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
October 5, 2022

**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 proposed a hemispheric box model considering only the saline process. In this paper, we consider both thermal and saline processes in the box model and employ mixed boundary conditions, so as to include more realistic physics. Generally, the thermal process has a stabilizing effect on the system, and additional physics, such as enhanced subpolar mixing or a nonlinear relation between the AMOC and meridional density gradient, is still needed to realize a self-sustained oscillation. Specifically, the thermal process exerts mainly two effects on the system: shortening marginally the multicentennial oscillation period of the AMOC, and stabilizing the oscillating system and subpolar stratification, which are contributed by the fast surface temperature restoring, the negative temperature-advection feedback and subpolar temperature stratification, respectively. The oscillation properties are controlled by the balance of destabilizing salinity advection and stabilizing temperature advection. Different from salinity-only situation, the enhanced subpolar mixing in the current situation makes the system more unstable. Weaker meridional temperature gradient and stronger meridional salinity gradient can lead to weaker temperature-advection feedback and stronger salt-advection feedback, and thus a longer AMOC oscillation period with less stability at multicentennial timescale, which might be expected in the future due to more intense high-latitude warming and freshwater hosing.

Keywords: AMOC, Self-sustained oscillation, Temperature feedbacks, Box model

1. Introduction

In our first publication on the multicentennial oscillation of the Atlantic meridional overturning circulation (AMOC) (Li and Yang 2022, hereafter LY22), we used a 4-box model to study the self-sustained multicentennial AMOC oscillation, which only considered salinity equations. Here, we expand our horizon to explore theoretically the multicentennial oscillation of the AMOC by including both thermal and saline processes. Thus, the term AMOC in this paper refers to its thermohaline circulation portion that is controlled by both temperature and salinity. It is well recognized that the AMOC is a crucial regulator for the North Atlantic and likely the global climate over a wide range of timescales (Chabaud et al. 2014; Muir and Fedorov 2015; Zhang et al. 2019). Low-order theoretical models are essentials for understanding the AMOC dynamics. To isolate the most fundamental dynamics, only salinity equations were kept in the theoretical model of LY22, paralleling a few other studies (Rahmstorf 1996; Cimatoribus et al. 2014; Sévellec and Fedorov 2014). However, such treatment is unphysical to an extent since all the temperature-related effects were excluded.

In ocean-only theoretical models of the AMOC, mixed boundary conditions (Haney 1971) are often employed; that is, sea surface temperature (SST) is restored to a prescribed value following the Newtonian law, while sea surface salinity (SSS) is forced by the surface freshwater flux. As seen in quite a few studies, the negative temperature-advection feedback (Stommel 1961; Walin 1985) and the positive restoring-advection feedback (Griffies and Tziperman 1995, hereafter GT95; Scott et al. 1999; Colin de Verdière 2010) are included in their temperature equations. The former works as follows: commencing with an initial positive AMOC perturbation, the elevated poleward heat transport reduces subpolar density, therefore restraining deep-water formation, which is followed by the slowdown of the AMOC. The latter works as an opponent to the former: an enhanced AMOC increases the subpolar SST, leading to restore the warming itself, and the reduced positive SST anomaly hinders the AMOC slowdown. The restoring-advection feedback strengthens as the restoring process (timescale) speeds up (shortens). However, this feedback merely offsets, but never overruns the temperature-advection feedback even under extremely strong restoring. Hence, the net effect of temperature feedbacks is to stabilize the system. It has also been illustrated in other studies that the oceanic thermal process influences the stability of AMOC system (Zhang et al. 1993; Marotzke and Stone 1995; Rahmstorf and Willebrand 1995; Marotzke 1996). Consequently, adding temperature equations to the salinity-only model of LY22 should be more realistic for the multicentennial AMOC oscillation.

Usually, the thermal process is faster than the saline process due to the fast SST restoring, whose timescale is not more than a few years (Pierce 1996). It is thus intuitive that including this fast process might shorten the multicentennial period. However, Schmidt and Mysak (1996) considered that such fast restoring removes high-frequency anomalies and therefore might prolong the multicentennial period. They then stated that such lengthening is not so obvious; and their main focus was on the system's stability, leaving this question unanswered. Therefore, it would be intriguing to see the role of temperature in influencing AMOC oscillation timescale. An in-depth understanding of temperature effects in shaping multicentennial AMOC oscillation should also provide us insight about how the AMOC will respond to future climate change.

In this study, we extend the 4-box salinity-only model in LY22 to a temperature-salinity one; and we employ mixed boundary conditions. The conciseness of this model enables stability analysis. We aim to unravel the effects induced by thermal process. We then work on realization of a self-sustained oscillation with bounding terms affiliated, in order to test whether the advection process dominates the eigenmode. Motivated by the difference in period and stability between our model and other studies, we also examine the sensitivity of eigenmode to model parameters controlling the model geometry, flow properties and feedback processes.

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. In section 3, the role of temperature equations is analyzed. In section 4, we test two ways for realizing a self-sustained oscillation, and elucidate the critical role of advection process. In section 5, the sensitivities of eigenmode's period and stability to model parameters are examined. Summary and discussion are presented in section 6.

2. Box model

a. Model formulae

The model we use 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 flows through the boxes in a clockwise sense. Excluding multi-equilibria, we do not discuss possibility of a reversed AMOC cell. Analogous box models have been widely used (Joyce 1991; Huang et al. 1992; GT95). In the 4-box salinity-only model (hereafter 4S) of LY22, only salinity equations were used to obtain analytical solutions. In the 4TS model, we have:

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

It is an advection-dominated box model, with mixed boundary conditions where Haney-style restoring for SST (Haney 1971) and surface freshwater flux for SSS are adopted. V_i, T_i and S_i are the volume, temperature and salinity in each box. Dots over variables denote their temporal derivatives. q stands for AMOC strength. F_w represents the surface freshwater flux, which actually takes the form of virtual salt flux. T_1^* and T_2^* correspond to the restoring temperatures for boxes 1 and 2, respectively. The Newtonian restoring coefficient τ is also the reciprocal of the restoring timescale for T_1 and T_2.

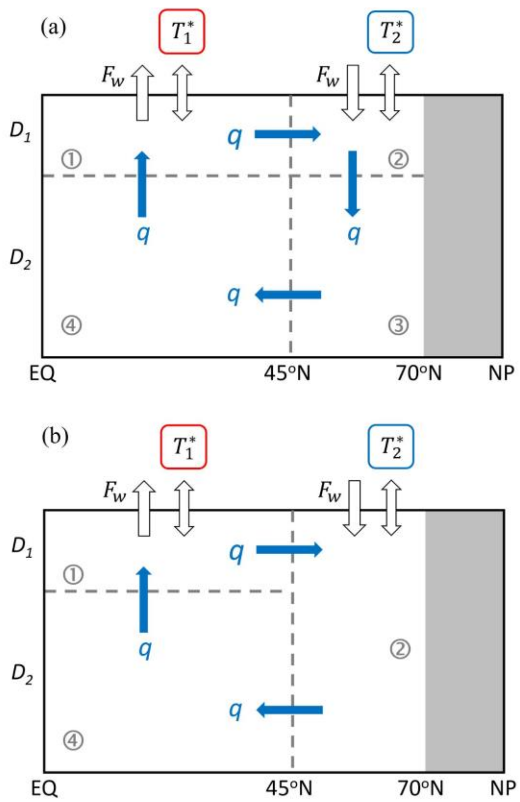

FIG. 1. Schematic diagrams of temperature-salinity box models. (a) The 4-box model. (b) The 3-box model reduced from the 4-box one. The circled numbers ①, ②, ③, and ④ denote the ocean boxes. Boxes 1 and 4 stand for the upper and lower tropical oceans, respectively, while boxes 2 and 3 stand for the upper and lower subpolar oceans, respectively. D_1 and D_2 represent the upper and lower ocean depths, respectively. The net freshwater flux out of (into) the tropical (subpolar) ocean is represented by F_w. T_1^* (T_2^*) is the restoring temperature of the tropical (subpolar) ocean. q represents the AMOC.

The equilibrium solutions at steady state can be easily written as:

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

$$\overline{S_1} = F_w/\overline{q} + \overline{S_2}, \qquad \overline{S_2} = \overline{S_3} = \overline{S_4} \tag{2b}$$

Variables with overbar denote their equilibrium values. Similar to previous studies (Winton and Sarachik 1993; Cessi 1994; Roebber 1995), we choose $\overline{q} = 10\ Sv$. Following LY22, the upper and total ocean depths are 500 and 4000 m, respectively. We set $\overline{S_2}$ and F_w to 33.5 psu and $25.0\ psu \cdot Sv$, respectively, leading to $\overline{S_1} = 36\ psu$. Straightway, T_1^* and T_2^* are considered to be close to the averaged realistic SSTs in the tropical and subpolar regions; thus we have 25℃ and 7℃, respectively. The setting of τ uses the idea of Bretherton (1982), and the restoring timescale $1/\tau$ is represented as follows:

$$1/\tau = \frac{\rho_w c \Delta z A}{\kappa_0 A} = \frac{\rho_w c \Delta z}{\kappa_0} \tag{3}$$

Here, A is the area of ocean surface. ρ_w and c are the typical density and specific heat of seawater, set to 1027 $kg \cdot m^{-3}$ and 3850 $J/(kg \cdot ℃)$, respectively. A value of 30 m is given to the thickness of the surface layer Δz (not the upper ocean). κ_0 is a restoring coefficient. Bretherton (1982) stated that κ_0 should be small when averaged over the whole globe, while it is larger if a smaller area is considered. In view of our one-hemisphere configuration, it is reasonable to choose $\kappa_0 = 4\ W/(m^2 \cdot ℃)$, yielding a restoring timescale of almost one year in our box model. For convenience, we set $\tau = 3.171 \times 10^{-8}\ s^{-1}$, corresponding to a precise 1-year restoring timescale.

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

$$q = \overline{q} + q' \tag{4a}$$

$$q' = q'_T + q'_S = \lambda \Delta\rho'_T + \lambda \Delta\rho'_S = \lambda \Delta\rho' \tag{4b}$$

where

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

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

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

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. D_1 and D correspond to the upper and total ocean depths, 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.

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, lower oceans and the entire ocean	$500, 3500, 4000\ m$
T_1^*, T_2^*	Restoring temperatures of boxes 1 and 2	25℃, 7℃
τ	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	$10\ Sv\ (10^6\ m^3\ s^{-1})$
F_w	Surface freshwater flux	$25.0\ psu \cdot Sv$
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:

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

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

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

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

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

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

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

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

In LY22, we assumed 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. (4c-e) become,

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

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

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

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

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

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

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

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

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

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

b. Stability analysis

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

TABLE 2. Eigenvalues ($10^{-10}\ s^{-1}$) for the 4TS and 4S models using 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 (Fig. 3a). That is, the e-folding time changes from positive 1025 years in the 4S model to negative 576 years in the 4TS model, and the period changes from 340 years to about 300 years. This seems to suggest that the thermal processes have a stabilizing effect on the system, and tend to shorten the oscillation period slightly (Fig. 3a). 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.

The stability of the box model system is strongly dependent on the linear closure parameter λ, i.e., the sensitivity of the AMOC to the meridional density gradient as formulated in Eq. (4b). 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 of the 4TS and 3TS models. Figure 2 shows the dependence of real and imaginary parts of the oscillatory mode on λ. The results from the 4S and 3S models in LY22 are also plotted in Fig. 2 for comparison. The intersections between line $y = 0$ and the stability diagrams of each model, $(\lambda_C, 0)$, $(\lambda_1, 0)$ and $(\lambda_2, 0)$, correspond to the instability threshold (Fig. 2b), the lower and upper limits for the existence of the imaginary parts (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 potentials exist, suggested by the corresponding positive or negative real parts (Fig. 2b). When $\lambda_1 < \lambda < \lambda_2$, the systems exhibit oscillatory behavior because of the presence of the imaginary parts (Fig. 2a). With the increase of λ, the models have the tendency to change from a damped oscillation to a growing oscillation. In comparison with the 4S and 3S models of LY22, it appears that including the temperature equations in the system has at least two consequences:

(a) An acceleration of the oscillation, 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 for the system, evidenced by the higher 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).

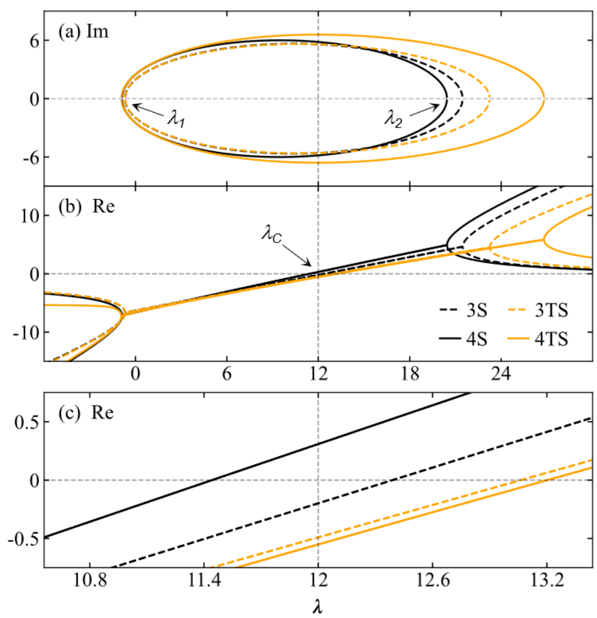

FIG. 2. Dependences of (a) imaginary parts and (b) real parts of the oscillatory mode on λ in the 4TS (solid orange curves), 3TS (dashed orange curves), 4S (solid black curves), and 3S (dashed black curves) models. (c) is the magnified version of (b) near line $y = 0$. Results of the 4S and 3S models 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 gray 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 the mathematical fundamentals, showing how the oscillatory behaviors of the system change when the thermal process is included. Physical insight into why the oscillation changes will be deliberated next.

3. Effects of temperature equations

a. Temperature feedbacks

There are mainly two feedbacks between the thermal process and the AMOC: the negative temperature-advection feedback and the positive restoring-advection feedback. Let us illustrate them using box 2 (Fig. 3b). Starting with a positive perturbation of q', the anomalous advection $q'\left(\overline{T_1} - \overline{T_2}\right)$ transports more warm water northward, T_2' is increased and thus $\Delta\rho_T'$ is lowered, causing a decrease in q_T'. This is the negative temperature-advection feedback, which can be further illustrated by the lead/lag correlation between $q'\left(\overline{T_1} - \overline{T_2}\right)$ and q_T': the former leads the latter by about $\pi/4$ with a negative correlation near 1.0 (Fig. 3e, orange curve). However, the increased T_2' also triggers a relaxation via the anomalous restoring $-V_2\tau T_2'$, whose strength is proportional to the restoring coefficient τ. This limits the growth of the positive T_2' itself, bounding the decreases of $\Delta\rho_T'$ and q_T'. This is the positive restoring-advection feedback, which is also illustrated clearly in Fig. 3e (green curve): $-V_2\tau T_2'$ leads q_T' by about $\pi/4$ with a positive correlation near 1.0. These two feedbacks are local ones, which have the comparable amplitude and can offset each other mostly (Fig. 3b). There is a third feedback coming from $\overline{q}(T_1' - T_2')$ of Eq. (5b). This term is related to the remote response T_1', and is a weak positive feedback (Figs. 3b, e).

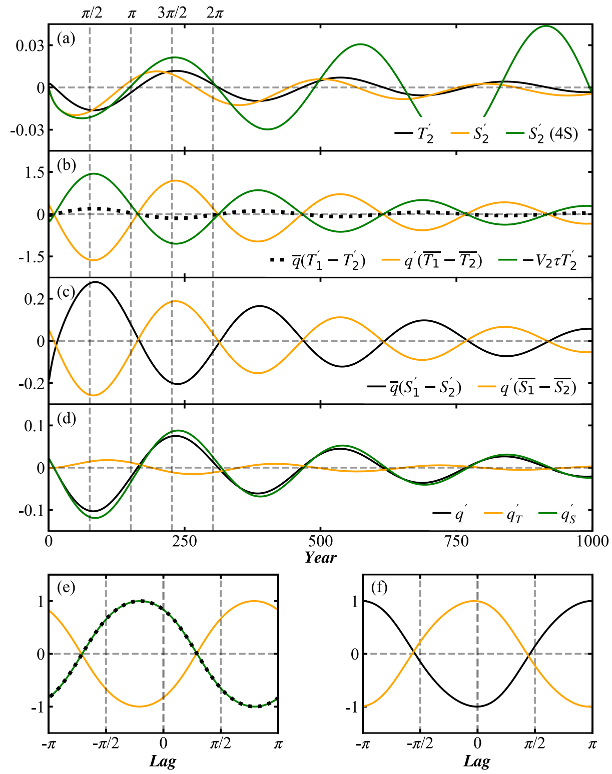

FIG. 3. (a) Damped and growing oscillations in the 4TS and 4S models using the standard parameters in Table 1. Black, orange and green curves are the time series of T_2' (units: °C), S_2' (units: psu) in the 4TS model and S_2' in the 4S model, respectively. (b) Time series for temperature terms (units: $Sv\cdot$°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); (d) time series for q', q_T' and q_S' (units: Sv) in the 4TS model. The vertical dashed gray lines in (a)-(d) mark the locations of π/2, π, 3π/2, and 2π of the period (302 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 $-V_2\tau T_2'$ (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), the negative lag represents q' lags the other terms.

The salt-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; and the former is much bigger than the latter (Fig. 3d), suggesting that the salt-advection feedback has more remarkable effect on the AMOC than the temperature-advection feedback does. In addition, although introducing the fast-restoring processes leads to an obvious time lag between the thermal process and the AMOC (Fig. 3e), there is almost no time lag between the saline process and the AMOC (Fig. 3f), further suggesting the deterministic role of the saline process in the multicentennial oscillation of the AMOC.

Results shown in Fig. 3 are obtained from forward numerical integration of Eq. (5) with the standard parameters 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 $S' = T' = 0$. The integration time step is 7.2 days, and the total integration length exceeds 10000 years, but only the very front parts are shown. Throughout this paper, the same numerical method is employed for all experiments; and annual mean data is used for analysis.

b. Role of restoring feedback

It is the restoring feedback in the temperature equations that changes the oscillatory behavior of the system. To understand this better, let us first examine how the restoring timescale affects the temperature-advection feedback. Based on Eq. (2a), we have,

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

This depicts a larger τ (or a shorter restoring timescale) causes a larger $\overline{T_1} - \overline{T_2}$, thus stronger advection $q'\left(\overline{T_1} - \overline{T_2}\right)$. However, since the negative temperature-advection feedback is realized through an increase in T_2', which is in turn limited by stronger $-V_2 \tau T_2'$, the restoring-advection feedback tends to always offset mostly the temperature-advection feedback, regardless of the restoring strength (Fig. 3b).

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

τ	Eigenvalues ($10^{-10}\ s^{-1}$)	E-folding time (Years)	Period (Years)
$\tau_0 = 0$	$0.31 \pm 5.83i$	1025	341
$\tau_1 = (5\ year)^{-1}$	$-1.48 \pm 7.30i$	-215	273
$\tau_2 = (1\ year)^{-1}$	$-0.55 \pm 6.59i$	-576	302
$\tau_3 = (0.25\ year)^{-1}$	$0.036 \pm 6.09i$	8830	327

$\tau_4 = (1\, month)^{-1}$	$0.21 \pm 5.92i$	1492	336
$\tau_5 = (1\, minute)^{-1}$	$0.31 \pm 5.83i$	1025	341

There are two extreme situations. When $\tau \to 0$ or $\tau \to \infty$, that is, the restoring timescale for SST goes to either infinity or zero, the oscillatory eigenmode ($0.31 \pm 5.83i$) in the 4TS model is identical to that in the 4S model (Table 4). Under these two situations, the thermal process has no effect on the AMOC oscillation, and the 4TS model is practically reduced to the 4S model. In the situation with $\tau \to 0$, 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. Now, $\overline{T_1} = \overline{T_2} = \overline{T_3} = \overline{T_4}$ based on Eqs. (2) and (8). There is no temperature-advection feedback ($q'\left(\overline{T_1} - \overline{T_2}\right) = 0$) anywhere, so the system is totally controlled by the saline process. In the situation with $\tau \to \infty$, $\overline{T_1} - \overline{T_2} = T_1^* - T_2^*$. The extremely strong restoring kills any temperature perturbations immediately, which makes $T_1' = T_2' = 0$ and the 4TS system is equivalent to a system without active thermal process, so that only the saline process matters to the oscillatory behavior. In summary, under these two extreme situations, the results from linear stability analysis suggest the oscillations of salinity and AMOC are identical to those of the 4S model in LY22; as a result, the corresponding figures (Figs. 2-3 when $\tau \to 0$ or $\tau \to \infty$) are not shown.

The dependences of imaginary and real parts of the oscillatory mode on τ in the 4TS model are shown in Fig. 4 (solid orange curves). Under a reasonable range of τ (from τ_1 to τ_4; Table 4), the oscillatory behavior of the 4TS model can vary from a damped oscillation to a weakly growing oscillation (Fig. 4b, solid orange curve). This is because the positive restoring-advection feedback becomes stronger as the restoring timescale gets shorter; and the system changes from under-compensation to overcompensation, to the negative temperature-advection feedback. 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 τ and λ, temperature-induced q_T' is always opposite to salinity-induced q_S', so that the total q' is always smaller, i.e., 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. Practically, since a reasonable restoring timescale is always much shorter than the multicentennial timescale, the effect of restoring timescale on the oscillation period of the system can be neglected.

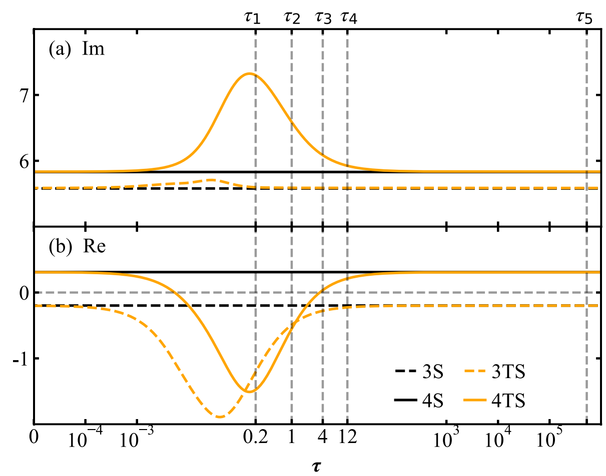

FIG. 4. Dependence of (a) positive imaginary parts and (b) real parts of the oscillatory mode on τ (units: year^{-1}) in the 4TS model (solid orange curve) and the 3TS model (dashed orange curve). The units of the ordinate are 10^{-10} s^{-1}. The vertical dashed gray lines from left to right denote the situations under τ_1, τ_2, τ_3, τ_4, and τ_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 τ. 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 process, the change of stability from the 4TS to 3TS model is not that obvious (Fig. 4). To better understand the stability change, we should first recognize that whether the temperature and salinity anomalies stay in the subpolar upper or lower ocean does not influence the meridional density gradient due to the vertically weighted volume-averaged treatment. However, the time consumed in transporting temperature and salinity anomalies from the upper to lower ocean 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 stratifications in the 3TS model have destabilizing and stabilizing effects, respectively, on the oscillation of the system.

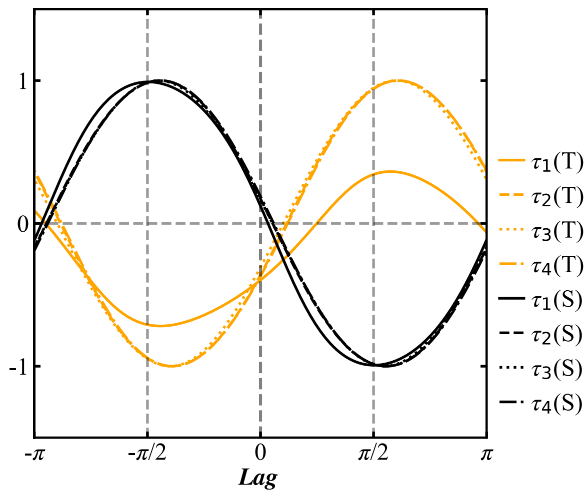

FIG. 5. Lead/lag correlation coefficients between $T_2' - T_3'$ and q_T' (orange curves) and between $S_2' - S_3'$ and q_S' (black curves) in the 4TS model under different τ. The negative lag represents q' lags the other terms. Here, λ is set to 12 $Sv \cdot kg^{-1}\, m^3$. The values of the other parameters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

In the 4TS model, the subpolar temperature stratification $T_2' - T_3'$ leads q_T' by about $\frac{\pi}{2}$ with a negative correlation, while the subpolar salinity stratification $S_2' - S_3'$ leads q_S' by $\frac{\pi}{2}$ with a positive correlation (Fig. 5). Moreover, at lag 0, $T_2' - T_3'$ ($S_2' - S_3'$) also has negative (positive) correlation with q_T' (q_S'). These correlation relationships do not rely on the temperature restoring coefficient τ. These confirm that the existences of subpolar temperature and salinity related stratifications have stabilizing and destabilizing effects, respectively, on the system. However, whether the total subpolar buoyancy stratification plays as a stabilizing or destabilizing role depends on τ. When τ lies in the range of about several years (from τ_1 to τ_2; Fig. 4b), the subpolar buoyancy stratification plays as a stabilizing factor since the stabilizing effect of temperature stratification overcomes the destabilizing effect of salinity stratification. When τ is too small or too large, the temperature effect becomes weaker while the salinity effect is not influenced. Hence, the temperature stratification no longer overcomes the salinity stratification; and the 3TS model is more stable than the 4TS model, that is, including extreme mixing in the subpolar ocean can stabilize the system, as deliberated in LY22. We conclude that under realistic ranges of the parameters, the 4TS model can be more stable than the 3TS model, due to the stabilizing effect of subpolar temperature 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 mixing process is added explicitly in the subpolar ocean, to realize a self-sustained oscillation. There is also an alternative way to realize a self-sustained oscillation as shown in Rivin and Tziperman (1997) (hereafter RT97), in which a nonlinear relationship between the AMOC strength and meridional density gradient is employed. Here, we want to emphasize that a self-sustained oscillation should first satisfy the instability criterion detailed in LY22, that is, $\lambda > \lambda_C$, depicting that the AMOC should be sensitive enough to the perturbation of meridional density gradient, and the intrinsic oscillatory mode is an unstable mode. When $\lambda \le \lambda_C$, the oscillatory mode itself is a decayed or neutral mode; and any additional mixing or nonlinear processes will make the oscillation even more decayed.

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,

$$V_2\dot{T_2'} = q'\left(\overline{T_1} - \overline{T_2}\right) + \overline{q}(T_1' - T_2') - k_m(T_2' - T_3') - V_2\tau T_2' \tag{9a}$$

$$V_3\dot{T_3'} = \overline{q}(T_2' - T_3') + k_m(T_2' - T_3') \tag{9b}$$

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

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

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

$$k_m = \kappa q'^2 \tag{9e}$$

Here, κ (units: $m^{-3}s$) is a positive constant. We set it 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 the subpolar upper-ocean anomalies. Detailed physics of the enhanced mixing process was discussed in LY22.

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. Here, $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$; and 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, which enters the lower ocean through anomalous mixings $-k_m(T_2' - T_3')$ and $-k_m(S_2' - S_3')$ (Figs. 6b, c, solid orange curves). In turn, further growth (decrease) of q' is restrained. Beware that the temperature and salinity mixing-advection feedbacks have destabilizing and stabilizing effects on q', respectively (Fig. 6d). Their combined effect on subpolar density is to stabilize q'. In summary, including enhanced mixing in the subpolar ocean can well establish a self-sustained oscillation, which can be seen more clearly in the phase diagram of T_2' vs S_2' (Fig. 6e, orange curve); that is, a limit cycle is formed eventually.

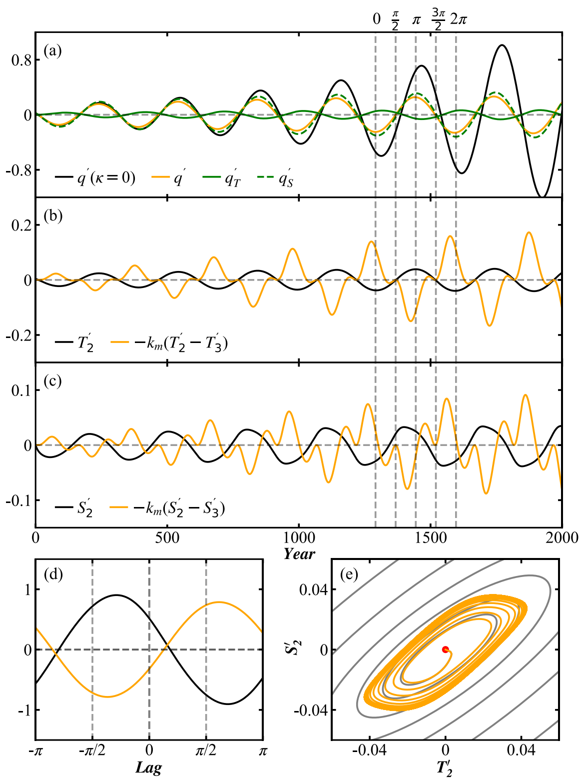

FIG. 6. Oscillations under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$. (a) Time series for q' (solid black curve) under $\kappa = 0$, q' (solid orange curve), $q_T{}'$ (solid green curve) and q_S' (dashed green curve) under $\kappa = 1 \times 10^{-4}\ m^{-3} s$ (units: Sv). (b) Time series for T_2' (solid 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), $-k_m(S_2' - S_3')$ and q_S' (solid orange curve). (e) T_2'-S_2' phase space diagrams 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 vertical dashed gray line in (a), (b) and (c) marks a holonomic oscillation period under $\kappa = 1 \times 10^{-4}\ m^{-3} s$. The values of the other parameters are the same as those listed in Table 1.

The oscillation period under $\kappa = 1 \times 10^{-4}\ m^{-3} s$ is 300 years, quite close to the 306-year analytical period under $\kappa = 0$. The period is insensitive to the value of κ, indicating that it is dominated by advection instead of the mixing process. The κ chosen here is one order smaller than the value used in LY22, suggesting that even a small bounding from subpolar vertical mixing can lead to self-sustained oscillation.

b. Self-sustained oscillation with nonlinear AMOC-density relation

It has been prevailing to set the sensitivity of the AMOC in low-order models to be linearly proportional to the meridional density gradient (Stommel 1961; GT95; Zhao et al. 2016; Shi and Yang 2021). As an alternative, we use a nonlinear relation analogous to Cessi (1994) and RT97 to introduce a degree of nonlinearity in this study. Now, Eq. (4b) becomes,

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

At $k = 1$, Eq. (10) is reduced to the linear Eq. (4b); and the system exhibits growing oscillation under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$ (Figs. 6a, 7a, black curves). 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 gray lines and the abscissa axis in Fig. 7c mark the upper and lower limits for $\Delta\rho'$ during the integration. As $\Delta\rho'$ grows, the nonlinear bounding effect of Eq. (10) gradually emerges, limiting the fluctuation tendency of q'. The bounding manifested as the difference between the solid and orange curves is very small (Fig. 7c). Hence, even a tiny degree of internal nonlinearity from the AMOC-meridional density gradient relation can lead to self-sustained oscillation. The period here is 303 years, hardly deviated from the 306-year eigen period of the linear system, reflecting again the robustness of the advection-dominated eigenmode.

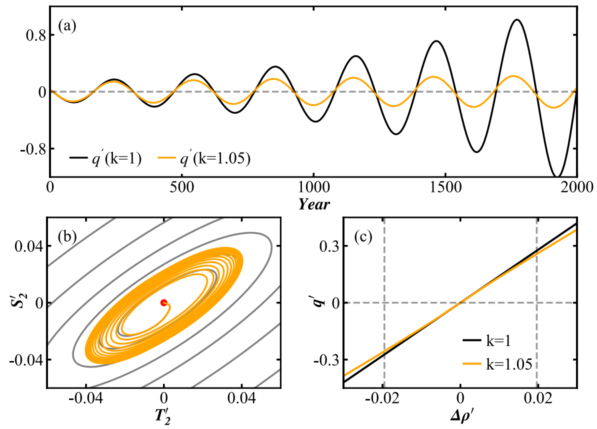

FIG. 7. Oscillations under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$. (a) Time series for q' (units: Sv) under $k = 1$ (black curve) and $k = 1.05$ (orange curve). (b) T_2'-S_2' phase space diagrams for years 1-10000. The red dot represents the initial location of T_2' and S_2'. Black curve is for $k = 1$, and orange curve, for $k = 1.05$. (c) Variations 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 gray 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

In theoretical models, model parameters can be tuned to control oscillation properties. GT95 identified a multidecadal oscillation in their 4TS model, whose period is far shorter than our multicentennial period. RT97 used a 3TS model, and also identified a multi-decadal mode. The system stability in RT97's model is far lower than ours. Since both GT95 and RT97 also employed mixed boundary conditions, and their core dynamics are all advection feedbacks, such differences in eigenmodes are likely to originate from parameter choices. Previously studies usually tuned parameters to study multi-equilibria problems (Colin de Verdière et al. 2006; Colin de Verdière 2007; Sévellec et al. 2010; Sévellec and Fedorov 2014). In single-equilibrium oscillation studies, the parameters effects on the eigenmode have not been widely heeded, which will be addressed next via numerical stability analyses.

a. Effect of basin geometry

Basin geometry can affect both the period and e-folding time of the system eigenmode (Fig. 8). 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). 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 stars in Fig. 8. In GT95, the fractions of the upper and subpolar ocean boxes are both 1/11, falling in the lower left corner of Fig. 8 (denoted by the green star), with a period less than 200 years if $\overline{q}$ is set to 10 Sv. Actually, the $\overline{q}$ in GT95 was set to a higher value of about 17 Sv, representing a much faster overturning rate; thus, the period is further shortened to the century scale. Consequently, it is reasonable to deduce that the multi-decadal period in GT95 is not at odds with our multicentennial period. Popularity of multi-decadal phenomena back then might account for their choice of model parameters.

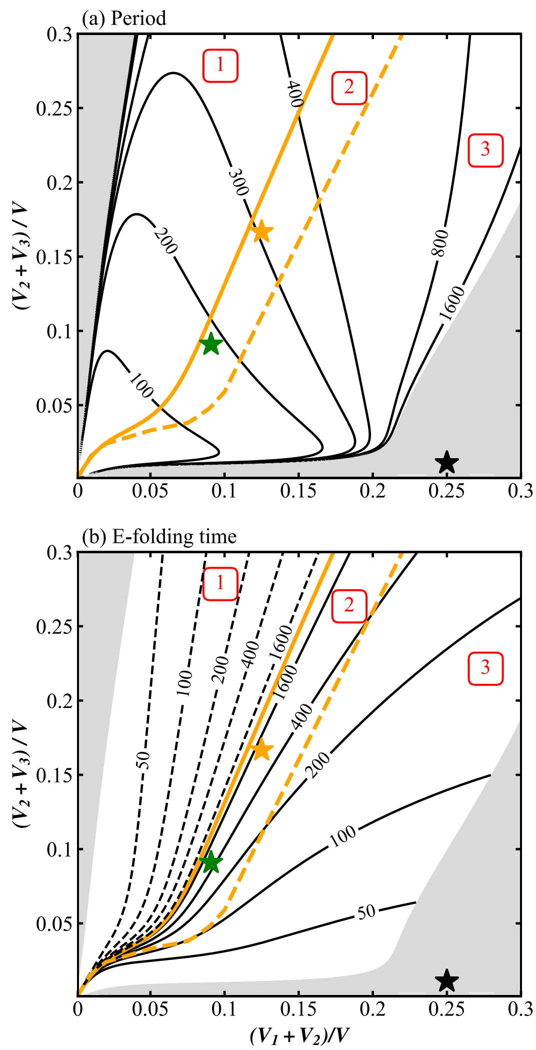

FIG. 8. Sensitivity of (a) period (units: years) and (b) e-folding time (units: years) of the eigenmode to subpolar ocean fraction $(V_2 + V_3)/V$ and upper ocean fraction $(V_1 + V_2)/V$ under $\lambda = 14\ Sv \cdot kg^{-1}\ m^3$. The orange star denotes the mode with standard values used in this work. The green and black stars denote the standard values used in GT95 and RT97, respectively. The solid orange curve is both the stability threshold and the lower limit of probability for self-sustained oscillation. The dashed orange curve is the upper limit of

probability for self-sustained oscillation. The light gray areas correspond to purely damped or growing regime without the imaginary part. The oscillatory mode is damped in region 1, potentially self-sustained in region 2 if bounding terms are affiliated, and growing in region 3. The values of the other parameters are the same as those listed in Table 1.

In Fig. 8b, a higher $(V_1 + V_2)/V$ and a lower $(V_2 + V_3)/V$ are linked to lower stability. This explains why the mode in RT97 can be easily unstable even under a low AMOC sensitivity (equivalent to $\lambda = 5.7\ Sv \cdot kg^{-1}\ m^3$ in this paper). The basin geometry of RT97 is denoted by the black star in Fig. 8. Their high-latitude box stands for a small deep-water formation region instead of the subpolar region, so it was set to only around 1/100 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. The light gray areas in Fig. 8 denote purely damped or growing region without oscillatory potentials. The solid orange curve partitions the stable and unstable regions, making itself also the lower limit for a possible self-sustained oscillation. The oscillatory mode is damped in region 1 because of negative real part, and is growing due to positive real part in regions 2 and 3 (Fig. 8b). Nevertheless, we find that only in region 2 could the self-sustained oscillation take place if bounding terms are affiliated, through a large number of numerical experiments (not shown). No self-sustained oscillation is able to exist in region 3, which is separated from region 2 by the dashed orange curve. Compared to Fig. 9 in LY22, the probability for a self-sustained oscillation is increased in the 4TS model, since the area of region 2 is larger than that in the 4S model of LY22.

b. Effect of mean flow

Given the meridional density gradient, the total AMOC strength q is determined by its sensitivity λ to the meridional density gradient 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. 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 density gradient, contributing to a less stable system. A decreased $\overline{q}$ weakens the equilibrium advection terms $\overline{q}(T_1' - T_2')$ and $\overline{q}(S_1' - S_2')$, therefore 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 advection under a smaller $\overline{q}$ is to make the system more unstable.

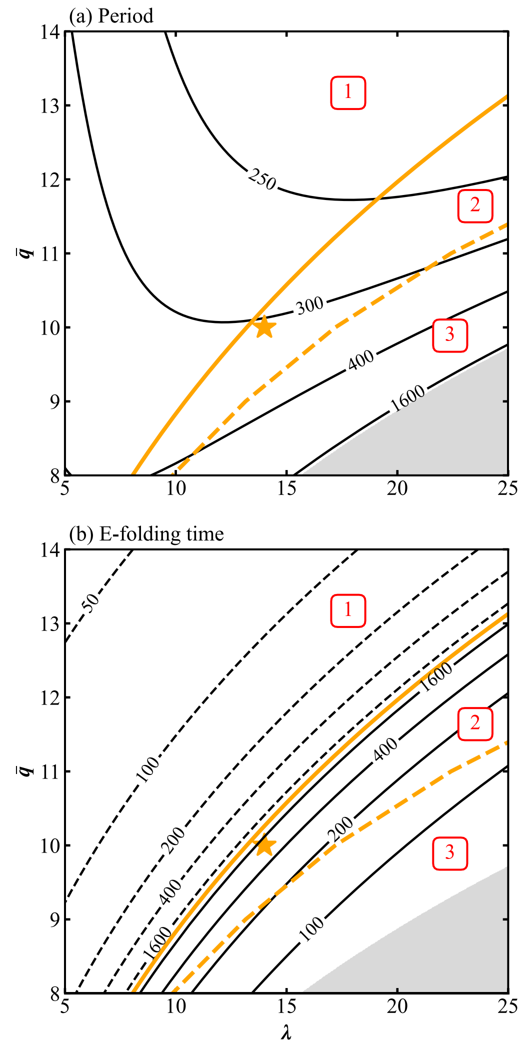

FIG. 9. Same as Fig. 8, but the ordinate and abscissa correspond to the equilibrium AMOC strength $\overline{q}$ (units: Sv) and the linear closure coefficient λ (units: $Sv \cdot kg^{-1}\ m^3$), respectively.

c. *Effect of boundary conditions*

Mixed boundary conditions are adopted in the 4TS model, where T_1^* and T_2^* control the surface heat flux while F_w controls the surface virtual salt flux. The meridional restoring temperature gradient $T_1^* - T_2^*$ influences the system, while the exact values of T_1^* and T_2^* have no effect. Figure 10a shows that the period shortens marginally with the increase of $T_1^* - T_2^*$, but exhibits a decrease-to-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 $\overline{T_1} - \overline{T_2}$ lowers as $T_1^* - T_2^*$ decreases; therefore, the temperature effects are hampered due to the weaker negative temperature-advection feedback. We illustrated in section 3.1 that a faster restoring denoted by a smaller $1/\tau$ also limits the temperature effects. Hence, the temperature effects are promoted by the increases in $T_1^* - T_2^*$ and $1/\tau$, followed by a more stable system with a shorter period. It can also be seen from Eq. (2b) that a larger F_w increases $q'\left(\overline{S_1} - \overline{S_2}\right)$, so the destabilizing positive salt-advection feedback is reinforced, which is consistent with the finding of Sévellec et al. (2006).

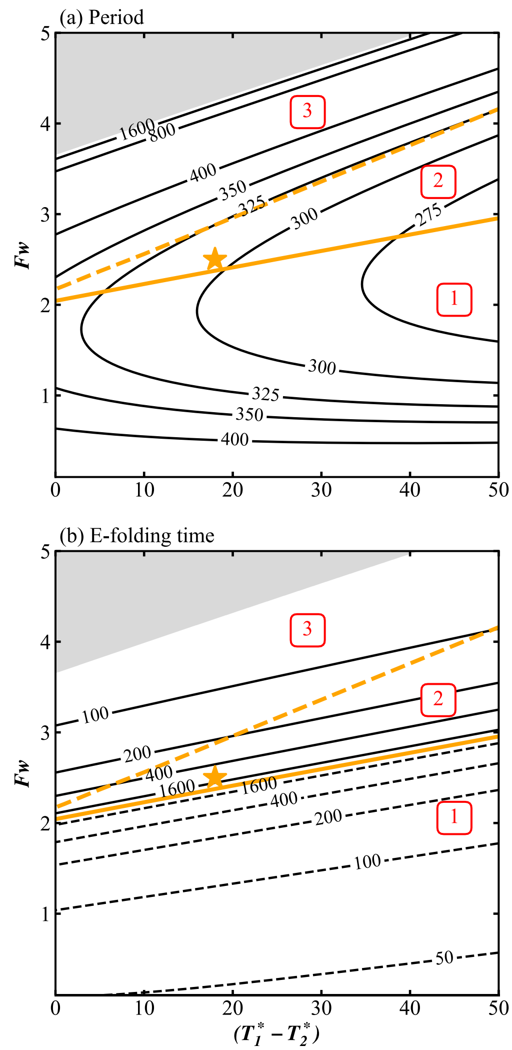

FIG. 10. Same as Fig. 8, but the ordinate and abscissa correspond to the freshwater flux F_w (units: 10 $psu \cdot Sv$) and the meridional restoring temperature gradient $T_1^* - T_2^*$ (units: °C), respectively.

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 thermal process includes the negative temperature-advection feedback and positive restoring-advection feedback. The latter never overruns the former; thus, the resultant temperature feedback is negative. Including the thermal process leads to an acceleration of oscillation because of the fast thermal-restoring process, a stabilization force for the system because of the negative temperature-advection feedback and a portion of stabilization for the subpolar stratification due to temperature stratification.

Similar to LY22, bounding processes are needed to realize a self-sustained oscillation, which can be either enhanced subpolar mixing or weak nonlinearity in the AMOC-meridional density gradient relation. Multicentennial eigenmode is robust regardless of these bounding processes, because the oscillatory eigenmode is fundamentally determined by advection processes. Only a tiny magnitude of the bounding process is able to realize the self-sustained oscillation. Same as in LY22, the effects of nonlinear temperature and salinity advection terms on the self-sustained oscillation are trivial and can be safely neglected (figure not shown), and the external stochastic forcing can also excite a sustainable multicentennial oscillation (figure not shown). Compared to the 4S model in LY22, the probability for a self-sustained oscillation in the 4TS model is much increased with temperature equations added.

Stability analyses reveal that the period and stability of the oscillatory eigenmode are sensitive to model geometry, flow properties and boundary conditions. Generally, smaller subpolar and upper oceans tend to shorten the period. Larger subpolar ocean and smaller upper ocean have stabilizing effects on the system. A stronger AMOC shortens the period due to the faster overturning rate, stabilizing the system through balancing the positive salt-advection feedback more quickly. Higher AMOC sensitivity to the meridional density gradient makes the system less stable. Increasing surface freshwater flux energizes the destabilizing salt-advection feedback, and lowers the system stability, because the background meridional salinity gradient will be stronger. It also lengthens the period of the system because more time is needed to consume the stronger background salinity gradient. Larger meridional restoring temperature gradient strengthens the thermal process; thus, it shortens the period and increases the system stability.

The box model is highly idealized, aimed at providing heuristic understanding of the multicentennial AMOC oscillation. 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 (Vellinga and Wu 2004; 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 study suggests that a self-sustained oscillation can appear as long as a tiny magnitude of nonlinearity or additional mixing is included, which is easy to realize in the realistic ocean. Thereby, we conclude that even with random components removed completely, self-sustained multicentennial oscillation has a good chance to exist in high-order models.

The core of the oscillation mechanism here is the advection process, consistent with many previous studies (Mikolajewicz and Maier-Reimer 1990; Winton and Sarachik 1993; Drijfhout et al. 1996; Delworth and Zeng 2012). Sensitivity of period to model geometry is observed not only in our theoretical model but also in higher complexity models (Weaver and Sarachik 1991; Drijfhout et al. 1996; Delworth and Zeng 2012). Since the flow rate and route affect the overturning rate, a more precise simulation of AMOC structure and a finer model resolution are likely to improve the simulation of AMOC oscillation. Although the boundary conditions in our model influence the eigenmode, the essence for such impact is climate feedbacks. It inspires us that a better representation of climate feedbacks in high-order models may improve their performances.

The warming and freshwater hosing in the North Atlantic will reduce the meridional temperature gradient while enhance the meridional salinity gradient, hampering the negative temperature-advection feedback and strengthening the positive salt-advection feedback. On one hand, this implies that the AMOC might march gradually toward (not necessarily reach) the collapse state (Gregory et al. 2005; Sévellec et al. 2017; Dai 2022), since its stability is likely to reduce, as revealed in this paper. On the other hand, this also implies that the period for the multicentennial timescale portion of the AMOC oscillation is likely to be lengthened in the future, which has not gained attention yet. However, the portion with decadal to multi-decadal periods of the AMOC 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 the multicentennial AMOC period would change in the future, since the global warming might occur on the background of a multicentennial oscillation.

This theoretical study can be improved in several aspects. The one-hemisphere configuration singles out only North Atlantic advection, and contributions from other ocean basins are not considered. Extending the one-hemisphere model into an inter-hemisphere one as 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 natural feedbacks are reserved in our model because of the ocean-only configuration and mixed boundary conditions. Adding more feedbacks, such as 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 authenticity 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 and no datasets were generated during the current study.

References

Bretherton, F. P., 1982: Ocean climate modeling. *Prog. Oceanogr.*, **11,** 93-129.

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. Y. Liu, S. Q. Zhang, W. Liu, L. N. Dong, P. Liu, and H. L. Li, 2016: Reduced interdecadal variability of Atlantic meridional overturning circulation under global warming. *Proc. Natl. Acad. Sci. USA*, **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.

Colin de Verdière, A., 2010: The instability of the thermohaline circulation in a low-order model. *J. Phys. Oceanogr.*, **40,** 757-773.

Colin de Verdière, A., M. Ben Jelloul, and F. Sévellec, 2006: Bifurcation structure of thermohaline millennial oscillations. *J. Climate*, **19,** 5777-5795.

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. 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**.

Drijfhout, S., C. Heinze, M. Latif, and E. MaierReimer, 1996: Mean circulation and internal variability in an ocean primitive equation model. *J. Phys. Oceanogr.*, **26,** 559-580.

Gregory, J. M., and Coauthors, 2005: A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO2 concentration. *Geophys. Res. Lett.*, **32,** 5.

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

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

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. *Journal of Geophysical Research-Oceans*, **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. 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., 1996: Analysis of thermohaline feedbacks, in Decadal Climate Variability: Dynamics and Predictability, D. L. T. Anderson and J. Willebrand, eds., NATO ASI Series, Series I**,** 333-378.

Marotzke, J., and P. H. Stone, 1995: Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *J. Phys. Oceanogr.*, **25,** 1350-1364.

Mikolajewicz, U., and E. Maier-Reimer, 1990: Internal secular variability in an ocean general circulation model. *Climate Dyn.*, **4,** 145-156.

Muir, L. C., and A. V. Fedorov, 2015: How the AMOC affects ocean temperatures on decadal to centennial timescales: the North Atlantic versus an interhemispheric seesaw. *Climate Dyn.*, **45,** 151-160.

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.

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.

Sévellec, F., T. Huck, and C. d. V. A, 2010: From centennial to millennial oscillation of the thermohaline circulation. *J. Mar. Res.*, **68,** 723-742.

Sévellec, F., A. V. Fedorov, and W. Liu, 2017: Arctic sea-ice decline weakens the Atlantic meridional overturning circulation. *Nat. Climate Change*, **7,** 604-610.

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.

Shi, J. Q., and H. J. Yang, 2021: Bjerknes compensation in a coupled global box model. *Climate Dyn.*, **57,** 3569-3582.

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.

Walin, G., 1985: The thermohaline circulation and the control of ice ages. *Palaeogeogr. Palaeocl. Palaeoecol.*, **50,** 323-332.

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.

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., and Coauthors, 2019: A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Rev. Geophys.*, **57,** 316-375.

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.

Zhao, Y. Y., H. J. Yang, and Z. Y. Liu, 2016: Assessing Bjerknes compensation for climate variability and its time-scale dependence. *J. Climate*, **29,** 5501-5512.