The Stability of Thermohaline Circulation in a Two-Box Model*

YOUNG-GYU PARK

MIT/WHOI Joint Program in Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 31 March 1997, in final form 4 February 1999)

ABSTRACT

In Stommel's simple two-box model, which has provided an insight on the thermohaline circulation and climate instability mechanisms, a linear mass transport was used. However, a scaling law based on geostrophy and advective–diffusive heat balance suggests a nonlinear mass transport relation for the oceans. By including this nonlinear mass transport relation to Stommel's box model, it is possible to study the effects of the thermocline, which was not considered before, on the stability of the thermohaline circulation while keeping the simplicity of Stommel's box model. The results were compared with those obtained with the traditional model using a linear mass transport relation. The thermal mode circulation of the nonlinear model is significantly more stable than that of the linear model, suggesting the thermohaline catastrophe is less likely to occur in the present North Atlantic if the thermocline is considered. In the nonlinear model, the circulation removes density anomalies rapidly so that significantly higher haline forcing is needed to initiate the thermohaline catastrophe. A linear stability analysis shows that negative feedback from the mass transport law has the strongest effect on the stability within a parameter range relevant for the present North Atlantic. The analysis also shows that freshwater flux parameterization does not have significant effect on the stability excluding artificial stability due to the details of the salinity restoring boundary condition.

1. Introduction

The sea surface thermal forcing, which acts in the direction of the present thermohaline circulation in the North Atlantic, is opposite to the haline forcing from the freshwater flux. The strong negative feedback between the sea surface temperature and the surface heat flux removes changes in the sea surface temperature rapidly. However, the freshwater flux, which arises from the local imbalance between precipitation and evaporation, is independent of the sea surface salinity. The difference in the boundary conditions for temperature and salinity may give rise to multiple equilibria of the thermohaline circulation under identical boundary conditions (see Weaver and Hughes 1992; Marotzke 1994; Whitehead 1995 for reviews). Therefore, the thermohaline circulation can switch from one equilibrium to another rapidly (thermohaline catastrophe) if the thermal or haline forcing is perturbed. Studies ranging from simple box models to fully developed primitive-equation numerical models have demonstrated the thermohaline catastrophe.

It has been suggested that the mechanism behind the thermohaline catastrophe is a large-scale advective process (Walin 1985), but the effect this advective process has on stability of the thermohaline circulation has been neglected. Instead, many studies have been focused on boundary conditions for salinity, that is, the parameterization of the air–sea freshwater exchanges and their effects. In Stommel's (1961) two-box model, a *linear mass transport relation* was assumed, and the strength of the circulation Ψ through a capillary pipe connecting two boxes,

$$\Psi \sim \Delta\rho,$$

where $\Delta\rho$ is the density difference between the two adjacent boxes. This linear relation has been used in most box models, which suggest that if freshwater flux to the Northern North Atlantic increases by a small amount, the present thermohaline circulation of North Atlantic would be reversed (Huang et al. 1992).

A simple scaling analysis based on geostrophy and advective–diffusive buoyancy balance in the thermocline (Bryan and Cox 1967; Welander 1971) suggests a *nonlinear mass transport relation*,

$$\Psi \sim \Delta\rho^{1/3},$$

for the oceanic thermohaline circulation. The details of the scaling law and studies supporting the scaling law

* Woods Hole Oceanographic Institution, Contribution Number 9441.

Corresponding author address: Dr. Young-Gyu Park, Center for Ocean–Land–Atmosphere Studies, Institute of Global Environment and Society, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705-3106.
E-mail: ypark@cola.iges.org

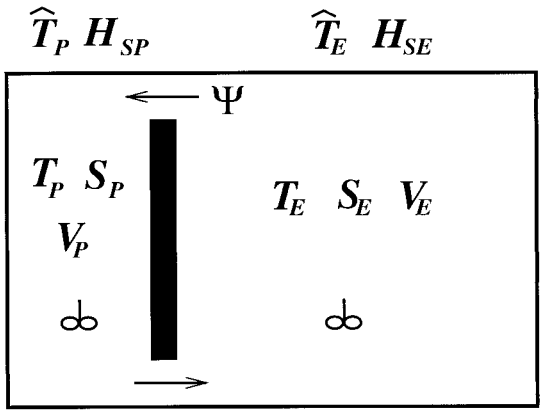

FIG. 1. Configuration of a two-box model. The temperature of each box T_i, where $i = P$ and E, are restored to $\hat{T}_i$. The salinity of each box is S_i and H_S represents virtual salt flux from the freshwater flux parameterization. Volume transport Ψ is positive for counterclockwise circulation (thermal mode or high latitude sinking) and negative for clockwise circulation (haline mode circulation or low latitude sinking).

and the difference between the linear and nonlinear relations are described in section 2.

This scaling law does not add any real complexity to Stommel's classical two-box thermohaline circulation model, but it does change some of the model prediction in an interesting way. In order to compare the effect of mass transport on the stability with that of salinity boundary conditions, both a restoring condition and an interactive condition by Nakamura et al. (1994) are considered in the freshwater flux parameterization. In the interactive case, evaporation ($\mathcal{E}$) minus precipitation ($\mathcal{P}$) depend on the meridional temperature gradient, and a fixed value of water flux is a special case.

The configuration of a two-box model and the scaling law due to Bryan and Cox (1967) are described in section 2. In section 3, the results of the box model are compared to those of a box model using Stommel's linear mass transport relation. Discussions and summaries are given in section 4.

2. A two-box model

The classical model, based on Stommel (1961), consists of a polar box and an equatorial box as shown in Fig. 1. The equations of temperature and salinity in each box, using the notation of Thual and McWilliams (1992), are

$$\begin{aligned} V_P\dot{T}_P &= C_{TP}(\hat{T}_P - T_P) + |\Psi|(T_E - T_P) \\ V_P\dot{S}_P &= V_P H_{SP} + |\Psi|(S_E - S_P) \\ V_E\dot{T}_E &= C_{TE}(\hat{T}_E - T_E) + |\Psi|(T_P - T_E) \\ V_E\dot{S}_E &= V_E H_{SE} + |\Psi|(S_P - S_E). \end{aligned} \tag{1}$$

The subscripts P and E denote quantities for the polar and equatorial boxes, respectively. The restoring condition is used for temperature boundary condition and the quantities with a caret ($\hat{T}_i$) denotes reference temperature. The restoring timescale for temperature $\tau_T = V_i/C_{Ti}$, where the volume of a box is V_i. Virtual salt flux H_{Si}, where $i = E$ or P, is from the commonly used restoring condition or the interactive condition by Nakamura et al. (1994). Since the parameterization of freshwater flux is not the goal of this study, each method is described in appendix A.

A scaling law

A scaling law describing meridional mass and heat transports due to thermohaline circulation is proposed by Bryan and Cox (1967) and is equivalent to the classical thermocline theory by Robinson and Stommel (1959). The thermal wind relation $f\upsilon_z \approx -g\rho_x/\rho_o$ within the thermocline yields

$$V \sim \frac{g\Delta\rho\delta_T}{\rho_o f L} \tag{2}$$

if we assume a linear relation between meridional and zonal density gradient. Here V is meridional velocity scale, δ_T is the thickness scale of the thermocline, L is the width scale of the ocean. In Hughes and Weaver (1994), the meridional difference of the depth-integrated pressure ΔP is compared to the strength of meridional mass transport Ψ. The scale of pressure variation due to buoyancy forcing is $g\delta_T\Delta\rho$, so $\Delta P \sim g\delta_T^2\Delta\rho$. If there is a linear relation between the zonal and the meridional pressure gradients and if geostrophy is valid, we would get a linear relation between ΔP and Ψ, the depth-integrated meridional velocity, as Hughes and Weaver (1994) found.

The advective–diffusive buoyancy balance $w\rho_z \approx \kappa\rho_{zz}$ (Munk 1966) gives

$$W \sim \frac{\kappa}{\delta_T}, \tag{3}$$

where W is the vertical velocity scale and κ is the vertical diffusivity. From the continuity equation $\upsilon_y \approx w_z$,

$$W \sim \frac{V\delta_T}{L}. \tag{4}$$

The meridional mass transport Ψ and buoyancy transport B are

$$\Psi \approx V\delta_T \sim \left(\frac{g\kappa^2L^4}{\rho_o f}\right)^{1/3}\Delta\rho^{1/3}$$

and

$$B \approx \Psi\Delta\rho \sim \left(\frac{g\kappa^2L^4}{\rho_o f}\right)^{1/3}\Delta\rho^{4/3}. \tag{5}$$

Bryan (1987) and Colin de Verdière (1988) studied the dependence of the scaling law on the vertical dif-

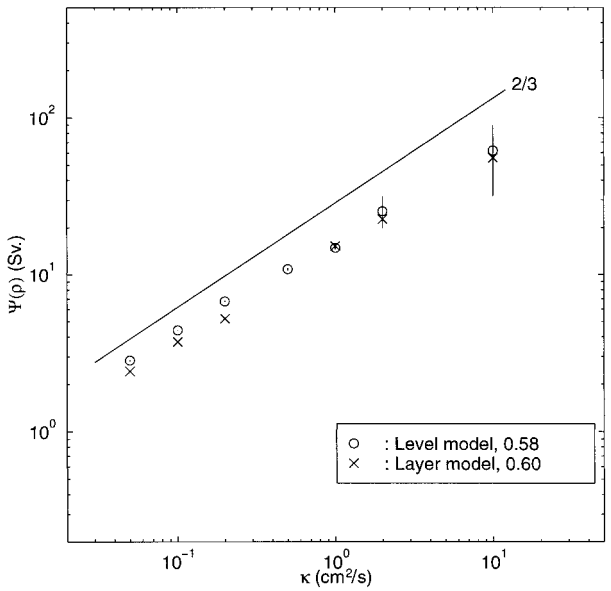

FIG. 2. Comparison of the vertical diffusivity κ and $\max(\Psi(\rho))$ using results from a depth level coordinate model (GFDL MOM 2) and an isopycnal layer model designed by Hallberg (1996). The results excerpted from Park and Bryan (2000).

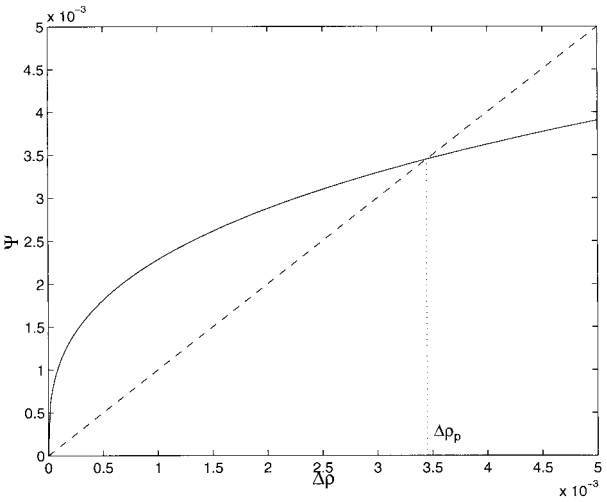

FIG. 3. Meridional mass transport Ψ vs $\Delta\rho$. The solid curve is for the nonlinear model Ψ_G, and the dashed line is for the linear model Ψ_F. The value of the present North Atlantic is shown as $\Delta\rho_p$.

fusivity κ, using three-dimensional numerical models. They found that the meridional heat transports follow the scaling law reasonably well, but the maximum of the meridional mass transport (Bryan 1987) or velocity scale from total kinetic energy over the entire basin (Colin de Verdière 1988) does not. Through a series of numerical experiments, Park and Bryan (2000) found that there is a significant amount of recirculation occurring on slanted isopycnal surfaces in the northern region of weak stratification. The meridional overturning circulation defined along depth level surfaces $[\Psi(z) \equiv \int_W^E \int_{\text{botom}}^{\text{top}} v\, dz\, dx]$ contains recirculation that cannot contribute to the meridional buoyancy transport. The contribution of the recirculation becomes stronger toward the polar wall where the strength of the overturning circulation was estimated in Bryan (1987).

Using two numerical models Park and Bryan (2000) show that meridional overturning circulation defined along isopycnal surfaces $[\Psi(\rho) = \int_E^W \int_{\rho_{\text{bottom}(x,\theta)}}^{\rho} (\partial z/\partial\rho')\, v(\theta,\ \rho')\, d\rho'\, dx]$ follows the scaling law closely as shown in Fig. 2, in which $\max(\Psi(\rho))$ and κ are compared. In their study, the power dependence of $\max(\Psi(z))$ on κ, which is estimated from the same results used in the figure, is 0.45 and does not follow the scaling law. In Winton (1996), the strength of the meridional overturning circulation is estimated away from the northern region of weak stratification and the effect of the recirculation on the scaling was avoided; he obtained results consistent with the scaling law. Park and Bryan (2000) also show that the discrepancy in Colin de Verdière (1988) is due to fact that the thickness of the thermocline is not considered in his scaling analysis of the kinetic energy. The good agreement between the scaling law and the numerical experiment in Park and Bryan (2000) supports the assumptions used in derivation of the scaling law. The results of a laboratory experiment, in which $\Delta\rho$ is varied, performed by Park and Whitehead (1999) further verify the heat transport relation based on the scaling law as in Eq. (5).

The volume transport in this study is either from the scaling law [Eq. (5)] or from Stommel's linear relation, so Ψ is one of the following:

$$\Psi \equiv \begin{cases} \Psi_G = C_G[\alpha(T_E - T_P) - \beta(S_E - S_P)]^{1/3} & \text{nonlinear} \\ \Psi_F = C_F[\alpha(T_E - T_P) - \beta(S_E - S_P)] & \text{linear,} \end{cases} \tag{6}$$

where α is the thermal expansion coefficient, β is the salt expansion coefficient, and C_G and C_F are constants whose values are set to yield $\Psi = 10$ Sv (Sv $\equiv 10^6$ m^3 s^{-1}) when $\Delta\rho = \Delta\rho_p$ ($\Delta T = 20$°C and $\Delta S = 2$ ppt), which represents the meridional sea surface density gradient of the present North Atlantic. If we neglect the advective–diffusive buoyancy balance in the scaling law, there would be no thermocline and the thickness scale of thermocline δ_T becomes irrelevant. In Eq. (2) δ_T can be ignored and $\Psi \sim V \sim \Delta\rho$. If thermodynamics is ignored, the scaling law yields a linear mass transport relation as in Stommel (1959).

In Fig. 3, the two mass transport laws are compared. Both scaling laws give the same mass transport, by definition, when $\Delta\rho = \Delta\rho_p$. When the haline forcing becomes larger than the present value so that $\Delta\rho < \Delta\rho_p$, we find that $\Psi_G/\Psi_F > 1$. Thus, the meridional heat and salt transports become relatively greater in a model with the nonlinear mass transport law. The main result is that the nonlinear relation makes the model less sensitive to the forcing than the classical model except at very low values of $\Delta\rho < \Delta\rho_c$.

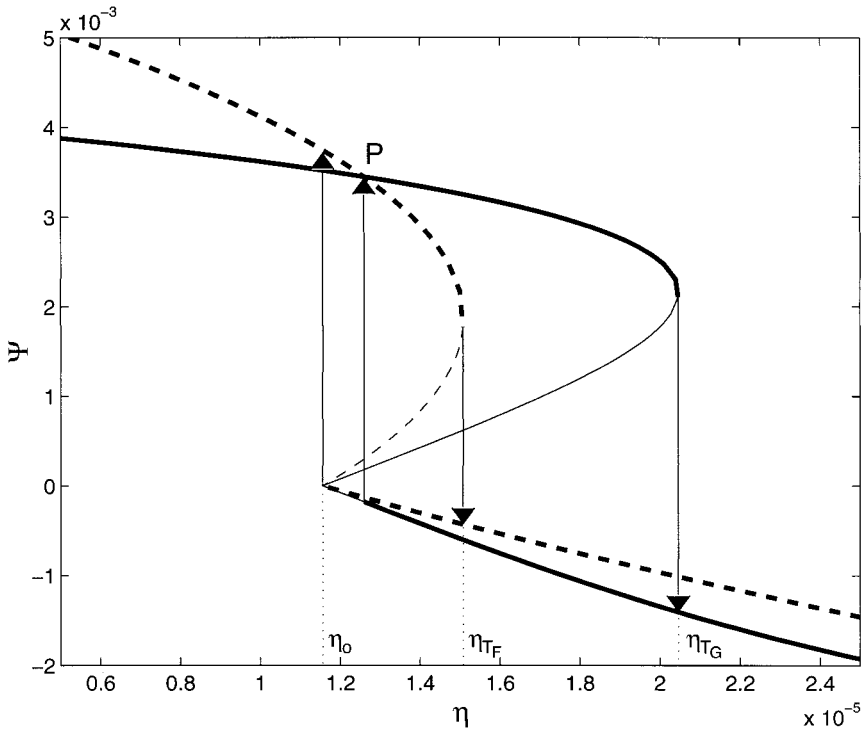

FIG. 4. Mass transport Ψ vs haline forcing η, with $\zeta = 0.005\,77$ and $\tau = 1$. The solid curves are for the nonlinear model and the dotted curves are for the linear model. The lower branches ($\Psi < 0$) are for the haline modes and the upper branches are for the stable thermal mode in each case. The middle branches in each case are for the unstable thermal modes. Thick curves represent stable solutions in each case. The arrows indicate the direction of change in each case. Here, $V = V_E + V_P = 5000$ km $\times$ 5000 km $\times$ 4 km, $V_P = 0.1V$, $V_E = 0.9V$, $\alpha = 2 \times 10^{-4}$ °C^{-1}, $\beta = 8 \times 10^{-4}$ ppt^{-1}, $\tau = 1$ ($\tau_T = 100$ days).

Following Thual and McWilliams (1992), the scaling factors for the variables are

time:	$\tau_{T_o} = 100$ days
mass transport:	$(1/V_P + 1/V_E)^{-1}/\tau_{T_o}$
temperature:	$(1/V_P + 1/V_E)^{-1}/\alpha\tau_{To}C_F$
salinity:	$(1/V_P + 1/V_E)^{-1}/\beta\tau_{To}C_F$.

The mass transport Ψ is determined by $\Delta\rho = \alpha(T_E - T_P) - \beta(S_E - S_P)$, so $\Theta = T_E - T_P$ and $\Sigma = S_E - S_P$ are more convenient variables than T and S. The dimensionless equations are then

$$\dot{\Theta} = \zeta - \Theta(\tau^{-1} + |\Psi|),$$
$$\dot{\Sigma} = H_S - \Sigma|\Psi|, \tag{7}$$

$$\Psi = \begin{cases} \lambda(\Theta - \Sigma)^{1/3} = \lambda\Delta\rho^{1/3} & \text{nonlinear model} \\ \Theta - \Sigma = \Delta\rho & \text{linear model} \end{cases} \tag{8}$$

$$H_S = \begin{cases} \eta - \tau^{-1}\xi\Sigma & \text{restoring condition} \\ F_n\Theta^n & \text{interactive condition,} \end{cases} \tag{9}$$

where $\lambda = C_G C_F^{1/3}[(1/V_P + 1/V_E)\tau_T]^{2/3}$. The thermal forcing $\zeta = \hat{T}_E - \hat{T}_P$, and the haline forcing $\eta = \xi(\hat{S}_E - \hat{S}_P)$ (or F_n and n); $\xi = \tau_T/\tau_S$ and $\tau = \tau_T/\tau_{T_o}$ are the control parameters of the model. The multiple equilibria occur when $\xi < 1$. Thual and McWilliams (1992) showed that when $\xi = 0.002$ a linear two-box model and a two-dimensional model show similar bifurcation structures. In this study, we fix $\xi = 0.002$, and the effect of varying ξ was not studied.

3. Results

a. Restoring salinity boundary condition

In Fig. 4, hysteresis diagrams from models with the restoring salinity boundary condition are presented. In the linear model, the present thermal-model circulation (point P) becomes unstable and switches to a haline model circulation if the haline forcing increases by 20%. This is similar to earlier box models using the linear mass transport relation such as Huang et al. (1992), in which H_S is from ($\mathcal{E} - \mathcal{P}$). In contrast, in the nonlinear model 60% increase in the haline forcing is required for the instability of the present thermal-model circulation. In this model it is less likely that a small increase in the haline forcing (or freshwater flux) would cause the catastrophic transition. Results with different thermal forcing show the same tendency.

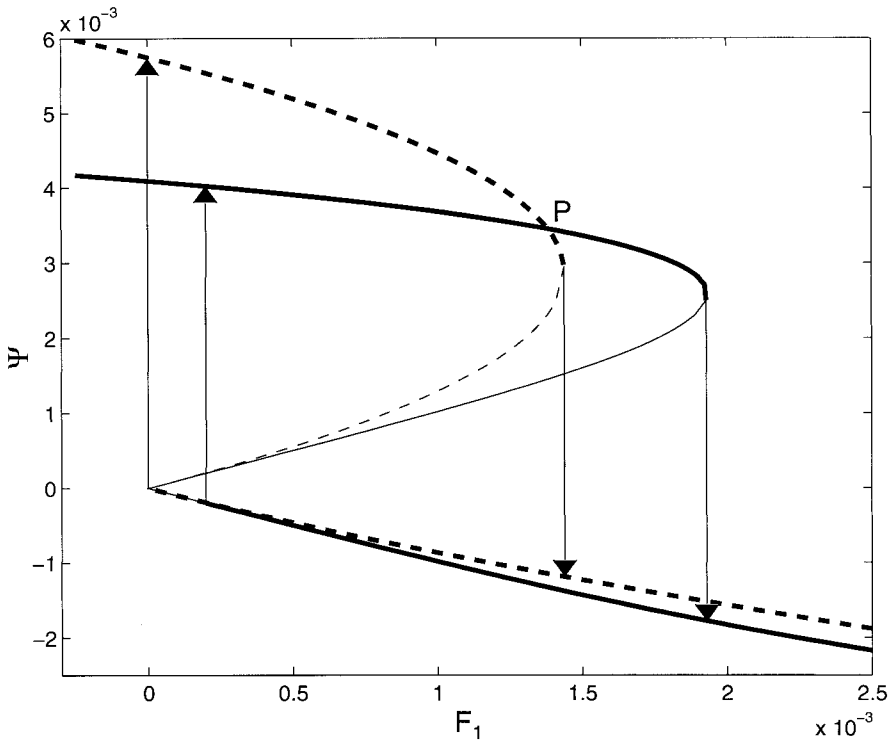

FIG. 5. Mass transport Ψ vs F_1 with $\tau = 1$. See Fig. 4 for the explanation.

Joyce (1991) used a two-hemisphere model with the linear mass transport law. When an asymmetric sinusoidal perturbation was applied to the haline forcing, oscillation between a thermal mode and a haline mode with a frequency different from that of the perturbation occurred. He suggested that this oscillation might be related to the glacial oscillation. Thual and McWilliams (1992) showed that, when $\xi \ll 1$, the circulation in a complicated box model can be treated as a linear superposition of a two-box model, so Joyce's (1991) calculation can be compared to that of the present study. The oscillations suggest that the amplitude of the perturbation in the boundary condition is comparable to the width of the multiple equilibria region of his model. Since a nonlinear model shows a significantly wider multiple equilibria region than that of a linear model, a version of Joyce's model with the nonlinear scaling law is less likely to show oscillations between a thermal mode and a haline mode under the same perturbation.

b. Interactive salinity boundary condition

In Fig. 5, hysteresis diagrams from models with the interactive boundary condition are presented with $n = 1$ as in Marotzke and Stone (1995) and considering F_1 to be an external parameter. The results are qualitatively similar to those with the restoring salinity boundary condition. The former apparently, however, shows much wider multiple equilibria regions because the lower bound of the multiple equilibria regions is at the vanishing haline forcing. When $\tau_S \to \infty$ so $\xi \to 0$, the lower bound of the multiple equilibria region $\eta_0 = \xi\zeta$ in Fig. 4 approaches $\eta = 0$ (where haline forcing vanishes). If the salinity restoring boundary condition is used properly as suggested by Welander (1986), salinity boundary conditions do not affect the lower catastrophic transition point.

The effect of the positive feedback due to the interactive condition is prominent in the linear model, so a 5% increase in haline forcing makes the present thermal mode circulation (point P in Fig. 5) unstable. In the nonlinear model the positive feedback is less important so that the present state is still quite far away from the transition point. We can clearly see that the mass transport relation has a stronger effect on the stability of a thermal mode circulation than the parameterization of the air–sea freshwater exchange. This comparison is, however, qualitative, so a linear stability analysis is performed in the next section. Since we are more interested in the transition from a thermal mode to a haline mode, the stability analysis was done with a thermal mode circulation.

c. Linear stability analysis

Apply a small perturbation (Θ', Σ') to an equilibrium thermal mode equilibrium circulation $(\overline{\Theta}, \overline{\Sigma})$ without changing forcing. In the nonlinear model with the restoring salinity boundary condition, $\overline{\Theta}$ and $\overline{\Sigma}$ satisfy

$\zeta - \overline{\Theta}(\tau^{-1} + \overline{\Psi}) = 0$ and $\eta - \overline{\Sigma}(\tau^{-1}\xi + \overline{\Psi}) = 0$.

The equations for the perturbation variables from Eq. (7) in matrix form are

$$\begin{pmatrix} \dot{\Theta}' \\ \dot{\Sigma}' \end{pmatrix} = \mathbf{A}\begin{pmatrix} \Theta' \\ \Sigma' \end{pmatrix}. \tag{10}$$

Here

$$\mathbf{A} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with

$$a = -\left(\tau^{-1} + \overline{\Psi} + \frac{1}{3}\lambda^3\overline{\Theta\Psi}^{-2}\right), \qquad b = \frac{1}{3}\lambda^3\overline{\Theta\Psi}^{-2},$$

$$c = -\frac{1}{3}\lambda^3\overline{\Sigma\Psi}^{-2},$$

$$d = -\left(\tau^{-1}\xi + \overline{\Psi} - \frac{1}{3}\lambda^3\overline{\Sigma\Psi}^{-2}\right), \tag{11}$$

where

$$\overline{\Psi} = \lambda(\overline{\Theta} - \overline{\Sigma})^{1/3} \quad \text{and} \quad \Psi' = \frac{1}{3}\lambda^3\overline{\Psi}^{-2}(\Theta' - \Sigma').$$

In appendix B, $\mathbf{A}$ is given for models with a different mass transport law or salinity boundary condition.

In all models considered here, the sum of the two eigenvalues of a model satisfies trace($\mathbf{A}$) < 0. Therefore, if the product of the two eigenvalues is positive, that is, det($\mathbf{A}$) $= ad - bc > 0$, then the eigenvalues are negative and the equilibrium solution of the model is stable. If not, one of the eigenvalues is positive and the equilibrium solution is unstable.

When $\tau < 10$ (or $\tau_T < 1000$ days), $a \approx -\tau^{-1} \gg b$, c, d in all cases considered here. Thus, a thermal mode circulation is stable if

$$\det(\mathbf{A}) \approx -\tau^{-1}d > 0, \quad \text{so} \quad d < 0. \tag{12}$$

As in Walin (1985), assume for a moment that the effect of the salinity restoring boundary condition $-\tau^{-1}\xi$ is negligible. The stability condition becomes independent of the salinity boundary conditions and is completely determined by the advection related properties. The equilibrium circulation ($-\overline{\Psi}$) removes salinity anomalies, so it stabilizes the circulation. The interaction between the circulation induced by the salinity anomalies and the meridional salinity gradient ($\lambda^3\overline{\Sigma\Psi}^{-2}$ in the nonlinear models or $\overline{\Sigma}$ in the linear models) is the *only* significant positive feedback; the salinity anomalies weaken circulation and intensity Σ that prefers a haline mode.

The stability condition yields that the nonlinear thermal mode is stable when the density ratio R, the ratio of the meridional salinity difference $\overline{\Sigma}$ to the meridional temperature difference $\overline{\Theta}$, satisfies

$$R \equiv \frac{\overline{\Sigma}}{\overline{\Theta}} < R_C = \frac{3}{4}, \tag{13}$$

regardless of the salinity boundary conditions. One can easily show that, when $\Psi \sim \Delta\rho^{1/n}$, $R_C = n/(n+1)$. Thus, the linear thermal mode is stable when

$$R < R_C = \frac{1}{2}, \tag{14}$$

as Walin (1985) and Marotzke (1990) show. Tziperman et al.'s (1994) linear four-box model shows that the critical value would decrease to about 0.45 mainly due to the increase in the resolution. Such a small difference may not be significant in such simple models, especially when we are to apply the result to more complicated systems such as the oceans or two- or three-dimensional numerical models. The nonlinear model, however, yields a critical value that is 50% larger than that of the linear one.

It is easy, in fact, to find R that satisfies det($\mathbf{A}$) > 0 for each model when τ varies. In Fig. 6, a curve (R_C), which satisfies det($\mathbf{A}$) $= 0$, is presented for several different cases. To the left of a curve det($\mathbf{A}$) > 0, so the model is stable; to the right det($\mathbf{A}$) < 0, so the model is unstable. Since the strong negative feedback between the sea surface temperature and the atmosphere does not allow much change in temperature, $\overline{\Theta} = \Theta_o = 0.005\,77$. Note that the same equilibrium solutions are compared so that the boundary forcing is not necessarily the same for each case. In the figure, the results from the models with the interactive feedback in which $H_S \sim F_{3.5}\Theta^{3.5}$ as in Nakamura et al. (1994) are included (curves $g1$ and $f1$).

The models with the nonlinear mass transport law show greater stability than those with the linear law when the same salinity boundary conditions are used. The largest contribution to det($\mathbf{A}$) is from advection related properties [$-\tau^{-1}d$ for the model with the interactive condition or $-\tau^{-1}(d + \tau^{-1}\xi)$ for the model with the restoring salinity condition] so that the change in the mass transport law makes the largest change in R_C. The faster circulation in the nonlinear model weakens the meridional salinity gradient and removes the salinity anomalies more effectively. The faster circulation, furthermore, prevents the water in the polar (equatorial) box from obtaining (losing) enough freshwater to build up a strong salinity gradient. Stronger haline forcing is required to reverse the thermal-mode circulation.

The R_C with the restoring salinity boundary condition ($f4$ and $g4$) is larger than that with ($\mathcal{E} - \mathcal{P}$) or the interactive condition, whereas the earlier results [Eqs. (13) and (14)] are independent of salinity boundary conditions. This is because the term $\tau^{-1}\xi$ ($=100$ days/τ_S) is neglected in d in obtaining the results. In the models with the interactive condition $\tau^{-1}\xi$ is absent so the R_C are not significantly different from the previous results (0.5 for the linear models and 0.75 for the nonlinear models).

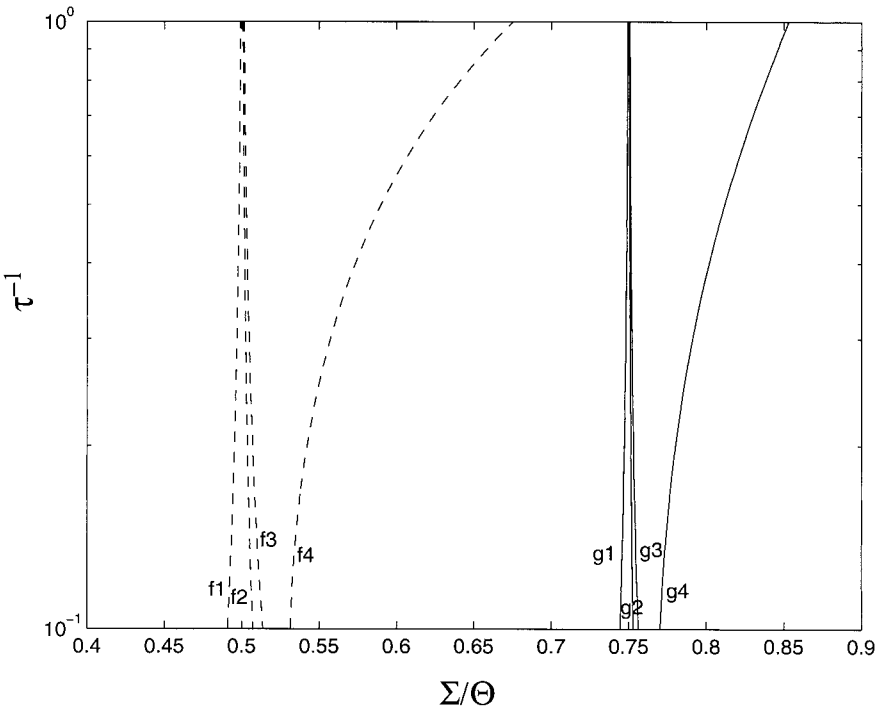

FIG. 6. Stability diagram. A curve denotes R_C that makes det($\mathbf{A}$) = 0 for a certain model. The results from the models with the nonlinear scaling law are labeled using the symbol ''g,'' and those with the linear scaling laws are labeled using ''f.'' The results of the models with the interactive condition with $n = 3.5$ are labeled using ''1'' and with $n = 1$ using ''2,'' those with a fixed value of $(\mathcal{E} - \mathcal{P})$ (or the interactive condition with $n = 0$) using ''3,'' and those with the restoring boundary condition using ''4.'' Here $\xi = 0.002$ and $F_1 = 1.38 \times 10^{-3}$.

In the models with the restoring condition, the R_C decline as the effect of salinity restoring becomes weaker, in other words, τ_S (or τ since $\xi = \tau_T/\tau_S$ is fixed) increases. In this study $\tau_S = 500\tau_T$ so it is about two orders of magnitude larger than commonly used values in other studies, but the effect is still significant. If the artificial stability due to the restoring condition is excluded, the freshwater flux parameterization does not have a significant effect on R_C, as evident in Eq. (12) and Fig. 6; the curves $g1$, $g2$, and $g3$ (or $f1$, $f2$, and $f3$) are not significantly different from each other within a parameter range comparable to that of the present ocean.

4. Discussion and conclusions

Using a two-box model, the effect of mass transport on the stability of a thermal mode (high latitude sinking) circulation has been studied. The equilibrium circulation tries to remove temperature and salinity (or density) anomalies and stabilize the circulation. On the other hand, salinity anomalies try to weaken the circulation and intensify the meridional salinity gradient; this is the strongest destabilizing process. In a model with nonlinear transport, the circulation becomes relatively stronger than that in the linear model as density anomalies intensify. The stronger circulation in the nonlinear model reduces the meridional salinity gradient and removes the anomalies effectively. Thus, nonlinear models show significantly greater stability, irrespective of the freshwater flux parameterization.

During the last glacial period, thermohaline forcing was weaker than that of the present interglacial period. The chemical properties of Greenland ice cores, however, suggest that the climate of the glacial time could be more variable than that of the present interglacial period. (See Fig. 4 of Dansgaard et al. 1982.) The nonlinear mass transport relation (Fig. 3) becomes more sensitive to forcing as the forcing weakens. When $\Delta\rho < \Delta\rho_c$, a small change in $\Delta\rho$ would cause relatively larger and more rapid variation in the circulation and climate than that would do the present circulation as the ice cores suggest.

Two-dimensional (or zonally averaged) numerical models have been widely used to study the stability of thermohaline circulation. While averaging the equation of motion, information on the zonal structure, which drives a meridional circulation, is lost. To close the equation, some assumptions were made. For example, in Marotzke et al. (1988), a balance between the Fickian diffusion of the meridional momentum and the meridional pressure gradient,

$$A_* \overline{v}_{zz} = \overline{p}_y,$$

was assumed with a large value of viscosity A_* adopted to reduce the transport to the present oceanic value. The configuration of the model is the same as that of Rossby's (1965) thermally driven circulation in a nonrotating frame, which yields a scaling law different from the nonlinear one. A similar parameterization is also used in other two-dimensional thermohaline circulation models such as Thual and McWilliams (1992) and Quon and Ghil (1992). Therefore, the results of those models may be different from those of oceans with a nonlinear mass transport relation.

In Wright and Stocker's (1991) two-dimensional model, the parameterization for the east–west pressure gradient is equivalent to a balance between Rayleigh damping of the meridional momentum and the meridional pressure gradient. Therefore,

$$rv \sim p_y,$$

where r is a damping coefficient, and the velocity scale is linear function of the meridional pressure gradient. In the scaling law described in section 2, the velocity scale is also a linear function of the meridional pressure gradient because geostrophy and a linear relation between the zonal and meridional pressure gradient are assumed ($fv \sim p_x \sim p_y$). As Winton (1996) shows the transport relation from a two-dimensional model with Rayleigh damping shows the same power dependence on $\Delta\rho$ as the nonlinear mass transport relation [Eq. (5a)].

The thermal mode circulation with a restoring salinity boundary condition is significantly more stable than that with other salinity boundary conditions considered. The effect, which cannot exist in the real world, disappears as $\tau_S \rightarrow \infty$ so that the boundary condition for salinity becomes independent of the surface salinity. In this study, $\tau_S = \tau_T/0.002$ so it is about two orders of magnitude smaller than those in similar studies but the effect is still significant. Excluding the artificial stability, salinity boundary conditions and freshwater flux parameterization do not have a significant effect on the stability of the thermal mode circulation.

In many two- and three-dimensional numerical studies (for example Marotzke et al. 1988; Rahmstorf 1995), the model flow is initialized using a restoring salinity boundary condition with $1 < \xi < 0.1$. After the circulation reaches an equilibrium state, virtual salt fluxes at the sea surface are estimated. When the restoring boundary condition is replaced by the virtual salt flux diagnosed, the circulation in some cases shows rapid transition to a different state. It has been argued that such a transition indicates that the thermohaline circulation is unstable under mixed boundary conditions (a restoring boundary condition for temperature and a flux boundary condition for salinity). In those studies, the salinity boundary condition supplies an artificial stability during spinup, which is removed when the salinity boundary condition is changed. In Tziperman et al. (1994), a spinup state with longer τ_S does not become unstable while one with shorter τ_S does become unstable upon switching to mixed boundary conditions. It is not clear whether the transition in the thermohaline circulation is caused by the external removal of the artificial stability or by an intrinsic instability in the circulation.

In Rahmstorf (1995), hysteresis from a global general circulation model compares closely to that of Stommel's two-box model, which uses a linear mass transport relation. Therefore, Rahmstorf's thermohaline circulation is significantly less stable than that in the nonlinear model. He used the common method of switching the salinity boundary condition; so it is unclear how much of his hysteresis is related to the stability of the thermohaline circulation.

The models and the real oceans are so different that one can ask what is the proper choice of the tuning parameters (e.g., $\Delta\rho_p$ and Ψ_p) and how do the stability and the bifurcation structure depend on it. Although the choice made in this study is just one realization of the real ocean, it is well within the correct order of magnitude. Furthermore, $\Psi_G/\Psi_F > 1$ if $\Delta\rho < \Delta\rho_p$, irrespective of the choice of the tuning parameters. The thermal mode of a model with nonlinear mass transport is always significantly more stable than that of a model with linear mass transport.

Although we cannot apply the model to the oceans directly, we can roughly estimate the stability parameter of the oceans. The surface temperature and salinity distribution of the North Atlantic give $\overline{\Theta}_{\mathrm{NA}} \approx 20°\mathrm{C}$ and $\overline{\Sigma}_{\mathrm{NA}} \approx 2$ ppt so $R_{\mathrm{NA}} \approx 0.4$. In the North Pacific, $\overline{\Theta}_{\mathrm{NP}} \approx 20°\mathrm{C}$ and $\overline{\Sigma}_{\mathrm{NP}} \approx 3$ ppt so $R_{\mathrm{NP}} \approx 0.6$. If we use the linear model, both stability parameters are close to the critical value 0.5, so we may conclude that both oceans are close to catastrophic transition or they are in a haline mode as Walin (1985) and Huang et al. (1992) argue. If we use the nonlinear model, R_{NA} is significantly smaller than the critical value 0.75, so we may conclude that the thermohaline circulation of the North Atlantic is stable to a small change in freshwater flux. In the case of the North Pacific, R_{NP} is closer to the critical value, so we may conclude that the North Pacific is close to the transition or in a haline mode. (During a haline mode $R > 1$ always regardless of model configuration, so we can easily see the limitation of the model.) We also may argue that this is consistent with Warren (1983) who showed that the freshwater flux to the northern North Pacific prevents deep-water mass formation, so the direction of the thermohaline circulation in the Pacific is opposite to that of the Atlantic.

Acknowledgments. The author thanks Drs. J. A. Whitehead, K. Helfrich, G. Flierl, R. X. Huang, J. Marotzke, N. Hogg, A. Rogerson, K. Bryan, and S. Griffies for their valuable comments and suggestions. This study has been funded by NSF Grant OCE92-01464 and Korean Government Overseas Scholarship Grant. A part of the manuscript was prepared while the author was at the Atmospheric and Oceanic Sciences Program, GFDL/Princeton University.

APPENDIX A

Freshwater Flux Parameterization

The virtual salt flux from evaporation ($\mathcal{E}$) minus precipitation ($\mathcal{P}$) is

$$H_{SP} = (\mathcal{E} - \mathcal{P})\frac{\overline{S}}{d} \quad \text{and} \quad H_{SE} = -H_{SP},$$

where d is the depth of a box and $\overline{S}$ is the mean salinity.

The restoring boundary condition is

$$H_{Si} = \frac{\hat{S}_i - S_i}{\tau_S}.$$

Here $i = P$ or E, $\hat{S}$ is a reference salinity, and τ_S is a restoring timescale for salinity. In the oceans freshwater leaves or enters the surface so that the freshwater flux condition is physical (Huang 1993). However, it is shown that when $\tau_S > \tau_T$, the thermohaline catastrophe is possible under restoring conditions for temperature and salinity (Stommel 1961). It is also shown that, when $\tau_S \to \infty$ and $\hat{S} \to \infty$ with a finite $\hat{S}/\tau_S$, the virtual salt flux is weakly influenced by salinity so that a restoring boundary condition can be used as a first approximation to a flux boundary condition (Welander 1986).

In the interactive condition by Nakamura et al. (1994),[1] a freshwater flux depends on the meridional temperature gradient, so

$$H_{SP} = -\frac{1}{2}F_n(T_E - T_P)^n, \qquad n > 0 \quad \text{and} \quad H_{SN} = -H_{SE}.$$

Here F_n is a moisture transfer coefficient. Nakamura et al. (1994) show that the condition makes a thermal mode circulation less stable compared to one with a virtual flux from fixed ($\mathcal{E} - \mathcal{P}$) or a restoring condition. When $n = 0$, the virtual salt flux from the interactive condition is independent of sea temperature and salinity, so it becomes the same as that from fixed ($\mathcal{E} - \mathcal{P}$).

APPENDIX B

Stability Matrices

If we follow the procedure explained in section 3, the perturbation equation satisfies Eq. (10), while $\mathbf{A}$ for each case is as follows:

frictional model with restoring

$$\mathbf{A} = \begin{pmatrix} -\tau^{-1} - \overline{\Psi} - \overline{\Theta} & \overline{\Theta} \\ -\overline{\Sigma\Psi} & -\tau^{-1}\xi - \overline{\Psi} + \overline{\Sigma} \end{pmatrix}$$

geostrophic model with interactive condition

$$\mathbf{A} = \begin{pmatrix} -\tau^{-1} - \overline{\Psi} - \frac{1}{3}\lambda^3\overline{\Theta\Psi}^{-2} & \frac{1}{3}\lambda^3\overline{\Theta\Psi}^{-2} \\ F_1 - \frac{1}{3}\lambda^3\overline{\Sigma\Psi}^{-2} & -\overline{\Psi} + \frac{1}{3}\lambda^3\overline{\Sigma\Psi}^{-2} \end{pmatrix}$$

frictional model with interactive condition

$$\mathbf{A} = \begin{pmatrix} -\tau^{-1} - \overline{\Psi} - \overline{\Theta} & \overline{\Theta} \\ F_1 - \overline{\Sigma\Psi} & -\overline{\Psi} + \overline{\Sigma} \end{pmatrix}. \qquad \text{(B1)}$$

When $H_S = F_n\Theta^n$ in the interactive condition, from the condition

$$F_1\Theta_p = F_n\Theta_p^n,$$

we obtain

$$H_S' = nF_n\overline{\Theta}^{n-1}\Theta' = nF_1(\overline{\Theta}/\Theta_p)^{n-1}\Theta',$$

where Θ_p is the meridional temperature gradient of the present model ocean. By replacing F_1 with $nF_1(\overline{\Theta}/\Theta_p)^{n-1}$, we can obtain results for any n.

[1] In Nakamura et al. (1994), it is called the ''eddy moisture transport–thermohaline circulation'' feedback. Since the condition includes the interaction between the sea surface temperature and the freshwater flux, it can be called an ''interactive'' salinity boundary condition (J. Marotzke 1996, personal communication).

REFERENCES

Bryan, F., 1987: Parameter sensitivity of primitive equation ocean general circulation models. *J. Phys. Oceanogr.*, **17,** 970–985.

Bryan, K., and M. D. Cox, 1967: A numerical investigation of the oceanic general circulation. *Tellus,* **19,** 54–80.

Colin de Verdière, A., 1988: Buoyancy driven planetary flows. *J. Mar. Res.*, **46,** 215–265.

Dansgaard, W., H. B. Clausen, N. Gundestrup, C. U. Hammer, S. F. Johnsen, P. H. Kristinsdottir, and N. Reeh, 1982: A new Greenland Deep Ice Core. *Nature,* **218,** 1273–1277.

Hallberg, R., 1996: Buoyancy-driven circulation in an ocean basin with isopycnals intersecting the sloping boundary. *J. Phys. Oceanogr.,* **26,** 913–940.

Huang, R. X., 1993: Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation. *J. Phys. Oceanogr.,* **23,** 2428–2446.

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

Hughes, T. M. C., and A. J. Weaver, 1994: Multiple equilibria of an asymmetric two-basin ocean model. *J. Phys. Oceanogr.,* **24,** 619–637.

Joyce, T. M., 1991: Thermohaline catastrophe in a simple four-box model of the ocean climate. *J. Geophys. Res.,* **96,** 20 393–20 402.

Marotzke, J., 1990: Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. thesis, Institut für Meereskund, University of Kiel, Kiel, Germany, 126 pp.

——, 1994: Ocean models in climate problems. *Ocean Processes in Climate Dynamics: Global and Mediterranean Examples,* P. Malanotte-Rizzoli and A. R. Robinson, Eds., Kluwer, 79–109.

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

——, P. Welander, and J. Willebrand, 1988: Instability and multiple steady states in a meridional-plane model. *Tellus,* **40,** 162–172.

Munk, W. H., 1966: Abyssal recipes. *Deep-Sea Res.,* **13,** 707–730.

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

Park, Y.-G., and J. A. Whitehead, 1999: Rotating convection driven by differential bottom heating. *J. Phys. Oceanogr.,* **29,** 1208–1220.

——, and K. Bryan, 2000: Comparison of thermally driven circulations from a depth coordinate model and an isopycnal layer model. Part I: A scaling law—Sensitivity to vertical diffusivity. *J. Phys. Oceanogr.,* in press.

Quon, C., and M. Ghil, 1992: Multiple equilibria in the thermosolutal convection due to salt-flux boundary conditions. *J. Fluid Mech.,* **245,** 449–483.

Rahmstorf, S., 1995: Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature,* **378,** 145–149.

Robinson, A., and H. Stommel, 1959: The oceanic thermocline and the associated thermohaline circulation. *Tellus,* **2,** 295–308.

Rossby, H. T., 1965: On the thermal convection driven by non-uniform heating from below: an experimental study. *Deep-Sea Res.,* **12,** 9–16.

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

Thual, O., and J. C. McWilliams, 1992: The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models. *Geophys. Astrophys. Fluid Dyn.,* **64,** 67–95.

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

Walin, G., 1985: The thermohaline circulation and the control of ice age. *Palaeoclimatol. Palaeoecol.,* **50,** 323–32.

Warren, B. A., 1983: Why there is no deep water formed in the North Pacific? *J. Mar. Res.,* **41,** 327–347.

Weaver, A. J., and T. M. C. Hughes, 1992: Stability and variability of the thermohaline circulation and its link to climate. *Trends Phys. Oceanogr.,* **1,** 15–70.

Welander, P., 1971: The thermocline problem. *Philos. Trans. Roy. Soc. London,* **270A,** 415–421.

——, 1986: Thermocline effects in the ocean circulation and related simple model. *Large-Scale Transport Processes in Oceans and Atmosphere,* J. Willebrand and D. L. T. Anderson, Eds., Reidel, 163–200.

Whitehead, J. A., 1995: Thermohaline ocean processes and models. *Annu. Rev. Fluid Mech.,* **27,** 89–113.

Winton, M., 1996: The role of horizontal boundaries in parameter sensitivity and decadal-scale variability of coarse-resolution ocean general circulation models. *J. Phys. Oceanogr.,* **26,** 289–304.

Wright, D. G., and T. F. Stocker, 1991: A zonally averaged ocean model for the thermohaline circulation. Part I: Model development and flow dynamics. *J. Phys. Oceanogr.,* **21,** 1713–1724.