q'_T (solid black curve), and for $-k_m(S'_2 - S'_3)$ and q'_S (solid orange curve). (e) T'_2-S'_2 phase space diagram for years 1-10000. The red dot represents the initial location of T'_2 and S'_2. Black curve is for $\kappa = 0$, and orange curve, for $\kappa = 1 \times 10^{-4}\ m^{-3}s$. The values of the other parameters are the same as those listed in Table 1.

b. Self-sustained oscillation with nonlinear relation between AMOC anomaly and meridional difference of density anomaly

A prominent feature of the subpolar vertical mixing is its unique nonlinearity in such a linear system. A nonlinear relation between AMOC anomaly and the meridional difference of density anomaly was adopted in Cessi (1994) and in RT97. Here, we replace the linear relation with a nonlinear one as the substitute nonlinearity for the subpolar vertical mixing introduced in section 4a. To this end, we set Eq. (3b) to,

$$q' = \begin{cases} \lambda\rho_{cri}\left[k\left(\left[\frac{\Delta\rho'}{\rho_{cri}}\right]^{\frac{1}{k}} - 1\right) + 1\right], & if\ \Delta\rho' > \rho_{cri} \\ \lambda\Delta\rho' & if\ -\rho_{cri} < \Delta\rho' < \rho_{cri} \\ -\lambda\rho_{cri}\left[k\left(\left[-\frac{\Delta\rho'}{\rho_{cri}}\right]^{\frac{1}{k}} - 1\right) + 1\right], & if\ \Delta\rho' < -\rho_{cri} \end{cases} \quad (10)$$

It takes the form of Eq. (3) of RT97. Note that there is no enhanced vertical mixing in the system at present. At $k = 1$, Eq. (10) is reduced to the linear Eq. (3b); and the system exhibits growing but purely linear oscillation under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$ (Figs. 6a, 7a, black curves). When k is larger than 1, a nonlinear restraining effect will be introduced when $\Delta\rho'$ becomes larger (lower) than a given threshold ρ_{cri} ($-\rho_{cri}$), making q' increasingly insensitive to $\Delta\rho'$ as $\Delta\rho'$ grows (decreases). For example, if $k = 1.05$ with $\rho_{cri} = 0.002\ kg/m^3$, a small degree of nonlinearity (Fig. 7c, orange curve) will be introduced into the linear system. The self-sustained oscillation is then realized (Fig. 7a, orange curve), and a limit cycle is achieved (Fig. 7b, orange curve). The intersections between the vertical dashed lines and the abscissa axis in Fig. 7c mark the upper and lower limits for $\Delta\rho'$ during the integration. The restraining effect manifested as the difference between the black and orange curves is very small (Fig. 7c). Hence, even a tiny internal nonlinearity between AMOC anomaly and the meridional difference of density anomaly can lead to a self-sustained oscillation, reflecting that the self-sustained oscillation mechanism in the 4TS model does not have to be bounded together with subpolar vertical mixing. Even in the 3TS model under the same parameters, including this nonlinear relation also led to a self-sustained oscillation (figure not shown). Here, at the heart of the transition from a linearly growing oscillation into a nonlinear self-sustained one, the nonlinear restraining term matters, which can take the form as subpolar vertical mixing, or a nonlinear relation between AMOC anomaly and meridional difference of density anomaly, or even in a form of other nonlinear processes. The self-sustained AMOC oscillation mechanism can be concluded as a combination of a linearly growing oscillation dominated by linear advection and a nonlinear restraining effect. The self-sustained oscillation mechanism in LY22 becomes more explicit, and is shown to be not only consistent with but also advanced by this current study.

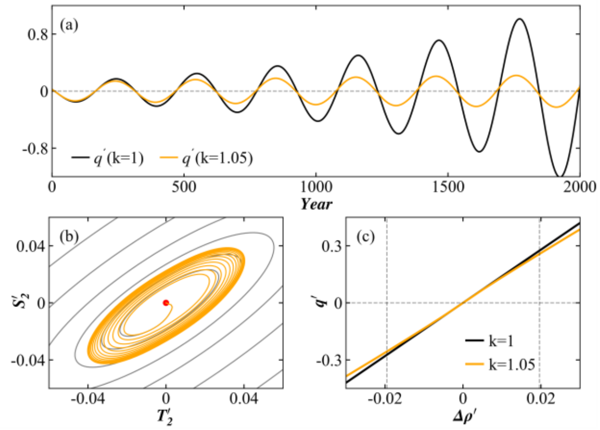

FIG. 7. Oscillations under $\lambda = 14\, Sv \cdot kg^{-1}\, m^3$. (a) Time series for q' (units: Sv) under $k = 1$ (black curve) and $k = 1.05$ (orange curve). (b) T_2'-S_2' phase space diagram for years 1-10000. The red dot represents the initial location of T_2' and S_2'. Black curve is for $k = 1$, and orange curve, for $k = 1.05$. (c) Variation of q' with $\Delta\rho'$ (units: kg/m^3) under $k = 1$ (black curve) and $k = 1.05$ (orange curve). The intersections between the vertical dashed lines and the abscissa axis mark the upper and lower limits for $\Delta\rho'$ during the integration. The values of the other parameters are the same as those listed in Table 1.

5. Eigenmode sensitivity

a. Effects of basin geometry

Basin geometry can affect both the period and e-folding time of the 4TS model's eigenmode (Fig. 8). The standard geometry is $(V_1 + V_2)/V = 1/8$ and $(V_2 + V_3)/V = 1/6$ in this paper, denoted by the orange star in Fig. 8. The gray areas in Fig. 8 denote purely damped or growing regions without oscillatory potential. The oscillatory mode is damped in region 1 because of negative real part; and it is growing in region 2 due to positive real part (Fig. 8b). The orange curve partitions regions 1 and 2, making itself also the lower limit for a possible self-sustained oscillation. The eigen period increases roughly monotonously with the increases of both the subpolar ocean fraction $(V_2 + V_3)/V$ and the upper ocean fraction $(V_1 + V_2)/V$ (Fig. 8a), while the stability of the eigenmode increases with a decrease in $(V_1 + V_2)/V$ and an increase in $(V_2 + V_3)/V$ (Fig. 8b). This suggests that it is possible for different periods and stabilities of the eigenmode to be found. Different considerations for upper and northern box volumes are possible, potentially leading to diversified periods and stabilities. In GT95, the fractions of the upper and subpolar ocean boxes are both 1/11, falling in the lower left corner of Fig. 8, with a period shorter than 200 years if $\overline{q}$ is set to 10 Sv. Actually, $\overline{q}$ in GT95 was set to a larger value of about 17 Sv, representing a much faster overturning rate; thus, the period is further shortened to the century timescale. Consequently, it is reasonable to deduce that the multidecadal period in GT95 is not at odds with the multicentennial period here. Popularity of studying multidecadal phenomena back then might account for their choice of model parameters.

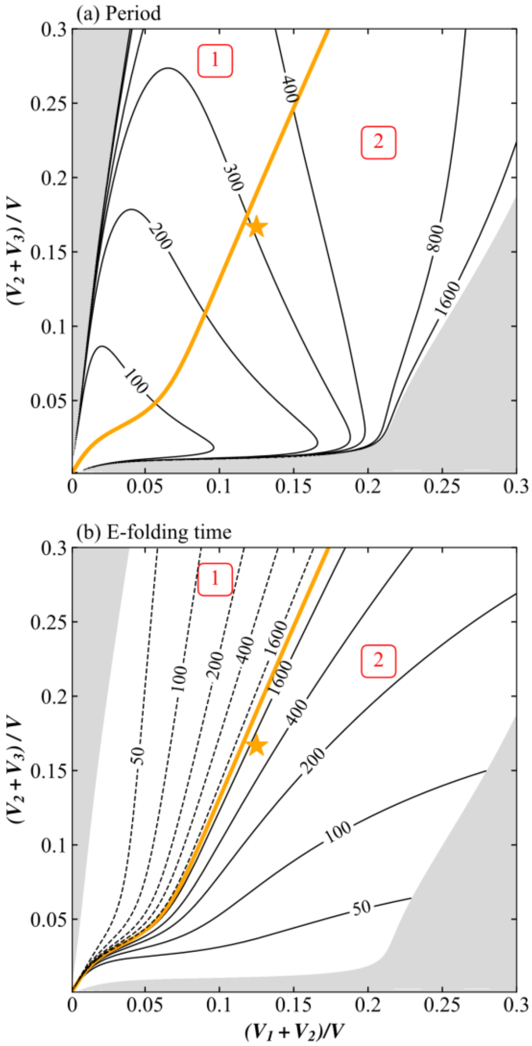

FIG. 8. Sensitivity of (a) period (units: years) and (b) e-folding time (units: years) of the eigenmode in the 4TS model to subpolar ocean fraction $(V_2 + V_3)/V$ and upper ocean fraction $(V_1 + V_2)/V$ under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$. The values of the other parameters are the same as those listed in Table 1. The orange star denotes the mode with standard basin geometry $(V_1 + V_2)/V = 1/8$ and $(V_2 + V_3)/V = 1/6$. The orange curve is both the stability threshold and the lower limit of probability for self-sustained oscillation in the 4TS model. The oscillatory mode is damped in region 1 and growing in region 2, partitioned by the orange curve. The gray area corresponds to purely damped or growing regime without the imaginary part in the 4TS model. The dashed (solid) contours in (b) show the damped (growing) regime of the oscillation.

We can also explain why the mode in RT97 (although it is a 3-box model) can be easily unstable even under a low AMOC sensitivity (equivalent to $\lambda = 5.7\ Sv \cdot kg^{-1}\ m^3$ in this paper). Their high-latitude box represents a small deep-water formation region instead of the subpolar region, so it was set to only around 1/100 of the volume of the entire ocean basin. However, their upper ocean is as large as 1/4 of the entire ocean basin. Therefore, the low stability seen in RT97 owes to their volume configuration, according to our stability analyses.

b. Effects of flow properties

Given the meridional difference of density anomaly, the total AMOC strength q is determined by its sensitivity λ to the meridional difference of density anomaly and the equilibrium strength $\overline{q}$. The period decreases monotonically as $\overline{q}$ increases (Fig. 9a), reflecting that a faster overturning leads to a shorter oscillation period. However, we should note that the oscillation in our model is not strictly paced by the transport of anomalies around the depth-latitude plane by the mean flow, contrary to the group of loop model studies (Mysak et al. 1993; Winton and Sarachik 1993; Sévellec et al. 2006). In our model, when warmer (more saline) water is transported from box 1 to box 2 [this weakens (strengthens) the AMOC], cold (fresh) anomaly will simultaneously be advected by $q'\left(\overline{T_4} - \overline{T_1}\right)$ [$q'\left(\overline{S_4} - \overline{S_1}\right)$] from box 4 to box 1, without waiting for anomaly to be transported back around the boxes. Larger λ and smaller $\overline{q}$ are both destabilizing factors (Fig. 9b). A larger λ results in more intense fluctuation of q' under the same perturbation of meridional difference of density anomaly, contributing to a less stable system. A decreased $\overline{q}$ weakens the equilibrium advection terms $\overline{q}(T_1' - T_2')$ and $\overline{q}(S_1' - S_2')$, thus limiting their destabilizing and stabilizing effects, respectively. The latter is more evident due to the dominant role that salinity plays in establishing AMOC variability. The combined effect of temperature and salinity advections under a smaller $\overline{q}$ is to make the system more unstable.

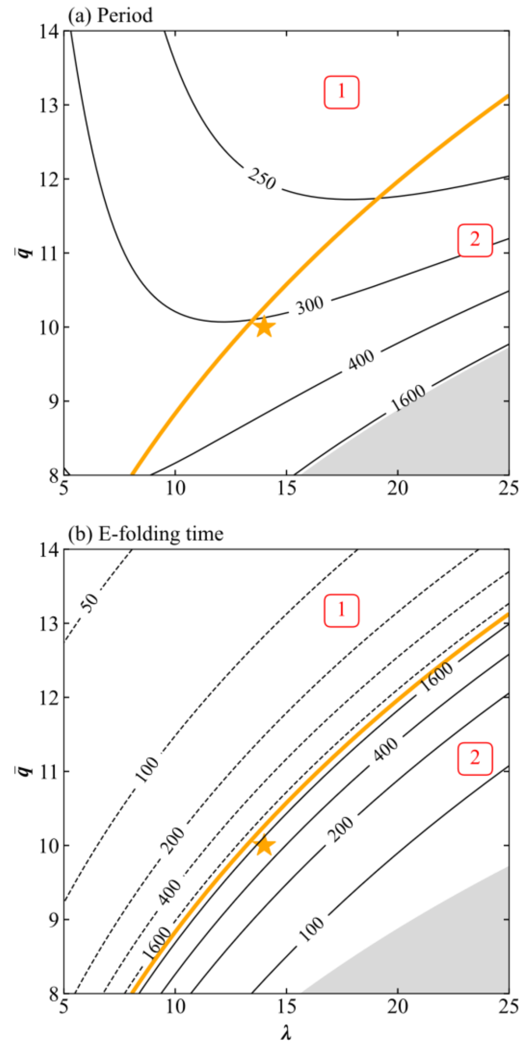

FIG. 9. Same as Fig. 8, but the ordinate and abscissa correspond to the equilibrium AMOC strength $\overline{q}$ (units: Sv) and linear closure coefficient λ (units: $Sv \cdot kg^{-1}\, m^3$), respectively. The orange star denotes the mode with $\lambda = 14\, Sv \cdot kg^{-1}\, m^3$ and the standard value $\overline{q} = 10\, Sv$.

c. Effects of mean meridional difference of equilibrium SST and SSS

From Eqs. (2b) and (8), we see, when keeping other parameters fixed, $T_1^* - T_2^*$ controls $\overline{T_1} - \overline{T_2}$ while F_w controls the $\overline{S_1} - \overline{S_2}$. Since it is $\overline{T_1} - \overline{T_2}$ and $\overline{S_1} - \overline{S_2}$ instead of the individual values for $\overline{T_1}$, $\overline{T_2}$, $\overline{S_1}$, or $\overline{S_2}$ that influence the system, we should examine the eigenmode sensitivity to $\overline{T_1} - \overline{T_2}$ and $\overline{S_1} - \overline{S_2}$ through altering $T_1^* - T_2^*$ and F_w. Note that although the standard F_w in Table 1 is calculated through the given $\overline{S_1}$, $\overline{S_2}$, and $\overline{q}$, we can artificially change $\overline{S_1}$, or $\overline{S_2}$, or both, to obtain a new F_w, which actually has the same effect as changing F_w. Figure 10a shows that the period shortens marginally with the increase of $T_1^* - T_2^*$, but exhibits a decrease-then-increase tendency as F_w grows.

Smaller $T_1^* - T_2^*$ and larger F_w all lead to lower stability (Fig. 10b). From Eq. (8), we derive that $\overline{T_1} - \overline{T_2}$ lower as $T_1^* - T_2^*$ decreases; therefore, the temperature effects are hampered due to the weaker negative temperature advection feedback, leading to a longer period and lower stability for the system. As for salinity, it can be seen from Eq. (2b) that a larger F_w increases $q'\left(\overline{S_1} - \overline{S_2}\right)$, so the destabilizing positive salinity advection feedback is reinforced, which is consistent with the finding of Sévellec et al. (2006).

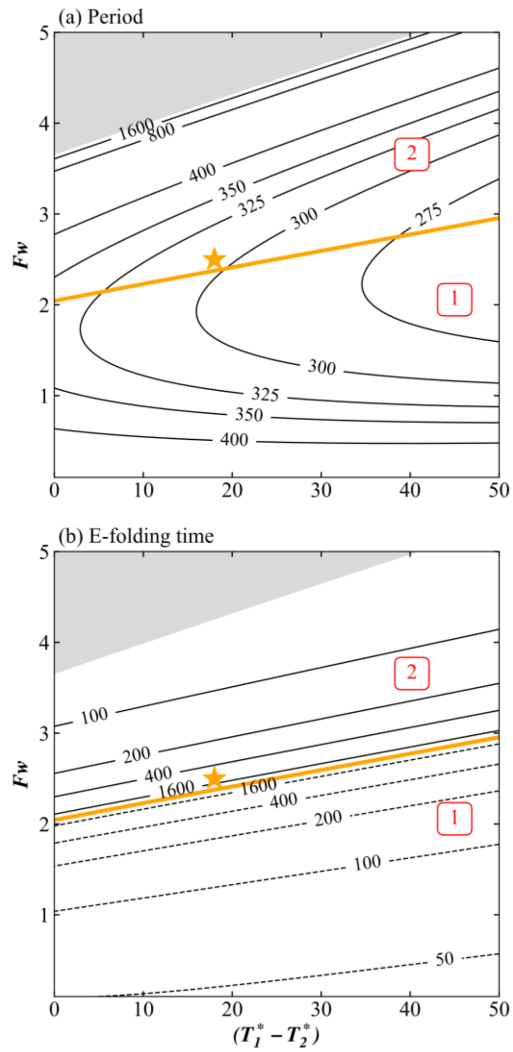

FIG. 10. Same as Fig. 8, but the ordinate and abscissa correspond to F_w (units: $10\ psu \cdot Sv$) and the meridional difference of restoring temperature $T_1^* - T_2^*$ (units: °C), respectively. The orange star denotes the mode with standard parameter values $T_1^* - T_2^* = 18$ °C and $F_w = 25.0\ psu \cdot Sv$.

6. Summary and discussion

As the second part of our theoretical studies on AMOC multicentennial variability, this study complements LY22 by including temperature equations in the box model. Mixed boundary conditions are employed for surface temperature and salinity. The added thermal processes consist of the negative temperature advection feedback and the positive restoring advection feedback. The latter is driven by and never overpowers the former; thus, the resultant temperature feedback is negative as a whole. Including the thermal processes leads to an increase of oscillation frequency because of the fast thermal-restoring process, a stabilization force for the system because of the negative temperature advection feedback, and a second stabilizing effect for the subpolar stratification due to the stabilizing subpolar temperature stratification.

With the overall stabilizing thermal processes included, additional processes are needed to turn a linearly growing oscillation into a self-sustained oscillation. The enhanced subpolar vertical mixing raised in LY22 is able to realize a self-sustained oscillation in the 4TS model here. The subpolar vertical mixing can be overall destabilizing, but is not at odds with LY22, since the subpolar temperature mixing in this work is destabilizing, and can overcome the stabilizing subpolar salinity mixing under realistic surface temperature restoring timescale. Additionally, this destabilizing subpolar total vertical mixing reflects that it is its nonlinear restraining effect instead of the linear stabilizing/destabilizing effect that leads to the self-sustained oscillation, further complementing and advancing the theory of LY22. Furthermore, we show that the nonlinear restraining process is not confined to subpolar vertical mixing, but can also take the form of a nonlinear relation between the AMOC anomaly and the meridional difference of density anomaly, or even in other forms. The magnitude of such nonlinearity does not have to be large. Basically, we can generalize the self-sustained oscillation mechanism in both LY22 and this paper as: “a combination of a linearly growing oscillation dominated by linear advection and a nonlinear restraining effect,” which is a more essential expression of the LY22 mechanism proposed in their introduction, namely, “a combination of salinity advection and enhanced mixing.”

Stability analyses reveal that the period and stability of the oscillatory eigenmode are sensitive to model geometry, flow properties, and meridional difference of mean SST and SSS. Generally, smaller subpolar and upper ocean boxes tend to shorten the oscillation period; larger subpolar ocean and smaller upper ocean boxes have stabilizing effects on the system. A stronger mean AMOC shortens the period due to the faster overturning rate, stabilizing the system through enhancing the stabilizing mean advection of meridional difference of salinity anomaly, which dominates over the destabilizing mean advection of meridional difference of temperature anomaly. Higher sensitivity of AMOC anomaly to the meridional difference of density anomaly makes the system less stable. Increasing surface freshwater flux energizes the destabilizing salinity advection feedback and lowers the system stability, because the background meridional difference of mean salinity will be stronger. Larger meridional difference of restoring temperature strengthens the thermal processes; thus, it shortens the period and increases system stability.

The box model is highly idealized, aimed at providing heuristic understanding of the multicentennial AMOC oscillation. The difference in system behaviors before and after the permission of temperature variation suggests that too few physics involved in theoretical models may lead to unrealistic eigenmode. This work can also help us understand the prevalence of centennial-to-multicentennial AMOC oscillations found in a few pre-industrial control runs using high-order models (Park and Latif 2008; Delworth and Zeng 2012; Yang et al. 2015; Jiang et al. 2021). Whether the multicentennial AMOC oscillations found in Earth system models are self-sustained or stochastically-sustained is obscured by their intricate model physics. However, our box-model study suggests that a self-sustained oscillation can appear as long as a tiny nonlinearity is included, which is present in some form in the real ocean. Thereby, we conclude that self-sustained multicentennial AMOC oscillation has a good chance to exist in high-order models. A more precise mean AMOC strength simulated in high-order models may improve the simulation of multicentennial AMOC oscillation, especially its period. For instance, the difference in the oscillation period between Jiang et al. (2021) and Meccia et al. (2022) (~200 years vs ~150 years), both utilizing the NEMO 3.6 ocean component, is believed to derive from their different mean AMOC strengths (10.8 Sv vs 16.3 Sv). Our stability analyses also suggest that if the strength of mean AMOC decreases in the future, it is likely that the stability and period of multicentennial AMOC oscillation may be lowered and lengthened, respectively.

The more intense warming at northern high latitude in the context of global warming is also likely to result in more freshwater hosing in the subpolar North Atlantic (Serreze and Barry 2011; Dai 2022), reducing meridional SST difference while enhancing meridional SSS difference. Therefore, the negative temperature advection feedback is weakened, while the positive salinity advection feedback is strengthened. This implies that the multicentennial AMOC oscillation may be less stable; this also implies that the period of the multicentennial AMOC oscillation is likely to be lengthened in the future, which has not gained attention yet. However, the period of decadal-to-multidecadal

AMOC oscillation is believed to be shortened under global warming scenario based on Rossby wave dynamics (Cheng et al. 2016; Ma et al. 2021). As global warming persists, more attention should be paid to how multicentennial AMOC oscillation period would change in the future, since global warming may occur on the background of a multicentennial oscillation.

Finally, this theoretical study can be improved in several aspects. The one-hemisphere configuration singles out only North Atlantic advection, and contributions from the other ocean basins are not considered. Extending the one-hemisphere model into an inter-hemisphere one as that in Scott et al. (1999) and in Lucarini and Stone (2005), or incoraporating the Arctic Ocean as in Lambert et al. (2016), may provide more insightful results. Too few feedbacks are included in our model because of the ocean-only configuration and mixed boundary conditions. Adding more feedbacks, such as the meridional mositure transport feedback (Tziperman and Gildor 2002), wind forcing feedback (Sherriff-Tadano and Abe-Ouchi 2020), and sea ice feedback (Jayne and Marotzke 1999), should improve the simulation of stability and other characteristics of AMOC oscillation.

Acknowledgement: This research is jointly supported by the NSF of China (Nos. 42230403, 41725021, and 91737204) and by the foundation at the Shanghai Frontiers Science Centre of Atmosphere-Ocean Interaction of Fudan University.

Data Availability Statement:

This is a theory-based article; thus no datasets were generated.

References

Ayache, M., D. Swingedouw, Y. Mary, F. Eynaud, and C. Colin, 2018: Multi-centennial variability of the AMOC over the Holocene: A new reconstruction based on multiple proxy-derived SST records. *Global Planet. Change*, **170**, 172-189

Cessi, P., 1994: A simple box model of stochastically forced thermohaline flow. *J. Phys. Oceanogr.*, **24**, 1911-1920

Chabaud, L., M. F. S. Goni, S. Desprat, and L. Rossignol, 2014: Land-sea climatic variability in the eastern North Atlantic subtropical region over the last 14,200 years: Atmospheric and oceanic processes at different timescales. *Holocene*, **24**, 787-797

Cheng, J., Z. Liu, S. Zhang, W. Liu, L. Dong, P. Liu, and H. Li, 2016: Reduced interdecadal variability of Atlantic meridional overturning circulation under global warming. *Proc Natl Acad Sci U S A*, **113**, 3175-3178

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

Colin de Verdière, A., 2007: A simple model of millennial oscillations of the thermohaline circulation. *J. Phys. Oceanogr.*, **37**, 1142-1155

Dai, A. G., 2022: Arctic amplification is the main cause of the Atlantic meridional overturning circulation weakening under large CO2 increases. *Climate Dyn.*, **58**, 3243-3259

Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model. *J. Climate*, **6**, 1993-2011

Delworth, T. L., and F. R. Zeng, 2012: Multicentennial variability of the Atlantic meridional overturning circulation and its climatic influence in a 4000 year simulation of the GFDL CM2.1 climate model. *Geophys. Res. Lett.*, **39**

Dong, B. W., and R. T. Sutton, 2005: Mechanism of interdecadal thermohaline circulation variability in a coupled ocean-atmosphere GCM. *J. Climate*, **18**, 1117-1135

Griffies, S. M., and E. Tziperman, 1995: A linear thermohaline oscillator driven by stochastic atmospheric forcing. *J. Climate*, **8**, 2440-2453

Grootes, P. M., M. Stuiver, J. W. C. White, S. Johnsen, and J. Jouzel, 1993: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature*, **366**, 552-554

Hall, I. R., G. G. Bianchi, and J. R. Evans, 2004: Centennial to millennial scale Holocene climate-deep water linkage in the North Atlantic. *Quat. Sci. Rev.*, **23**, 1529-1536

Haney, R. L., 1971: Surface thermal boundary condition for ocean circulation models. *J. Phys. Oceanogr.*, **1**, 241-248

Huang, R. X., and W. K. Dewar, 1996: Haline circulation: Bifurcation and chaos. *J. Phys. Oceanogr.*, **26**, 2093-2106

Huang, R. X., J. R. Luyten, and H. M. Stommel, 1992: Multiple equilibrium states in combined thermal and saline circulation. *J. Phys. Oceanogr.*, **22**, 231-246

Jayne, S. R., and J. Marotzke, 1999: A destabilizing thermohaline circulation-atmosphere-sea ice feedback. *J. Climate*, **12**, 642-651

Jiang, W. M., G. Gastineau, and F. Codron, 2021: Multicentennial variability driven by salinity exchanges between the Atlantic and the Arctic Ocean in a coupled climate model. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002366

Joyce, T. M., 1991: Thermohaline catastrophe in a simple four-box model of the ocean climate. *J. Geophys. Res.*, **96**, 20393-20402

Lambert, E., T. Eldevik, and P. M. Haugan, 2016: How northern freshwater input can stabilise thermohaline circulation. *Tellus Ser. A-Dyn. Meteorol. Oceanol.*, **68**, 15

Li, Y., and H. J. Yang, 2022: A theory for self-sustained multicentennial oscillation of the Atlantic meridional overturning circulation. *J. Climate*, **35**, 5883-5896

Lucarini, V., and P. H. Stone, 2005: Thermohaline circulation stability: A box model study. Part I: Uncoupled model. *J. Climate*, **18**, 501-513

Ma, X. F., W. Liu, N. J. Burls, C. L. Chen, J. Cheng, G. Huang, and X. C. Li, 2021: Evolving AMOC multidecadal variability under different CO2 forcings. *Climate Dyn.*, **57**, 593-610

Marotzke, J., and J. Willebrand, 1991: Multiple equilibria of the global thermohaline circulation. *J. Phys. Oceanogr.*, **21**, 1372-1385

Martin, T., W. Park, and M. Latif, 2013: Multi-centennial variability controlled by Southern Ocean convection in the Kiel Climate Model. *Climate Dyn.*, **40**, 2005-2022

——, 2015: Southern Ocean forcing of the North Atlantic at multi-centennial time scales in the Kiel Climate Model. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, **114**, 39-48

McCarthy, G. D., and Coauthors, 2015: Measuring the Atlantic meridional overturning circulation at 26 degrees N. *Prog. Oceanogr.*, **130**, 91-111

Meccia, V. L., R. Fuentes-Franco, P. Davini, K. Bellomo, F. Fabiano, S. T. Yang, and J. von Hardenberg, 2022: Internal multi-centennial variability of the Atlantic meridional overturning circulation simulated by EC-Earth3. *Climate Dyn.*, 18

Miettinen, A., D. Divine, N. Koc, F. Godtliebsen, and I. R. Hall, 2012: Multicentennial variability of the sea surface temperature gradient across the subpolar North Atlantic over the last 2.8 kyr. *J. Climate*, **25**, 4205-4219

Mysak, L. A., T. F. Stocker, and F. Huang, 1993: Century-scale variability in a randomly forced, two-dimensional thermohaline ocean circulation model. *Climate Dyn.*, **8**, 103-116

Nakamura, M., P. H. Stone, and J. Marotzke, 1994: Destabilization of the thermohaline circulation by atmospheric eddy transports. *J. Climate*, **7**, 1870-1882

Oppo, D. W., J. F. McManus, and J. L. Cullen, 2003: Palaeo-oceanography: Deepwater variability in the Holocene epoch. *Nature*, **422**, 277

Otto-Bliesner, B. L., and Coauthors, 2017: The PMIP4 contribution to CMIP6-Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations. *Geosci. Model Dev.*, **10**, 3979-4003

Park, W., and M. Latif, 2008: Multidecadal and multicentennial variability of the meridional overturning circulation. *Geophys. Res. Lett.*, **35**

Pierce, D. W., 1996: Reducing phase and amplitude errors in restoring boundary conditions. *J. Phys. Oceanogr.*, **26**, 1552-1560

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

Rahmstorf, S., and J. Willebrand, 1995: The role of temperature feedback in stabilizing the thermohaline circulation. *J. Phys. Oceanogr.*, **25**, 787-805

Rivin, I., and E. Tziperman, 1997: Linear versus self-sustained interdecadal thermohaline variability in a coupled box model. *J. Phys. Oceanogr.*, **27**, 1216-1232

Roebber, P. J., 1995: Climate variability in a low-order coupled atmosphere-ocean model. *Tellus Ser. A-Dyn. Meteorol. Oceanol.*, **47**, 473-494

Schmidt, G. A., and L. A. Mysak, 1996: The stability of a zonally averaged thermohaline circulation model. *Tellus Ser. A-Dyn. Meteorol. Oceanol.*, **48**, 158-178

Scott, J. R., J. Marotzke, and P. H. Stone, 1999: Interhemispheric thermohaline circulation in a coupled box model. *J. Phys. Oceanogr.*, **29**, 351-365

Serreze, M. C., and R. G. Barry, 2011: Processes and impacts of Arctic amplification: A research synthesis. *Global Planet. Change*, **77**, 85-96

Sévellec, F., and A. V. Fedorov, 2014: Millennial variability in an idealized ocean model: Predicting the AMOC regime shifts. *J. Climate*, **27**, 3551-3564

Sévellec, F., T. Huck, and M. Ben Jelloul, 2006: On the mechanism of centennial thermohaline oscillations. *J. Mar. Res.*, **64**, 355-392

Sherriff-Tadano, S., and A. Abe-Ouchi, 2020: Roles of sea ice-surface wind feedback in maintaining the glacial Atlantic meridional overturning circulation and climate. *J. Climate*, **33**, 3001-3018

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

Tziperman, E., and H. Gildor, 2002: The stabilization of the thermohaline circulation by the temperature-precipitation feedback. *J. Phys. Oceanogr.*, **32**, 2707-2714

Vellinga, M., and P. L. Wu, 2004: Low-latitude freshwater influence on centennial variability of the Atlantic thermohaline circulation. *J. Climate*, **17**, 4498-4511

Weaver, A. J., and E. S. Sarachik, 1991: Evidence for decadal variability in an ocean general-circulation model - An advective mechanism *Atmos. Ocean*, **29**, 197-231

Wei, X., and R. Zhang, 2022: A simple conceptual model for the self-sustained multidecadal AMOC variability. *Geophys. Res. Lett.*, **49**, 11

Welander, P., 1982: A simple heat salt oscillator. *Dyn. Atmos. Oceans*, **6**, 233-242

Winton, M., and E. S. Sarachik, 1993: Thermohaline oscillations induced by strong steady salinity forcing of ocean general-circulation models. *J. Phys. Oceanogr.*, **23**, 1389-1410

Yang, H. J., Q. Li, K. Wang, Y. Sun, and D. X. Sun, 2015: Decomposing the meridional heat transport in the climate system. *Climate Dyn.*, **44**, 2751-2768

Zhang, R., M. Follows, and J. Marshall, 2002: Mechanisms of thermohaline mode switching with application to warm equable climates. *J. Climate*, **15**, 2056-2072

Zhang, S., R. J. Greatbatch, and C. A. Lin, 1993: A reexamination of the polar halocline catastrophe and implications for coupled ocean atmosphere modeling. *J. Phys. Oceanogr.*, **23**, 287-299